Competing risks regression for clustered data with covariate-dependent censoring

Manoj Khanal[a], Soyoung Kim[a,b] and Kwang Woo Ahn[a,b]

[a]Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, 53226, Wisconsin,USA
[b]CIBMTR, Medical College of Wisconsin Clinical Cancer Center, 9200 W. Wisconsin Avenue, Milwaukee, 53226, Wisconsin,USA

ARTICLE HISTORY
Compiled March 19, 2023

ABSTRACT
Competing risks data in clinical trial or observational studies often suffer from cluster effects such as center effects and matched pairs design. The proportional subdistribution hazards (PSH) model is one of the most widely used methods for competing risks data analyses. However, the current literature on the PSH model for clustered competing risks data is limited to covariate-independent censoring and the unstratified model. In practice, competing risks data often face covariate-dependent censoring and have the non-PSH structure. Thus, we propose a marginal stratified PSH model with covariate-adjusted censoring weight for clustered competing risks data. We use a marginal stratified proportional hazards model to estimate the survival probability of censoring by taking clusters and non-proportional hazards structure into account. Our simulation results show that, in the presence of covariate-dependent censoring, the parameter estimates of the proposed method are unbiased with approximate 95% coverage rates. We apply the proposed method to stem cell transplant data of leukemia patients to evaluate the clinical implications of donor-recipient HLA matching on chronic graft-versus-host disease.

KEYWORDS
Proportional subdistribution hazards model, Covariate dependent censoring, Stratified model, Competing risks regression

1. Introduction

A competing risk is an event which precludes the primary event of interest. Competing risks data often suffer from cluster effects such as study center effects and matched pair designs. For example, Pidala et al. (2014) studied patients who received unrelated donor allogeneic hematopoietic cell transplantation to treat blood cancer. Their primary interest was investigating the clinical impact of donor-recipeint HLA matching on right-censored outcomes. One of the main outcomes was chronic graft-versus-host disease (GVHD), where death without experiencing chronic GVHD was a competing risk. As in Section 5, it had a significant study center effect. They fitted the proportional cause-specific hazards model for chronic GVHD. However, the cause-specific hazards model does not evaluate direct covariate effects on cumulative incidence, which

CONTACT Kwang Woo Ahn. Email: kwooahn@mcw.edu

is often of interest for a competing risks data analysis.

There is rich literature which evaluates direct covariate effects on cumulative incidence for clustered competing risks data. Scheike et al. (2010) studied a semiparametric random effects model based on the direct binomial modeling, where the marginal cumulative incidence functions follow a generalized semiparametric additive model. Logan et al. (2011) proposed a pseudovalue approach based on the jackknife estimator and a generalized estimating equation. Ruan and Gray (2008) proposed a non-parametric multiple imputation method. Zhou et al. (2012) extended the proportional subdistribution hazards (PSH) model of Fine and Gray (1999) to clustered competing risks data. However, the asymptotic results of all of these methods are limited to covariate-independent censoring. And Zhou et al. (2012) is limited to the PSH structure. In practice, the censoring distribution may depend on some covariates and the PSH assumption may not hold. Addressing these limitations under the stratified PSH model is crucial in practice because many clinicians and investigators widely use the PSH model of Fine and Gray (1999) for competing risks data analysis.

Therefore, we study a stratified PSH model with covariate-adjusted censoring weight for clustered data in this article. The survival probability of censoring is estimated using the marginal stratified proportional hazards model. In practice, stratification variables for the stratified PSH model and the stratified proportional hazards model for censoring may not be the same. Thus, we allow stratification variables for both models to be different. Simulation studies of the existing literature on the PSH model (Fine and Gray, 1999; Zhou et al., 2012) focused on parameter estimation for covariates and did not perform simulation studies on baseline cumulative subdistribution hazards. Our simulations consider the estimation of both parameters for covariates and baseline cumulative subdistribution hazard functions. The proposed competing risks regression can be implemented using R package '**adjSURVCI**' (Khanal and Ahn, 2021). We apply our method to Pidala et al. (2014) which studied leukemia patients who received stem cell transplant to evaluate the clinical impact of donor-recipient HLA matching on chronic GVHD.

2. Notations and a marginal stratified PSH model

We define notations first. Assume the total number of clusters is n, n_j is the number of clusters in stratum j, S is the fixed total number of strata, n_{ij} is the number of observations within cluster i and stratum j and $n = \sum_{j=1}^{S} n_j$, where n_{ij} is bounded. In what follows, the subscripts i, j, k correspond to the k^{th} subject in the i^{th} cluster and j^{th} stratum for $i = 1, \ldots, n$, $j = 1, \ldots, S$, and $k = 1, \ldots, n_{ij}$. Let T_{ijk} and C_{ijk} be the failure time and the censoring time, respectively. Denote $\epsilon_{ijk} \in \{1, \ldots, l\}$ as a cause of failure. We assume $\epsilon_{ijk} = 1$ to be the main cause of interest. Let $\boldsymbol{Z}_{ijk} = (Z_{1ijk}, \ldots, Z_{pijk})^\intercal$ be a $p \times 1$ vector of covariates. Let $\Delta_{ijk} = I(T_{ijk} \leq C_{ijk})$ be the event indicator. We observe $\{X_{ijk} = \min(T_{ijk}, C_{ijk}), \Delta_{ijk}, \Delta_{ijk}\epsilon_{ijk}, \boldsymbol{Z}_{ijk}; i = 1, \ldots, n, j = 1, \ldots, S, k = 1, \ldots, n_{ij}\}$, where X_{ijk} is the observed time. Define $\psi_{ijk} = \Delta_{ijk}\epsilon_{ijk} \in \{0, 1, \ldots, l\}$. Let $\boldsymbol{T}_i = \{T_{ijk}; j = 1, \ldots, n_S, k = 1, \ldots, n_{ij}\}$, with $\boldsymbol{C}_i$, $\boldsymbol{Z}_i$, $\boldsymbol{\Delta}_i$ and $\boldsymbol{\epsilon}_i$ defined similarly. We assume $(\boldsymbol{T}_i, \boldsymbol{\epsilon}_i)$ to be independent of $\boldsymbol{C}_i$ given $\boldsymbol{Z}_i$. Assume that $(\boldsymbol{T}_i, \boldsymbol{C}_i, \boldsymbol{Z}_i, \boldsymbol{\Delta}_i, \boldsymbol{\epsilon}_i)$ are independent and identically distributed across clusters $i = 1, \ldots, n$. The study period is $[0, \tau]$. Let $N^1_{ijk}(t) = I(T_{ijk} \leq t; \psi_{ijk} = 1)$ be the counting process associated with Cause 1 and $Y^1_{ijk}(t) = 1 - N^1_{ijk}(t^-)$ be the risk process. Similarly, let $N^C_{ijk}(t) = I(X_{ijk} \leq t; \psi_{ijk} = 0)$ and $Y^C_{ijk}(t) = I(X_{ijk} \geq t)$ be the counting process and the at-risk indicator associated with censoring, respectively. Let $\Phi_{ij} = 1$ if the j^{th} stratum has at least one subject from cluster i; otherwise $\Phi_{ij} = 0$.

We assume the censoring distribution follows the stratified proportional hazards model (Cox, 1972):

$$\lambda_{Cj}(t, \boldsymbol{Z}_{ijk}) = \lambda_{C0j}(t) \exp(\boldsymbol{\gamma}_0^{\intercal} \boldsymbol{Z}_{ijk}),\ j = 1, \ldots, S, \tag{1}$$

where $\lambda_{C0j}(t)$ is an unspecified baseline hazard function and $\boldsymbol{\gamma}_0$ is the true parameter vector for censoring. Define

$$S_{Cj}^{(r)}(\boldsymbol{\gamma}, t) = \frac{1}{n_j} \sum_{i=1}^{n} \Phi_{ij} \sum_{k=1}^{n_{ij}} Y_{ijk}^{C}(t) \boldsymbol{Z}_{ijk}^{\bigotimes r} \exp(\boldsymbol{\gamma}^{\intercal} \boldsymbol{Z}_{ijk}),\ j = 1, \ldots, S,$$

for $r = 0, 1, 2$ such that $\boldsymbol{a}^{\bigotimes r} = 1$, $\boldsymbol{a}^{\bigotimes 1} = \boldsymbol{a}$ and $\boldsymbol{a}^{\bigotimes 2} = \boldsymbol{a}\boldsymbol{a}^{\intercal}$ for $\boldsymbol{a}$. The score function for $\boldsymbol{\gamma}$ is

$$\boldsymbol{U}_C(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \sum_{j=1}^{S} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^{\tau} \left\{ \boldsymbol{Z}_{ijk} - \frac{S_{Cj}^{(1)}(\boldsymbol{\gamma}, u)}{S_{Cj}^{(0)}(\boldsymbol{\gamma}, u)} \right\} dN_{ijk}^{C}(u) = \mathbf{0}. \tag{2}$$

Let $\widehat{\boldsymbol{\gamma}}$ be the estimator for $\boldsymbol{\gamma}_0$ from (2). The Breslow-type estimator of $\Lambda_{C0j}(t) = \int_0^t \lambda_{C0j}(t)$ is

$$\widehat{\Lambda}_{C0j}(t) = \frac{1}{n_j} \sum_{i=1}^{n} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^t \frac{dN_{ijk}^{C}(s)}{S_{Cj}^{(0)}(\widehat{\boldsymbol{\gamma}}, s)}.$$

Thus, the estimated survival probability of the censoring distribution is $\widehat{G}_{Cj}(t \mid \boldsymbol{Z}_{ijk}) = \exp\{-\widehat{\Lambda}_{C0j}(t) \exp(\widehat{\boldsymbol{\gamma}}^{\intercal} \boldsymbol{Z}_{ijk})\}$.

The covariate-adjusted censoring weight function is $\widehat{w}_{ijk}^{COX}(t) = I(C_{ijk} \geq X_{ijk} \wedge t)\widehat{G}_{C,j}(t; \boldsymbol{Z}_{ijk})/\widehat{G}_{C,j}(X_{ijk} \wedge t; \boldsymbol{Z}_{ijk})$. For $j = 1, \ldots, S$, define

$$S_{COX,j}^{(r)}(\boldsymbol{\beta}, t) = \frac{1}{n_j} \sum_{i=1}^{n} \Phi_{ij} \sum_{k=1}^{n_{ij}} \widehat{w}_{ijk}^{COX}(t) Y_{ijk}^{1}(t) \boldsymbol{Z}_{ijk}^{\bigotimes r} \exp(\boldsymbol{\beta}^{\intercal} \boldsymbol{Z}_{ijk}).$$

Our main interest is to evaluate the direct effects of covariates on the cumulative incidence function of Cause 1, $F_{1j}(t \mid \boldsymbol{Z}_{ijk}) = P(T_{ijk} \leq t, \epsilon_{ijk} = 1 \mid \boldsymbol{Z}_{ijk})$. The stratified PSH model for Cause 1 is

$$\lambda_{1j}(t \mid \boldsymbol{Z}_{ijk}) = \lambda_{10j}(t) \exp(\boldsymbol{\beta}_0^{\intercal} \boldsymbol{Z}_{ijk}),\ j = 1, \ldots, S, \tag{3}$$

where $\lambda_{10j}(t)$ is an unspecified baseline subdistribution hazard function and $\boldsymbol{\beta}_0$ is the true parameter vector. We estimate $\boldsymbol{\beta}_0$ in (3) by setting the following score equation equal to zero:

$$
\begin{aligned}
&\boldsymbol{U}_{COX}(\boldsymbol{\beta}) \\
&= \sum_{i=1}^{n} \left[\sum_{j=1}^{S} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^{\tau} \left\{ \boldsymbol{Z}_{ijk} - \frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}, u)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}, u)} \right\} \widehat{w}^{COX}_{ijk}(u) dN^1_{ijk}(u) \right]. \qquad (4)
\end{aligned}
$$

Denote the estimator of $\boldsymbol{\beta}_0$ as $\widehat{\boldsymbol{\beta}}$. The Breslow-type estimator of the cumulative baseline subdistribution hazard function of Cause 1, $\Lambda_{10j}(t) = \int_0^t \lambda_{10j}(s)$, for stratum j is

$$
\widehat{\Lambda}_{10j}(t) = \frac{1}{n_j} \int_0^t \frac{1}{S^{(0)}_{COX,j}(\widehat{\boldsymbol{\beta}}, s)} \sum_{i=1}^{n} \Phi_{ij} \sum_{k=1}^{n_{ij}} \widehat{w}^{COX}_{ijk}(s) dN^1_{ijk}(s).
$$

3. Asymptotic Properties

We establish the asymptotic properties of the proposed estimators in this section. Define $M^1_{ijk}(t) = N^1_{ijk}(t) - \int_0^t Y^1_{ijk}(u) e^{\boldsymbol{\beta}_0^{\intercal} \boldsymbol{Z}_{ijk}} d\Lambda_{10j}(u)$ and $M^C_{ijk}(t) = N^C_{ijk}(t) - \int_0^t Y_{ijk}(u) e^{\boldsymbol{\gamma}^{\intercal} \boldsymbol{Z}_{ijk}} d\Lambda_{C0j}(u)$ for Cause 1 and censoring processes, respectively. We assume the following conditions:

(A1) $P(T_{ijk} > \tau) > 0$ and $P(C_{ijk} > \tau) > 0$ for all i, j and k.
(A2) The covariates $\boldsymbol{Z}$ are bounded almost surely.
(A3) $\int_0^{\tau} \lambda_{C0j}(t) dt < \infty$ for $j = 1, \ldots, S$.
(A4) $-1/n \times \partial \boldsymbol{U}_C(\boldsymbol{\gamma}_0)/\partial \boldsymbol{\gamma}_0^{\intercal}$ converges to $\boldsymbol{\Omega}_C(\gamma_0)$ which is positive definite.
(A5) $\int_0^{\tau} \lambda_{10j}(t) dt < \infty$ for $j = 1, \ldots, S$.
(A6) $-1/n \times \partial \boldsymbol{U}_{COX}(\boldsymbol{\beta}_0)/\partial \beta_0^{\intercal}$ converges to $\boldsymbol{\Omega}(\beta_0)$ which is positive definite.
(A7) There exists a neighborhood $\mathcal{B}_C$ of $\boldsymbol{\gamma}_0$ and $s^{(r)}_{Cj}$ defined on $\mathcal{B}_C \times [0, \tau]$ such that for $r = 0, 1, 2$; $\sup_{t \in [0,\tau], \boldsymbol{\gamma} \in \mathcal{B}_C} \| S^{(r)}_{Cj}(\boldsymbol{\gamma}, t) - s^{(r)}_{Cj}(\boldsymbol{\gamma}, t) \| \xrightarrow{P} 0$ and $s^{(r)}_{Cj}(\boldsymbol{\gamma}, t) = E\left[S^{(r)}_{Cj}(\boldsymbol{\gamma}, t) \right]$.
(A8) $S^{(r)}_{Cj}(\boldsymbol{\gamma}, t), r = 0, 1, 2$ are continuous functions of $\boldsymbol{\gamma} \in \mathcal{B}_C$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B}_C \times [0, \tau]$. $s^{(0)}_{Cj}(\boldsymbol{\gamma}, t)$ is bounded away from zero on $\mathcal{B}_C \times [0, \tau]$.
(A9) There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ and $s^{(r)}_{COX,j}$ defined on $\mathcal{B} \times [0, \tau]$ such that for $r = 0, 1, 2$; $\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}} \| S^{(r)}_{COX,j}(\boldsymbol{\beta}, t) - s^{(r)}_{COX,j}(\boldsymbol{\beta}, t) \| \xrightarrow{P} 0$ and $s^{(r)}_{COX,j}(\boldsymbol{\beta}, t) = E\left[S^{(r)}_{COX,j}(\boldsymbol{\beta}, t) \right]$.
(A10) $S^{(r)}_{COX,j}(\boldsymbol{\beta}, t)$ are continuous functions of $\boldsymbol{\beta} \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$. $s^{(0)}_{COX,j}(\boldsymbol{\beta}, t)$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$.
(A11) $p_j = n_j/n \to \pi_j$ as n and n_j go to infinity.

These conditions are the standard regularity conditions for the stratified PSH model for Cause 1 and the stratified proportional hazards model for censoring similarly to Kim et al. (2020).

We have the following theorem for $\widehat{\boldsymbol{\beta}}$.

Theorem 3.1. *Under Assumptions (A1)-(A11),* $\widehat{\boldsymbol{\beta}}$ *converges in probability to* $\boldsymbol{\beta}_0$ *and*

$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *converges in distribution to a zero-mean normal distribution with covariance matrix* $\boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1}$, *where*

$$\boldsymbol{\Sigma} = E\left\{\left(\boldsymbol{\eta}_{1..}^{COX} + \boldsymbol{\psi}_{1..}^{COX}\right)^{\bigotimes 2}\right\},$$

$$\boldsymbol{\eta}_{i..}^{COX} = \sum_{j=1}^{S} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^{\tau} \left\{ \boldsymbol{Z}_{ijk} - \frac{s_{COX,j}^{(1)}(\boldsymbol{\beta}_0, u)}{s_{COX,j}^{(0)}(\boldsymbol{\beta}_0, u)} \right\} w_{ijk}^{COX}(u) dM_{ijk}^{1}(u),$$

$$\boldsymbol{\psi}_{i..}^{COX} = \sum_{j=1}^{S} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^{\tau} \boldsymbol{q}_{ijk}^{(1)}(u) dM_{ijk}^{C}(u).$$

The expression for $\boldsymbol{q}_{ijk}^{(1)}(u)$ and its proof are provided in the Appendix is provided in the Supplementary Materials. We consider the same strata for events from Cause 1 and censoring for a mathematical simplicity in Theorem 3.1. This avoids the abuse of complicated notations and subscripts. However, similarly to Kim et al. (2020), one can show the asymptotic results of Theorem 3.1 even when strata for models for Cause 1 and censoring are different using similar arguments to the proof of Theorem 3.1. In summary, when strata for the models for Cause 1 and censoring are different, one can fit the stratified proportional hazards model using strata for censoring, estimate all censoring-related terms for each subject, and plug them into the asymptotic formula of Theorem 3.1 as described in Kim et al. (2020). We conduct a simulation study with different strata between the model for Cause 1 and the model for censoring in Table 6 in Section 4. The results show little bias and the empirical coverage rates close to 95%. We assume the same strata for the following theorems for brevity. However, the same argument for different strata between Cause 1 and censoring models can be applied to them.

Theorem 3.2. *Under Assumptions (A1)-(A11),* $\widehat{\Lambda}_{10j}(t)$ *is a consistent estimator for* $\Lambda_{10j}(t)$ *for* $t \in [0, \tau]$ *and* $\sqrt{n_j}\{\widehat{\Lambda}_{10j}(t) - \Lambda_{10j}(t)\}$ *converges weakly to a zero mean Gaussian process with the variance* $\Sigma_{\Lambda_{10j}}(t) = E\{W_{\Lambda,ij}(t)\}$, *where*

$$W_{\Lambda,ij}(t) = \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^t \frac{w_{ijk}^{COX}(s) dM_{ijk}^{1}(s)}{s_{COX,j}^{(0)}(\boldsymbol{\beta}_0, s)} + \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^t q_{ijk}^{(2)}(s,t) dM_{ijk}^{C}(s)$$
$$- \pi_j (\boldsymbol{\eta}_{i..}^{COX} + \boldsymbol{\psi}_{i..}^{COX})^{\intercal} \boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1} \int_0^t \left(\frac{s_{COX,j}^{(1)}(\boldsymbol{\beta}_0, s)}{s_{COX,j}^{(0)}(\boldsymbol{\beta}_0, s)} \right) d\Lambda_{10j}^{COX}(s).$$

The expression for $q_{ijk}^{(2)}(s,t)$ and the proof of Theorem 3.2 are provided in the Appendix and Supplementary Materials, respectively.

For a given value of covariate vector $\boldsymbol{Z}$, the predicted CIF of Cause 1 for the j^{th} stratum can be estimated by $\widehat{F}_{1j}(t \mid \boldsymbol{Z}) = 1 - \exp\left\{-\widehat{\Lambda}_{10j}(t)\exp(\widehat{\boldsymbol{\beta}}^{\intercal}\boldsymbol{Z})\right\}$.

Theorem 3.3. $\sqrt{n_j}\{\widehat{F}_{1j}(t \mid \boldsymbol{Z}) - F_{1j}(t \mid \boldsymbol{Z})\}$ *converges weakly to a zero mean Gaussian process and asymptotic variance which can be consistently estimated by* $\widehat{\Sigma}_{F_{1j}}(t)$.

The expression for $\widehat{\Sigma}_{F_{1j}}(t)$ and its proof are provided in the Appendix and Supplementary Materials, respectively.

The asymptotic variances of the proposed estimators for $\boldsymbol{\beta}$ and $\Lambda_{10j}(t)$ when using Kaplan-Meier estimates for censoring are provided in the Supplementary Materials.

4. Simulation

We conduct simulation studies using R package **adjSURVCI** (Khanal and Ahn, 2021). We consider two weight functions: covariate-dependent weight $\widehat{w}_{ijk}^{COX}(t)$ and covariate-independent weight $\widehat{w}_{ijk}^{KM}(t) = I(C_{ijk} \geq X_{ijk} \wedge t)\widehat{G}_{C,j}(t)/\widehat{G}_{C,j}(X_{ijk} \wedge t)$, where $\widehat{G}_{C,j}(\cdot)$ is a Kaplan-Meier estimator for the survival probability of censoring for stratum j. We consider two competing risks with Cause 1 being the main event of interest and 2 strata. Similarly to Logan et al. (2011), we generate clustered competing risks data for Causes 1, 2, and censoring for stratum j using

$$F_{1j}(t \mid \omega, \mathbf{Z}) = 1 - [1 - p\,(1 - \exp\{-\rho_j t\})]^{\omega e^{\boldsymbol{\beta}_*^\intercal \mathbf{z}}},$$

$$F_{2j}(t \mid \omega, \mathbf{Z}) = (1-p)^{\omega e^{\boldsymbol{\beta}_*^\intercal \mathbf{z}}} \left[1 - \exp\left\{-\rho_j t e^{\boldsymbol{\kappa}^\intercal \mathbf{Z}}\right\}\right],$$

$$G_j(t \mid \omega_c, \mathbf{Z}) = \exp\left\{-t\rho_{c_j}\omega_c e^{\boldsymbol{\gamma}_*^\intercal \boldsymbol{Z}}\right\},$$

where random effects ω and ω_c are generated from a positive stable frailty distribution with parameter α. If $\alpha = 1$, all subjects are independent while $\alpha = 0.5, 0.25$ result in clustered event times with a smaller value of α indicating a stronger correlation within a cluster. By Logan et al. (2011), the true covariate effects associated with Cause 1 and censoring are $\boldsymbol{\beta}_0 = \alpha\boldsymbol{\beta}_*$ and $\boldsymbol{\gamma}_0 = \alpha\boldsymbol{\gamma}_*$, respectively. We set $\boldsymbol{\beta}_0 = (0.5, -0.5, 0.5)^\intercal, \rho_1 = 1, \rho_2 = 2$, and $\boldsymbol{\kappa} = (2.5, 2.5, 2.5)^\intercal$. Thus, $\boldsymbol{\beta}_*$'s are $(2, -2, 2)^\intercal, (1, -1, 1)^\intercal$, and $(0.5, -0.5, 0.5)^\intercal$ for $\alpha = 0.25, 0.5$, and 1, respectively. For covariate-dependent censoring (CDC) and covariate-independent censoring (CIC), $\boldsymbol{\gamma}_0 = (2.5, 2.5, -3)^\intercal$ and $\boldsymbol{\gamma}_0 = (0, 0, 0)^\intercal$, respectively. Table 1 shows the other parameter values for the simulation study. With these values, it results in approximately 30% censoring, 40% Cause 1, and 30% Cause 2. We consider $n = 200, 400$, and 800. In

Table 1. True parameter values

Censoring	α	p	ρ_{c_1}	ρ_{c_2}
CDC	1.00	0.60	1.40	0.70
	0.50	0.50	1.20	0.50
	0.25	0.40	1.00	0.30
CIC	1.00	0.57	1.40	0.40
	0.50	0.45	1.20	0.50
	0.25	0.20	0.80	0.30

practice, all subjects of some clusters may belong to one stratum and other clusters may have subjects from various strata. To consider such scenarios, we generate data such that the first 25% of the clusters only contain the first stratum whereas the last 25% contain the second stratum only. The middle 50% of the clusters contain both strata. In any cluster, each strata has 2 observations. Thus, the first 25% and the last 25% of clusters consist of 2 subjects within a cluster while the middel 50% of clusters have 4 subjects within a cluster. Two continuous and one binary covariates are considered. For the k^{th} observation in cluster i and stratum j, we generate $Z_{1,ijk} \sim N(0,1)$ and $Z_{3,ijk} \sim Bern(0.7)$. For cluster i and stratum j, $Z_{2,ijk} \sim U(0,1)$ is a cluster-level covariate, so all $Z_{2,ijk}$'s in cluster i and stratum j have the same value. We conduct 5000 replicates for each setting.

Table 2 summarizes the simulation results. For CDC, the proposed method with $\widehat{w}_{ijk}^{COX}(t)$ has little bias and the coverage rates get closer to 95% as n increases. On the other hand, the proposed method with $\widehat{w}_{ijk}^{KM}(t)$ suffer from biased estimates and coverage rates much less than 95%. For CIC, the proposed method with $\widehat{w}_{ijk}^{COX}(t)$ performs as good as that with $\widehat{w}_{ijk}^{KM}(t)$. For both methods, the parameter estimates for Cause 1 are unbiased. As n increases, the coverage rates get closer to 95%. And the standard deviations of both estimates are very close.

We also conduct a simulation for the cumulative baseline subdistribution hazard function at two different time points $t = 0.1$ and 0.7 for both strata. Table 3 includes the simulation results. For CDC, the proposed method with $\widehat{w}_{ijk}^{COX}(t)$ has unbiased estimates in both strata and its coverage rates approach 95% as n increases. However, the stratified PSH model with $\widehat{w}_{ijk}^{KM}(t)$ results in biased estimates and coverage rates less than 95%. Furthermore, for CIC, the estimates of both the proposed methods with $\widehat{w}_{ijk}^{COX}(t)$ and $\widehat{w}_{ijk}^{KM}(t)$ are unbiased and their coverage rates are approximately 95%. And the standard deviations of both estimates are very close.

Next, we perform a simulation study under the same settings as those for Tables 2 and 3, but we ignore clusters and instead treat clustered data as independent data. Tables 4 and 5 report the results. The parameter estimates are the same as those for Tables 2 and 3 because we fit the marginal model. When $\alpha = 1$, the coverage rates for all parameters are close to 95%. When $\alpha = 0.25$ or 0.5, ignoring clusters leads to coverage rates much less than 95% in both tables.

Next, we conduct a simulation study for CDC when strata for censoring and Cause 1 are different. We generate event times for Cause 1 based on the setting for Table 2. However, we create two strata for censoring independent of those for Cause 1 using a Bernoulli distribution with probability 0.5 for each cluster. Table 6 shows the simulation results. The estimates of the proposed method with $\widehat{w}_{ijk}^{COX}(t)$ are unbiased and their coverage rates approach 95% as n increases. However, the proposed method with $\widehat{w}_{ijk}^{KM}(t)$ suffer from biased estimates and low coverage rates.

Finally, we conducted a simulation study to examine the performance of the proposed method when the model for censoring is mis-specified. The parameter setting for Table 2 is used to generate times for Causes 1 and 2. However, correlated censoring times are generated using the following accelerated failure time model for observation k in cluster i and stratum j.

$$\log(C_{ijk}) = \mu_j + \boldsymbol{\gamma}_0^{\intercal} \boldsymbol{Z}_{ijk} + \mathcal{W}_i,\ i = 1, \ldots, n, j = 1, \ldots, S, k = 1, \ldots, n_{ij}.$$

The first 25% and last 25% of the clusters only contain two observations. Thus, we consider

$$\mathcal{W}_i \sim MVN\left(\mathbf{0}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right),$$

where MVN is the multivariate normal distribution. However, the middle 50% of the clusters contain 4 observations. For those clusters, we assume

$$\mathcal{W}_i \sim MVN \left(\mathbf{0}, \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{bmatrix} \right).$$

The number of strata for censoring is 2 and $\boldsymbol{\gamma}_0 = (2.5, 2.5, -3)^\intercal$. Table 7 presents the other parameter values for the AFT model. With these values, it results in approximately 30% censoring, 40% Cause 1, and 30% Cause 2. Table 8 provides the simulation results. The proposed method with $\widehat{w}_{ijk}^{COX}(t)$ has little bias and the coverage rates close to 95%. The estimates of the proposed method with $\widehat{w}_{ijk}^{KM}(t)$ are biased and their coverage rates are less than 95%. The proposed method with $\widehat{w}_{ijk}^{COX}(t)$ is robust against mis-specified modeling of censoring under this limited setting.

5. Real data analysis

We apply our method to the data set of Pidala et al. (2014) to investigate the impact of donor-recipient HLA matching on chronic GVHD of acute lymphoblastic leukemia patients who received unrelated donor allogeneic hematopoietic cell transplantation, where death without experiencing chronic GVHD is a competing event. The variables considered in this application are GVHD prophylaxis (FK506 + (MTX or MMF or steroids) + other , FK506 + other, CsA + MTX + other, CsA + other(No MTX), T-cell depletion and Other); HLA matching (8/8, 7/8 and 6/8); year of transplant (1999 − 2002, 2003 − 2006 and 2007 − 2011); graft type (bone marrow and peripheral blood); in vivo T-cell depletion (no and yes) and total body irradiation (no and yes). The total sample size is 2200. Table 9 shows patient characteristics for the 2200 patients. There are 451 censored patients, 879 patients experienced chronic GVHD and 870 patients died without experiencing chronic GVHD.

First, we check the proportional hazard assumption for censoring by examining the coefficient of variable $\times \log(t)$ under the marginal PH model for each variable. Year of transplant does not satisfy the proportional hazard assumption with p-value < 0.001. Therefore, we stratify year of transplant for the proportional hazards model for censoring. We also include the other variables in Table 9 in modeling the censoring distribution. Next, we check the PSH assumption for each variable by examining the coefficient of variable $\times \log(t)$ based on the proposed model. GVHD prophylaxis does not satisfy the PSH assumption with p-value < 0.001. Hence, we stratify it for the PSH model for chronic GVHD. We also check a study center effect using the score test of homogeneity (Commenges and Andersen, 1995) and it was significant with p-value < 0.001. The level of significance is set at 0.05. Then, we fit four different stratified PSH model as follows:

(1) Proposed method with $\widehat{w}_{ijk}^{COX}(t)$;
(2) Proposed method with $\widehat{w}_{ijk}^{COX}(t)$, but ignores clusters and instead treat the data as independent data;
(3) Proposed method with $\widehat{w}_{ijk}^{KM}(t)$.
(4) Proposed method with $\widehat{w}_{ijk}^{KM}(t)$, but ignores clusters and instead treat the data as independent data;

Table 10 show the analysis results. The results for the models with $\widehat{w}_{ijk}^{COX}(t)$ and $\widehat{w}_{ijk}^{KM}(t)$ are similar because the magnitude of parameter estimates for censoring is modest and thus covariate-dependent censoring does not impact on the analysis result much. In general, the standard errors of the models ignoring clusters are smaller than those of the models accounting for clusters. In all models, HLA matching is not significant. Year of transplant is significant using the proposed method to account for clusters; however when ignoring clusters, it is not significant. Furthermore, total body irradiation effect is not significant using the proposed method to account for clusters, but ignoring clusters leads to a statistical significance. Verneris et al. (2015) studied leukemia patients who received 8/8 or 7/8 HLA matched unrelated donor transplants from 1999-2011. Interestingly, their analysis results show year of transplant was significant, but total body irradiation was not significant for chronic GVHD, which match our analysis results based on the proposed model accounting for clusters.

Table 2. Simulation results for parameter estimation for Cause 1

				$\widehat{w}_{ijk}^{COX}(t)$			$\widehat{w}_{ijk}^{KM}(t)$		
Scenario	α	n	$\boldsymbol{\beta}_0$	Bias	SE(SD)	CP	Bias	SE(SD)	CP
CDC	0.25	200	β_{01}	-0.003	0.089(0.090)	0.946	0.051	0.086(0.089)	0.894
			β_{02}	-0.001	0.356(0.356)	0.950	0.075	0.355(0.356)	0.944
			β_{03}	-0.015	0.209(0.211)	0.950	-0.050	0.207(0.213)	0.945
		400	β_{01}	-0.001	0.063(0.063)	0.945	0.054	0.061(0.062)	0.849
			β_{02}	0.003	0.252(0.252)	0.948	0.079	0.251(0.253)	0.939
			β_{03}	-0.007	0.146(0.147)	0.951	-0.043	0.146(0.148)	0.944
		800	β_{01}	0.000	0.044(0.044)	0.952	0.056	0.043(0.043)	0.732
			β_{02}	0.004	0.178(0.180)	0.947	0.081	0.178(0.180)	0.922
			β_{03}	-0.002	0.103(0.102)	0.950	-0.039	0.103(0.102)	0.939
	0.5	200	β_{01}	0.000	0.087(0.088)	0.943	0.111	0.079(0.082)	0.697
			β_{02}	-0.005	0.332(0.336)	0.942	0.140	0.328(0.334)	0.927
			β_{03}	-0.016	0.210(0.216)	0.943	-0.168	0.209(0.219)	0.885
		400	β_{01}	0.001	0.061(0.061)	0.950	0.114	0.056(0.056)	0.464
			β_{02}	0.005	0.234(0.235)	0.948	0.151	0.232(0.233)	0.899
			β_{03}	-0.007	0.147(0.150)	0.944	-0.160	0.148(0.152)	0.827
		800	β_{01}	0.000	0.043(0.043)	0.951	0.114	0.040(0.040)	0.190
			β_{02}	0.003	0.165(0.165)	0.949	0.152	0.164(0.164)	0.853
			β_{03}	-0.003	0.103(0.104)	0.949	-0.157	0.104(0.105)	0.685
	1	200	β_{01}	0.002	0.085(0.085)	0.948	0.171	0.073(0.074)	0.357
			β_{02}	0.007	0.239(0.237)	0.951	0.203	0.229(0.230)	0.858
			β_{03}	-0.021	0.211(0.206)	0.958	-0.445	0.207(0.207)	0.421
		400	β_{01}	-0.001	0.060(0.059)	0.950	0.170	0.051(0.052)	0.094
			β_{02}	0.006	0.169(0.167)	0.951	0.205	0.162(0.162)	0.755
			β_{03}	-0.007	0.147(0.149)	0.948	-0.435	0.146(0.149)	0.134
		800	β_{01}	0.001	0.042(0.042)	0.951	0.173	0.036(0.037)	0.005
			β_{02}	0.001	0.119(0.120)	0.951	0.201	0.114(0.115)	0.580
			β_{03}	-0.003	0.103(0.103)	0.948	-0.431	0.103(0.103)	0.009
CIC	0.25	200	β_{01}	-0.004	0.074(0.076)	0.943	-0.004	0.074(0.077)	0.941
			β_{02}	-0.003	0.367(0.370)	0.951	-0.002	0.367(0.371)	0.950
			β_{03}	-0.005	0.157(0.158)	0.948	-0.006	0.156(0.158)	0.947
		400	β_{01}	-0.002	0.052(0.053)	0.948	-0.002	0.052(0.053)	0.945
			β_{02}	0.002	0.259(0.259)	0.950	0.002	0.259(0.260)	0.949
			β_{03}	-0.003	0.110(0.111)	0.950	-0.004	0.110(0.111)	0.951
		800	β_{01}	0.000	0.037(0.037)	0.953	0.000	0.037(0.037)	0.954
			β_{02}	0.005	0.183(0.186)	0.944	0.005	0.183(0.186)	0.945
			β_{03}	-0.001	0.078(0.076)	0.954	-0.001	0.078(0.077)	0.953
	0.5	200	β_{01}	-0.004	0.070(0.072)	0.946	-0.004	0.070(0.072)	0.944
			β_{02}	-0.007	0.328(0.330)	0.947	-0.007	0.329(0.330)	0.948
			β_{03}	-0.006	0.151(0.154)	0.946	-0.007	0.151(0.156)	0.943
		400	β_{01}	-0.001	0.049(0.050)	0.946	-0.001	0.050(0.050)	0.945
			β_{02}	0.000	0.231(0.232)	0.946	0.000	0.232(0.233)	0.947
			β_{03}	-0.003	0.106(0.107)	0.950	-0.003	0.107(0.108)	0.947
		800	β_{01}	0.000	0.035(0.035)	0.951	0.000	0.035(0.035)	0.949
			β_{02}	0.002	0.163(0.163)	0.953	0.002	0.164(0.163)	0.953
			β_{03}	-0.001	0.075(0.075)	0.948	-0.002	0.075(0.075)	0.947
	1	200	β_{01}	-0.003	0.066(0.066)	0.948	-0.003	0.066(0.066)	0.946
			β_{02}	0.002	0.230(0.237)	0.939	0.002	0.231(0.239)	0.939
			β_{03}	-0.007	0.150(0.149)	0.952	-0.008	0.150(0.150)	0.950
		400	β_{01}	-0.002	0.047(0.047)	0.946	-0.002	0.047(0.048)	0.942
			β_{02}	0.004	0.163(0.164)	0.945	0.004	0.164(0.167)	0.945
			β_{03}	-0.004	0.105(0.107)	0.946	-0.004	0.106(0.108)	0.944
		800	β_{01}	0.000	0.033(0.033)	0.952	0.000	0.033(0.033)	0.950
			β_{02}	-0.001	0.115(0.116)	0.946	-0.002	0.116(0.118)	0.947
			β_{03}	-0.001	0.074(0.075)	0.949	-0.001	0.075(0.076)	0.951

Table 3. Simulation results for cumulative baseline subdistribution hazard estimation at $t = 0.1$ and 0.7

					$\widehat{w}^{COX}_{ijk}(t)$			$\widehat{w}^{KM}_{ijk}(t)$		
Scenario	α	n	t	Stratum	Bias	SE(SD)	CP	Bias	SE(SD)	CP
CDC	0.25	200	0.1	0	-0.007	0.126(0.128)	0.932	-0.001	0.124(0.127)	0.924
				1	-0.007	0.144(0.144)	0.940	-0.001	0.141(0.143)	0.929
			0.7	0	-0.011	0.190(0.190)	0.938	0.008	0.183(0.187)	0.921
				1	-0.012	0.209(0.209)	0.938	0.008	0.202(0.205)	0.924
		400	0.1	0	-0.005	0.088(0.090)	0.940	0.002	0.087(0.089)	0.929
				1	-0.005	0.101(0.103)	0.939	0.002	0.100(0.103)	0.929
			0.7	0	-0.008	0.133(0.135)	0.941	0.013	0.129(0.132)	0.924
				1	-0.008	0.146(0.149)	0.943	0.013	0.142(0.145)	0.930
		800	0.1	0	-0.003	0.062(0.063)	0.942	0.004	0.061(0.062)	0.933
				1	-0.003	0.071(0.072)	0.943	0.003	0.071(0.071)	0.936
			0.7	0	-0.005	0.093(0.094)	0.945	0.016	0.091(0.092)	0.928
				1	-0.005	0.103(0.104)	0.946	0.015	0.100(0.102)	0.931
	0.5	200	0.1	0	-0.001	0.064(0.067)	0.913	0.016	0.060(0.064)	0.865
				1	-0.002	0.084(0.087)	0.920	0.019	0.080(0.083)	0.880
			0.7	0	-0.004	0.145(0.150)	0.923	0.052	0.131(0.138)	0.852
				1	-0.003	0.178(0.184)	0.929	0.061	0.162(0.169)	0.857
		400	0.1	0	-0.001	0.046(0.047)	0.934	0.016	0.043(0.044)	0.881
				1	-0.002	0.060(0.061)	0.936	0.019	0.057(0.058)	0.890
			0.7	0	-0.005	0.102(0.103)	0.938	0.053	0.093(0.095)	0.851
				1	-0.005	0.125(0.127)	0.937	0.061	0.115(0.117)	0.858
		800	0.1	0	-0.001	0.032(0.033)	0.942	0.017	0.030(0.031)	0.870
				1	-0.001	0.042(0.043)	0.941	0.020	0.040(0.041)	0.880
			0.7	0	-0.003	0.072(0.073)	0.943	0.055	0.066(0.066)	0.817
				1	-0.004	0.089(0.090)	0.942	0.062	0.081(0.082)	0.834
	1	200	0.1	0	0.000	0.018(0.019)	0.913	0.014	0.015(0.015)	0.730
				1	0.000	0.030(0.030)	0.930	0.026	0.025(0.025)	0.715
			0.7	0	-0.001	0.088(0.088)	0.937	0.099	0.065(0.066)	0.592
				1	-0.001	0.136(0.136)	0.938	0.156	0.103(0.104)	0.598
		400	0.1	0	0.000	0.013(0.013)	0.938	0.014	0.010(0.011)	0.654
				1	-0.001	0.021(0.022)	0.938	0.025	0.017(0.018)	0.632
			0.7	0	-0.003	0.062(0.064)	0.938	0.098	0.046(0.048)	0.424
				1	-0.004	0.096(0.099)	0.938	0.155	0.073(0.075)	0.423
		800	0.1	0	0.000	0.009(0.009)	0.947	0.014	0.007(0.007)	0.504
				1	0.000	0.015(0.015)	0.948	0.026	0.012(0.012)	0.438
			0.7	0	-0.001	0.044(0.044)	0.945	0.100	0.032(0.033)	0.179
				1	-0.001	0.067(0.067)	0.948	0.158	0.051(0.051)	0.168
CIC	0.25	200	0.1	0	-0.005	0.098(0.100)	0.934	-0.005	0.098(0.100)	0.933
				1	-0.007	0.112(0.113)	0.937	-0.007	0.112(0.113)	0.937
			0.7	0	-0.009	0.146(0.148)	0.938	-0.009	0.146(0.148)	0.937
				1	-0.010	0.158(0.161)	0.938	-0.010	0.158(0.161)	0.938
		400	0.1	0	-0.004	0.069(0.070)	0.942	-0.004	0.069(0.070)	0.942
				1	-0.004	0.079(0.079)	0.941	-0.004	0.079(0.079)	0.940
			0.7	0	-0.006	0.103(0.103)	0.945	-0.006	0.103(0.103)	0.944
				1	-0.006	0.111(0.112)	0.941	-0.006	0.111(0.112)	0.939
		800	0.1	0	-0.003	0.049(0.049)	0.946	-0.002	0.049(0.049)	0.946
				1	-0.003	0.056(0.057)	0.944	-0.003	0.056(0.057)	0.944
			0.7	0	-0.004	0.072(0.072)	0.947	-0.004	0.072(0.072)	0.946
				1	-0.004	0.079(0.079)	0.943	-0.004	0.078(0.079)	0.943
	0.5	200	0.1	0	-0.001	0.055(0.056)	0.923	-0.001	0.055(0.056)	0.922
				1	-0.002	0.071(0.073)	0.935	-0.002	0.071(0.073)	0.933
			0.7	0	-0.004	0.121(0.124)	0.932	-0.004	0.121(0.124)	0.930
				1	-0.004	0.146(0.148)	0.937	-0.004	0.146(0.148)	0.937
		400	0.1	0	-0.001	0.039(0.039)	0.941	-0.001	0.039(0.039)	0.939
				1	-0.001	0.050(0.051)	0.941	-0.001	0.050(0.051)	0.942
			0.7	0	-0.003	0.085(0.086)	0.940	-0.003	0.085(0.086)	0.940
				1	-0.004	0.103(0.104)	0.941	-0.003	0.103(0.104)	0.943
		800	0.1	0	-0.001	0.027(0.027)	0.946	-0.001	0.027(0.027)	0.944
				1	-0.001	0.035(0.036)	0.942	-0.001	0.036(0.036)	0.941
			0.7	0	-0.002	0.060(0.060)	0.946	-0.002	0.060(0.060)	0.946
				1	-0.002	0.073(0.073)	0.950	-0.002	0.073(0.073)	0.950
	1	200	0.1	0	0.000	0.016(0.016)	0.918	0.000	0.016(0.016)	0.920
				1	0.000	0.025(0.025)	0.928	0.000	0.025(0.026)	0.930
			0.7	0	-0.001	0.070(0.071)	0.933	-0.001	0.070(0.071)	0.934
				1	-0.003	0.103(0.104)	0.945	-0.003	0.103(0.104)	0.942
		400	0.1	0	0.000	0.011(0.011)	0.936	0.000	0.011(0.011)	0.933
				1	-0.001	0.018(0.018)	0.939	-0.001	0.018(0.018)	0.940
			0.7	0	-0.001	0.050(0.050)	0.941	-0.001	0.050(0.050)	0.939
				1	-0.003	0.073(0.075)	0.941	-0.003	0.073(0.075)	0.940
		800	0.1	0	0.000	0.008(0.008)	0.946	0.000	0.008(0.008)	0.947
				1	0.000	0.013(0.012)	0.946	0.000	0.013(0.012)	0.950
			0.7	0	-0.001	0.035(0.035)	0.948	-0.001	0.035(0.035)	0.946
				1	0.000	0.051(0.052)	0.946	0.000	0.052(0.052)	0.945

Table 4. Simulation results for parameter estimation when ignoring clusters

				$\widehat{w}_{ijk}^{COX}(t)$		$\widehat{w}_{ijk}^{KM}(t)$	
Scenario	α	n	$\boldsymbol{\beta}_0$	SE(SD)	CP	SE(SD)	CP
CDC	0.25	200	β_{01}	0.083(0.090)	0.931	0.081(0.089)	0.874
			β_{02}	0.230(0.356)	0.796	0.230(0.356)	0.790
			β_{03}	0.204(0.211)	0.947	0.201(0.213)	0.943
		400	β_{01}	0.058(0.063)	0.930	0.057(0.062)	0.818
			β_{02}	0.162(0.252)	0.792	0.162(0.253)	0.773
			β_{03}	0.142(0.147)	0.944	0.141(0.148)	0.939
		800	β_{01}	0.041(0.044)	0.932	0.040(0.043)	0.697
			β_{02}	0.114(0.180)	0.786	0.114(0.180)	0.750
			β_{03}	0.100(0.102)	0.945	0.099(0.102)	0.932
	0.5	200	β_{01}	0.084(0.088)	0.933	0.077(0.082)	0.675
			β_{02}	0.232(0.336)	0.827	0.228(0.334)	0.776
			β_{03}	0.208(0.216)	0.944	0.205(0.219)	0.881
		400	β_{01}	0.059(0.061)	0.938	0.054(0.056)	0.441
			β_{02}	0.163(0.235)	0.823	0.160(0.233)	0.739
			β_{03}	0.145(0.150)	0.944	0.144(0.152)	0.817
		800	β_{01}	0.041(0.043)	0.940	0.038(0.040)	0.173
			β_{02}	0.115(0.165)	0.832	0.113(0.164)	0.652
			β_{03}	0.101(0.104)	0.945	0.101(0.105)	0.670
	1	200	β_{01}	0.085(0.085)	0.947	0.073(0.074)	0.356
			β_{02}	0.232(0.237)	0.944	0.222(0.230)	0.848
			β_{03}	0.211(0.206)	0.958	0.207(0.207)	0.422
		400	β_{01}	0.060(0.059)	0.951	0.051(0.052)	0.092
			β_{02}	0.163(0.167)	0.944	0.157(0.162)	0.736
			β_{03}	0.147(0.149)	0.950	0.145(0.149)	0.132
		800	β_{01}	0.042(0.042)	0.950	0.036(0.037)	0.005
			β_{02}	0.115(0.120)	0.944	0.110(0.115)	0.552
			β_{03}	0.103(0.103)	0.947	0.102(0.103)	0.009
CIC	0.25	200	β_{01}	0.067(0.076)	0.915	0.067(0.077)	0.913
			β_{02}	0.225(0.370)	0.771	0.225(0.371)	0.771
			β_{03}	0.154(0.158)	0.947	0.153(0.158)	0.946
		400	β_{01}	0.047(0.053)	0.920	0.047(0.053)	0.918
			β_{02}	0.158(0.259)	0.769	0.158(0.260)	0.770
			β_{03}	0.108(0.111)	0.946	0.108(0.111)	0.946
		800	β_{01}	0.033(0.037)	0.930	0.033(0.037)	0.929
			β_{02}	0.111(0.186)	0.755	0.112(0.186)	0.759
			β_{03}	0.076(0.076)	0.949	0.076(0.077)	0.948
	0.5	200	β_{01}	0.066(0.072)	0.929	0.066(0.072)	0.929
			β_{02}	0.219(0.330)	0.806	0.220(0.330)	0.807
			β_{03}	0.149(0.154)	0.945	0.150(0.156)	0.944
		400	β_{01}	0.046(0.050)	0.929	0.046(0.050)	0.929
			β_{02}	0.153(0.232)	0.806	0.154(0.233)	0.811
			β_{03}	0.105(0.107)	0.948	0.105(0.108)	0.944
		800	β_{01}	0.033(0.035)	0.934	0.033(0.035)	0.937
			β_{02}	0.108(0.163)	0.803	0.109(0.163)	0.802
			β_{03}	0.074(0.075)	0.944	0.074(0.075)	0.942
	1	200	β_{01}	0.067(0.066)	0.949	0.067(0.066)	0.949
			β_{02}	0.220(0.237)	0.932	0.222(0.239)	0.930
			β_{03}	0.150(0.149)	0.954	0.151(0.150)	0.951
		400	β_{01}	0.047(0.047)	0.947	0.047(0.048)	0.944
			β_{02}	0.155(0.164)	0.933	0.156(0.167)	0.935
			β_{03}	0.106(0.107)	0.949	0.106(0.108)	0.948
		800	β_{01}	0.033(0.033)	0.950	0.033(0.033)	0.950
			β_{02}	0.109(0.116)	0.936	0.110(0.118)	0.936
			β_{03}	0.075(0.075)	0.950	0.075(0.076)	0.950

Table 5. Simulation results for cumulative baseline subdistribution hazard estimation at $t = 0.1$ and 0.7 when ignoring clusters

					$\widehat{w}_{ijk}^{COX}(t)$		$\widehat{w}_{ijk}^{KM}(t)$	
Scenario	α	n	t	Stratum	SE(SD)	CP	SE(SD)	CP
CDC	0.25	200	0.1	0	0.103(0.128)	0.881	0.100(0.127)	0.870
				1	0.118(0.144)	0.887	0.115(0.143)	0.880
			0.7	0	0.157(0.190)	0.892	0.151(0.187)	0.874
				1	0.173(0.209)	0.891	0.167(0.205)	0.877
		400	0.1	0	0.072(0.090)	0.880	0.071(0.089)	0.871
				1	0.082(0.103)	0.878	0.081(0.103)	0.868
			0.7	0	0.110(0.135)	0.891	0.106(0.132)	0.868
				1	0.121(0.149)	0.889	0.117(0.145)	0.869
		800	0.1	0	0.050(0.063)	0.887	0.050(0.062)	0.876
				1	0.058(0.072)	0.886	0.057(0.071)	0.879
			0.7	0	0.077(0.094)	0.893	0.075(0.092)	0.871
				1	0.085(0.104)	0.890	0.083(0.102)	0.872
	0.5	200	0.1	0	0.054(0.067)	0.870	0.050(0.064)	0.819
				1	0.071(0.087)	0.882	0.067(0.083)	0.833
			0.7	0	0.125(0.150)	0.886	0.112(0.138)	0.806
				1	0.154(0.184)	0.889	0.139(0.169)	0.814
		400	0.1	0	0.038(0.047)	0.889	0.036(0.044)	0.825
				1	0.050(0.061)	0.891	0.047(0.058)	0.830
			0.7	0	0.088(0.103)	0.898	0.079(0.095)	0.796
				1	0.108(0.127)	0.899	0.098(0.117)	0.809
		800	0.1	0	0.027(0.033)	0.889	0.025(0.031)	0.801
				1	0.035(0.043)	0.896	0.033(0.041)	0.818
			0.7	0	0.062(0.073)	0.897	0.056(0.066)	0.744
				1	0.076(0.090)	0.901	0.069(0.082)	0.763
	1	200	0.1	0	0.018(0.019)	0.914	0.014(0.015)	0.727
				1	0.030(0.030)	0.931	0.024(0.025)	0.711
			0.7	0	0.087(0.088)	0.936	0.064(0.066)	0.586
				1	0.135(0.136)	0.938	0.102(0.104)	0.594
		400	0.1	0	0.013(0.013)	0.936	0.010(0.011)	0.646
				1	0.021(0.021)	0.936	0.017(0.018)	0.628
			0.7	0	0.062(0.064)	0.935	0.045(0.048)	0.413
				1	0.096(0.099)	0.937	0.072(0.075)	0.420
		800	0.1	0	0.009(0.009)	0.946	0.007(0.007)	0.497
				1	0.015(0.015)	0.947	0.012(0.012)	0.428
			0.7	0	0.043(0.044)	0.944	0.033(0.032)	0.171
				1	0.067(0.067)	0.946	0.051(0.051)	0.164
CIC	0.25	200	0.1	0	0.074(0.100)	0.851	0.074(0.100)	0.849
				1	0.084(0.113)	0.858	0.084(0.113)	0.858
			0.7	0	0.111(0.148)	0.860	0.111(0.148)	0.861
				1	0.120(0.161)	0.858	0.120(0.161)	0.858
		400	0.1	0	0.052(0.070)	0.850	0.052(0.070)	0.850
				1	0.059(0.079)	0.858	0.059(0.079)	0.858
			0.7	0	0.078(0.103)	0.863	0.078(0.103)	0.862
				1	0.084(0.112)	0.864	0.084(0.112)	0.864
		800	0.1	0	0.037(0.049)	0.857	0.037(0.049)	0.859
				1	0.042(0.057)	0.854	0.042(0.057)	0.853
			0.7	0	0.055(0.072)	0.864	0.055(0.072)	0.864
				1	0.059(0.079)	0.861	0.059(0.079)	0.860
	0.5	200	0.1	0	0.043(0.056)	0.860	0.043(0.056)	0.858
				1	0.056(0.073)	0.866	0.056(0.073)	0.863
			0.7	0	0.098(0.124)	0.873	0.098(0.124)	0.871
				1	0.117(0.148)	0.876	0.117(0.148)	0.873
		400	0.1	0	0.030(0.039)	0.873	0.030(0.039)	0.876
				1	0.040(0.051)	0.872	0.040(0.051)	0.875
			0.7	0	0.069(0.086)	0.881	0.069(0.086)	0.876
				1	0.083(0.104)	0.880	0.083(0.104)	0.879
		800	0.1	0	0.021(0.027)	0.877	0.022(0.027)	0.879
				1	0.028(0.036)	0.873	0.028(0.036)	0.873
			0.7	0	0.049(0.060)	0.884	0.049(0.060)	0.885
				1	0.058(0.073)	0.884	0.058(0.073)	0.883
	1	200	0.1	0	0.015(0.016)	0.917	0.015(0.016)	0.919
				1	0.025(0.025)	0.928	0.025(0.026)	0.929
			0.7	0	0.069(0.071)	0.933	0.069(0.071)	0.932
				1	0.102(0.104)	0.944	0.102(0.104)	0.942
		400	0.1	0	0.011(0.011)	0.936	0.011(0.011)	0.933
				1	0.018(0.018)	0.937	0.018(0.018)	0.938
			0.7	0	0.049(0.050)	0.940	0.049(0.050)	0.938
				1	0.072(0.075)	0.940	0.072(0.075)	0.937
		800	0.1	0	0.008(0.008)	0.943	0.008(0.008)	0.943
				1	0.012(0.012)	0.944	0.012(0.012)	0.947
			0.7	0	0.035(0.035)	0.945	0.035(0.035)	0.944
				1	0.051(0.052)	0.943	0.051(0.052)	0.942

Table 6. Simulation results for parameter estimates of Cause 1, when strata for Cause 1 and censoring models are different

			$\widehat{w}_{ijk}^{COX}(t)$			$\widehat{w}_{ijk}^{KM}(t)$		
α	n	$\boldsymbol{\beta}_0$	Bias	SE(SD)	CP	Bias	SE(SD)	CP
0.25	200	β_{01}	-0.004	0.089(0.090)	0.943	0.050	0.085(0.088)	0.900
		β_{02}	0.011	0.356(0.364)	0.946	0.085	0.355(0.364)	0.938
		β_{03}	-0.013	0.209(0.206)	0.957	-0.049	0.206(0.207)	0.951
	400	β_{01}	-0.001	0.063(0.063)	0.951	0.054	0.061(0.062)	0.842
		β_{02}	0.012	0.251(0.254)	0.949	0.088	0.251(0.254)	0.938
		β_{03}	-0.006	0.146(0.146)	0.947	-0.042	0.145(0.146)	0.943
	800	β_{01}	-0.001	0.044(0.044)	0.954	0.055	0.043(0.043)	0.742
		β_{02}	0.009	0.177(0.178)	0.946	0.086	0.177(0.178)	0.926
		β_{03}	-0.003	0.103(0.103)	0.947	-0.039	0.103(0.104)	0.934
0.5	200	β_{01}	0.002	0.087(0.086)	0.948	0.111	0.079(0.080)	0.695
		β_{02}	0.009	0.330(0.336)	0.945	0.151	0.327(0.332)	0.918
		β_{03}	-0.016	0.211(0.213)	0.948	-0.166	0.209(0.216)	0.889
	400	β_{01}	0.001	0.061(0.061)	0.955	0.113	0.056(0.057)	0.470
		β_{02}	0.003	0.233(0.236)	0.946	0.147	0.231(0.235)	0.902
		β_{03}	-0.005	0.147(0.146)	0.951	-0.157	0.147(0.149)	0.836
	800	β_{01}	0.000	0.043(0.043)	0.953	0.113	0.040(0.040)	0.185
		β_{02}	0.008	0.164(0.167)	0.947	0.155	0.163(0.167)	0.836
		β_{03}	-0.004	0.103(0.102)	0.953	-0.158	0.104(0.104)	0.690
1	200	β_{01}	0.004	0.084(0.085)	0.947	0.173	0.073(0.075)	0.345
		β_{02}	0.015	0.230(0.232)	0.948	0.210	0.220(0.223)	0.834
		β_{03}	-0.013	0.211(0.211)	0.953	-0.443	0.206(0.214)	0.432
	400	β_{01}	0.002	0.060(0.057)	0.959	0.173	0.051(0.051)	0.082
		β_{02}	0.006	0.162(0.166)	0.945	0.204	0.156(0.160)	0.736
		β_{03}	-0.008	0.147(0.143)	0.960	-0.442	0.145(0.144)	0.118
	800	β_{01}	0.002	0.042(0.042)	0.950	0.174	0.036(0.037)	0.002
		β_{02}	0.005	0.115(0.114)	0.952	0.204	0.110(0.111)	0.542
		β_{03}	-0.004	0.103(0.102)	0.955	-0.438	0.103(0.102)	0.003

Table 7. True parameter values of the AFT model for censoring

Censoring	α	μ_1	μ_2	p
AFT	1.00	0.70	2.00	0.52
	0.50	-0.50	2.00	0.38
	0.25	-2.00	2.00	0.15

Table 8. Simulation results of parameter estimates for Cause 1 when censoring follows AFT model

			$\widehat{w}_{ijk}^{COX}(t)$			$\widehat{w}_{ijk}^{KM}(t)$		
α	n	$\boldsymbol{\beta}_0$	Bias	SE(SD)	CP	Bias	SE(SD)	CP
0.25	200	β_{01}	-0.008	0.079(0.080)	0.944	-0.059	0.077(0.079)	0.880
		β_{02}	0.002	0.352(0.355)	0.948	-0.077	0.347(0.350)	0.939
		β_{03}	-0.001	0.150(0.153)	0.944	0.042	0.143(0.146)	0.931
	400	β_{01}	-0.006	0.055(0.056)	0.949	-0.058	0.054(0.055)	0.821
		β_{02}	0.000	0.248(0.253)	0.944	-0.081	0.245(0.249)	0.936
		β_{03}	0.003	0.106(0.106)	0.949	0.047	0.101(0.101)	0.927
	800	β_{01}	-0.004	0.039(0.039)	0.950	-0.057	0.038(0.038)	0.697
		β_{02}	0.004	0.175(0.174)	0.950	-0.077	0.173(0.172)	0.926
		β_{03}	0.004	0.075(0.075)	0.948	0.049	0.071(0.072)	0.894
0.5	200	β_{01}	-0.012	0.081(0.082)	0.947	-0.100	0.076(0.079)	0.747
		β_{02}	-0.002	0.322(0.326)	0.948	-0.135	0.312(0.318)	0.927
		β_{03}	0.004	0.146(0.150)	0.941	0.102	0.136(0.140)	0.871
	400	β_{01}	-0.009	0.056(0.057)	0.947	-0.098	0.054(0.055)	0.565
		β_{02}	-0.003	0.227(0.228)	0.947	-0.139	0.221(0.222)	0.905
		β_{03}	0.004	0.103(0.102)	0.952	0.103	0.096(0.096)	0.807
	800	β_{01}	-0.005	0.040(0.040)	0.947	-0.095	0.038(0.038)	0.300
		β_{02}	-0.004	0.161(0.159)	0.952	-0.140	0.156(0.155)	0.853
		β_{03}	0.004	0.073(0.073)	0.954	0.105	0.068(0.068)	0.652
1	200	β_{01}	-0.012	0.082(0.082)	0.949	-0.143	0.072(0.075)	0.499
		β_{02}	-0.002	0.234(0.238)	0.942	-0.196	0.221(0.227)	0.847
		β_{03}	0.007	0.143(0.145)	0.944	0.171	0.132(0.135)	0.731
	400	β_{01}	-0.009	0.057(0.057)	0.949	-0.142	0.051(0.052)	0.201
		β_{02}	-0.007	0.165(0.163)	0.951	-0.202	0.156(0.155)	0.752
		β_{03}	0.009	0.101(0.101)	0.953	0.172	0.093(0.093)	0.534
	800	β_{01}	-0.007	0.040(0.040)	0.953	-0.140	0.036(0.036)	0.025
		β_{02}	-0.003	0.116(0.116)	0.948	-0.198	0.110(0.111)	0.564
		β_{03}	0.009	0.071(0.071)	0.948	0.172	0.066(0.065)	0.254

Table 9. Patient characteristics

	8/8	7/8	6/8	Total
Year of transplant				
1999 - 2002	237	131	64	432
2003 - 2006	497	223	60	780
2007 - 2011	702	259	27	988
Graft type				
Peripheral blood	759	297	65	1121
Bone marrow	677	316	86	1079
In vivo T cell depletion				
No	1087	400	92	1579
Yes	349	213	59	621
Total body irradiation				
No	166	69	4	239
Yes	1270	544	147	1961
GVHD prophylaxis				
FK506 + (MTX or MMF or steroids) + other	779	300	58	1137
FK506 + other	85	21	4	110
CsA + MTX + other	407	187	50	644
CsA + other(No MTX)	48	23	6	77
T-cell depletion	95	63	31	189
Other	22	19	2	43

Table 10.: Real data analysis

	$\widehat{w}_{ijk}^{COX}(t)$ accounting for clusters		$\widehat{w}_{ijk}^{COX}(t)$ ignoring clusters		$\widehat{w}_{ijk}^{KM}(t)$ accounting for clusters		$\widehat{w}_{ijk}^{KM}(t)$ ignoring clusters	
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
HLA matching								
8/8 (ref)	-	0.382*	-	0.413*	-	0.384*	-	0.411*
7/8	0.069 (0.077)	0.369	0.068 (0.078)	0.380	0.068 (0.076)	0.371	0.068 (0.077)	0.379
6/8	-0.133 (0.224)	0.554	-0.133 (0.157)	0.399	-0.133 (0.224)	0.554	-0.133 (0.157)	0.396
Year of transplant								
1999 - 2002 (ref)	-	0.030*	-	0.053*	-	0.036*	-	0.061*
2003 - 2006	0.075 (0.097)	0.442	0.074 (0.103)	0.474	0.073 (0.098)	0.453	0.073 (0.103)	0.480
2007 - 2011	-0.109 (0.110)	0.319	-0.110 (0.106)	0.297	-0.106 (0.110)	0.333	-0.106 (0.106)	0.314
Graft type								
Peripheral blood (ref)	-	<0.001*	-	<0.001*	-	<0.001*	-	<0.001*
Bone marrow	-0.387 (0.097)	<0.001	-0.387 (0.076)	<0.001	-0.386 (0.097)	<0.001	-0.386 (0.076)	<0.001
In vivo T cell depletion								
No(ref)	-	<0.001*	-	<0.001*	-	<0.001*	-	<0.001*
Yes	-0.527 (0.106)	<0.001	-0.526 (0.086)	<0.001	-0.526 (0.106)	<0.001	-0.526 (0.086)	<0.001
Total body irradiation								
No (ref)	-	0.076*	-	0.021	-	0.074*	-	0.020*
Yes	0.280 (0.158)	0.076	0.280 (0.121)	0.021	0.282 (0.158)	0.074	0.282 (0.121)	0.020

SE is the standard error and * denotes the overall p-value.

6. Conclusion

We have studied when the number of strata is finite, but the number of clusters increases. Zhou et al. (2011) considered highly stratified data which have a large number of strata compared with the strata sizes. In this case, one needs to consider that the number of clusters for each stratum is finite, but the number of strata increases. As Zhou et al. (2011) discussed, the asymptotic variance estimator could be unstable for modest sample sizes and bootstrap for clustered data may need to be implemented. Investigating highly stratified data would be an interesting future research question. Recently Mao and Lin (2017) and Bellach et al. (2019) proposed semiparametric transformation models for independent competing risks data. They considered a general class of functions for the cumulative subdistribution hazard function to handle possibly non-proportional subdistribution hazards structure. The PSH model and proportional subdistribution odds model can be handled as special cases. Although their method is flexible, when the transformation function is not trivial, it may be hard to interpret parameter estimates. Nonetheless, developing transformation models for clustered competing risks data would be a worthy future topic. The proposed method is a marginal model. Another popular method to handle clustered data is a random effects model. Developing a frailty model would be another important future topic.

Supplementary information

The Supplementary Materials provide proofs for Theorems 3.1 to 3.3 and the asymptotic variances of the proposed estimators for $\boldsymbol{\beta}$ and $\Lambda_{10j}(t)$ when using Kaplan-Meier estimates for censoring.

Acknowledgements

This work was supported in part by the Advancing a Healthier Wisconsin Endowment (Project # 5520569) and the US National Cancer Institute (U24CA076518).

Appendices

$$
\begin{aligned}
\boldsymbol{q}^{(1)}_{ijk}(u) = &- \lim_{n_j\to\infty}\Bigg(\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{s=X_{i'jk'}}^{\tau}\left\{\boldsymbol{Z}_{i'jk'}-\frac{s^{(1)}_{COX,j}(\boldsymbol{\beta}_0,s)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\right\}\\
&\times w^{COX}_{i'jk'}(s)\frac{\exp(\boldsymbol{\gamma}_0^{\intercal}\boldsymbol{Z}_{i'jk'})I(u\le s)}{s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,u)}dM^1_{i'jk'}(s)\\
&+\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\int_{s=X_{i'j'k'}}^{\tau}\left\{\boldsymbol{Z}_{i'j'k'}-\frac{s^{(1)}_{COX,j}(\boldsymbol{\beta}_0,s)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\right\}w^{COX}_{i'j'k'}(s)\\
&\times \boldsymbol{h}^{\intercal}_{Cj}(s,X_{i'j'k'},\boldsymbol{Z}_{i'j'k'})\boldsymbol{\Omega}_C^{-1}\left\{\boldsymbol{Z}_{ijk}-\frac{s^{(1)}_{Cj}(\boldsymbol{\gamma}_0,u)}{s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,u)}\right\}dM^1_{i'j'k'}(s)\Bigg),
\end{aligned}
$$

$$
\boldsymbol{h}_{Cj}(t,u,\boldsymbol{Z}) = e^{\boldsymbol{\gamma}_0^T\boldsymbol{Z}}\int_{s=u}^{t}\left\{\boldsymbol{Z}-\frac{s^{(1)}_{Cj}(\boldsymbol{\gamma}_0,s)}{s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,s)}\right\}d\Lambda_{C0j}(s),
$$

$$
\begin{aligned}
q^{(2)}_{ijk}(s,t) = &- \lim_{n_j\to\infty}\Bigg[\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{\nu=X_{i'jk'}}^{t}\frac{\exp\{\boldsymbol{\gamma}_0^{\intercal}\boldsymbol{Z}_{i'jk'}\}I(s\le\nu)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,\nu)s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,s)}\\
&\times w^{COX}_{i'jk'}(\nu)dM_{i'jk'}(\nu)+\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\frac{1}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,\nu)}\\
&\times \boldsymbol{h}^{\intercal}_{Cj}(\nu,X_{i'j'k'},\boldsymbol{Z}_{i'j'k'})\boldsymbol{\Omega}_C^{-1}\left\{\boldsymbol{Z}_{ijk}-\frac{s^{(1)}_{Cj}(\boldsymbol{\gamma}_0,s)}{s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,s)}\right\}dM^1_{i'j'k'}(\nu)\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\Sigma}_{F_{1j}}(t) = &\frac{1}{n_j}\left\{\exp\left(\widehat{\boldsymbol{\beta}}^{\intercal}\boldsymbol{Z}_0-\widehat{\Lambda}_{10j}(t)\exp(\widehat{\boldsymbol{\beta}}^{\intercal}\boldsymbol{Z}_0)\right)\right\}^2\sum_{i=1}^{n}\Bigg\{\widehat{W}^{(1)}_{ij.}(t)+\widehat{W}^{(2)}_{ij.}(t)\\
&+p_j\widehat{\Lambda}_{10j}(t)\boldsymbol{Z}_0^{\intercal}\left(\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}})\right)^{-1}\left(\widehat{\boldsymbol{\eta}}^{COX}_{i..}+\widehat{\boldsymbol{\psi}}^{COX}_{i..}\right)\Bigg\}^2,
\end{aligned}
$$

$$
\widehat{W}^{(1)}_{ij.}(t) = -\pi_j(\widehat{\boldsymbol{\eta}}^{COX}_{i..}+\widehat{\boldsymbol{\psi}}^{COX}_{i..})^{\intercal}\left\{\mathcal{I}_1(\widehat{\boldsymbol{\beta}})/n\right\}^{-1}\int_0^t\left(\frac{S^{(1)}_{COX,j}(\widehat{\boldsymbol{\beta}},s)}{S^{(0)}_{COX,j}(\widehat{\boldsymbol{\beta}},s)}\right)d\widehat{\Lambda}^{COX}_{10j}(s),
$$

$$
\widehat{W}^{(2)}_{ij.}(t) = \Phi_{ij}\sum_{k=1}^{n_{ij}}\left\{\int_0^t\frac{\widehat{w}^{COX}_{ijk}(s)d\widehat{M}^1_{ijk}(s)}{S^{(0)}_{COX,j}(\widehat{\boldsymbol{\beta}},s)}+\int_0^t\widehat{q}^{(2)}_{ijk}(s,t)d\widehat{M}^C_{ijk}(s)\right\}.
$$

The estimates $\widehat{\boldsymbol{q}}^{(1)}_{ijk}(u)$, $\widehat{\boldsymbol{h}}_{Cj}(t,u,\boldsymbol{Z})$, $\widehat{q}^{(2)}_{ijk}(s,t)$, $\widehat{\boldsymbol{\eta}}^{COX}_{i..}$ and $\widehat{\boldsymbol{\psi}}^{COX}_{i..}$ are provided in the Supplementary Materials.

References

Joseph Pidala, Stephanie J Lee, Kwang Woo Ahn, Stephen Spellman, Hai-Lin Wang, Mahmoud Aljurf, Medhat Askar, Jason Dehn, Marcelo Fernandez Viña, Alois Gratwohl, et al. Nonpermissive HLA-DPB1 mismatch increases mortality after myeloablative unrelated allogeneic hematopoietic cell transplantation. *Blood, The Journal of the American Society of Hematology*, 124(16):2596–2606, 2014.

Thomas H Scheike, Yanqing Sun, Mei-Jie Zhang, and Tina Kold Jensen. A semiparametric random effects model for multivariate competing risks data. *Biometrika*, 97 (1):133–145, 2010.

Brent R Logan, Mei-Jie Zhang, and John P Klein. Marginal models for clustered time-to-event data with competing risks using pseudovalues. *Biometrics*, 67(1):1–7, 2011.

Ping K Ruan and Robert J Gray. Analyses of cumulative incidence functions via nonparametric multiple imputation. *Statistics in Medicine*, 27(27):5709–5724, 2008.

Bingqing Zhou, Jason Fine, Aurelien Latouche, and Myriam Labopin. Competing risks regression for clustered data. *Biostatistics*, 13(3):371–383, 2012.

Jason P Fine and Robert J Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

Manoj Khanal and Kwang Woo Ahn. *adjSURVCI: Parameter and Adjusted Probability Estimation for Right-Censored Data*, 2021. URL `https://CRAN.R-project.org/package=adjSURVCI`. R package version 1.0.

David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

Soyoung Kim, Yayun Xu, Mei-Jie Zhang, and Kwang-Woo Ahn. Stratified proportional subdistribution hazards model with covariate-adjusted censoring weight for case-cohort studies. *Scandinavian Journal of Statistics*, 47(4):1222–1242, 2020.

Daniel Commenges and Per Kragh Andersen. Score test of homogeneity for survival data. *Lifetime Data Analysis*, 1(2):145–156, 1995.

Michael R Verneris, Stephanie J Lee, Kwang Woo Ahn, Hai-Lin Wang, Minoo Battiwalla, Yoshihiro Inamoto, Marcelo A Fernandez-Vina, James Gajewski, Joseph Pidala, Reinhold Munker, et al. Hla mismatch is associated with worse outcomes after unrelated donor reduced-intensity conditioning hematopoietic cell transplantation: an analysis from the center for international blood and marrow transplant research. *Biology of Blood and Marrow Transplantation*, 21(10):1783–1789, 2015.

Bingqing Zhou, Aurelien Latouche, Vanderson Rocha, and Jason Fine. Competing risks regression for stratified data. *Biometrics*, 67(2):661–670, 2011.

Lu Mao and DY Lin. Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 79(2):573, 2017.

Anna Bellach, Michael R Kosorok, Ludger Rüschendorf, and Jason P Fine. Weighted NPMLE for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 114(525):259–270, 2019.

Supplementary Materials for Competing risks regression for clustered data with covariate-dependent censoring

Manoj Khanal[a], Soyoung Kim[a,b] and Kwang Woo Ahn[a,b]

[a]Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, 53226, Wisconsin,USA
[b]CIBMTR, Medical College of Wisconsin Clinical Cancer Center, 9200 W. Wisconsin Avenue, Milwaukee, 53226, Wisconsin,USA

ARTICLE HISTORY
Compiled March 19, 2023

ABSTRACT
We provide proofs for Theorems 1 to 3. We also provide the asymptotic variances of the proposed estimators for $\boldsymbol{\beta}$ and $\Lambda_{10j}(t)$ when using Kaplan-Meier estimates for censoring.

1. Supplementary Materials

1.1. Proof of Theorem 1

Using similar arguments as in Kim et al. (2020), we have

$$\begin{aligned}
&\widehat{G}_{Cj}(t \mid \boldsymbol{Z}_{ijk}) - G_{Cj}(t \mid \boldsymbol{Z}_{ijk}) \\
&= -\frac{G_{Cj}(t \mid \boldsymbol{Z}_{ijk})}{n_j} \int_{u=0}^{\tau} \sum_{i'=1}^{n} \Phi_{i'j} \sum_{k'=1}^{n_{i'j}} \frac{e^{\gamma_0^\intercal \boldsymbol{Z}_{ijk}} I(u \le t)}{s_{Cj}^{(r)}(\boldsymbol{\gamma}_0, u)} dM_{i'jk'}^{C}(u) \\
&- \frac{G_{Cj}(t \mid \boldsymbol{Z}_{ijk})}{n_j} \int_{u=0}^{\tau} \boldsymbol{h}_{Cj}^\intercal(t, 0, \boldsymbol{Z}_{ijk}) \boldsymbol{\Omega}_C^{-1} \sum_{i'=1}^{n} \sum_{j'=1}^{S} \Phi_{i'j'} \\
&\times \sum_{k'=1}^{n_{i'j'}} \left\{ \boldsymbol{Z}_{i'j'k'} - \frac{s_{Cj'}^{(1)}(\boldsymbol{\beta}_{0C}, s)}{s_{Cj'}^{(0)}(\boldsymbol{\beta}_{0C}, s)} \right\} dM_{i'j'k'}^{C}(u) + o_p(n_j^{-1/2}),
\end{aligned} \tag{1}$$

where

$$\boldsymbol{h}_{Cj}(t, u, \boldsymbol{Z}) = e^{\gamma_0^\intercal \boldsymbol{Z}} \int_{s=u}^{t} \left\{ \boldsymbol{Z} - \frac{s_{Cj}^{(1)}(\boldsymbol{\gamma}_0, s)}{s_{Cj}^{(0)}(\boldsymbol{\gamma}_0, s)} \right\} d\Lambda_{C0j}(s).$$

CONTACT Kwang Woo Ahn. Email: kwooahn@mcw.edu

Using (1), we have

$$
\begin{aligned}
&\frac{\widehat{G}_{Cj}(t \mid \boldsymbol{Z}_{ijk})}{\widehat{G}_{Cj}(X_{ijk} \wedge t \mid \boldsymbol{Z}_{ijk})} - \frac{G_{Cj}(t \mid \boldsymbol{Z}_{ijk})}{G_{Cj}(X_{ijk} \wedge t \mid \boldsymbol{Z}_{ijk})} \\
&= I(X_{ijk} < t)\frac{G_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})\{\widehat{G}_{Cj}(t \mid \boldsymbol{Z}_{ijk}) - G_{Cj}(t \mid \boldsymbol{Z}_{ijk})\}}{\widehat{G}_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})G_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})} \\
&- \frac{G_{Cj}(t \mid \boldsymbol{Z}_{ijk})\{\widehat{G}_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk}) - G_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})\}}{\widehat{G}_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})G_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})} \\
&= -I(X_{ijk} < t)\frac{G_{Cj}(t \mid \boldsymbol{Z}_{ijk})}{G_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})}\left(\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{s=0}^{\tau}\exp(\boldsymbol{\gamma}_0^\intercal \boldsymbol{Z}_{ijk})\right. \\
&\times \frac{I(X_{ijk} < s \le t)}{s_{Cj}^{(0)}(\boldsymbol{\gamma}_0, s)}dM_{i'jk'}^{C}(s) + \frac{1}{n_j}\sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\int_{s=0}^{\tau} \\
&\left.\boldsymbol{h}_{Cj}^\intercal(t, X_{ijk}, \boldsymbol{Z}_{ijk})\boldsymbol{\Omega}_C^{-1}\left\{\boldsymbol{Z}_{i'j'k'} - \frac{s_{Cj'}^{(1)}(\boldsymbol{\gamma}_0, s)}{s_{Cj'}^{(0)}(\boldsymbol{\gamma}_0, s)}\right\}dM_{i'j'k'}^{C}(s)\right) + o_p(n_j^{-1/2}).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\widehat{w}_{ijk}^{COX}(t) - w_{ijk}^{COX}(t) \\
&= -I(C_{ijk} \ge T_{ijk} \wedge t)I(X_{ijk} < t)\frac{G_{Cj}(t \mid \boldsymbol{Z}_{ijk})}{G_{Cj}(X_{ijk} \mid \boldsymbol{Z}_{ijk})} \\
&\frac{1}{n_j}\left(\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{s=0}^{\tau}\frac{\exp(\boldsymbol{\gamma}_0^T \boldsymbol{Z}_{ijk})I(X_{ijk} < s \le t)}{s_{Cj}^{(0)}(\boldsymbol{\gamma}_0, s)}dM_{i'jk'}^{C}(s)\right. \\
&+ \sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\int_{s=0}^{\tau}\boldsymbol{h}_{Cj}^\intercal(t, X_{ijk}, \boldsymbol{Z}_{ijk})\boldsymbol{\Omega}_C^{-1} \\
&\left.\times\left\{\boldsymbol{Z}_{i'j'k'} - \frac{s_{Cj'}^{(1)}(\boldsymbol{\gamma}_0, s)}{s_{Cj'}^{(0)}(\boldsymbol{\gamma}_0, s)}\right\}dM_{i'j'k'}^{C}(s)\right) + o_p(n_j^{-1/2}).
\end{aligned} \tag{2}
$$

Denote $\boldsymbol{U}_{n,COX}(\boldsymbol{\beta}) = 1/n \times \boldsymbol{U}_{COX}(\boldsymbol{\beta})$. By Foutz (1977), if

(C1) $\partial \boldsymbol{U}_{n,COX}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^\intercal$ exists and is continuous in an open neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$,
(C2) $-\partial \boldsymbol{U}_{n,COX}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}_0^\intercal$ converges to $\Omega(\boldsymbol{\beta}_0)$ which is positive definite,
(C3) $\boldsymbol{U}_{n,COX}(\boldsymbol{\beta})$ converges to 0 in probability,

then, $\widehat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}_0$ in probability. Clearly $\partial \boldsymbol{U}_{n,COX}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^\intercal$ exists and $S_{COX,j}^{(r)}(\boldsymbol{\beta}, t)$ is continuous for $r = 0, 1, 2$ by (A9). Therefore, (C1) is satisfied. Condition (C2) is satisfied by (A6). Now we need to show (C3). By (A10), the score function evaluated at $\boldsymbol{\beta}_0$ is

$$
\begin{aligned}
&\boldsymbol{U}_{COX}(\boldsymbol{\beta}_0)\\
&= \sum_{i=1}^{n}\left[\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\left\{\boldsymbol{Z}_{ijk} - \frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}_0,u)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,u)}\right\}\widehat{w}^{COX}_{ijk}(u)dM^1_{ijk}(u)\right]\\
&= \sum_{i=1}^{n}\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\left\{\boldsymbol{Z}_{ijk} - \frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}_0,u)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,u)}\right\}w^{COX}_{ijk}(u)dM^1_{ijk}(u)\\
&+ \sum_{i=1}^{n}\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\left\{\boldsymbol{Z}_{ijk} - \frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}_0,u)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,u)}\right\}\left\{\widehat{w}^{COX}_{ijk}(u) - w^{COX}_{ijk}(u)\right\}dM^1_{ijk}(u) \qquad (3)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\left\{\boldsymbol{Z}_{ijk} - \frac{s^{(1)}_{COX,j}(\boldsymbol{\beta}_0,u)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,u)}\right\}w^{COX}_{ijk}(u)dM^1_{ijk}(u)\\
&+ \sum_{i=1}^{n}\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\left\{\boldsymbol{Z}_{ijk} - \frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}_0,u)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,u)}\right\}\left\{\widehat{w}^{COX}_{ijk}(u) - w^{COX}_{ijk}(u)\right\}dM^1_{ijk}(u)\\
&+ o_P(n^{1/2}).
\end{aligned}
$$

Using (2), the second term of the last equation of (3) is asymptotically equivalent to

$$
\sum_{i=1}^{n}\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\boldsymbol{q}^{(1)}_{ijk}(t)dM^C_{ijk}(t),
$$

where

$$
\begin{aligned}
\boldsymbol{q}^{(1)}_{ijk}(u) = -\lim_{n_j\to\infty}&\Bigg(\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{s=X_{i'jk'}}^{\tau}\left\{\boldsymbol{Z}_{i'jk'} - \frac{s^{(1)}_{COX,j}(\boldsymbol{\beta}_0,s)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\right\}\\
&\times w^{COX}_{i'jk'}(s)\frac{\exp(\boldsymbol{\gamma}_0^{\intercal}\boldsymbol{Z}_{i'jk'})I(u\le s)}{s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,u)}dM^1_{i'jk'}(s)\\
&+\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\int_{s=X_{i'j'k'}}^{\tau}\left\{\boldsymbol{Z}_{i'j'k'} - \frac{s^{(1)}_{COX,j}(\boldsymbol{\beta}_0,s)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\right\}w^{COX}_{i'j'k'}(s)\\
&\times \boldsymbol{h}^{\intercal}_{Cj}(s,X_{i'j'k'},\boldsymbol{Z}_{i'j'k'})\boldsymbol{\Omega}_C^{-1}\left\{\boldsymbol{Z}_{ijk} - \frac{s^{(1)}_{Cj}(\boldsymbol{\gamma}_0,u)}{s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,u)}\right\}dM^1_{i'j'k'}(s)\Bigg).
\end{aligned}
$$

Therefore, (3) becomes

$$
\frac{1}{\sqrt{n}}\boldsymbol{U}_{COX}(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\boldsymbol{\eta}^{COX}_{i\cdot\cdot} + \boldsymbol{\psi}^{COX}_{i\cdot\cdot}\right) + o_P(1),
$$

where

$$
\boldsymbol{\eta}^{COX}_{i\cdot\cdot} = \sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\left\{\boldsymbol{Z}_{ijk} - \frac{s^{(1)}_{COX,j}(\boldsymbol{\beta}_0,u)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,u)}\right\}w^{COX}_{ijk}(u)dM^1_{ijk}(u),
$$

$$
\boldsymbol{\psi}^{COX}_{i\cdot\cdot} = \sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}\boldsymbol{q}^{(1)}_{ijk}(u)dM^C_{ijk}(u).
$$

Since $\boldsymbol{\eta}_{i..}^{COX}$ and $\boldsymbol{\psi}_{i..}^{COX}$ are i.i.d. zero mean variables, $\boldsymbol{U}_{n,COX}(\boldsymbol{\beta})$ converges to 0 in probability by the law of large numbers. Hence, (C3) is satisfied. Therefore, $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$. Assume, $\boldsymbol{\beta}^*$ lies between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. By the mean value theorem,

$$\boldsymbol{U}_{COX}(\boldsymbol{\beta}_0) = \frac{\partial \boldsymbol{U}_{COX}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0} \mid_{\boldsymbol{\beta}^*} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathcal{I}_1(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

where $\mathcal{I}_1(\boldsymbol{\beta}^*) = -\partial \boldsymbol{U}_{COX}(\boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}_0 \mid_{\boldsymbol{\beta}^*}$. Therefore,

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \{n^{-1}\mathcal{I}_1(\boldsymbol{\beta}^*)\}^{-1} n^{-1/2} \boldsymbol{U}_{COX}(\boldsymbol{\beta}_0).$$

Since $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}_0\| \leq \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|$ implies $\boldsymbol{\beta}^* \xrightarrow{p} \boldsymbol{\beta}_0$, and $n^{-1}\mathcal{I}_1(\boldsymbol{\beta}_0) \to \boldsymbol{\Omega}(\boldsymbol{\beta}_0)$ by (A6),

$$\begin{aligned} n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \{\boldsymbol{\Omega}(\boldsymbol{\beta}_0)\}^{-1} n^{-1/2} \boldsymbol{U}_{COX}(\boldsymbol{\beta}_0) + o_p(1) \\ &= \{\boldsymbol{\Omega}(\boldsymbol{\beta}_0)\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\boldsymbol{\eta}_{i..}^{COX} + \boldsymbol{\psi}_{i..}^{COX} \right) + o_P(1). \end{aligned} \tag{4}$$

By the multivariate central limit theorem, $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\{\boldsymbol{\Omega}(\boldsymbol{\beta}_0)\}^{-1} \boldsymbol{\Sigma} \{\boldsymbol{\Omega}(\beta_0)\}^{-1},$$

where

$$\boldsymbol{\Sigma} = E\left\{ \left(\boldsymbol{\eta}_{1..}^{COX} + \boldsymbol{\psi}_{1..}^{COX} \right)^{\otimes 2} \right\}.$$

The estimator of the covariance matrix is

$$\left\{ \frac{\mathcal{I}_1(\widehat{\boldsymbol{\beta}})}{n} \right\}^{-1} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\boldsymbol{\eta}}_{i..}^{COX} + \widehat{\boldsymbol{\psi}}_{i..}^{COX} \right)^{\otimes 2} \left\{ \frac{\mathcal{I}_1(\widehat{\boldsymbol{\beta}})}{n} \right\}^{-1},$$

where

$$\begin{aligned} \mathcal{I}_1(\widehat{\boldsymbol{\beta}}) &= \sum_{i=1}^{n} \sum_{j=1}^{S} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^{\tau} V_{COX,j}(\widehat{\boldsymbol{\beta}}; t) \widehat{w}_{ijk}^{COX}(t) dN_{ijk}^1(t), \\ V_{COX,j}(\widehat{\boldsymbol{\beta}}; t) &= \frac{S_{COX,j}^{(2)}(\widehat{\boldsymbol{\beta}}, t)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}}, t)} - \left(\frac{S_{COX,j}^{(2)}(\widehat{\boldsymbol{\beta}}, t)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}}, t)} \right)^{\otimes 2}, \\ \widehat{\boldsymbol{\eta}}_{i..}^{COX} &= \sum_{j=1}^{S} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^{\tau} \left\{ \boldsymbol{Z}_{ijk} - \frac{S_{COX,j}^{(1)}(\widehat{\boldsymbol{\beta}}, u)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}}, u)} \right\} \widehat{w}_{ijk}^{COX}(u) d\widehat{M}_{ijk}^1(u), \\ d\widehat{M}_{ijk}^1(t) &= dN_{ijk}^1(t) - \int_0^t Y_{ijk}^1(u) e^{\widehat{\boldsymbol{\beta}}^{\intercal} \boldsymbol{Z}_{ijk}} d\widehat{\Lambda}_{10j}(u), \\ d\widehat{\Lambda}_{10j}(t) &= \frac{1}{n_j} \frac{1}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}}, t)} \sum_{i=1}^{n} \Phi_{ij} \sum_{k=1}^{n_{ij}} \widehat{w}_{ijk}^{COX}(t) dN_{ijk}^1(t), \\ \widehat{\boldsymbol{\psi}}_{i..}^{COX} &= \sum_{j=1}^{S} \Phi_{ij} \sum_{k=1}^{n_{ij}} \int_0^{\tau} \widehat{\boldsymbol{q}}_{ijk}^{(1)}(u) d\widehat{M}_{ijk}^C(u), \end{aligned}$$

where

$$\widehat{\boldsymbol{q}}_{ijk}^{(1)}(u) = -\lim_{n_j\to\infty}\left(\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{s=X_{i'jk'}}^{\tau}\left\{\boldsymbol{Z}_{i'jk'} - \frac{S_{COX,j}^{(1)}(\widehat{\boldsymbol{\beta}},s)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},s)}\right\}\right.$$

$$\times\, w_{i'jk'}^{COX}(s)\frac{\exp\{\widehat{\boldsymbol{\gamma}}^{\intercal}\boldsymbol{Z}_{i'jk'}\}I(u\le s)}{s_{Cj}^{(0)}(\boldsymbol{\gamma}_0,u)}d\widehat{M}_{i'jk'}^{1}(s)$$

$$+\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\int_{s=X_{i'j'k'}}^{\tau}\left\{\boldsymbol{Z}_{i'j'k'} - \frac{S_{COX,j}^{(1)}(\widehat{\boldsymbol{\beta}},s)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},s)}\right\}\widehat{w}_{i'j'k'}^{COX}(s)$$

$$\left.\times\,\widehat{\boldsymbol{h}}_{Cj}^{\intercal}(s,X_{i'j'k'},\boldsymbol{Z}_{i'j'k'})\widehat{\boldsymbol{\Omega}}_C^{-1}\left\{\boldsymbol{Z}_{ijk} - \frac{S_{Cj}^{(1)}(\widehat{\boldsymbol{\gamma}},u)}{S_{Cj}^{(0)}(\widehat{\boldsymbol{\gamma}},u)}\right\}d\widehat{M}_{i'j'k'}^{1}(s)\right),$$

$$\widehat{\boldsymbol{h}}_{Cj}(t,u,\boldsymbol{Z}) = e^{\widehat{\boldsymbol{\gamma}}^{\intercal}\boldsymbol{Z}}\int_{s=u}^{t}\left\{\boldsymbol{Z} - \frac{S_{Cj}^{(1)}(\widehat{\boldsymbol{\gamma}},s)}{S_{Cj}^{(0)}(\widehat{\boldsymbol{\gamma}},s)}\right\}d\widehat{\Lambda}_{C0j}(s),$$

$$\widehat{\boldsymbol{\Omega}}_C = \sum_{i=1}^{n}\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{\tau}V_{Cj}(\widehat{\boldsymbol{\gamma}};t)dN_{ijk}^{C}(t),$$

$$V_{Cj}(\widehat{\boldsymbol{\gamma}},t) = \frac{S_{Cj}^{(2)}(\widehat{\boldsymbol{\beta}},t)}{S_{Cj}^{(0)}(\widehat{\boldsymbol{\beta}},t)} - \left(\frac{S_{Cj}^{(2)}(\widehat{\boldsymbol{\beta}},t)}{S_{Cj}^{(0)}(\widehat{\boldsymbol{\beta}},t)}\right)^{\otimes 2},$$

$$d\widehat{M}_{ijk}^{C}(t) = dN_{ijk}^{C}(t) - \int_0^{t}Y_{ijk}^{C}(u)e^{\widehat{\boldsymbol{\gamma}}^{\intercal}\boldsymbol{Z}_{ijk}}d\widehat{\Lambda}_{C0j}(u),$$

$$d\widehat{\Lambda}_{C0j}(t) = \frac{1}{n_j}\frac{1}{S_{Cj}^{(0)}(\widehat{\boldsymbol{\gamma}},t)}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}dN_{ijk}^{C}(t).$$

Proof of Theorem 2

After some algebra, we have

$$\sqrt{n_j}\left\{\widehat{\Lambda}_{10j}(t) - \Lambda_{10j}(t)\right\}$$

$$= \frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{t}\left\{\frac{1}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},s)} - \frac{1}{S_{COX,j}^{(0)}(\boldsymbol{\beta}_0,s)}\right\}\widehat{w}_{ijk}^{COX}(s)dN_{ijk}^{1}(s) \tag{5}$$

$$+\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^{t}\frac{\widehat{w}_{ijk}^{COX}(s)dM_{ijk}^{1}(s)}{S_{COX,j}^{(0)}(\boldsymbol{\beta}_0,s)}.$$

By the consistency of $\widehat{\boldsymbol{\beta}}$ and (A10), (5) converges to zero in probability uniformly in $t\in[0,\tau]$. Using mean value theorem $\boldsymbol{\beta}^*\in[\widehat{\boldsymbol{\beta}},\boldsymbol{\beta}_0]$,

$$\frac{\partial S_{COX,j}^{(0)}(\boldsymbol{\beta},s)^{-1}}{\partial\boldsymbol{\beta}}\mid_{\boldsymbol{\beta}^*} = \frac{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},s)^{-1} - S_{COX,j}^{(0)}(\boldsymbol{\beta}_0,s)^{-1}}{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0}.$$

Therefore,

$$\frac{1}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},s)} - \frac{1}{S_{COX,j}^{(0)}(\boldsymbol{\beta}_0,s)} = -(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\frac{S_{COX,j}^{(1)}(\boldsymbol{\beta}^*,s)}{S_{COX,j}^{(0)}(\boldsymbol{\beta}^*,s)^2}. \tag{6}$$

Plugging (6) into (5),

$$\begin{aligned}
&\sqrt{n_j}\left\{\widehat{\Lambda}_{10j}(t)-\Lambda_{10j}(t)\right\}\\
&=\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^t -(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}^*,s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}^*,s)^2}\widehat{w}^{COX}_{ijk}(s)dN^1_{ijk}(s)\\
&+\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^t\frac{\widehat{w}^{COX}_{ijk}(s)dM^1_{ijk}(s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}.
\end{aligned}\tag{7}$$

We have $\boldsymbol{\beta}^*\xrightarrow{P}\boldsymbol{\beta}_0$. In addition, $S^{(0)}_{COX,j}(\boldsymbol{\beta}^*,s)$ and $S^{(1)}_{COX,j}(\boldsymbol{\beta}^*,s)$ are bounded away from zero. Therefore, the first term of (7) can be written as

$$\begin{aligned}
&-\int_0^t\left(\frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}_0,s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\right)^{\intercal}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\frac{\widehat{w}^{COX}_{ijk}(s)dN^1_{ijk}(s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\\
&+o_P(1)\\
&=-\int_0^t\left(\frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}_0,s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\right)^{\intercal}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\sqrt{n_j}d\Lambda_{10j}(s)+o_P(1).
\end{aligned}\tag{8}$$

Using (4), (8) can be written as

$$-\int_0^t\left(\frac{S^{(1)}_{COX,j}(\boldsymbol{\beta}_0,s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\right)^{\intercal}\{\boldsymbol{\Omega}(\boldsymbol{\beta}_0)\}^{-1}\frac{\sqrt{n_j}}{n}\sum_{i=1}^{n}\left(\boldsymbol{\eta}^{COX}_{i\cdot\cdot}+\boldsymbol{\psi}^{COX}_{i\cdot\cdot}\right)d\Lambda_{10j}(s)+o_P(1).\tag{9}$$

Now, consider the second term of (7),

$$\begin{aligned}
&\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^t\frac{\widehat{w}^{COX}_{ijk}(s)dM^1_{ijk}(s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\\
&=\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^t\frac{w^{COX}_{ijk}(s)dM^1_{ijk}(s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\\
&+\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^t\frac{\left\{\widehat{w}^{COX}_{ijk}(s)-w^{COX}_{ijk}(s)\right\}dM^1_{ijk}(s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}.
\end{aligned}\tag{10}$$

Using similar arguments as before, (10) becomes

$$\begin{aligned}
&\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^t\frac{w^{COX}_{ijk}(s)dM^1_{ijk}(s)}{S^{(0)}_{COX,j}(\boldsymbol{\beta}_0,s)}\\
&+\frac{1}{\sqrt{n_j}}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^t q^{(2)}_{ijk}(s,t)dM^C_{ijk}(s)+o_P(1),
\end{aligned}\tag{11}$$

where

$$\begin{aligned}
q^{(2)}_{ijk}(s,t)=-\lim_{n_j\to\infty}&\Bigg[\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{\nu=X_{i'jk'}}^{t}\frac{\exp\{\boldsymbol{\gamma}_0^{\intercal}\boldsymbol{Z}_{i'jk'}\}I(s\le\nu)}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,\nu)s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,s)}\\
&\times w^{COX}_{i'jk'}(\nu)dM_{i'jk'}(\nu)+\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\frac{1}{s^{(0)}_{COX,j}(\boldsymbol{\beta}_0,\nu)}\\
&\times\boldsymbol{h}^{\intercal}_{Cj}(\nu,X_{i'j'k'},\boldsymbol{Z}_{i'j'k'})\boldsymbol{\Omega}_C^{-1}\left\{\boldsymbol{Z}_{ijk}-\frac{s^{(1)}_{Cj}(\boldsymbol{\gamma}_0,s)}{s^{(0)}_{Cj}(\boldsymbol{\gamma}_0,s)}\right\}dM^1_{i'j'k'}(\nu)\Bigg].
\end{aligned}$$

Using (9) and (11),

$$\sqrt{n_j}\left\{\widehat{\Lambda}_{10j}(t)-\Lambda_{10j}(t)\right\}=\frac{1}{\sqrt{n_j}}\left\{\sum_{i=1}^{n}\left(W_{ij.}^{(1)}(t)+W_{ij.}^{(2)}(t)\right)\right\}+o_P(1), \tag{12}$$

where

$$W_{ij.}^{(1)}(t)=-\pi_j(\boldsymbol{\eta}_{i..}^{COX}+\boldsymbol{\psi}_{i..}^{COX})^{\intercal}\{\boldsymbol{\Omega}(\boldsymbol{\beta}_0)\}^{-1}\int_0^t\left(\frac{s_{COX,j}^{(1)}(\boldsymbol{\beta}_0,s)}{s_{COX,j}^{(0)}(\boldsymbol{\beta}_0,s)}\right)d\Lambda_{10j}^{COX}(s),$$

$$W_{ij.}^{(2)}(t)=\Phi_{ij}\sum_{k=1}^{n_j}\left\{\int_0^t\frac{w_{ijk}^{COX}(s)dM_{ijk}^{1}(s)}{s_{COX,j}^{(0)}(\boldsymbol{\beta}_0,s)}+\int_0^t q_{ijk}^{(2)}(s,t)dM_{ijk}^{C}(s)\right\}.$$

Because n_{ij} is bounded, by a similar argument to the proof of Theorem 3 of Spiekerman and Lin (1998) using the uniform metric, we can show the tightness of $\sqrt{n_j}\left\{\widehat{\Lambda}_{10j}(t)-\Lambda_{10j}(t)\right\}$. Thus, $\sqrt{n_j}\left\{\widehat{\Lambda}_{10j}(t)-\Lambda_{10j}(t)\right\}$ converges weakly to a mean zero Gaussian process with variance $E\left\{\left(W_{1j.}^{(1)}+W_{2j.}^{(2)}\right)^2\right\}$ which can be estimated by

$$\frac{1}{n_j}\sum_{i=1}^{n}\left\{\widehat{W}_{ij.}^{(1)}(t)+\widehat{W}_{ij.}^{(2)}(t)\right\}^2,$$

where

$$\widehat{W}_{ij.}^{(1)}(t)=-\pi_j(\widehat{\boldsymbol{\eta}}_{i..}^{COX}+\widehat{\boldsymbol{\psi}}_{i..}^{COX})^{\intercal}\left\{\mathcal{I}_1(\widehat{\boldsymbol{\beta}})/n\right\}^{-1}\int_0^t\left(\frac{S_{COX,j}^{(1)}(\widehat{\boldsymbol{\beta}},s)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},s)}\right)d\widehat{\Lambda}_{10j}^{COX}(s),$$

$$\widehat{W}_{ij.}^{(2)}(t)=\Phi_{ij}\sum_{k=1}^{n_{ij}}\left\{\int_0^t\frac{\widehat{w}_{ijk}^{COX}(s)d\widehat{M}_{ijk}^{1}(s)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},s)}+\int_0^t\widehat{q}_{ijk}^{(2)}(s,t)d\widehat{M}_{ijk}^{C}(s)\right\},$$

$$\begin{aligned}\widehat{q}_{ijk}^{(2)}(s,t)=-\lim_{n_j\to\infty}\Bigg[&\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j}}\int_{\nu=X_{i'jk'}}^{t}\frac{\exp\{\widehat{\boldsymbol{\gamma}}^{\intercal}\boldsymbol{Z}_{i'jk'}\}I(s\le\nu)}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},\nu)S_{Cj}^{(0)}(\widehat{\boldsymbol{\gamma}},s)}\\ &\times\widehat{w}_{i'jk'}^{COX}(\nu)d\widehat{M}_{i'jk'}(\nu)+\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{S}\Phi_{i'j'}\sum_{k'=1}^{n_{i'j'}}\frac{1}{S_{COX,j}^{(0)}(\widehat{\boldsymbol{\beta}},\nu)}\\ &\times\widehat{\boldsymbol{h}}_{Cj}^{\intercal}(\nu,X_{i'j'k'},\boldsymbol{Z}_{i'j'k'})(\widehat{\boldsymbol{\Omega}}_C/n)^{-1}\left\{\boldsymbol{Z}_{ijk}-\frac{S_{Cj}^{(1)}(\widehat{\boldsymbol{\gamma}},s)}{S_{Cj}^{(0)}(\widehat{\boldsymbol{\gamma}},s)}\right\}\widehat{w}_{i'jk'}^{COX}(\nu)d\widehat{M}_{i'j'k'}^{1}(\nu)\Bigg].\end{aligned}$$

Proof of Theorem 3

We have

$$\begin{aligned}&\sqrt{n_j}\{\widehat{\Lambda}_{10j}(t)\exp(\widehat{\boldsymbol{\beta}}^{\intercal}\boldsymbol{Z}_0)-\Lambda_{10j}(t)\exp(\boldsymbol{\beta}_0^{\intercal}\boldsymbol{Z}_0)\}\\ &=\sqrt{n_j}\{\widehat{\Lambda}_{10j}(t)\exp(\boldsymbol{\beta}_0^{\intercal}\boldsymbol{Z}_0)-\Lambda_{10j}(t)\exp(\boldsymbol{\beta}_0^{\intercal}\boldsymbol{Z}_0)\}\\ &+\sqrt{n_j}\{\widehat{\Lambda}_{10j}(t)\exp(\widehat{\boldsymbol{\beta}}^{\intercal}\boldsymbol{Z}_0)-\widehat{\Lambda}_{10j}(t)\exp(\boldsymbol{\beta}_0^{\intercal}\boldsymbol{Z}_0)\}.\end{aligned} \tag{13}$$

Consider the first term of (13) and using (12),

$$
\begin{aligned}
&\sqrt{n_j}\exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{Z}_0)\{\widehat{\Lambda}_{10j}(t) - \Lambda_{10j}(t)\} \\
&= \frac{1}{\sqrt{n_j}}\exp(\boldsymbol{\beta}^\intercal \boldsymbol{Z}_0)\left\{\sum_{i=1}^{n}\left(W_{ij.}^{(1)}(t) + W_{ij.}^{(2)}(t)\right)\right\} + o_P(1).
\end{aligned}
$$

Now consider the second term of (13) and using the functional delta method,

$$
\begin{aligned}
&\sqrt{n_j}\{\widehat{\Lambda}_{10j}(t)\exp(\widehat{\boldsymbol{\beta}}^\intercal \boldsymbol{Z}_0) - \widehat{\Lambda}_{10j}(t)\exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{Z}_0)\} \\
&= \sqrt{n_j}\widehat{\Lambda}_{10j}(t)\{\exp(\widehat{\boldsymbol{\beta}}^\intercal \boldsymbol{Z}_0) - \exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{Z}_0)\} \\
&= \sqrt{n_j}\widehat{\Lambda}_{10j}(t)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{Z}_0)\boldsymbol{Z}_0^\intercal\{\boldsymbol{\Omega}(\boldsymbol{\beta}_0)\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\eta}_{i..}^{COX} + \boldsymbol{\psi}_{i..}^{COX}\right) + o_P(1).
\end{aligned}
$$

Therefore, the variance of $\sqrt{n_j}\{\widehat{\Lambda}_{10j}(t)\exp(\widehat{\boldsymbol{\beta}}^\intercal \boldsymbol{Z}_0) - \Lambda_{10j}(t)\exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{Z}_0)\}$ can be estimated by

$$
\begin{aligned}
&\frac{1}{n_j}\left\{\exp(\widehat{\boldsymbol{\beta}}^\intercal \boldsymbol{Z}_0)\right\}^2\sum_{i=1}^{n}\left\{\widehat{W}_{ij.}^{(1)}(t) + \widehat{W}_{ij.}^{(2)}(t) + p_j\widehat{\Lambda}_{10j}(t)\boldsymbol{Z}_0^\intercal\left(\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}})\right)^{-1}\right. \\
&\left.\left(\widehat{\boldsymbol{\eta}}_{i..}^{COX} + \widehat{\boldsymbol{\psi}}_{i..}^{COX}\right)\right\}^2.
\end{aligned}
$$

Hence, using the functional delta method, the estimated variance $\widehat{\Sigma}_{F_{1j}}(t)$ of $\sqrt{n_j}\{\widehat{F}_{1j}(t \mid \boldsymbol{Z}_0) - F_{1j}(t \mid \boldsymbol{Z}_0)\}$ is

$$
\begin{aligned}
&\frac{1}{n_j}\left\{\exp\left(\widehat{\boldsymbol{\beta}}^\intercal \boldsymbol{Z}_0 - \widehat{\Lambda}_{10j}(t)\exp(\widehat{\boldsymbol{\beta}}^\intercal \boldsymbol{Z}_0)\right)\right\}^2\sum_{i=1}^{n}\left\{\widehat{W}_{ij.}^{(1)}(t) + \widehat{W}_{ij.}^{(2)}(t)\right. \\
&\left. + p_j\widehat{\Lambda}_{10j}(t)\boldsymbol{Z}_0^\intercal\left(\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}})\right)^{-1}\left(\widehat{\boldsymbol{\eta}}_{i..}^{COX} + \widehat{\boldsymbol{\psi}}_{i..}^{COX}\right)\right\}^2.
\end{aligned}
$$

Asymptotics using Kaplan-Meier estimates for censoring

The consistency and the asymptotic normality of $\widehat{\boldsymbol{\beta}}^{KM}$ and $\widehat{\Lambda}_{10j}^{KM}(t)$ using $\widehat{w}_{ijk}^{KM}$ can be shown similarly to the proofs of Theorems 1 and 2. Thus, we only provide the final asymptotic formula. We have $n^{1/2}(\widehat{\boldsymbol{\beta}}^{KM} - \boldsymbol{\beta}_0)$ converges to a normal distribution with mean $\mathbf{0}$ and covariance matrix

$$
\{\boldsymbol{\Omega}_{KM}(\boldsymbol{\beta}_0)\}^{-1}\boldsymbol{\Sigma}_{KM}\{\boldsymbol{\Omega}_{KM}(\beta_0)\}^{-1},
$$

where

$$
\boldsymbol{\Sigma}_{KM} = \sum_{i=1}^{n}E\left\{\left(\boldsymbol{\eta}_{i..}^{KM} + \boldsymbol{\psi}_{i..}^{KM}\right)^{\otimes 2}\right\}.
$$

The estimate of the covariance matrix is

$$
\left\{\frac{\mathcal{I}_{KM}(\widehat{\boldsymbol{\beta}}^{KM})}{n}\right\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{\eta}}_{i..}^{KM} + \widehat{\boldsymbol{\psi}}_{i..}^{KM}\right)^{\otimes 2}\left\{\frac{\mathcal{I}_{KM}(\widehat{\boldsymbol{\beta}}^{KM})}{n}\right\}^{-1},
$$

where

$$\mathcal{I}_{KM}(\widehat{\boldsymbol{\beta}}^{KM}) = \sum_{i=1}^{n}\sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^\tau V_{KM,j}(\widehat{\boldsymbol{\beta}}^{KM},t)\widehat{w}_{ijk}^{KM}(t)dN_{ijk}^1(t),$$

$$V_{KM,j}(\widehat{\boldsymbol{\beta}}^{KM},t) = \frac{S_{KM,j}^{(2)}(\widehat{\boldsymbol{\beta}}^{KM},t)}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}}^{KM},t)} - \left(\frac{S_{KM,j}^{(2)}(\widehat{\boldsymbol{\beta}}^{KM},t)}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}}^{KM},t)}\right)^{\otimes 2},$$

$$S_{KM,j}^{(r)}(\widehat{\boldsymbol{\beta}}^{KM},t) = \frac{1}{n_j}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\widehat{w}_{ijk}^{KM}(t)Y_{ijk}^1(t)\boldsymbol{Z}_{ijk}^{\otimes r}\exp(\widehat{\boldsymbol{\beta}}^{KM\intercal}\boldsymbol{z}_{ijk}), \quad r=0,1,2,$$

$$\widehat{\boldsymbol{\eta}}_{i..}^{KM} = \sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^\tau\left\{\boldsymbol{Z}_{ijk} - \frac{S_{KM,j}^{(1)}(\widehat{\boldsymbol{\beta}}^{KM},u)}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}}^{KM},u)}\right\}\widehat{w}_{ijk}^{KM}(u)d\widehat{M}_{KM,ijk}^1(u),$$

$$d\widehat{M}_{KM,ijk}^1(t) = dN_{ijk}^1(t) - \int_0^t Y_{ijk}^1(u)e^{\widehat{\boldsymbol{\beta}}^{KM\intercal}\boldsymbol{z}_{ijk}}d\widehat{\Lambda}_{10j}^{KM}(u),$$

$$d\widehat{\Lambda}_{10j}^{KM}(t) = \frac{1}{n_j}\frac{1}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}},t)}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\widehat{w}_{ijk}^{KM}(t)dN_{ijk}^1(t),$$

$$\widehat{\boldsymbol{\psi}}_{i..}^{KM} = \sum_{j=1}^{S}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_0^\tau\frac{\widehat{\boldsymbol{q}}_{KM,j}^{(1)}(u)}{\zeta_j(u)}d\widehat{M}_{KM,ijk}^C(u),$$

$$\widehat{\boldsymbol{q}}_{KM,j}^{(1)}(u) = -\lim_{n_j\to\infty}\left(\frac{1}{n_j}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}\int_{s=X_{ijk}}^{\tau}\left\{\boldsymbol{Z}_{ijk} - \frac{S_{KM,j}^{(1)}(\widehat{\boldsymbol{\beta}}^{KM},s)}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}}^{KM},s)}\right\}\right.$$
$$\left.\widehat{w}_{i'jk'}^{KM}(t)d\widehat{M}_{KM,i'jk'}^1(s)I(s\geq u>T_{i'jk'})\right),$$

$$d\widehat{M}_{KM,ijk}^C(t) = dN_{ijk}^C(t) - \int_0^t Y_{ijk}^1(u)d\widehat{\Lambda}_{KM,10j}^C(u),$$

$$\zeta_j(u) = \lim_{n\to\infty}\frac{1}{n_j}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}Y_{ijk}^1(u),$$

$$d\widehat{\Lambda}_{C0j}(t) = \frac{1}{n_j}\frac{1}{\zeta_j(u)}\sum_{i=1}^{n}\Phi_{ij}\sum_{k=1}^{n_{ij}}dN_{ijk}^C(t).$$

Now, we provide the asymptotic formula for cumulative baseline hazard function. $\widehat{\Lambda}_{10j}^{KM}(t)$ is a consistent estimator for $\Lambda_{10j}^{KM}(t)$ for $t\in[0,\tau]$ and $\sqrt{n_j}\{\widehat{\Lambda}_{10j}^{KM}(t) - \Lambda_{10j}^{KM}(t)\}$ converges weakly to a zero mean Gaussian process with the variance $\Sigma_{\Lambda_{10j}^{KM}}(t) = E\left\{\left(W_{KM,ij.}^{(1)}(t) + W_{KM,ij.}^{(2)}(t)\right)^2\right\}$, where

$$W_{KM,ij.}^{(1)}(t) = -\pi_j(\boldsymbol{\eta}_{i..}^{KM} + \boldsymbol{\psi}_{i..}^{KM})^\intercal\boldsymbol{\Omega}_{KM}(\boldsymbol{\beta}_0)^{-1}\int_0^t\left(\frac{s_{KM,j}^{(1)}(\boldsymbol{\beta}_0,s)}{s_{KM,j}^{(0)}(\boldsymbol{\beta}_0,s)}\right)$$
$$d\Lambda_{10j}^{KM}(s),$$

$$W_{KM,ij.}^{(2)}(t) = \Phi_{ij}\sum_{k=1}^{n_{ij}}\left\{\int_0^t\frac{w_{ijk}^{KM}(s)dM_{KM,ijk}^1(s)}{s_{KM,j}^{(0)}(\boldsymbol{\beta}_0,s)}\right.$$
$$\left.+\int_0^t q_{KM,j}^{(2)}(s,t)dM_{KM,ijk}^C(s)\right\}.$$

The variance can be estimated by

$$\widehat{\Sigma}_{\Lambda_{10j}^{KM}}(t) = \frac{1}{n_j}\sum_{i=1}^{n}\left\{\widehat{W}_{KM,ij.}^{(1)}(t) + \widehat{W}_{KM,ij.}^{(2)}(t)\right\}^2,$$

where

$$\begin{aligned}
\widehat{W}_{KM,ij.}^{(1)}(t) &= -\pi_j(\widehat{\boldsymbol{\eta}}_{i..}^{KM} + \widehat{\boldsymbol{\psi}}_{i..}^{KM})^\intercal(\mathcal{I}_{KM}(\widehat{\boldsymbol{\beta}}^{KM})/n)^{-1}\int_0^t\left(\frac{S_{KM,j}^{(1)}(\widehat{\boldsymbol{\beta}}^{KM}, s)}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}}^{KM}, s)}\right) \\
&\quad d\widehat{\Lambda}_{10j}^{KM}(s), \\
\widehat{W}_{KM,ij.}^{(2)}(t) &= \Phi_{ij}\sum_{k=1}^{n_{ij}}\left\{\int_0^t \frac{\widehat{w}_{ijk}^{KM}(s)d\widehat{M}_{KM,ijk}^{1}(s)}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}}^{KM}, s)}\right. \\
&\quad \left. + \int_0^t \widehat{q}_{KM,j}^{(2)}(s,t)d\widehat{M}_{KM,ijk}^{C}(s)\right\}, \\
\widehat{q}_{KM,j}^{(2)}(s,t) &= \lim_{n_j\to\infty}\left[-\frac{1}{n_j}\sum_{i'=1}^{n}\Phi_{i'j}\sum_{k'=1}^{n_{i'j'}}\frac{\widehat{w}_{i'jk'}^{KM}(\nu)}{S_{KM,j}^{(0)}(\widehat{\boldsymbol{\beta}}^{KM}, \nu)}d\widehat{M}_{i'jk'}^{1}(\nu)\right].
\end{aligned}$$

References

Soyoung Kim, Yayun Xu, Mei-Jie Zhang, and Kwang-Woo Ahn. Stratified proportional subdistribution hazards model with covariate-adjusted censoring weight for case-cohort studies. *Scandinavian Journal of Statistics*, 47(4):1222–1242, 2020.

Robert V Foutz. On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72(357):147–148, 1977.

C. F. Spiekerman and D. Y. Lin. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93(443):1164–1175, 1998.