

Proportional Subdistribution Hazards Model for Competing Risks in Case-Cohort Studies

Adane Fekadu Wogu¹, Shanshan Zhao², Hazel Bogan Nichols³, Jianwen Cai^{4,*}

¹Department of Biostatistics & Informatics, University of Colorado Anschutz Medical Campus, Aurora, USA

²Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, USA

³Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, USA

⁴Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, USA

Email address:

adane.wogu@cuanschutz.edu (A. F. Wogu), shanshan.zhao@nih.gov (Shanshan Zhao), hazel.nichols@unc.edu (H. B. Nichols),

cai@bios.unc.edu (Jianwen Cai)

*Corresponding author

To cite this article:

Adane Fekadu Wogu, Shanshan Zhao, Hazel Bogan Nichols, Jianwen Cai. Proportional Subdistribution Hazards Model for Competing Risks in Case-Cohort Studies. *American Journal of Applied Mathematics*. Vol. 9, No. 5, 2021, pp. 165-185. doi: 10.11648/j.ajam.20210905.12

Received: June 15, 2021; **Accepted:** July 6, 2021; **Published:** September 9, 2021

Abstract: Competing risks refer to the situation where there are multiple causes of failure and the occurrence of one type of event prohibits the occurrence of the other types of event or alters the chance to observe them. In large cohort studies with long-term follow-up, there are often competing risks. When the failure events are rare, or the information on certain risk factors is difficult or costly to measure for the full cohort, a case-cohort study design can be a desirable approach. In this paper, we consider a semiparametric proportional subdistribution hazards model in the presence of competing risks in case-cohort studies. The subdistribution hazards function, unlike the cause-specific hazards function, gives the advantage of outlining the marginal probability of a particular type of event. We propose estimating equations based on inverse probability weighting techniques for the estimation of the model parameters. In the estimation methods, we considered a weighted availability indicator to properly account for the case-cohort sampling scheme. We also proposed a Breslow-type estimator for the cumulative baseline subdistribution hazard function. The resulting estimators are shown, using empirical processes and martingale properties, to be consistent and asymptotically normally distributed. The performance of the proposed methods in finite samples are examined through simulation studies by considering different levels of censoring and event of interest percentages. The simulation results from the different scenarios suggest that the parameter estimates are reasonably close to the true values of the respective parameters in the model. Finally, the proposed estimation methods are applied to a case-cohort sample from the Sister Study, in which we illustrated the proposed methods by studying the association between selected CpGs and invasive breast cancer in the presence of ductal carcinoma *in situ* as competing risk.

Keywords: Case-cohort Study, Competing Risks, Inverse Probability of Censoring Weight, Subdistribution Hazard, Weighted Estimating Equation

1. Introduction

Large epidemiologic cohort studies that require the follow-up of thousands of subjects for a prolonged period of time can generally be expensive as data collection from participating subjects is resource-demanding. Further, when the disease under investigation is rare, large number of subjects must be enrolled to ensure adequate number of failures are observed during the follow-up period. Case-cohort study design [1] is an

efficient alternative to reduce cost and achieve the same goal as a cohort study. A case-cohort study includes a random sample of the entire cohort, named subcohort, plus all subjects who develop the disease of interest regardless of being selected in the random subcohort or not.

Conventional statistical methods for survival data consider that there is only one possible event to occur and define the outcome as the time until the occurrence of that particular event of interest. However, in many studies, it is possible that

there may be other events which ‘compete’ with the event of interest, so that the occurrence of such an event precludes or modifies the chance of the occurrence of the primary event. For instance, in the Sister Study, a prospective observational cohort study designed to assess environmental and genetic risk factors for breast cancer [2], participants can develop either invasive breast cancer or ductal carcinoma in situ (DCIS), a pre-invasive condition of breast cancer. Oftentimes, patients with DCIS receive aggressive treatments such as lumpectomy, hormonal therapy or radiation therapy, which greatly reduce their chance of getting subsequent invasive breast cancer [3]. Thus, we consider DCIS as a competing risk for invasive breast cancer.

A standard approach for competing risks data involves modeling the cause-specific hazard functions of the different competing events under the proportional hazards assumption [4]. Cause-specific hazard function for an event is defined as the instantaneous rate of the event in subjects who are currently event free (i.e., in subjects who have not yet experienced any type of events). The cause-specific hazard function is very useful when the interest is in, for example, understanding disease etiology or comparison of cause-specific hazards across groups. However, the cause-specific hazard function does not have a direct interpretation in terms of survival probabilities for the particular event type [5-7]. Instead, the subdistribution hazard, introduced by Fine & Gray [5], is defined as the instantaneous risk of failure from the event of interest in subjects who have not yet experienced this event of interest (i.e., subjects who have not yet experienced any type of event and subjects who developed another type of event prior to this time point), and it is directly related to the survival probability for the particular event type. Hence, modeling the subdistribution hazards is of greater interest when the focus is on estimating the actual risk and prognosis or if one is interested in the overall effect of covariates on the incidence of the event of a particular type [8, 9].

Despite the considerable literature on statistical methods for case-cohort study [1, 10-16], methods for case-cohort study in the presence of competing risks have been limited. Sørensen & Andersen [17] considered proportional cause-specific hazards model, where they generalized the pseudolikelihood approach proposed by Prentice [1] and Self & Prentice [10] for a single event to competing risks setting. Sun et al. [18] followed a similar approach for additive cause-specific hazards model. Wolkewitz et al. [19] implemented Prentice’s [1] pseudolikelihood approach for cause-specific hazards in the analyses of a real dataset under a case-cohort design. There has not been much work for examining the subdistribution hazard in the presence of competing risks in case-cohort designs.

In this paper, we consider a proportional subdistribution hazards model in the presence of competing risks in case-cohort studies and examine a weighted estimating equation approach for parameter estimation. In Section 2, we introduce the data structure of a case-cohort study design in the presence of competing risks and the proportional hazards model for the subdistribution of a competing risk. Methods for parameter estimation are presented in Section 3. We establish the

asymptotic properties of the proposed estimation procedures in Section 4 and investigate the performance of the proposed methods in finite samples through simulation studies in Section 5. We apply the proposed methods to analyze a case-cohort sample data from the Sister Study in Section 6. We conclude with some final remarks in Section 7. Outline of the proofs of the theorems are available in the Appendix.

2. Data Structure and Model

2.1. Data Structure with Competing Risks

Let n be the total number of independent subjects, indexed by $i = 1, \dots, n$, in the entire cohort. The follow-up period is $[0, \tau]$, where τ is the study end time. Let T_i and C_i be the potential failure and censoring times, respectively, and $\varepsilon_i \in (1, \dots, K)$ be the cause of failure for subject i . Without loss of generality, we denote the event of interest as ‘cause 1’ ($\varepsilon_i = 1$) and refer to it as ‘cause of interest’ or ‘event of interest’ interchangeably. Let $\mathbf{Z}_i(t)$ be a $p \times 1$ possibly time-dependent covariates vector of bounded variation for subject i at time t . Note that the time-dependent covariates are “external” in the sense that they are not affected by the disease process [6, Chapter 6]. From now on, we assume that the cause of failure ε_i is observable whenever T_i is observed. Considering the entire cohort, for right-censored data we observe $\{X_i = \min(T_i, C_i), \Delta_i = I(T_i \leq C_i), \Delta_i \varepsilon_i, \mathbf{Z}_i(t)\}$, $i = 1, \dots, n$, and assume that they are independent and identically distributed.

Let $N_i(t)$ and $Y_i(t)$, $0 \leq t \leq \tau$, denote the counting and “at risk” processes for the event of interest for subject i . Specifically, $N_i(t) = I(T_i \leq t, \varepsilon_i = 1)$ has right-continuous paths taking value 0 if subject i has not experienced the event of interest prior to time t , and value 1 otherwise. The process $Y_i(t) = 1 - N_i(t-)$ has left-continuous sample paths taking value 1 if subject i has (i) neither failed from the cause of interest nor censored by time t or (ii) already failed by time t from a cause other than the cause of interest, and value 0 otherwise. Therefore, the risk set at a specific time point includes not only subjects who have neither been censored nor failed from any cause but also subjects who have already failed from other causes than the cause of interest [5]. Subjects who experienced other event remain in the risk set indefinitely as long as they have not failed from the cause of interest. Without loss of generality, we assume that there are only two types of events or causes of failure ($K = 2$). This does not create any restriction on the generalization of the results, since when there are more than two types of event, all event types other than the event of interest can be combined into one “other” category and considered as the competing event [20].

2.2. Case-Cohort Sampling Design in the Presence of Competing Risks

In a case-cohort study, we sample a random subcohort and all subjects who failed from the event of interest regardless of whether they are in the selected subcohort. Suppose there are

m subjects in the subcohort. Covariate information $\mathbf{Z}_i(t)$ can be decomposed into two parts as $\mathbf{Z}_i(t) = \{\mathbf{Z}_i^C(t), \mathbf{Z}_i^E(t)\}$ where $\mathbf{Z}_i^C(t)$ are available on the entire cohort and $\mathbf{Z}_i^E(t)$ are the expensive covariates which are only available on: (1) the subjects who experienced the event of interest, i.e., $\Delta_i \varepsilon_i = 1$, regardless of whether they were selected into the subcohort or not, and (2) all of the subcohort members regardless of their disease status. We considered simple random sampling (SRS) method for the selection of the subcohort. Let ξ_i be an indicator for subject i being selected into the subcohort. When the subcohort is selected using SRS with fixed sample size, ξ_1, \dots, ξ_n are correlated. Let $P(\xi_i = 1) = \alpha \in (0, 1]$ be the sampling probability for the subcohort and assume that $\tilde{\alpha} = m/n$ converges to α in probability as $n \rightarrow \infty$. Thus, in a case-cohort study, the observable data for subject i is $\{X_i, \Delta_i, \Delta_i \varepsilon_i, \xi_i, \mathbf{Z}_i^C(t), \mathbf{Z}_i^E(t)\}$ if $\xi_i = 1$ or $\Delta_i \varepsilon_i = 1$ and is $\{X_i, \Delta_i, \Delta_i \varepsilon_i, \xi_i, \mathbf{Z}_i^C(t)\}$ if $\xi_i = 0$ and $\Delta_i \varepsilon_i \neq 1$.

2.3. The Proportional Subdistribution Hazards Model

Without loss of generality, we denote the event of interest as ‘event 1’ and we are interested in modeling the cumulative incidence function for this event conditional on the covariates, i.e., $F_1(t|\mathbf{Z}(t)) = P(T \leq t, \varepsilon = 1|\mathbf{Z}(t))$. The corresponding subdistribution hazard is $\lambda_1(t|\mathbf{Z}) = -\partial \log\{1 - F_1(t|\mathbf{Z})\}/\partial t$. We consider a proportional subdistribution hazards model [5]

$$\lambda_1(t|\mathbf{Z}) = \lambda_{1.0}(t) \exp\{\beta_0^T \mathbf{Z}(t)\}, \quad (1)$$

where $\lambda_{1.0}(t) \geq 0$ is an unspecified baseline subdistribution hazard function and β_0 is a $p \times 1$ vector of fixed and unknown parameters. We further denote $M_i^1(t) = N_i(t) - \int_0^t Y_i(u) \lambda_{1.0}(u) \exp\{\beta_0^T \mathbf{Z}_i(u)\} du$, which satisfies the definition of a martingale under the filtration $\mathcal{F}^1(t) = \sigma\{N_i(u), Y_i(u), \mathbf{Z}_i(u), u \leq t, i = 1, \dots, n\}$.

3. Estimation

In the presence of right censoring with the full cohort data, Fine & Gray [5] used the inverse probability censoring weighting (IPCW) techniques [21] to construct estimating equations for the parameters in model (1). Here, we assume that (T, ε) and C are independent; and C is independently distributed with $P(C \geq t) = G(t)$. The common cumulative hazard of the censoring distribution is given by $\Lambda^c(t) = -\log G(t)$. Further, let $\hat{G}(\cdot)$ be the Kaplan-Meier estimator of $G(\cdot)$ which can be calculated based on $\{X_i, 1 - \Delta_i, i = 1, \dots, n\}$. Note that if C is dependent on some covariates in the model, the IPCW methods can still be applied with some modifications. Specifically, if C depends on discrete covariates, $G(\cdot)$ can be estimated by Kaplan-Meier estimator for each combination of the covariates. On the other hand, if C depends on continuous covariates, $G(\cdot)$ can be estimated through modeling the hazard function of the censoring time with the covariates. Fine & Gray [5] considered a time-dependent weight function, $\omega_i(t) = I(C_i \geq T_i \wedge$

$t) \hat{G}(t) / \hat{G}(X_i \wedge t)$, $i = 1, \dots, n$, and their proposed estimator is the solution to the weighted score equation,

$$U_F(\beta) = \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(u) - \bar{\mathbf{Z}}_F(\beta, u)\} \omega_i(u) dN_i(u), \quad (2)$$

where $\bar{\mathbf{Z}}_F(\beta, t) = \hat{\mathbf{S}}_F^{(1)}(\beta, t) / \hat{\mathbf{S}}_F^{(0)}(\beta, t)$, and

$$\hat{\mathbf{S}}_F^{(d)}(\beta, t) = n^{-1} \sum_{i=1}^n \omega_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\}$$

for $d = 0, 1, 2$.

Only the subjects who failed from the cause of interest contribute to the summation in (2) and the other subjects affect the score function $U_F(\cdot)$ only through the at-risk covariate average $\bar{\mathbf{Z}}_F(\cdot)$. Iterative methods, such as the Newton-Raphson or Fisher Scoring, can be used to solve the system of equations $U_F(\beta) = 0$ for β .

Under the case-cohort study design, because $\mathbf{Z}(\cdot)$ is not observed on every subject, the score function in (2) cannot be calculated directly. We consider an approximation for (2) based on the sampling proportions. Specifically, we consider the following weighted estimating equation:

$$U_C(\beta) = \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(u) - \hat{\mathbf{Z}}_C(\beta, u)\} \omega_i(u) dN_i(u) = 0, \quad (3)$$

where

$$\hat{\mathbf{Z}}_C(\beta, t) = \hat{\mathbf{S}}_C^{(1)}(\beta, t) / \hat{\mathbf{S}}_C^{(0)}(\beta, t),$$

$$\hat{\mathbf{S}}_C^{(d)}(\beta, t) = n^{-1} \sum_{i=1}^n \rho_i(t) \omega_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\}$$

for $d=0, 1, 2$,

$$\rho_i(t) = \Delta_i I(\varepsilon_i = 1) + (1 - \Delta_i I(\varepsilon_i = 1)) \xi_i / \hat{\alpha}(t),$$

and $\hat{\alpha}(t)$ is the ratio of the number of subjects in the sub-cohort who are at risk for the event cause of interest at time t to the number of subjects in the entire cohort who are at risk for the event cause of interest at time t . That is,

$$\hat{\alpha}(t) = \frac{\sum_{i=1}^n \xi_i (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t)}{\sum_{i=1}^n (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t)}.$$

Note that if time-dependency is not needed, $\hat{\alpha}(\cdot)$ can be replaced by the inverse of the sampling fraction, m/n . In the score function $U_C(\beta)$ in (3), $\hat{\mathbf{Z}}_C(t)$ includes not only the subcohort members but also all the failures from the cause of interest which occurred outside of the subcohort. The estimator $\hat{\beta}$ is the solution to $U_C(\beta) = 0$.

For the cumulative baseline subdistribution hazard function, $\Lambda_{1.0}(t) = \int_0^t \lambda_{1.0}(u) du$, we propose Breslow-type of estimator:

$$\hat{\Lambda}_{1.0}(t) = n^{-1} \sum_{i=1}^n \int_0^t \frac{1}{\hat{\mathbf{S}}_C^{(0)}(\hat{\beta}, u)} \omega_i(u) dN_i(u), \quad (4)$$

where $\widehat{S}_C^{(0)}(\widehat{\beta}, t)$ is $\widehat{S}_C^{(0)}(\beta, t)$ evaluated at $\widehat{\beta}$. The cumulative incidence at time t for an individual with covariate values $\mathbf{Z} = \mathbf{z}_0$ can be estimated by $\widehat{F}_1(t|\mathbf{z}_0) = 1 - \exp\{-\widehat{\Lambda}_1(t|\mathbf{z}_0)\}$ where

$$\begin{aligned}\widehat{\Lambda}_1(t|\mathbf{z}_0) &= \int_0^t \exp\{\widehat{\beta}^T \mathbf{z}_0(u)\} d\widehat{\Lambda}_{1.0}(u) \\ &= n^{-1} \sum_{i=1}^n \int_0^t \frac{\exp\{\widehat{\beta}^T \mathbf{z}_0(u)\}}{\widehat{S}_C^{(0)}(\widehat{\beta}, u)} \omega_i(u) dN_i(u). \quad (5)\end{aligned}$$

The consistency of $\widehat{\beta}$ for β_0 and uniform convergence of $\widehat{\Lambda}_{1.0}(t)$ to $\Lambda_{1.0}(t)$, both summarized in the next section,

$$\begin{aligned}\mathbf{S}^{(d)}(\beta, t) &= n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\} \quad \text{for } d = 0, 1, 2, \\ \mathbf{s}^{(d)}(\beta, t) &= \lim_{n \rightarrow \infty} \mathbf{S}^{(d)}(\beta, t) \quad \text{for } d = 0, 1, 2, \quad e(\beta, t) = \mathbf{s}^{(1)}(\beta, t) / \mathbf{s}^{(0)}(\beta, t), \\ \mathbf{v}(\beta, t) &= \frac{\mathbf{s}^{(2)}(\beta, t) \mathbf{s}^{(0)}(\beta, t) - \mathbf{s}^{(1)}(\beta, t)^{\otimes 2}}{\mathbf{s}^{(0)}(\beta, t)^2}, \quad \bar{\mathbf{Z}}_i(\beta, t) = \mathbf{Z}_i(t) - e(\beta, t), \\ \Omega(\beta) &= \int_0^\tau \left\{ \frac{\mathbf{s}^{(2)}(\beta, t)}{\mathbf{s}^{(0)}(\beta, t)} - e(\beta, t)^{\otimes 2} \right\} \mathbf{s}^{(0)}(\beta, t) d\Lambda_{1.0}(t) = \int_0^\tau \mathbf{v}(\beta, t) \mathbf{s}^{(0)}(\beta, t) d\Lambda_{1.0}(t), \\ M_i^c(t) &= I(X_i \leq t, \Delta_i = 0) - \int_0^t I(X_i \geq u) d\Lambda^c(u), \\ \pi(u) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n I(X_i \geq u), \quad \tilde{\omega}_i(t) = I(C_i \geq T_i \wedge t) G(t) / G(X_i \wedge t), \\ \phi(t) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t) = E\{(1 - \Delta_1 I(\varepsilon_1 = 1)) Y_1(t)\}, \\ \eta_i(\beta) &= \int_0^\tau \bar{\mathbf{Z}}_i(\beta, t) \tilde{\omega}_i(t) dM_i^1(t), \quad \bar{\rho}_i = \Delta_i I(\varepsilon_i = 1) + (1 - \Delta_i I(\varepsilon_i = 1)) \xi_i / \tilde{\alpha}, \\ \mathbf{r}_i^{(d)}(\beta, t) &= (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t) \left\{ \tilde{\omega}_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\} - \frac{\mathbf{g}^{(d)}(\beta, t)}{\phi(t)} \right\} \quad \text{for } d = 0, 1, \\ \mathbf{g}^{(d)}(\beta, t) &= E\{(1 - \Delta_j I(\varepsilon_j = 1)) \tilde{\omega}_j(t) Y_j(t) \mathbf{Z}_j(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_j(t)\}\} \quad \text{for } d = 0, 1, \\ \mu_i(\beta) &= \int_0^\tau [\mathbf{r}_i^{(1)}(\beta, t) - e(\beta, t) \mathbf{r}_i^{(0)}(\beta, t)] d\Lambda_{1.0}(t), \\ \mathbf{p}(\beta, u, t) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n \frac{1}{\mathbf{s}^{(0)}(\beta, t)} \rho_j(t) \tilde{\omega}_j(t) Y_j(t) \exp\{\beta^T \mathbf{Z}_j(t)\} [\mathbf{Z}_j(t) - e(\beta, t)] I(t \geq u > X_j) \\ \mathbf{q}(\beta, u) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{p}(\beta, u, t) \tilde{\omega}_i(t) dN_i(t), \quad \text{and} \quad \psi_i(\beta) = \int_0^\tau \frac{\mathbf{q}(\beta, u)}{\pi(u)} dM_i^c(u).\end{aligned} \quad (6)$$

We assume the following regularity conditions:

- (a) $\{X_i, \Delta_i, \Delta_i \varepsilon_i, \mathbf{Z}_i\}$ are independent and identically distributed for $i = 1, \dots, n$.
- (b) [Finite Interval]. $\int_0^\tau \lambda_{1.0}(t) dt < \infty$.
- (c) $Z_{ij}(\cdot)$, for $i = 1, \dots, n$; $j = 1, \dots, m$, have bounded total variations. That is, for the j^{th} component of \mathbf{Z}_i , say Z_{ij} , $j = 1, \dots, m$, we have $|Z_{ij}(0)| + \int_0^\tau |dZ_{ij}(t)| \leq D_m < \infty$, almost surely for some constant D_m .
- (d) $P((1 - \Delta)Y(\tau) = 1) > 0$.

ensure the uniform convergence of $\widehat{F}_1(t|\mathbf{z}_0)$ to $F_1(t|\mathbf{z}_0)$ on a suitably chosen interval. Moreover, using similar techniques as in Fine & Gray [5], analytic as well as simulation-based confidence intervals and confidence bands can be constructed subsequently.

4. Asymptotic Properties

In this section, we establish the asymptotic properties of the proposed estimates. For convenience, we define some additional notations.

- (e) [Asymptotic stability]. There exists a compact neighborhood \mathcal{B} of β_0 , and continuous functions $\mathbf{s}^{(0)}$, $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$ (scalar, vector and matrix, respectively) of $\beta \in \mathcal{B} \times [0, \tau]$ such that:

$$i) \lim_{n \rightarrow \infty} \sup_{\substack{t \in [0, \tau] \\ \beta \in \mathcal{B}}} \|\mathbf{S}^{(d)}(\beta, t) - \mathbf{s}^{(d)}(\beta, t)\| \xrightarrow{p} 0.$$

- ii) There exists a positive definite matrix $\Sigma(\beta_0)$ such that $\mathbf{var}\left\{\sum_{i=1}^n [(\eta_i(\beta_0) + \psi_i(\beta_0))^{\otimes 2} + (1 - \alpha)\alpha^{-1}(\mu_i(\beta_0))^{\otimes 2}]\right\} \xrightarrow{p} \Sigma(\beta_0)$.

(f) [Asymptotic regularity conditions].

- i) For all $\beta \in \mathcal{B}$, where \mathcal{B} is a compact set of \mathcal{R}^p with nonempty interior, and $t \in [0, \tau]$, $s^{(d)}(\cdot, t)$ (for $d = 0, 1, 2$) are continuous functions of $\beta \in \mathcal{B}$, uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$.

- ii) The scalar $s^{(0)}(\beta, t)$ is bounded away from zero and the covariance matrix $\Omega(\beta) = \int_0^\tau v(\beta, t) s^{(0)}(\beta, t) d\Lambda_{1.0}(t)$ is positive definite.

(g) [Alternative Lindeberg condition]. There exists $\delta > 0$ such that

$$\lim_{n \rightarrow \infty} n^{-1/2} \sup_{i, t} \|Z_i(t)\| Y_i(t) I\{\beta_0^T Z_i(t) > -\delta\} \xrightarrow{p} 0.$$

(h) [Nontrivial subcohort]. $\lim_{n \rightarrow \infty} \tilde{\alpha} \xrightarrow{p} \alpha$ for some constant $\alpha \in (0, 1)$, $\alpha = P(\xi_i = 1)$.

(i) [Asymptotic normality of subcohort averages at β_0]. For $i = 1, \dots, n$

- i) $\lim_{n \rightarrow \infty} n^{-1} \sup_{i, t} \exp\{\beta_0^T Z_i(t)\}^2 \xrightarrow{p} 0$, and
 ii) $\lim_{n \rightarrow \infty} n^{-1} \sup_{i, t} \|Z_i(t)\|^2 \exp\{\beta_0^T Z_i(t)\}^2 \xrightarrow{p} 0$.

$$\hat{\Omega} = n^{-1} \sum_{i=1}^n \left\{ \frac{\hat{S}_C^{(2)}(\hat{\beta}, X_i)}{\hat{S}_C^{(0)}(\hat{\beta}, X_i)} - \hat{Z}_C(\hat{\beta}, X_i)^{\otimes 2} \right\} \Delta_i I(\varepsilon_i = 1),$$

$$\hat{\Sigma} = n^{-1} \left\{ \sum_{i=1}^n \tilde{\rho}_i [\hat{\eta}_i + \hat{\psi}_i]^{\otimes 2} + (1 - \tilde{\alpha}) \tilde{\alpha}^{-1} \sum_{i=1}^n \tilde{\rho}_i (\hat{\mu}_i)^{\otimes 2} \right\},$$

$$\hat{\eta}_i = \{Z_i(X_i) - \hat{Z}_C(\hat{\beta}, X_i)\} \omega_i(X_i) \Delta_i I(\varepsilon_i = 1) - n^{-1} \sum_{j=1}^n \frac{\omega_j(X_j) \Delta_j I(\varepsilon_j = 1) \omega_i(X_j) Y_i(X_j) \exp\{\hat{\beta}^T Z_i(X_j)\}}{\hat{S}_C^{(0)}(\hat{\beta}, X_j)} \times \{Z_i(X_j) - \hat{Z}_C(\hat{\beta}, X_j)\},$$

$$\hat{\mu}_i = n^{-1} \sum_{j=1}^n \frac{\omega_j(X_j) \Delta_j I(\varepsilon_j = 1)}{\hat{S}_C^{(0)}(\hat{\beta}, X_j)} [\hat{r}_i^{(1)}(\hat{\beta}, X_j) - \hat{Z}_C(\hat{\beta}, X_j) \hat{r}_i^{(0)}(\hat{\beta}, X_j)],$$

$$\hat{r}_i^{(d)}(\hat{\beta}, X_j) = (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(X_j) \left\{ \omega_i(X_j) Z_i(X_j)^{\otimes d} \exp\{\hat{\beta}^T Z_i(X_j)\} - \frac{\hat{g}^{(d)}(\hat{\beta}, X_j)}{\hat{\phi}(X_j)} \right\} \text{ for } d = 0, 1,$$

$$\hat{g}^{(d)}(\hat{\beta}, X_j) = n^{-1} \sum_{k=1}^n (1 - \Delta_k I(\varepsilon_k = 1)) \frac{\xi_k}{\tilde{\alpha}} \omega_k(X_j) Y_k(X_j) Z_k(X_j)^{\otimes d} \exp\{\hat{\beta}^T Z_k(X_j)\} \text{ for } d = 0, 1,$$

$$\hat{\phi}(X_j) = n^{-1} \sum_{i=1}^n \tilde{\rho}_i ((1 - \Delta_i I(\varepsilon_i = 1)) Y_i(X_j)),$$

$$\hat{\psi}_i = \frac{\hat{q}(X_i)}{\hat{\pi}(X_i)} I(\Delta_i = 0) - n^{-1} \sum_{l=1}^n \left(\frac{\hat{q}(X_l)}{\hat{\pi}(X_l)^2} \right) \rho_l(X_l) I(X_i \geq X_l) I(\Delta_l = 0),$$

$$\hat{q}(X_i) = n^{-1} \sum_{j=1}^n \hat{p}(X_i, X_j) \omega_j(X_j) \Delta_j I(\varepsilon_j = 1), \quad \hat{\pi}(X_i) = n^{-1} \sum_{j=1}^n \rho_j(X_i) I(X_j \geq X_i), \text{ and}$$

$$\hat{p}(X_i, X_j) = n^{-1} \sum_{k=1}^n \frac{1}{\hat{S}_C^{(0)}(\hat{\beta}, X_j)} \tilde{\rho}_k \omega_k(X_j) Y_k(X_j) \exp\{\hat{\beta}^T Z_k(X_j)\} \left\{ Z_k(X_j) - \hat{Z}_C(\hat{\beta}, X_j) \right\} I(X_j \geq X_i > X_k).$$

The asymptotic properties are summarized in the following theorems. Note that these regularity conditions are similar to those of Andersen [23, Theorem 4.1] with additional conditions (e), (h) and (i) to accommodate the case-cohort sampling design.

Theorem 4.1 (Asymptotic properties of $\hat{\beta}$). Under Conditions (a)–(i), $\hat{\beta}$ is a consistent estimator of β_0 , i.e., $\hat{\beta} \xrightarrow{p} \beta_0$ as $n \rightarrow \infty$. Further, $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean zero and variance matrix $\Omega(\beta_0)^{-1} \Sigma(\beta_0) \Omega(\beta_0)^{-1}$, where $\Omega(\beta)$ is defined in (6) and $\Sigma(\beta) = E\{[\eta_i(\beta) + \psi_i(\beta)]^{\otimes 2}\} + (1 - \alpha) \alpha^{-1} E\{\mu_i(\beta)^{\otimes 2}\}$.

The consistency of $\hat{\beta}$ can be shown based on an extension of the Inverse Function Theorem [24] by verifying these four conditions: (I) $\partial U_C(\beta)/\partial \beta$ exists and is continuous in an open neighborhood \mathcal{B} of β_0 ; (II) $-n^{-1} \partial U_C(\beta_0)/\partial \beta_0^T$ is a positive definite (and hence invertible) with probability going to one as $n \rightarrow \infty$; (III) $-n^{-1} \partial U_C(\beta)/\partial \beta^T$ converges in probability to a fixed function, $\Omega(\beta)$, uniformly in an open neighborhood \mathcal{B} of β_0 ; and (IV) asymptotic unbiasedness of the estimating functions, $n^{-1} U_C(\beta) \xrightarrow{p} 0$. Verification of (I)–(III) are based on conditions (a)–(i) whereas that of (IV) is additionally based on some algebraic manipulations under empirical processes. The outlines of the proofs are available in the Appendix. The asymptotic variance of $\hat{\beta}$ can be consistently estimated by replacing each component by its empirical counterpart, i.e. $n^{-1} \hat{\Omega}^{-1} \hat{\Sigma} \hat{\Omega}^{-1}$ where

Theorem 4.2 (Asymptotic properties of $\hat{\Lambda}_{1.0}(\hat{\beta}, t)$). Under conditions (a)–(i) listed above, $\hat{\Lambda}_{1.0}(\hat{\beta}, t)$ is a consistent estimator of $\Lambda_{1.0}(t)$. Furthermore, for every $t \in [0, \tau]$,

$$\begin{aligned} \mathcal{W}_{1i}(t) &= \Omega(\beta_0)^{-1} [\eta_i(\beta_0) + \psi_i(\beta_0)] \mathbf{h}(t) + \int_0^t \frac{\mathbf{p}^*(\beta_0, s, t)}{\pi(s)} dM_i^c(s) + \int_0^t \frac{\tilde{\omega}_i(u) dM_i^1(u)}{s^{(0)}(\beta_0, u)}, \\ \mathcal{W}_{2i}(t) &= \left[\Omega(\beta_0)^{-1} \mu_i(\beta_0) \mathbf{h}(t) + \int_0^t r_i^{(0)}(\beta_0, u) \frac{d\Lambda_{1.0}(u)}{s^{(0)}(\beta_0, u)} \right], \quad \mathbf{h}(t) = - \int_0^t \frac{s^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} d\Lambda_{1.0}(u), \quad \text{and} \\ \mathbf{p}^*(\beta, s, t) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n \int_0^t \frac{1}{s^{(0)}(\beta, u)} \tilde{\rho}_j Y_j(u) \exp\{\beta^T \mathbf{Z}_j(u)\} \tilde{\omega}_j(u) I(u \geq s > X_j) d\Lambda_{1.0}(u). \end{aligned}$$

The proof follows the ideas in Kulich & Lin [25] and Kang & Cai [15] by decomposing $n^{\frac{1}{2}} \{\hat{\Lambda}_{1.0}(\hat{\beta}, t) - \Lambda_{1.0}(t)\}$ into three parts and then apply Taylor series expansion,

$n^{\frac{1}{2}} \{\hat{\Lambda}_{1.0}(\hat{\beta}, t) - \Lambda_{1.0}(t)\}$ converges weakly to a mean-zero Gaussian process with variance $\Sigma_{\Lambda_{1.0}}(t) = E\{\mathcal{W}_{11}(t)^{\otimes 2}\} + (1 - \alpha)\alpha^{-1} E\{\mathcal{W}_{21}(t)^{\otimes 2}\}$ where, for $i = 1, \dots, n$,

$$\hat{\Sigma}_{\Lambda_{1.0}}(t) = n^{-1} \left\{ \sum_{i=1}^n \tilde{\rho}_i \{\hat{\mathcal{W}}_{1i}(t)\}^{\otimes 2} + (1 - \tilde{\alpha}) \tilde{\alpha}^{-1} \sum_{i=1}^n \tilde{\rho}_i \{\hat{\mathcal{W}}_{2i}(t)\}^{\otimes 2} \right\},$$

where

$$\begin{aligned} \hat{\mathcal{W}}_{1i}(t) &= \hat{\mathbf{h}}(t)^T \hat{\Omega}^{-1} (\hat{\eta}_i + \hat{\psi}_i) + \int_0^t \frac{\hat{\mathbf{p}}^*(\hat{\beta}, s, t)}{\hat{\pi}(s)} d\hat{M}_i^c(s) + \int_0^t \frac{\omega_i(u) d\hat{M}_i^1(u)}{\hat{S}_C^{(0)}(\hat{\beta}, u)}, \\ \hat{\mathcal{W}}_{2i}(t) &= \hat{\mathbf{h}}(t)^T \hat{\Omega}^{-1} \hat{\mu}_i + \int_0^t \hat{r}_i^{(0)}(\hat{\beta}, u) \frac{d\hat{\Lambda}_{1.0}(u)}{\hat{S}_C^{(0)}(\hat{\beta}, u)}, \quad \hat{\mathbf{h}}(t) = - \int_0^t \frac{\hat{S}_C^{(1)}(\hat{\beta}, u)}{\hat{S}_C^{(0)}(\hat{\beta}, u)} d\hat{\Lambda}_{1.0}(u), \\ \hat{M}_i^1(t) &= I(T_i \leq t, \varepsilon_i = 1) - \int_0^t \{1 - I(T_i \leq u, \varepsilon_i = 1)\} \exp\{\hat{\beta}^T \mathbf{Z}_i(u)\} d\hat{\Lambda}_{1.0}(u), \\ \hat{\mathbf{p}}^*(\hat{\beta}, s, t) &= n^{-1} \sum_{j=1}^n \int_0^t \frac{1}{\hat{S}_C^{(0)}(\hat{\beta}, u)} \tilde{\rho}_j Y_j(u) \exp\{\hat{\beta}^T \mathbf{Z}_j(u)\} \omega_j(u) I(u \geq s > X_j) d\hat{\Lambda}_{1.0}(u), \\ \hat{M}_i^c(t) &= I(X_i \leq t, \Delta_i = 0) - \int_0^t I(X_i \geq u) d\hat{\Lambda}^c(u), \quad \text{and} \quad \hat{\Lambda}^c(t) = \int_0^t \frac{\sum_i \rho_i(u) d\{I(X_i \leq u, \Delta_i = 0)\}}{\sum_i \rho_i(u) I(X_i \geq u)}. \end{aligned}$$

5. Simulation Studies

Simulation studies were conducted to assess the performance of the proposed estimators in finite samples. We considered two covariates, $\mathbf{Z}_i = (Z_{i1}, Z_{i2})$ and two causes of failures ($K = 2$), one is the event of interest ($\varepsilon = 1$) and the other is considered as a competing risk ($\varepsilon = 2$). The subdistribution for type 1 failure (event of interest) was given by

$$\begin{aligned} F_1(t|\mathbf{Z}_i) &= P(T_i \leq t, \varepsilon_i = 1|\mathbf{Z}_i) \\ &= 1 - [1 - p\{1 - \exp(-t)\}]^{\exp(\beta_{11}Z_{i1} + \beta_{12}Z_{i2})}, \end{aligned} \quad (7)$$

where $p = F_1(t = \infty|\mathbf{Z} = 0)$, $0 < p < 1$, is the baseline lifetime survival probability and $1 - (1 - p)^{\exp(\beta_{11}Z_{i1} + \beta_{12}Z_{i2})} = P(\varepsilon_i = 1|\mathbf{Z}_i)$ is the probability of the event of interest for an individual with covariate value \mathbf{Z}_i and p . Further, note that the subdistribution of failure time given in (7) is a unit exponential mixture with probability mass $1 - P(\varepsilon_i = 1|\mathbf{Z}_i = 0) = 1 - p$ at $t = \infty$ and based on the proportional subdistribution hazards model in (1) with baseline

hazard

$$\lambda_{1.0} = \frac{pe^{-t}}{1 - p(1 - e^{-t})}.$$

The subdistribution for type 2 failures (the competing event) is obtained by taking $P(\varepsilon_i = 2|\mathbf{Z}_i) = 1 - P(\varepsilon_i = 1|\mathbf{Z}_i)$ and considering an exponential distribution with rate $\exp(\beta_{21}Z_{i1} + \beta_{22}Z_{i2})$ for $P(T_i \leq t|\varepsilon_i = 2, \mathbf{Z}_i = \mathbf{Z}_i)$. Therefore,

$$\begin{aligned} F_2(t|\mathbf{Z}_i) &= P(T_i \leq t, \varepsilon_i = 2|\mathbf{Z}_i = \mathbf{Z}_i) \\ &= P(\varepsilon_i = 2|\mathbf{Z}_i = \mathbf{Z}_i) \cdot P(T_i \leq t|\varepsilon_i = 2, \mathbf{Z}_i = \mathbf{Z}_i) \\ &= (1 - p)^{\exp(\beta_{11}Z_{i1} + \beta_{12}Z_{i2})} \\ &\quad \times \{1 - \exp[-t \cdot \exp(\beta_{21}Z_{i1} + \beta_{22}Z_{i2})]\}. \end{aligned}$$

On average, in the absence of censoring, $P(\varepsilon_i = 1|\mathbf{Z}_i = \mathbf{Z})$ of individuals with covariate value \mathbf{Z} experience the event of interest whereas $P(\varepsilon_i = 2|\mathbf{Z}_i = \mathbf{Z})$ of individuals experience the competing event in the full cohort. Censoring times were generated from the uniform $[a, b]$ distribution with a and b chosen to depend on the desired percentage of censoring.

Table 1. Summary of simulation results for $(\hat{\beta}_{11}, \hat{\beta}_{12})$ when $(\beta_{11}, \beta_{12}) = (0.5, 0.5)$, $Z_1 \sim N(0, 1)$, $Z_2 \sim N(0, 1)$, $p = 0.3$, and $n = 4,000$.

Percent censored ¹	Events of interest %	Subcohort size	Ratio $n_d:n_{cc}-n_d$	β_{11}				β_{12}			
				$mean(\hat{\beta}_{11})$	\overline{SE}	$SSD(\hat{\beta}_{11})$	C.R.	$mean(\hat{\beta}_{11})$	\overline{SE}	$SSD(\hat{\beta}_{12})$	C.R.
80	6.5	278	1:1	0.509	0.105	0.113	0.954	0.510	0.105	0.108	0.944
		556	1:2	0.505	0.086	0.089	0.936	0.506	0.086	0.088	0.946
		834	1:3	0.503	0.078	0.081	0.940	0.504	0.077	0.081	0.938
90	3.2	131	1:1	0.512	0.161	0.177	0.934	0.509	0.160	0.175	0.942
		263	1:2	0.510	0.131	0.140	0.950	0.508	0.130	0.135	0.956
		394	1:3	0.507	0.118	0.125	0.950	0.508	0.117	0.121	0.942
95	1.7	70	1:1	0.517	0.231	0.276	0.920	0.516	0.230	0.270	0.914
		141	1:2	0.515	0.185	0.204	0.940	0.510	0.184	0.198	0.952
		211	1:3	0.511	0.161	0.181	0.930	0.509	0.167	0.172	0.942

Notations: $mean(\hat{\beta})$: the average of the parameter estimates; \overline{SE} : the average of the estimates of standard errors based on the proposed method; $SSD(\hat{\beta})$: the sample standard deviation of the parameter estimates; C.R.: the coverage rate of the nominal 95% C.Is. Censoring times were generated from $U[0.00, 0.40]$ for 80%, from $U[0.00, 0.18]$ for 90%, and from $U[0.00, 0.09]$ for 95%.

Table 2. Summary of simulation results for $(\hat{\beta}_{11}, \hat{\beta}_{12})$ when $(\beta_{11}, \beta_{12}) = (1.0, 0.5)$, $Z_1 \sim Bin(0.5)$, $Z_2 \sim N(0, 1)$, $p = 0.5$, and $n = 4,000$.

Percent censored ¹	Events of interest %	Subcohort size	Ratio $n_d:n_{cc}-n_d$	β_{11}				β_{12}			
				$mean(\hat{\beta}_{11})$	\overline{SE}	$SSD(\hat{\beta}_{11})$	C.R.	$mean(\hat{\beta}_{11})$	\overline{SE}	$SSD(\hat{\beta}_{12})$	C.R.
80	15.7	747	1:1	1.007	0.120	0.119	0.940	0.508	0.060	0.061	0.944
		1493	1:2	1.000	0.101	0.098	0.960	0.501	0.049	0.051	0.942
		2240	1:3	0.999	0.094	0.095	0.944	0.500	0.045	0.045	0.952
90	7.9	343	1:1	1.015	0.183	0.185	0.946	0.512	0.095	0.097	0.944
		692	1:2	1.012	0.153	0.147	0.962	0.510	0.076	0.081	0.940
		1038	1:3	0.999	0.143	0.151	0.940	0.502	0.069	0.073	0.936
95	4.0	167	1:1	1.039	0.270	0.272	0.944	0.534	0.144	0.142	0.930
		333	1:2	1.013	0.226	0.221	0.962	0.514	0.115	0.118	0.940
		500	1:3	1.011	0.210	0.220	0.942	0.505	0.103	0.110	0.936

Notations: $mean(\hat{\beta})$: the average of the parameter estimates; \overline{SE} : the average of the estimates of standard errors based on the proposed method; $SSD(\hat{\beta})$: the sample standard deviation of the parameter estimates; C.R.: the coverage rate of the nominal 95% C.Is. Censoring times were generated from $U[0.00, 0.38]$ for 80%, from $U[0.00, 0.17]$ for 90%, and from $U[0.00, 0.08]$ for 95%.

In the first scenario, we considered independent standard normal distributions for both covariates with parameter values $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (0.5, 0.5, -0.5, 0.5)$ and $p = 0.3$. In the second scenario, we considered a Bernoulli distribution with mean 0.5 for Z_1 and a standard normal distribution for Z_2 . The true parameter values were $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (1, 0.5, -1, 1)$ and $p = 0.5$. The full cohort sample size was 4,000 in both scenarios. We considered three different levels of censoring percentage: 95%, 90%, and 80%. Let n_{cc} denote the size for the case-cohort sample and n_d denote the number of events of interest (i.e., cases) in the entire cohort. Further, we note that ‘non-case’ refers to a subject who did not fail from the event of interest, thus it indicates a subject who has either never failed from any event or who failed from a competing event. For the case-cohort samples, we considered three average cases to non-cases ratios (i.e., $n_d : (n_{cc} - n_d)$) in the final case-cohort sample: 1:1, 1:2, and 1:3. This is achieved by increasing the size of the random subcohort. For example, suppose we have a full cohort of size 4,000 in which 80% of the subjects are censored, 6.5% experienced the event of interest, and 13.5% experienced the competing event. Thus we have about $4,000 \times 6.5\% = 260$ events of interest in the

cohort, and the ‘non-case’ percentage of subjects in the cohort is 93.5%. To achieve a 1:1 cases to non-cases ratio on average, the subcohort size should be about $260/0.935 \approx 278$, which will include about 260 non-cases and 18 cases. Including all cases outside of the subcohort, the final case-cohort sample will have $278 + (260 - 18) = 520$ subjects. Similarly, to achieve 1:2 cases to non-cases ratio, the subcohort size should be about $520/0.935 \approx 556$.

The simulation study results are summarized in Tables 1 (scenario 1) and 2 (scenario 2) for different case-cohort sample setups, i.e., under different censoring percentages, events of interest percentages and subcohort sizes, from full cohort sizes of 4,000. In the tables, *Events of interest %* and *Subcohort size* denote, respectively, the percentage of the events of interest observed in the entire cohort and the size of the randomly selected subcohort. A total of 500 samples were generated for each case-cohort scheme, under each scenario.

The simulation results suggest that, overall, the parameter estimates are reasonably close to the true values of the respective parameters. The results indicate that the proposed estimation methods provide estimates that have smaller bias and the estimated standard errors are closer to the empirical

standard deviations when the censoring percentages are lower. Considering the size of the subcohort in a case-cohort design, it was observed that, as the subcohort size increases, the parameter estimates tend to have smaller bias and lower standard errors. Moreover, the standard errors from the proposed methods are reasonably close to the empirical variances in both scenarios. The coverage rates of the nominal 95% confidence intervals for $\hat{\beta}$ were found to be in the range of 91–96%. The proposed methods perform well with different types of covariates.

6. Application to the Sister Study

The Sister Study is a long-term prospective cohort study examining the environmental and genetic risk factors for breast cancer and other health conditions, conducted by the National Institute of Environmental Health Sciences (NIEHS). Launched in July 2003, the Sister Study recruited a total of 50,884 women, aged 35–74 years, living in the United States and Puerto Rico, who had at least one sister diagnosed with breast cancer but did not have the disease themselves when they joined the study; and enrollment of the cohort was closed in March 2009. Baseline data collection included a comprehensive Computer Assisted Telephone Interview (CATI) covering a wide range of questions. Follow-up consisted of either a brief annual update on health status and contact information or, every two to three years, a detailed follow-up questionnaire. These follow-up questionnaires track changes in women's health, lifestyles, and environmental exposures over the course of the study. Study response rates were 91% for the third detailed follow-up (completed August 2016).

When a breast cancer or other incident cancer diagnosis is reported via the periodic follow-ups and through telephone calls, e-mails, or correspondence with the study helpdesk, participants are asked for permission to retrieve medical records, pathology reports and/or physician verification to complement self-reported data. To date, medical records have been obtained for > 80% of breast cancer diagnoses; the positive predictive value of a self-reported breast cancer is 99.4% [26]. The Sister Study data collection components and their details, the cohort enrollment and retention as well as other characteristics of the study are published elsewhere [2].

We are interested in investigating the association between DNA methylation signatures and invasive breast cancer risk in non-Hispanic white women because the majority of the women in the Sister Study cohort are non-Hispanic white. Because blood DNA methylations are expensive to measure, it was only available for a case-cohort sample which included: (1) 335 non-Hispanic white women who were diagnosed with incident breast cancer, i.e., either invasive breast cancer or ductal carcinoma *in situ* (DCIS), during the time interval between their blood draw during baseline data collection and May 2008, and (2) a random sample of 620 non-Hispanic white women drawn from the 29,026 participants enrolled in the study by June 2007. Forty-five of the 335 women who developed breast

cancer were included in the subcohort sample. Consequently, our analysis was performed on a case-cohort sample of 910 women. Details of DNA extraction and genome-wide DNA methylation profiling can be found in Xu et al. (2013) [27].

Our main event of interest is the incidence of invasive breast cancer. In the Sister Study, of those diagnosed with breast cancer, 72 were DCIS cases. DCIS is a pre-invasive condition of breast cancer, but without treatment, it could progress to invasive breast cancer over time. Treatment approaches for DCIS frequently include breast conserving surgery and radiation; some women also use hormone therapy or undergo mastectomy and may elect to also have a contralateral prophylactic mastectomy. Consequently, DCIS in women may preclude the observation of invasive breast cancer or greatly reduce the chance of invasive breast cancer [3]. Thus, we considered the DCIS incidents as competing events. The participating women in the case-cohort study are classified by their disease status as having no breast cancer ('no event'), invasive breast cancer, or DCIS.

We implemented the proposed methodology in this dataset to investigate the association between three CpGs, which were identified in a previous study based on the Sister Study [27], and invasive breast cancer (IBC) risk in the presence of DCIS as a competing risk. These three CpGs (and their gene symbols) are cg08287471 (NEK6), cg10237911 (ANKRA2), and cg22385477 (KM-HN-1). We fitted three separate proportional subdistribution hazards models:

$$\lambda_1(t|\mathbf{Z}) = \lambda_{1.0}(t) \exp\{\beta X + \gamma^T \mathbf{Z}\}, \quad (8)$$

where X represents an individual CpG (i.e., cg08287471, cg10237911, or cg22385477). We used the M-values of CpGs, which is more statistically valid and oftentimes recommended measurement, compared to the Beta-values, in differential methylation and other statistical analyses [28]. The vector \mathbf{Z} contains the adjustment variables: smoking status (current smoker, past smoker, or never smoked), BMI (normal/underweight, overweight, or obese), ever used hormone replacement therapy (HRT) (no or yes), and whether mother diagnosed with breast cancer (no or yes). We examined the distribution of the censoring time and found that it is dependent on HRT. Therefore to calculate the inverse probability weights, we used the Kaplan-Meier estimator for the survival function of the censoring time for each of the two HRT groups.

Table 3 shows the baseline characteristics of the subcohort members and of the full cohort participants in the Sister Study. The summary statistics for the full cohort are extracted from Sandler et al. [2] and White et al. [29]. The random subcohort was similar to the full cohort study with regard to the observed characteristics. The average age of the participants was around 55 years. Close to 20% of the participants had a mother with breast cancer and nearly half of the participants used menopausal hormone therapy in the past. Over half of participants of the study were either overweight or obese and just over half of the women never smoked cigarettes. The mean and standard deviations of the M-values of each of the three CpG probes for the subcohort are presented in Table 3.

This information is not available for the full cohort because the DNA methylation profiling and identification of CpG probes were carried out only on the case-cohort sample.

Table 3. Baseline characteristics of Sister Study (the subcohort and the full cohort).

Characteristics	Mean (SD) or N (%)	
	Subcohort (n = 620)	Full Cohort ^a (n = 50,884)
Age	55.1 (9.0)	55.4 (9.8)
Smoking status		
Current smoker	48 (7.7)	4,175 (8.2)
Past smoker	238 (38.4)	18,141 (35.7)
Never smoked	334 (53.9)	28,552 (56.1)
BMI		
Normal/Under.	265 (42.8)	19,438 (38.2)
Overweight	200 (32.3)	16,151 (31.8)
Obese	154 (24.9)	15,278 (30.0)
Ever used HRT		
No	328 (53.0)	27,793 (54.8)
Yes	291 (47.0)	22,932 (45.2)
Mother with BC		
No	480 (79.3)	39,834 (81.3)
Yes	125 (20.7)	9,135 (18.7)
Probe's M-value		
cg08287471	-2.67 (0.37)	—
cg10237911	-1.48 (0.29)	—
cg22385477	-2.57 (0.26)	—

Notations: *SD*: standard deviation; *BMI*: body mass index; *HRT*: hormone replacement therapy; *BC*: breast cancer. ^a Information were obtained from Sandler et al., 2017 and White et al., 2015.

Table 4. Estimated coefficients and standard errors in models for invasive breast cancer in the Sister Study.

Characteristics	$\hat{\beta}$	<i>S.E.</i>	<i>H.R.</i>	95% <i>C.I.</i>
Smoker: Past	0.237	0.444	0.789	(0.330, 1.884)
Smoker: Current	0.240	1.102	1.271	(0.147, 11.022)
BMI: Overweight	0.066	0.512	1.069	(0.392, 2.917)
BMI: Obese	0.212	0.551	1.236	(0.420, 3.636)
HRT	0.810	0.510	0.445	(0.164, 1.208)
Mother with BC	0.161	0.541	1.175	(0.407, 3.392)
cg08287471	1.000	0.638	0.368	(0.105, 1.284)
Smoker: Past	0.320	0.501	0.726	(0.272, 1.937)
Smoker: Current	0.169	0.989	1.184	(0.170, 8.232)
BMI: Overweight	0.030	0.514	0.970	(0.355, 2.655)
BMI: Obese	0.026	0.582	1.027	(0.328, 3.210)
HRT	0.793	0.478	0.453	(0.178, 1.154)
Mother with BC	0.182	0.543	1.200	(0.414, 3.477)
cg10237911	1.303	1.036	0.272	(0.036, 2.070)
Smoker: Past	0.264	0.443	0.768	(0.323, 1.828)
Smoker: Current	0.208	1.101	1.231	(0.142, 10.647)
BMI: Overweight	0.022	0.503	0.979	(0.365, 2.624)
BMI: Obese	0.075	0.567	1.078	(0.355, 3.275)
HRT	0.869	0.495	0.420	(0.159, 1.107)
Mother with BC	0.222	0.517	1.248	(0.453, 3.441)
cg22385477	1.474	0.822	0.229	(0.046, 1.146)

Notations: *S.E.*: standard error; *H.R.*: hazard ratio; *C.I.*: confidence interval of the hazard ratio.

We used age as the time-scale with left truncation at the entry age (age at study enrollment). The observable period of a woman is from the age at entry until the age of diagnosis of breast cancer or the age either at analysis cutoff date (May 15, 2008) or at loss to follow up. We implemented the IPCW technique proposed by Geskus [30] to accommodate the presence of left truncation in modeling the subdistribution hazard. Women with 'no event' had an average follow-up time of 1,695 days (range: 320–1,717 days), those with IBC had an average follow-up time of 524 days (range: 14–1,654 days), whereas women with DCIS had an average follow-up time of 429 days (range: 5–1,212 days). The results from the analyses of the proportional subdistribution hazards models (8) based on the proposed methods are presented in Table 4. The results from the tables suggest that the CpGs tend to be associated with lower risk of invasive breast cancer although the association is not statistically significant.

7. Concluding Remarks

We proposed a proportional subdistribution hazards model to handle competing risks in case-cohort studies. We used the inverse probability censoring weight (IPCW) to address the incompleteness of observations due to competing risks. In order to properly account for the case-cohort sampling scheme, we considered a weighted availability indicator. A Breslow-type estimator was proposed for the cumulative baseline subdistribution hazard function. The proposed score function reduces to the score function for the Cox proportional hazards model when there is no competing risk and the full cohort is observed. The asymptotic properties of the estimators were proved by using martingale properties and the proposed estimators were shown to be consistent and asymptotically normally distributed. Simulation studies were conducted to

investigate finite sample performance of the proposed methods under different settings. The Sister Study dataset was used to study the association between selected CpGs and invasive breast cancer in the presence of DCIS as competing risk for illustration.

Throughout our work, we considered a simple random sampling mechanism for the subcohort selection of the case-cohort design. The proposed methods can be extended to stratified case-cohort sampling [13] or generalized case-cohort sampling [31]. Extension to other types of time-to-event data such as interval-censored data is worthy of further investigation.

Conflict of Interest

The authors declare that they have no competing interests.

Data Accessibility

The Sister Study dataset used to support the findings of this study were supplied by NIEHS after the authors obtained approval. Requests for access to this dataset should be made to the Sister Study Steering Committee of NIEHS. The R codes implemented in simulation studies and the Sister Study data analyses can be obtained from the first author upon request.

Acknowledgements

This work is partially supported by NIH grants P01 CA 142538 (JC, AFW) and T32ES007018 (AFW), and by the Intramural Research Program of NIH, NIEHS (SZ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix

Outline of Proofs of Theorems

We define the following additional notations for convenience purposes.

$$\begin{aligned}\tilde{S}_F^{(d)}(\beta, t) &= n^{-1} \sum_{i=1}^n \tilde{\omega}_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\} \quad \text{for } d = 0, 1, 2, \\ \tilde{S}_C^{(d)}(\beta, t) &= n^{-1} \sum_{i=1}^n \rho_i(t) \tilde{\omega}_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\} \quad \text{for } d = 0, 1, 2, \\ \tilde{Z}_F(\beta, t) &= \tilde{S}_F^{(1)}(\beta, t) / \tilde{S}_F^{(0)}(\beta, t), \quad \text{and} \quad \tilde{Z}_C(\beta, t) = \tilde{S}_C^{(1)}(\beta, t) / \tilde{S}_C^{(0)}(\beta, t).\end{aligned}$$

The following lemmas will be used frequently in proving the theorems.

Lemma 7.1 (Proposition 1 of Self and Prentice (1988)). Let $\mathbf{W}_n = (W_{1n}, \dots, W_{nn})$ and $\boldsymbol{\xi}_n = (\xi_{1n}, \dots, \xi_{nn})$ be independent random variables such that:

- (i) $\boldsymbol{\xi}_n$ is a vector of ones and zeros, with m ones and $n - m$ zeros, each possible configuration of zeros and ones is equally likely and $m/n \rightarrow \alpha \in (0, 1)$.

(ii) For some scalar functions of \mathbf{W}_n , $f_{in}(\mathbf{W}_n)$, and for any $\epsilon > 0$,

$$n^{-1} \sum_{i=1}^n [f_{in}(\mathbf{W}_n) - f_{\cdot n}(\mathbf{W}_n)]^2 I_{(|f_{in}(\mathbf{W}_n) - f_{\cdot n}(\mathbf{W}_n)| > n^{1/2}\epsilon)} \xrightarrow{P} 0 \quad \text{and} \quad S_{fn}^2 \xrightarrow{P} \sigma_f^2,$$

$$\text{where } f_{\cdot n}(\mathbf{W}_n) = n^{-1} \sum_{i=1}^n f_{in}(\mathbf{W}_n) \text{ and } S_{fn}^2 = n^{-1} \sum_{i=1}^n [f_{in}(\mathbf{W}_n) - f_{\cdot n}(\mathbf{W}_n)]^2.$$

(iii) The scalar functions of \mathbf{W}_n , $g_n(\mathbf{W}_n)$, converge in distribution to a Gaussian random variable with mean 0 and variance σ_g^2 .

Then for $h_n(\mathbf{W}_n, \boldsymbol{\xi}_n) = n^{\frac{1}{2}} [m^{-1} \sum_{i=1}^n \xi_{in} f_{in}(\mathbf{W}_n) - f_{\cdot n}(\mathbf{W}_n)]$, $\{g_n(\mathbf{W}_n), h_n(\mathbf{W}_n, \boldsymbol{\xi}_n)\}$ converge in distribution to a bivariate Gaussian random variable with mean 0 and covariate matrix given by

$$\begin{pmatrix} \sigma_g^2 & 0 \\ 0 & (1 - \alpha)\alpha^{-1}\sigma_f^2 \end{pmatrix}.$$

Note that,

$$\begin{aligned} n^{\frac{1}{2}} [m^{-1} \sum_{i=1}^n \xi_{in} f_{in}(\mathbf{W}_n) - f_{\cdot n}(\mathbf{W}_n)] &= n^{-\frac{1}{2}} [\sum_{i=1}^n \frac{\xi_{in}}{m/n} f_{in}(\mathbf{W}_n) - \sum_{i=1}^n f_{in}(\mathbf{W}_n)] \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \left(\frac{\xi_{in}}{\alpha} - 1 \right) f_{in}(\mathbf{W}_n). \end{aligned}$$

Lemma 7.2 (Lemma 3 of Ni and Cai (2018)). Under Conditions (a)–(d), $n^{\frac{1}{2}} \{\hat{\alpha}^{-1}(t) - \tilde{\alpha}^{-1}\}$ converges to a zero-mean Gaussian process, which implies

$$\sup_{0 \leq t \leq \tau} \left| \hat{\alpha}^{-1}(t) - \tilde{\alpha}^{-1} \right| \xrightarrow{p} 0, \text{ and}$$

$$n^{\frac{1}{2}} \{\hat{\alpha}^{-1}(t) - \tilde{\alpha}^{-1}\} = \frac{1}{\tilde{\alpha} E\{(1 - \Delta_i I(\varepsilon_i = 1))Y_i(t)\}} \times n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t) + o_p(1).$$

Lemma 7.3. Under Conditions (a)–(e),

$$\begin{aligned} \text{(i)} \quad n^{-1} \sum_{i=1}^n \xi_i [\omega_i(t) Y_i(t) (1 - \Delta_i I(\varepsilon_i = 1)) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}] &\xrightarrow{p} \\ \alpha E[\tilde{\omega}_i(t) Y_i(t) (1 - \Delta_i I(\varepsilon_i = 1)) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}], \text{ and} \end{aligned}$$

$$\text{(ii)} \quad n^{-1} \sum_{i=1}^n (\xi_i/\tilde{\alpha} - 1) (1 - \Delta_i I(\varepsilon_i = 1)) \omega_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\} \xrightarrow{p} 0,$$

uniformly in $t \in [0, \tau]$ and $\boldsymbol{\beta} \in \mathcal{B}$ for $d = 0, 1, 2$.

Proof Note that, since $\hat{G}(t)$ is a consistent estimator of $G(t)$ (based on the asymptotic properties of the Kaplan-Meier estimator), $\omega_i(t)$ converges to $\tilde{\omega}_i(t)$, $n^{-1} \sum_{i=1}^n \xi_i [\omega_i(t) Y_i(t) (1 - \Delta_i I(\varepsilon_i = 1)) \times \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}]$ converges in probability to $n^{-1} \sum_{i=1}^n \xi_i [\tilde{\omega}_i(t) Y_i(t) (1 - \Delta_i I(\varepsilon_i = 1)) \mathbf{Z}_i(t)^{\otimes d} \times \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}]$ uniformly in $t \in [0, \tau]$ in (i). By the same argument, in (ii), $n^{-1} \sum_{i=1}^n (\xi_i/\tilde{\alpha} - 1) (1 - \Delta_i I(\varepsilon_i = 1)) \omega_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}$ converges in probability to $n^{-1} \sum_{i=1}^n (\xi_i/\tilde{\alpha} - 1) (1 - \Delta_i I(\varepsilon_i = 1)) \tilde{\omega}_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}$ uniformly in $t \in [0, \tau]$. Then by Condition (c), $Z_{ij}^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\} \tilde{\omega}_i(t) Y_i(t) (1 - \Delta_i I(\varepsilon_i = 1))$, $j = 1, \dots, p$, for $d = 0, 1, 2$ are of bounded total variation in $t \in [0, \tau]$ and $\boldsymbol{\beta} \in \mathcal{B}$. Then the respective convergences in probability of the terms in (i) and (ii) uniformly in $t \in [0, \tau]$ and $\boldsymbol{\beta} \in \mathcal{B}$ follows from Lemma 7.1.

Lemma 7.4. Under Conditions (a)–(e),

$$\sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \left\| \widehat{\mathbf{S}}_C^{(d)}(\beta, t) - \mathbf{S}^{(d)}(\beta, t) \right\| \xrightarrow{p} 0$$

for $d = 0, 1, 2$ and hence, as $n \rightarrow \infty$,

$$\sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \left\| \widehat{\mathbf{Z}}_C(\beta, t) - \mathbf{e}(\beta, t) \right\| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \left\| \mathbf{V}_C(\beta, t) - \mathbf{v}(\beta, t) \right\| \xrightarrow{p} 0,$$

where $\mathbf{V}_C(\beta, t) = \widehat{\mathbf{S}}_C^{(2)}(\beta, t) / \widehat{\mathbf{S}}_C^{(0)}(\beta, t) - \widehat{\mathbf{Z}}_C^{\otimes 2}(\beta, t)$.

Proof The proof is an extension of Appendix A.2 of Kulich and Lin (2004). We can write

$$\begin{aligned} \widehat{\mathbf{S}}_C^{(d)}(\beta, t) - \mathbf{S}^{(d)}(\beta, t) &= [\widehat{\mathbf{S}}_C^{(d)}(\beta, t) - \widehat{\mathbf{S}}_F^{(d)}(\beta, t)] + [\widehat{\mathbf{S}}_F^{(d)}(\beta, t) - \mathbf{S}^{(d)}(\beta, t)] \\ &= n^{-1} \sum_{i=1}^n (\rho_i(t) - 1) \omega_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\} + [\widehat{\mathbf{S}}_F^{(d)}(\beta, t) - \mathbf{S}^{(d)}(\beta, t)]. \end{aligned} \quad (9)$$

The second part of the right-hand side of (9) converges to zero by similar arguments in Appendix A of Fine and Gray (1999). Noting that $(\rho_i(t) - 1) = \left(\frac{\xi_i}{\tilde{\alpha}} - 1 \right) (1 - \Delta_i I(\varepsilon_i = 1)) - (\tilde{\alpha}^{-1} - \hat{\alpha}(t)^{-1}) \xi_i (1 - \Delta_i I(\varepsilon_i = 1))$, the first term in the right-hand side of (9) becomes

$$\widehat{\mathbf{S}}_C^{(d)}(\beta, t) - \widehat{\mathbf{S}}_F^{(d)}(\beta, t) = \mathbf{b}_1 - \mathbf{b}_2,$$

where

$$\begin{aligned} \mathbf{b}_1 &= n^{-1} \sum_{i=1}^n (\xi_i / \tilde{\alpha} - 1) (1 - \Delta_i I(\varepsilon_i = 1)) \omega_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\} \text{ and} \\ \mathbf{b}_2 &= (\tilde{\alpha}^{-1} - \hat{\alpha}(t)^{-1}) n^{-1} \sum_{i=1}^n (1 - \Delta_i I(\varepsilon_i = 1)) \xi_i \omega_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\}. \end{aligned} \quad (10)$$

$$\text{Thus, } \left\| \widehat{\mathbf{S}}_C^{(d)}(\beta, t) - \widehat{\mathbf{S}}_F^{(d)}(\beta, t) \right\| \leq \|\mathbf{b}_1\| + \|\mathbf{b}_2\|.$$

By Lemma 3, $\|\mathbf{b}_1\|$ converges to 0 in probability uniformly in $t \in [0, \tau]$ and $\beta \in \mathcal{B}$. For \mathbf{b}_2 , we have $\{\hat{\alpha}^{-1}(t) - \tilde{\alpha}^{-1}\}$ converges to 0 in probability uniformly in $t \in [0, \tau]$ by Lemma 7.2, together with (i) of Lemma 3, $\|\mathbf{b}_2\|$ converges to 0 in probability uniformly in $t \in [0, \tau]$ and $\beta \in \mathcal{B}$. This concludes that the first part of the right-hand side of (9) converges to 0 in probability uniformly in $t \in [0, \tau]$ and $\beta \in \mathcal{B}$. Consequently $\sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \left\| \widehat{\mathbf{S}}_C^{(d)}(\beta, t) - \mathbf{S}^{(d)}(\beta, t) \right\| \xrightarrow{p} 0$ for $d = 0, 1$. Furthermore,

both $E\{\widehat{\mathbf{S}}_C^{(0)}(\beta, t)\}$ and $E\{\widehat{\mathbf{S}}_F^{(0)}(\beta, t)\}$ are bounded away from zero in $\mathcal{B} \times [0, \tau]$, it follows that, as $n \rightarrow \infty$,

$$\sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \left\| \widehat{\mathbf{Z}}_C(\beta, t) - \mathbf{e}(\beta, t) \right\| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \left\| \mathbf{V}_C(\beta, t) - \mathbf{v}(\beta, t) \right\| \xrightarrow{p} 0.$$

Lemma 5 Under Conditions (a)–(d), we have

$$n^{\frac{1}{2}} \{ \widetilde{\mathbf{S}}_F^{(d)}(\beta_0, t) - \widetilde{\mathbf{S}}_C^{(d)}(\beta_0, t) \} = n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}} \right) \mathbf{r}_i^{(d)}(\beta_0, t) + o_p(1) \text{ for } d = 0, 1,$$

where

$$\mathbf{r}_i^{(d)}(\beta, t) = (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t) \left\{ \tilde{\omega}_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_i(t)\} - \frac{\mathbf{g}^{(d)}(\beta, t)}{\phi(t)} \right\},$$

and

$$\mathbf{g}^{(d)}(\beta_0, t) = E\{(1 - \Delta_j I(\varepsilon_j = 1)) \tilde{\omega}_j(t) Y_j(t) \mathbf{Z}_j(t)^{\otimes d} \exp\{\beta^T \mathbf{Z}_j(t)\}\} \text{ for } d = 0, 1.$$

Proof

$$\begin{aligned} n^{\frac{1}{2}} \{ \tilde{\mathbf{S}}_F^{(d)}(\boldsymbol{\beta}_0, t) - \tilde{\mathbf{S}}_C^{(d)}(\boldsymbol{\beta}_0, t) \} &= n^{-\frac{1}{2}} \sum_{i=1}^n (1 - \rho_i(t)) \tilde{\omega}_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i(t)\} \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) (1 - \Delta_i I(\varepsilon_i = 1)) \tilde{\omega}_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i(t)\} \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^n (\tilde{\alpha}^{-1} - \hat{\alpha}(t)^{-1}) (1 - \Delta_i I(\varepsilon_i = 1)) \xi_i \tilde{\omega}_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i(t)\}. \end{aligned}$$

From Lemma 7.2, $n^{\frac{1}{2}} \{ \hat{\alpha}^{-1}(t) - \tilde{\alpha}^{-1} \} = \frac{n^{-\frac{1}{2}}}{\tilde{\alpha} \phi(t)} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t) + o_p(1)$, where $\phi(t) = E\{(1 - \Delta_1 I(\varepsilon_1 = 1)) Y_1(t)\}$. Therefore,

$$\begin{aligned} n^{\frac{1}{2}} \{ \tilde{\mathbf{S}}_F^{(d)}(\boldsymbol{\beta}_0, t) - \tilde{\mathbf{S}}_C^{(d)}(\boldsymbol{\beta}_0, t) \} &= n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) (1 - \Delta_i I(\varepsilon_i = 1)) \tilde{\omega}_i(t) Y_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i(t)\} \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \left(\frac{\xi_i}{\tilde{\alpha}} - 1\right) (1 - \Delta_i I(\varepsilon_i = 1)) \frac{Y_i(t)}{\phi(t)} \\ &\quad \times \left\{ n^{-1} \sum_{j=1}^n (1 - \Delta_j I(\varepsilon_j = 1)) \frac{\xi_j}{\tilde{\alpha}} \tilde{\omega}_j(t) Y_j(t) \mathbf{Z}_j(t)^{\otimes d} \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_j(t)\} \right\} + o_p(1). \end{aligned}$$

Since $n^{-1} \sum_{j=1}^n (1 - \Delta_j I(\varepsilon_j = 1)) \frac{\xi_j}{\tilde{\alpha}} \tilde{\omega}_j(t) Y_j(t) \mathbf{Z}_j(t)^{\otimes d} \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_j(t)\}$ converges to $\mathbf{g}^{(d)}(\boldsymbol{\beta}_0, t)$ uniformly in $t \in [0, \tau]$ for $d = 0, 1$ by Lemma 3, we can show that

$$\begin{aligned} n^{\frac{1}{2}} \{ \tilde{\mathbf{S}}_F^{(d)}(\boldsymbol{\beta}_0, t) - \tilde{\mathbf{S}}_C^{(d)}(\boldsymbol{\beta}_0, t) \} &= n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) (1 - \Delta_i I(\varepsilon_i = 1)) Y_i(t) \\ &\quad \times \left\{ \tilde{\omega}_i(t) \mathbf{Z}_i(t)^{\otimes d} \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i(t)\} - \frac{\mathbf{g}^{(d)}(\boldsymbol{\beta}_0, t)}{\phi(t)} \right\} + o_p(1) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n (1 - (\xi_i/\tilde{\alpha})) \mathbf{r}_i^{(d)}(\boldsymbol{\beta}_0, t) + o_p(1). \end{aligned}$$

Proof of Theorem 1

Let $\mathbf{U}_C^*(\boldsymbol{\beta}) = n^{-1} \mathbf{U}_C(\boldsymbol{\beta})$. One can employ an extension of Foutz's theorem (Foutz, 1977) to prove the consistency of $\hat{\boldsymbol{\beta}}$ if the following conditions are established;

- (C1) $\partial \mathbf{U}_C^*(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^T$ exists and is continuous in an open neighborhood \mathcal{B} of $\boldsymbol{\beta}_0$,
- (C2) $-\partial \mathbf{U}_C^*(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^T|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is positive definite with probability going to one as $n \rightarrow \infty$,
- (C3) $\partial \mathbf{U}_C^*(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^T$ converges to $\boldsymbol{\Omega}$ uniformly in an open neighborhood \mathcal{B} of $\boldsymbol{\beta}_0$, and
- (C4) $\mathbf{U}_C^*(\boldsymbol{\beta}_0) \rightarrow 0$ in probability as $n \rightarrow \infty$.

We can show that

$$\begin{aligned} \frac{\partial \mathbf{U}_C^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} &= -n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\tilde{\mathbf{S}}_C^{(2)}(\boldsymbol{\beta}, t)}{\tilde{\mathbf{S}}_C^{(0)}(\boldsymbol{\beta}, t)} - \hat{\mathbf{Z}}_C^{\otimes 2}(\boldsymbol{\beta}, t) \right\} \tilde{\omega}_i(t) dN_i(t) \\ &= -n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{V}_C(\boldsymbol{\beta}, t) \omega_i(t) dN_i(t), \end{aligned}$$

since $\omega_i(t)dN_i(t) = \tilde{\omega}_i(t)dN_i(t)$. By the continuity of $\hat{S}_C^{(d)}(\beta, t)$ in $\beta \in \mathcal{B}$ for $d = 0, 1, 2$, condition (C1) is fulfilled. Since $dN_i(t) = dM_i^1(t) + Y_i(t) \exp\{\beta^T Z_i(t)\} \lambda_{1.0}(t)dt$ and by some algebraic manipulations, we have

$$\begin{aligned} \left\| \left(-\frac{\partial U_C^*(\beta)}{\partial \beta^T} \right) - \Omega(\beta) \right\| &\leq \left\| n^{-1} \sum_{i=1}^n \int_0^\tau \{V_C(\beta, t) - v(\beta, t)\} \tilde{\omega}_i(t) dN_i(t) \right\| \\ &+ \left\| n^{-1} \sum_{i=1}^n \int_0^\tau v(\beta, t) \tilde{\omega}_i(t) dM_i^1(t) \right\| \\ &+ \left\| \int_0^\tau v(\beta, t) \{ \tilde{S}_F^{(0)}(\beta, t) - s^{(0)}(\beta, t) \} \lambda_{1.0}(t) dt \right\|. \end{aligned} \quad (11)$$

We will show that each of the three terms on the right side of (11) uniformly converges to 0 for $\beta \in \mathcal{B}$. First, by Lemma ??, $\sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \|V_C(\beta, t) - v(\beta, t)\| \xrightarrow{p} 0$ as $n \rightarrow \infty$. Further, since $\hat{S}_C^{(d)}(\beta, t)$ for $d = 1, 2, 3$ are of bounded variations,

$$V_C(\beta, t) = \frac{\hat{S}_C^{(2)}(\beta, t) \hat{S}_C^{(0)}(\beta, t) - \hat{S}_C^{(1)}(\beta, t)^{\otimes 2}}{\hat{S}_C^{(0)}(\beta, t)^2} \quad (12)$$

is of bounded variation and can be expressed as the difference of two monotone functions. Second, let $\bar{N}(t) = \sum_{i=1}^n \tilde{\omega}_i(t) N_i(t)$ and, by Theorem I.1 (part (a)) of Andersen and Gill (1982) for the application of Lenglar's inequality, for all $\delta, \gamma > 0$,

$$P[n^{-1} \bar{N}(\tau) > \gamma] \leq \frac{\delta}{\gamma} + P\left[\int_0^\tau \tilde{S}_F^{(0)}(\beta, t) \lambda_{1.0}(t) dt > \delta\right]. \quad (13)$$

Under regularity conditions, $\sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \|\tilde{S}_F^{(0)}(\beta, t) - s^{(0)}(\beta, t)\| \xrightarrow{p} 0$ as $n \rightarrow \infty$.

Therefore, for $\delta > \int_0^\tau s^{(0)}(\beta, t) \lambda_{1.0}(t) dt$, $P\left[\int_0^\tau \tilde{S}_F^{(0)}(\beta, t) \lambda_{1.0}(t) dt > \delta\right] \rightarrow 0$ as $n \rightarrow \infty$. Consequently, $\lim_{\gamma \rightarrow \infty} \lim_{n \rightarrow \infty} P[n^{-1} \bar{N}(\tau) > \gamma] = 0$. Combining these results, by Lemma 1 of Lin *et al.* (2000), the first term on the right side of (11) converges to 0 in probability for $\beta \in \mathcal{B}$ as $n \rightarrow \infty$.

For the second term on the right side of (11), let $\bar{M}^1(t) = \sum_{i=1}^n \tilde{\omega}_i(t) M_i^1(t)$. By Theorem I.1 (part (b)) and (2.3) of Andersen and Gill (1982), for all $\delta, \gamma > 0$,

$$P\left[\sup_{0 \leq t \leq \tau} \left\| \int_0^\tau v(\beta, t) \frac{d\bar{M}^1(t)}{n} \right\| > \gamma\right] \leq \frac{\delta}{\gamma^2} + P\left[n^{-1} \int_0^\tau \{v_{ll'}(\beta, t)\}^2 \tilde{S}_F^{(0)}(\beta, t) \lambda_{1.0}(t) dt > \delta\right], \quad (14)$$

where ll' denotes the (l, l') element of the indicated matrix. By regularity conditions, the term

$$\int_0^\tau \{v_{ll'}(\beta, t)\}^2 \tilde{S}_F^{(0)}(\beta, t) \lambda_{1.0}(t) dt$$

is bounded in probability uniformly in $\beta \in \mathcal{B}$, hence the second term on the right side of the inequality in (14) uniformly converges in probability to zero for $\beta \in \mathcal{B}$ as $n \rightarrow \infty$. Since $\delta > 0$ may be taken arbitrarily small, it follows that $P\left[\sup_{0 \leq t \leq \tau} \left\| \int_0^\tau v(\beta, t) \frac{d\bar{M}^1(t)}{n} \right\| > \gamma\right]$ converges in probability to 0 for $\beta \in \mathcal{B}$. Consequently, the second term in (11) uniformly converges in probability to 0 for $\beta \in \mathcal{B}$ as $n \rightarrow \infty$.

Finally, under regularity conditions, the third term on the right side of (11) uniformly converges in probability to zero for $\beta \in \mathcal{B}$ as $n \rightarrow \infty$. Thus, $\left(-\frac{\partial U_C^*(\beta)}{\partial \beta^T} \right) - \Omega(\beta) \xrightarrow{p} 0$ uniformly for $\beta \in \mathcal{B}$ as $n \rightarrow \infty$ and hence (C2) and (C3) are fulfilled.

For (C4), we will investigate the asymptotic properties of $n^{-\frac{1}{2}}U_C(\beta_0)$. Note that

$$\begin{aligned} n^{-\frac{1}{2}}U_C(\beta_0) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \tilde{Z}_F(\beta_0, t)\} \tilde{\omega}_i(t) dN_i(t) \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \{\tilde{Z}_F(\beta_0, t) - \hat{Z}_C(\beta_0, t)\} \tilde{\omega}_i(t) dN_i(t), \end{aligned} \quad (15)$$

since $\tilde{\omega}_i(t) dN_i(t) = \omega_i(t) dN_i(t)$. The first term on the right-hand side of (15) is the pseudo score function of the full cohort data and, by the results in Fine and Gray (1999), it is asymptotically equivalent to $n^{-\frac{1}{2}} \sum_{i=1}^n \eta_i(\beta_0)$ where

$$\eta_i(\beta_0) = \int_0^\tau \{Z_i(t) - e(\beta, t)\} \tilde{\omega}_i(t) dM_i^1(t).$$

To explore the second term on the right-hand side of (15), first note that

$$\frac{\partial}{\partial \omega_j(t)} \{\tilde{Z}_F(\beta_0, t) - \hat{Z}_C(\beta_0, t)\} = -\frac{\rho_j(t) Y_j(t) \exp\{\beta_0^T Z_j(t)\}}{n \tilde{S}_C^{(0)}(\beta_0, t)} [Z_j(t) - \hat{Z}_C(\beta_0, t)].$$

Using this result, the first order Taylor series expansion of $\{\tilde{Z}_F(\beta_0, t) - \hat{Z}_C(\beta_0, t)\}$ with respect to $\omega_j(t)$ around $\tilde{\omega}_j(t)$ gives

$$\{\tilde{Z}_F(\beta_0, t) - \hat{Z}_C(\beta_0, t)\} \approx \{\tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t)\} - \sum_{j=1}^n R_j(\beta_0, t) \{\omega_j(t) - \tilde{\omega}_j(t)\},$$

where $R_j(\beta_0, t) = \frac{\rho_j(t) Y_j(t) \exp\{\beta_0^T Z_j(t)\}}{n \tilde{S}_C^{(0)}(\beta_0, t)} [Z_j(t) - \tilde{Z}_C(\beta_0, t)]$. Therefore, the second term on the right-hand side of (15) becomes

$$\begin{aligned} n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \{\tilde{Z}_F(\beta_0, t) - \hat{Z}_C(\beta_0, t)\} \tilde{\omega}_i(t) dN_i(t) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \{\tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t)\} \tilde{\omega}_i(t) dN_i(t) \\ &\quad - n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \sum_{j=1}^n R_j(\beta_0, t) \{\omega_j(t) - \tilde{\omega}_j(t)\} \tilde{\omega}_i(t) dN_i(t). \end{aligned} \quad (16)$$

Note that

$$\begin{aligned} \tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t) &= \frac{1}{\tilde{S}_F^{(0)}(\beta_0, t)} (\tilde{S}_F^{(1)}(\beta_0, t) - \tilde{S}_C^{(1)}(\beta_0, t)) \\ &\quad - \frac{\tilde{S}_C^{(1)}(\beta_0, t)}{\tilde{S}_F^{(0)}(\beta_0, t) \tilde{S}_C^{(0)}(\beta_0, t)} (\tilde{S}_F^{(0)}(\beta_0, t) - \tilde{S}_C^{(0)}(\beta_0, t)), \end{aligned}$$

and, from Lemma 5, we have

$$n^{\frac{1}{2}} \{\tilde{S}_F^{(d)}(\beta_0, t) - \tilde{S}_C^{(d)}(\beta_0, t)\} = n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) \mathbf{r}_i^{(d)}(\beta_0, t) + o_p(1) \text{ for } d = 0, 1.$$

Hence,

$$\begin{aligned} n^{\frac{1}{2}} \{\tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t)\} &= n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) \\ &\quad \times \left[\frac{1}{\tilde{S}_F^{(0)}(\beta_0, t)} \mathbf{r}_i^{(1)}(\beta_0, t) - \frac{\tilde{S}_C^{(1)}(\beta_0, t)}{\tilde{S}_F^{(0)}(\beta_0, t) \tilde{S}_C^{(0)}(\beta_0, t)} \mathbf{r}_i^{(0)}(\beta_0, t) \right] + o_p(1). \end{aligned} \quad (17)$$

Now, the first term on the right-hand side of (16) is

$$n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \{ \tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t) \} \tilde{\omega}_i(t) dN_i(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \{ \tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t) \} \tilde{\omega}_i(t) dM_i^1(t) \\ + n^{\frac{1}{2}} \int_0^\tau \{ \tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t) \} \tilde{S}_F^{(0)}(\beta_0, t) d\Lambda_{1.0}(t). \quad (18)$$

Under Conditions (a)–(e), we can show that $\sup_{\substack{0 \leq t \leq \tau \\ \beta \in \mathcal{B}}} \left\| \tilde{Z}_F(\beta, t) - \tilde{Z}_C(\beta, t) \right\| \xrightarrow{p} 0$ by the same arguments of Kulich and Lin

(2004) (Appendix A.2). Further, $n^{-\frac{1}{2}} \sum_{i=1}^n \tilde{\omega}_i(u) dM_i^1(u)$ converges weakly to a zero-mean Gaussian process. Therefore, the first term on the right-hand side of (18) converges in probability to 0 uniformly in $t \in [0, \tau]$ and $\beta \in \mathcal{B}$ by Lemma 1 of Lin *et al.* (2000). Based on (17), the second term on the right-hand side of (18) becomes

$$n^{\frac{1}{2}} \int_0^\tau \{ \tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t) \} \tilde{S}_F^{(0)}(\beta_0, t) d\Lambda_{1.0}(t) \\ = \int_0^\tau n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}} \right) \left[\frac{1}{\tilde{S}_F^{(0)}(\beta_0, t)} r_i^{(1)}(\beta_0, t) - \frac{\tilde{S}_C^{(1)}(\beta_0, t)}{\tilde{S}_F^{(0)}(\beta_0, t) \tilde{S}_C^{(0)}(\beta_0, t)} r_i^{(0)}(\beta_0, t) \right] \times \tilde{S}_F^{(0)}(\beta_0, t) d\Lambda_{1.0}(t) + o_p(1) \\ = n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}} \right) \int_0^\tau [r_i^{(1)}(\beta_0, t) - \tilde{Z}_C(\beta_0, t) r_i^{(0)}(\beta_0, t)] d\Lambda_{1.0}(t) + o_p(1).$$

Under the regularity conditions, $\tilde{S}_C^{(d)}(\beta_0, t)$ converges to $s^{(d)}(\beta_0, t)$ uniformly in $t \in [0, \tau]$ for $d = 0, 1, 2$ and hence $\tilde{Z}_C(\beta_0, t)$ converges to $e(\beta_0, t)$ uniformly in $t \in [0, \tau]$. Therefore,

$$n^{\frac{1}{2}} \int_0^\tau \{ \tilde{Z}_F(\beta_0, t) - \tilde{Z}_C(\beta_0, t) \} \tilde{S}_F^{(0)}(\beta_0, t) d\Lambda_{1.0}(t) \\ = n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}} \right) \int_0^\tau (r_i^{(1)}(\beta_0, t) - e(\beta_0, t) r_i^{(0)}(\beta_0, t)) d\Lambda_{1.0}(t) + o_p(1) \\ = n^{-\frac{1}{2}} \sum_{i=1}^n (1 - (\xi_i/\tilde{\alpha})) \mu_i(\beta_0) + o_p(1), \quad (19)$$

where $\mu_i(\beta_0) = \int_0^\tau [r_i^{(1)}(\beta_0, t) - e(\beta_0, t) r_i^{(0)}(\beta_0, t)] d\Lambda_{1.0}(t)$.

Now consider the second term on the right-hand side of (16), and call it $n^{-\frac{1}{2}} \varphi(\beta_0, \tau)$; i.e.,

$$n^{-\frac{1}{2}} \varphi(\beta_0, \tau) = -n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \sum_{j=1}^n \mathbf{R}_j(\beta_0, t) \{ \omega_j(t) - \tilde{\omega}_j(t) \} \tilde{\omega}_i(t) dN_i(t).$$

Since $\omega_j(t) - \tilde{\omega}_j(t) = -I(X_j < t) \tilde{\omega}_j(t) \int_{X_j}^t \frac{d\bar{M}^c(u)}{n\pi(u)} + o_p(1)$ based on Fine and Gray (1999), and hence

$$n^{-\frac{1}{2}} \varphi(\beta_0, \tau) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \sum_{j=1}^n \mathbf{R}_j(\beta_0, t) I(X_j < t) \tilde{\omega}_j(t) \int_{X_j}^t \frac{d\bar{M}^c(u)}{n\pi(u)} \tilde{\omega}_i(t) dN_i(t) + o_p(1) \\ = n^{-\frac{1}{2}} \sum_{k=1}^n \int_0^\tau \mathbf{Q}(\beta_0, u) \frac{dM_k^c(u)}{\pi(u)} + o_p(1),$$

where

$$\mathbf{Q}(\beta_0, u) = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \int_0^\tau \mathbf{R}_j(\beta_0, t) I(t \geq u > X_j) \tilde{\omega}_j(t) \tilde{\omega}_i(t) dN_i(t).$$

Further, it can be shown that $\sum_{j=1}^n \mathbf{R}_j(\beta_0, t) I(t \geq u > X_j) \tilde{\omega}_j(t)$ converges to $\mathbf{p}(\beta_0, u, t)$ uniformly in $t \in [0, \tau]$, where

$$\mathbf{p}(\beta_0, u, t) = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n \frac{I(t \geq u > X_j)}{s^{(0)}(\beta_0, t)} \tilde{\rho}_j \tilde{\omega}_j(t) Y_j(t) \exp\{\beta_0^T \mathbf{Z}_j(t)\} [\mathbf{Z}_j(t) - \mathbf{e}(\beta_0, t)],$$

noting that $(\rho_j(t) - \tilde{\rho}_j)$ converges in probability to 0 uniformly in $t \in [0, \tau]$ based on Lemma 7.2. Hence, by Lemma 1 of Lin *et al.* (2000), $\mathbf{Q}(\beta_0, u)$ can be shown to be asymptotically equivalent to $n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{p}(\beta_0, u, t) \tilde{\omega}_i(t) dN_i(t)$.

Finally, let $\mathbf{q}(\beta_0, u) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{p}(\beta_0, u, t) \tilde{\omega}_i(t) dN_i(t)$. Based on these results, we have

$$n^{-\frac{1}{2}} \boldsymbol{\varphi}(\beta_0, \tau) = n^{-\frac{1}{2}} \sum_{k=1}^n \int_0^\tau \frac{\mathbf{q}(\beta_0, u)}{\pi(u)} dM_k^c(u) + o_p(1) = n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{\psi}_i(\beta_0) + o_p(1),$$

where $\boldsymbol{\psi}_i(\beta_0) = \int_0^\tau \frac{\mathbf{q}(\beta_0, u)}{\pi(u)} dM_i^c(u)$.

Therefore, combining these results, we have

$$n^{-\frac{1}{2}} \mathbf{U}_C(\beta_0, t) = n^{-\frac{1}{2}} \sum_{i=1}^n [\boldsymbol{\eta}_i(\beta_0) + \boldsymbol{\psi}_i(\beta_0) + (1 - (\xi_i/\tilde{\alpha})) \boldsymbol{\mu}_i(\beta_0)] + o_p(1). \quad (20)$$

Following Lemma 7.1, $n^{-\frac{1}{2}} \mathbf{U}_C(\beta_0)$ is asymptotically normal with covariance matrix $\boldsymbol{\Sigma}(\beta_0) = E\{[\boldsymbol{\eta}_i(\beta_0) + \boldsymbol{\psi}_i(\beta_0)]^{\otimes 2}\} + (1 - \alpha)\alpha^{-1} E\{\boldsymbol{\mu}_i(\beta_0)^{\otimes 2}\}$. Further, by Taylor series expansion, $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\boldsymbol{\Omega}(\beta_0)^{-1} \boldsymbol{\Sigma}(\beta_0) \boldsymbol{\Omega}(\beta_0)^{-1}$ where $\boldsymbol{\Omega}(\beta)$ is as defined in Section 4.

Proof of Theorem 2

Using the martingale decomposition of $N_i(t)$ and after some algebraic manipulations for $n^{\frac{1}{2}}\{\hat{\Lambda}_{1.0}(\hat{\beta}, t) - \Lambda_{1.0}(t)\}$, we have

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{\Lambda}_{1.0}(\hat{\beta}, t) - \Lambda_{1.0}(t)\} &= n^{\frac{1}{2}} \sum_{i=1}^n \int_0^t \left(\frac{1}{n\hat{S}_C^{(0)}(\hat{\beta}, u)} - \frac{1}{n\hat{S}_C^{(0)}(\beta_0, u)} \right) \tilde{\omega}_i(u) dN_i(u) \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \left(\frac{1}{\hat{S}_C^{(0)}(\beta_0, u)} - \frac{1}{\tilde{S}_C^{(0)}(\beta_0, u)} \right) \tilde{\omega}_i(u) dN_i(u) \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{1}{\tilde{S}_C^{(0)}(\beta_0, u)} \tilde{\omega}_i(u) dM_i^1(u) \\ &\quad + n^{\frac{1}{2}} \int_0^t \left\{ \frac{\tilde{S}_F^{(0)}(\beta_0, u) - \tilde{S}_C^{(0)}(\beta_0, u)}{\tilde{S}_C^{(0)}(\beta_0, u)} \right\} d\Lambda_{1.0}(u), \end{aligned} \quad (21)$$

since $\omega_i(t) dN_i(t) = \tilde{\omega}_i(t) dN_i(t)$. Using the martingale decomposition of $N_i(t)$, the first term on the right-hand side of (21) can be written as

$$\begin{aligned} &n^{\frac{1}{2}} \sum_{i=1}^n \int_0^t \left(\frac{1}{n\hat{S}_C^{(0)}(\hat{\beta}, u)} - \frac{1}{n\hat{S}_C^{(0)}(\beta_0, u)} \right) \tilde{\omega}_i(u) dN_i(u) \\ &= n^{\frac{1}{2}} \sum_{i=1}^n \int_0^t \left(\frac{1}{n\hat{S}_C^{(0)}(\hat{\beta}, u)} - \frac{1}{n\hat{S}_C^{(0)}(\beta_0, u)} \right) \tilde{\omega}_i(u) dM_i^1(u) \\ &\quad + n^{\frac{1}{2}} \int_0^t \left(\frac{1}{\hat{S}_C^{(0)}(\hat{\beta}, u)} - \frac{1}{\hat{S}_C^{(0)}(\beta_0, u)} \right) S_F^{(0)}(\beta_0, u) d\Lambda_{1.0}(u). \end{aligned} \quad (22)$$

By Taylor series expansion of $\widehat{S}_C^{(0)}(\widehat{\beta}, u)^{-1}$ around β_0 , we have

$$\left(\frac{1}{n\widehat{S}_C^{(0)}(\widehat{\beta}, u)} - \frac{1}{n\widehat{S}_C^{(0)}(\beta_0, u)} \right) = -n^{-1} \frac{\widehat{S}_C^{(1)}(\beta^*, u)^T}{\widehat{S}_C^{(0)}(\beta^*, u)^2} (\widehat{\beta} - \beta_0), \quad (23)$$

where β^* lies on the line segment between β_0 and $\widehat{\beta}$. Thus the first term in (22) becomes

$$- \int_0^t \frac{\widehat{S}_C^{(1)}(\beta^*, u)^T}{\widehat{S}_C^{(0)}(\beta^*, u)^2} (\widehat{\beta} - \beta_0) \left(n^{-\frac{1}{2}} \sum_{i=1}^n \tilde{\omega}_i(u) dM_i^1(u) \right). \quad (24)$$

Since $\frac{\widehat{S}_C^{(1)}(\beta^*, u)^T}{\widehat{S}_C^{(0)}(\beta^*, u)^2}$ can be written as a sum of two monotone functions in $t \in [0, \tau]$, together with the weak convergence of $n^{-\frac{1}{2}} \sum_{i=1}^n \tilde{\omega}_i(u) dM_i^1(u)$ (Appendix 1 of Sun *et al.*, 2004), (24) converges in probability to 0 uniformly in $t \in [0, \tau]$. By similar expansion, the second term on the right-hand side of (22) is

$$-n^{\frac{1}{2}} \int_0^t \frac{\widehat{S}_C^{(1)}(\beta^*, u)^T}{\widehat{S}_C^{(0)}(\beta^*, u)^2} (\widehat{\beta} - \beta_0) S_F^{(0)}(\beta_0, u) d\Lambda_{1.0}(u). \quad (25)$$

By the consistency of $\widehat{\beta}$, the continuity of $\widehat{S}_C^{(1)}(\beta, t)$ and $\widehat{S}_C^{(0)}(\beta, t)$ and their uniform convergence, respectively, to $s^{(1)}(\beta, t)$ and $s^{(0)}(\beta, t)$, the uniform convergence of $S_F^{(0)}(\beta_0, t)$ to $s^{(0)}(\beta_0, t)$, as well as the boundedness of $\Lambda_{1.0}(t)$ on $[0, \tau]$, (25) can be written as

$$-n^{\frac{1}{2}} (\widehat{\beta} - \beta_0)^T \int_0^t e(\beta_0, u) d\Lambda_{1.0}(u) + o_p(1).$$

Using the first order Taylor expansion of $U_C(\widehat{\beta})$ around β_0 and the convergence of $\{-n^{-1} \partial U_C(\beta) / \partial \beta^T\}$ to $\Omega(\beta)$, we have $(\widehat{\beta} - \beta_0) = n^{-1} \Omega(\beta_0)^{-1} U_C(\beta_0) + o_p(n^{-1/2})$ and hence, using the results in (20) and letting $\mathbf{h}(t) = -\int_0^t e(\beta, u) d\Lambda_{1.0}(u)$, the first term on the right-hand side of (21) becomes $n^{-\frac{1}{2}} \Omega(\beta_0)^{-1} \sum_{i=1}^n [\eta_i(\beta_0) + \psi_i(\beta_0) + (1 - (\xi_i/\tilde{\alpha})) \mu_i(\beta_0)] \mathbf{h}(t) + o_p(1)$.

To explore the second term on the right-hand side of (21), applying the first order Taylor expansion of $\left(\frac{1}{\widehat{S}_C^{(0)}(\beta_0, u)} - \frac{1}{\widetilde{S}_C^{(0)}(\beta_0, u)} \right)$ with respect to $\omega_j(u)$ around $\tilde{\omega}_j(u)$ gives

$$\left(\frac{1}{\widehat{S}_C^{(0)}(\beta_0, u)} - \frac{1}{\widetilde{S}_C^{(0)}(\beta_0, u)} \right) = - \frac{\sum_{j=1}^n \rho_j(u) Y_j(u) \exp\{\beta_0^T \mathbf{Z}_j(u)\} (\omega_j(u) - \tilde{\omega}_j(u))}{n(\widehat{S}_C^{(0)}(\beta_0, u))^2}. \quad (26)$$

Since $\widehat{S}_C^{(0)}(\beta_0, t)^{-1}$ can be shown to converge uniformly to $s^{(0)}(\beta_0, t)^{-1}$ and is of bounded variation, it can be written as the sum of two monotone functions in t . Further,

$$\{\omega_j(t) - \tilde{\omega}_j(t)\} = -I(X_j < t) \tilde{\omega}_j(t) \int_{X_j}^t \frac{d\bar{M}^c(u)}{n\pi(u)} + o_p(1)$$

and hence the second term on the right-hand side of (21) can be shown to be

$$n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \left(\frac{1}{\widehat{S}_C^{(0)}(\beta_0, u)} - \frac{1}{\widetilde{S}_C^{(0)}(\beta_0, u)} \right) \tilde{\omega}_i(u) dN_i(u) = (\mathbf{PP1}) + (\mathbf{PP2}) + o_p(1),$$

where

$$(\mathbf{PP1}) = n^{-\frac{1}{2}} \int_0^t \left[\int_0^t \frac{\sum_{j=1}^n \rho_j(u) Y_j(u) \exp\{\beta_0^T \mathbf{Z}_j(u)\} \tilde{\omega}_j(u) I(X_j < s \leq u)}{n(s^{(0)}(\beta_0, u))^2} \times \left(n^{-1} \sum_{k=1}^n \tilde{\omega}_k(u) dM_k^1(u) \right) \right] \frac{d\bar{M}^c(s)}{\pi(s)},$$

and

$$(PP2) = n^{-\frac{1}{2}} \int_0^t \left[\int_0^u \frac{\sum_{j=1}^n \rho_j(u) Y_j(u) \exp\{\beta_0^T \mathbf{Z}_j(u)\} \tilde{\omega}_j(u) I(X_j < s \leq u)}{ns^{(0)}(\beta_0, u)} \right. \\ \left. \times \frac{\sum_{k=1}^n \tilde{\omega}_k(u) \{Y_k(u) \exp\{\beta_0^T \mathbf{Z}_k(u)\} d\Lambda_{1.0}(u)\}}{ns^{(0)}(\beta_0, u)} \right] \frac{d\bar{M}^c(s)}{\pi(s)}.$$

Note that, by extending Lemma 1 of Lin *et al.* (2000) and Appendix 2 of Sun *et al.*, (2004), $(PP1)$ converges in probability to 0 uniformly in $t \in [0, \tau]$ and $\beta \in \mathcal{B}$. On the other hand, under the regularity conditions, $\tilde{S}_F^{(0)}(\beta, t)$ uniformly converges to $s^{(0)}(\beta, t)$ and hence

$$(PP2) = n^{-\frac{1}{2}} \int_0^t \left[\int_0^u \frac{\sum_{j=1}^n \rho_j(u) Y_j(u) \exp\{\beta_0^T \mathbf{Z}_j(u)\} \tilde{\omega}_j(u) I(X_j < s \leq u)}{ns^{(0)}(\beta_0, u)} d\Lambda_{1.0}(u) \right] \frac{d\bar{M}^c(s)}{\pi(s)}.$$

Here also, note that $(\rho_j(t) - \tilde{\rho}_j)$ converges in probability to 0 uniformly in $t \in [0, \tau]$ based on Lemma 7.2 and let

$$\mathbf{p}^*(\beta_0, s, t) = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n \int_0^t \frac{I(X_j < s \leq u)}{s^{(0)}(\beta_0, u)} \tilde{\rho}_j Y_j(u) \exp\{\beta_0^T \mathbf{Z}_j(u)\} \tilde{\omega}_j(u) d\Lambda_{1.0}(u).$$

Thus, the second term on the right-hand side of (21) is

$$n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \left(\frac{1}{\tilde{S}_C^{(0)}(\beta_0, u)} - \frac{1}{\tilde{S}_C^{(0)}(\beta_0, u)} \right) \tilde{\omega}_i(u) dN_i(u) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{\mathbf{p}^*(\beta_0, s, t)}{\pi(s)} dM_i^c(s) + o_p(1).$$

For the third term on the right-hand side of (21), by the same arguments that followed (26), we can replace $\tilde{S}_C^{(0)}(\beta_0, t)^{-1}$ by its uniform limit $s^{(0)}(\beta_0, t)^{-1}$. Thus, it can be shown that the third term on the right-hand side of (21) is asymptotically equivalent to $\int_0^t \frac{1}{s^{(0)}(\beta_0, u)} \left(n^{-\frac{1}{2}} \sum_{i=1}^n \tilde{\omega}_i(u) dM_i^1(u) \right)$.

For the fourth term on the right-hand side of (21), by Lemma ??, we have

$$n^{\frac{1}{2}} \{ \tilde{S}_F^{(0)}(\beta_0, u) - \tilde{S}_C^{(0)}(\beta_0, u) \} = n^{-\frac{1}{2}} \sum_{i=1}^n (1 - (\xi_i/\tilde{\alpha})) r_i^{(0)}(\beta_0, u) + o_p(1)$$

and hence, with the uniform convergence of $\tilde{S}_C^{(0)}(\beta_0, t)^{-1}$ to $s^{(0)}(\beta_0, t)^{-1}$, the fourth term on the right-hand side of (21) becomes

$$n^{\frac{1}{2}} \int_0^t \left\{ \frac{\tilde{S}_F^{(0)}(\beta_0, u) - \tilde{S}_C^{(0)}(\beta_0, u)}{\tilde{S}_C^{(0)}(\beta_0, u)} \right\} d\Lambda_{1.0}(u) = n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}} \right) \int_0^t r_i^{(0)}(\beta_0, u) \frac{d\Lambda_{1.0}(u)}{s^{(0)}(\beta_0, u)} + o_p(1).$$

Combining these results, we have

$$n^{\frac{1}{2}} \{ \hat{\Lambda}_{1.0}(\hat{\beta}, t) - \Lambda_{1.0}(t) \} = n^{-\frac{1}{2}} \mathbf{\Omega}(\beta_0)^{-1} \sum_{i=1}^n [\boldsymbol{\eta}_i(\beta_0) + \boldsymbol{\psi}_i(\beta_0)] \mathbf{h}(t) \\ + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{\mathbf{p}^*(\beta_0, s, t)}{\pi(s)} dM_i^c(s) + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{\tilde{\omega}_i(u) dM_i^1(u)}{s^{(0)}(\beta_0, u)} \\ + n^{-\frac{1}{2}} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}} \right) \left[\mathbf{\Omega}(\beta_0)^{-1} \boldsymbol{\mu}_i(\beta_0) \mathbf{h}(t) + \int_0^t r_i^{(0)}(\beta_0, u) \frac{d\Lambda_{1.0}(u)}{s^{(0)}(\beta_0, u)} \right] + o_p(1)$$

Thus, $n^{\frac{1}{2}} \{ \hat{\Lambda}_{1.0}(\hat{\beta}, t) - \Lambda_{1.0}(t) \} = n^{-\frac{1}{2}} \sum_{i=1}^n \{ \mathcal{W}_{1i}(t) + (1 - (\xi_i/\tilde{\alpha})) \mathcal{W}_{2i}(t) \} + o_p(1)$ where

$$\mathcal{W}_{1i}(t) = \mathbf{\Omega}(\beta_0)^{-1} [\boldsymbol{\eta}_i(\beta_0) + \boldsymbol{\psi}_i(\beta_0)] \mathbf{h}(t) + \int_0^t \frac{\mathbf{p}^*(\beta_0, s, t)}{\pi(s)} dM_i^c(s) + \int_0^t \frac{\tilde{\omega}_i(u) dM_i^1(u)}{s^{(0)}(\beta_0, u)}$$

and

$$\mathcal{W}_{2i}(t) = \left[\boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\mu}_i(\boldsymbol{\beta}_0) \mathbf{h}(t) + \int_0^t r_i^{(0)}(\boldsymbol{\beta}_0, u) \frac{d\Lambda_{1.0}(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} \right],$$

for $i = 1, \dots, n$. Further, by Lemma 7.1, $n^{\frac{1}{2}} \{ \widehat{\Lambda}_{1.0}(\widehat{\boldsymbol{\beta}}, t) - \Lambda_{1.0}(t) \}$ converges to a mean-zero Gaussian process with covariance matrix

$$\boldsymbol{\Sigma}_{\Lambda_{1.0}}(t) = E\{\mathcal{W}_{11}(t)^{\otimes 2}\} + (1 - \alpha)\alpha^{-1}E\{\mathcal{W}_{21}(t)^{\otimes 2}\}.$$

This completes the proof.

References

- [1] Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73, 1-11.
- [2] Sandler, D. P., Hodgson, M. E., Deming-Halverson, S. L., Juras, P. S., D'Aloisio, A. A., Suarez, L. M., et al. (2017). The Sister Study cohort: Baseline methods and participant characteristics. *Environmental Health Perspectives*, 125, 127003.
- [3] Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V., and Sun, P. (2015). Breast cancer mortality after a diagnosis of ductal carcinoma in situ. *JAMA Oncology*, 1, 888-896.
- [4] Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V. Jr., Flournoy, N., Farewell, V., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34, 541-554.
- [5] 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.
- [6] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, Second Edition. John Wiley & Sons, New York.
- [7] Kim, H. T. (2007). Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical Cancer Research*, 13, 559-565.
- [8] Koller, M. T., Raatz, H., Steyerberg, E. W., and Wolbers, M. (2012). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine*, 31, 1089-1097.
- [9] Austin, P. C., Lee, D. S., and Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, 133, 601-609.
- [10] Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*, 16, 64-81.
- [11] Barlow, W. E., Ichikawa, L., Rosner, D., and Izumi, S. (1999). Analysis of case-cohort designs. *Journal of Clinical Epidemiology*, 52, 1165-1172.
- [12] Chen, K. and Lo, S. H. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika*, 86, 755-764.
- [13] Borgan, O., Langholz, B., Samuelsen, S. O., Goldstein, L., and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis*, 6, 39-58.
- [14] Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika*, 87, 73-87.
- [15] Kang, S. and Cai, J. (2009). Marginal hazards model for case-cohort studies with multiple disease outcomes. *Biometrika*, 96, 887-901.
- [16] Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics*, 65, 135-142.
- [17] Sørensen, P. and Andersen, P. K. (2000). Competing risks analysis of the case-cohort design. *Biometrika*, 87, 49-59.
- [18] Sun, J., Sun, L., and Flournoy, N. (2004). Additive hazards models for competing risks analysis of the case-cohort design. *Communications in Statistics*, 33, 351-366.
- [19] Wolkewitz, M., Palomar-Martinez, M., Olaechea-Astigarraga, P., Alvarez-Lerma, F., and Schumacher, M. (2016). A full competing risk analysis of hospital-acquired infections can easily be performed by a case-cohort approach. *Journal of Clinical Epidemiology*, 74, 187-193.
- [20] Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16, 1141-1154.
- [21] Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology: Methodological Issues*, eds. N. Jewell, K. Dietz, and V. Farewell. Birkhäuser, Boston, 24-33.
- [22] Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processed: A large sample study. *The Annals of Statistics*, 10, 1100-1120.
- [23] Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72, 147-148.

- [24] Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association*, 99, 832-844.
- [25] D'Aloisio, A. A., Nichols, H. B., Hodgson, M. E., Deming-Halverson, S. L. and Sandler, D. P. (2017). Validity of self-reported breast cancer characteristics in a nationwide cohort of women with a family history of breast cancer. *BMC Cancer*, 17, 692.
- [26] Xu, Z., Bolick, S. C., DeRoo, L. A., Weinberg, C. R., Sandler, D. P., and Taylor, J. A. (2013). Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *Journal of the National Cancer Institute*, 105, 694-700.
- [27] Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L. et al. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, 11, 587.
- [28] White, A. J., Nichols, H. B., Bradshaw, P. T., and Sandler, D. P. (2015). Overall and central adiposity and breast cancer risk in the sister study. *Cancer*, 121, 3700-3708.
- [29] 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.
- [30] Chen, K. (2001). Generalized case-cohort sampling. *Journal of the Royal Statistical Society, Series B*, 63, 791-809.