Received: 17 May 2021 | Revised: 29 August 2022 | Accepted: 31 August 2022

DOI: 10.1002/sim.9573

RESEARCH ARTICLE

Statistics in Medicine WILEY

Marginal semiparametric transformation models for clustered multivariate competing risks data

Yizeng He[1] | **Soyoung Kim**[1] | **Lu Mao**[2] | **Kwang Woo Ahn**[1]

[1]Division of Biostatistics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[2]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence
Kwang Woo Ahn, Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI, USA.
Email: kwooahn@mcw.edu

Funding information
Advancing a Healthier Wisconsin Endowment, Grant/Award Number: 5520569; National Cancer Institute, Grant/Award Number: U24CA076518; National Heart, Lung, and Blood Institute, Grant/Award Number: R01HL149875

Multivariate survival models are often used in studying multiple outcomes for right-censored data. However, the outcomes of interest often have competing risks, where standard multivariate survival models may lead to invalid inferences. For example, patients who had stem cell transplantation may experience multiple types of infections after transplant while reconstituting their immune system, where death without experiencing infections is a competing risk for infections. Such competing risks data often suffer from cluster effects due to a matched pair design or correlation within study centers. The cumulative incidence function (CIF) is widely used to summarize competing risks outcomes. Thus, it is often of interest to study direct covariate effects on the CIF. Most literature on clustered competing risks data analyses is limited to the univariate proportional subdistribution hazards model with inverse probability censoring weighting which requires correctly specifying the censoring distribution. We propose a marginal semiparametric transformation model for multivariate competing risks outcomes. The proposed model does not require modeling the censoring distribution, accommodates nonproportional subdistribution hazards structure, and provides a platform for joint inference of all causes and outcomes.

KEYWORDS
competing risks data, multivariate outcome, semiparametric transformation model

1 | INTRODUCTION

The need for multivariate outcome model for right-censored data arises when studying multiple correlated outcomes. For example, the Busselton Health Study[1,2] aims to evaluate the association between bivariate survival outcomes (coronary heart disease and stroke) and their risk factors. Such multivariate right-censored outcomes often have competing risks. Ballen et al[3] studied fungal and bacterial infections of acute leukemia patients who had hematopoietic cell transplantation (HCT). In their study, patients experienced multiple types of infections after transplant. Due to receiving extensive treatment procedures and the disease severity, patients may die before getting an infection. Thus, infections and death prior to infections are competing risks. We refer to such right-censored data with competing risks and multiple outcomes of interests as multivariate competing risks data in this article. Multivariate competing risks data occur when a subject can experience multiple time-to-event outcomes simultaneously, but each outcome has multiple competing causes.[4] Using multivariate survival models for multivariate competing risks data may lead to biased results due to informative censoring. When studying multiple infections after HCT, clinicians are often interested in evaluating drugs or treatments on

preventing patients from experiencing multiple infections. Thus, jointly modeling multiple infections would reveal the difference in treatment effects on various infections. In addition, competing risks data often have a cluster effect due to correlation within study centers,[3] matched pairs,[5] and/or families. Failing to adjust for a cluster effect leads to invalid statistical inference.

There is a rich literature on "univariate" competing risks analysis, which evaluates risk factors for a single outcome having competing causes, for example, fungal infection and death without experiencing a fungal infection in HCT studies. Prentice et al[6] proposed a proportional cause-specific hazards model. It estimates covariate effects on the rate of experiencing a specific cause for a patient who has not experienced any causes. It can also evaluate internal time-dependent covariate effects. However, it does not evaluate direct covariate effects on the cumulative incidence function (CIF), which is widely used to summarize competing risks outcomes. Thus, its analysis results are not necessarily consistent with cumulative incidence curves, which may cause confusion among clinicians. Therefore, estimating direct covariate effects on the CIF is often of interest. To do so, the Fine-Gray model, a direct binomial modeling, and a pseudo-value approach have been proposed for independent or clustered competing risks data.[7-13] These methods require modeling of the censoring distribution to compute the inverse probability censoring weight. Thus, if the censoring distribution is misspecified, the statistical inference could be invalid. In addition, joint inference on multiple causes is not provided. To overcome these limitations, Mao and Lin[14] proposed an efficient semiparametric transformation model based on the nonparametric maximum likelihood estimator for independent competing risks data. Their model directly examines covariate effects on the CIFs of all causes, does not require modeling of the censoring distribution, and jointly models all causes. Besides, it accommodates general transformation models including both proportional and nonproportional subdistribution hazards structures as special cases. Bellach et al[15] proposed a different semiparametric transformation model to directly model the CIF, but like many existing methods, it requires modeling of the censoring distribution and does not provide joint inference on multiple causes.

Despite a rich literature on the univariate competing risks analysis, the literature on joint modeling for multivariate competing risks data remains limited. Wohlfahrt et al[4] used a Poisson regression with piecewise constant baseline rates to fit cause-specific hazard rates for multivariate competing risks data. However, their Poisson regression model cannot accommodate for clustered multivariate competing risks data.

In this article, we propose marginal semiparametric transformation models to address the modeling need for clustered multivariate competing risks data, motivated by Ballen et al.[3] We generalize the semiparametric transformation models of Mao and Lin[14] to clustered multivariate competing risks data. Under the independence working assumption, we construct a pseudo-likelihood to estimate the model parameters. Like the generalized estimating equations (GEE),[16] the proposed method is valid and robust even when the independence working assumption does not hold. We propose a sandwich estimator for the covariance matrix of parameter estimates. We then conduct simulation studies to assess the performance of the proposed models, followed by the analysis of the motivating clinical study by Ballen et al.[3]

2 | METHOD

2.1 | Data, model, and pseudo-likelihood

Assume we have H different types of outcomes indexed by $h \in \{1, \dots, H\}$. Assume there are K causes $\epsilon \in \{1, \dots, K\}$ for each outcome type without loss of generality. For the infection study of Ballen et al,[3] $H = 2$: (i) fungal infection and death prior to fungal infection and (ii) bacterial infection and death prior to bacterial infection. And for each infection, there are two competing events ($K = 2$). Assume there are n clusters and cluster i has n_i individuals, with each individual indexed by $j \in \{1, \dots, n_i\}$. We assume K is fixed and n_i's are bounded. Let T_{ijh}, C_{ijh}, ϵ_{ijh}, and $\mathbf{Z}_{ijh} = (Z_{ijh1}, \dots, Z_{ijhp})^T$ be the event time, censoring time, cause of failure, and time-independent or time-dependent external covariate of individual j in cluster i for outcome type h, respectively, for $i = 1, \dots, n, j = 1, \dots, n_i$, and $h = 1, \dots, H$. We suppress $\mathbf{Z}_{ijh}$'s potential time dependence for simplicity.

Let $\mathbf{T}_i = \{T_{ijh}, j = 1, \dots, n_i, h = 1, \dots, H\}, \mathbf{C}_i = \{C_{ijh}, j = 1, \dots, n_i, h = 1, \dots, H\}, \boldsymbol{\epsilon}_i = \{\epsilon_{ijh}, j = 1, \dots, n_i, h = 1, \dots, H\}$, and $\mathbf{Z}_i = \{\mathbf{Z}_{ijh}, j = 1, \dots, n_i, h = 1, \dots, H\}$. Assume $(\mathbf{T}_i, \boldsymbol{\epsilon}_i, \mathbf{C}_i, \mathbf{Z}_i)$ are independent and identically distributed. We assume the $(\mathbf{T}_i, \boldsymbol{\epsilon}_i)$'s are independent of the $\mathbf{C}_i$'s given $\mathbf{Z}_i$ for $i = 1, \dots, n$. Let $X_{ijh} = T_{ijh} \wedge C_{ijh}$ be the observed time and $\Delta_{ijh} = I(T_{ijh} \le C_{ijh})\epsilon_{ijh}$, where $a \wedge b = \min(a, b)$. Let $F_{kh}(t|\mathbf{Z}_{ijh}) = P(T_{ijh} \le t, \epsilon_{ijh} = k|\mathbf{Z}_{ijh})$ be the cumulative incidence of cause k at time t given $\mathbf{Z}_{ijh}$ for outcome type h. Define the subdistribution hazard of cause k for outcome type h given

variable $\mathbf{Z}_{ijh}$ to be $\lambda_{kh}(t|\mathbf{Z}_{ijh}) = -d[\log\{1 - F_{kh}(t|\mathbf{Z}_{ijh})\}]/dt$. Then, $F_{kh}(t|\mathbf{Z}_{ijh}) = 1 - \exp\{-\Lambda_{kh}(t|\mathbf{Z}_{ijh})\}$, where $\Lambda_{kh}(t|\mathbf{Z}_{ijh}) = \int_0^t \lambda_{kh}(u|\mathbf{Z}_{ijh})du$. Similarly to Mao and Lin,[14] we consider

$$\Lambda_{kh}(t|\mathbf{Z}_{ijh}) = G_{kh}\left\{\int_0^t \exp(\boldsymbol{\beta}_{kh}^T\mathbf{Z}_{ijh})d\Lambda_{kh}(u)\right\}, \tag{1}$$

where $\boldsymbol{\beta}_{kh} = (\beta_{kh1}, \ldots, \beta_{khp})^T$ is a parameter vector, G_{kh} is a known increasing function, and $\Lambda_{kh}(t)$ is an arbitrary increasing function. The $\Lambda_{kh}(t)$'s are infinite dimensional parameters. The class of Box-Cox transformations or the class of logarithmic transformations can be candidates for G_{kh},[17] where the proportional subdistribution hazards model and the proportional subdistribution odds model are special cases by choosing $G_{kh}(x) = x$ and $G_{kh}(x) = \log(1 + x)$, respectively.

Define $\boldsymbol{\beta} = (\boldsymbol{\beta}_{11}^T, \ldots, \boldsymbol{\beta}_{KH}^T)^T$ and $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_{11}^T, \ldots, \boldsymbol{\Lambda}_{KH}^T)^T$, where $\boldsymbol{\Lambda}_{kh}$ consists of $\Lambda_{kh}(t)$'s. Under an independence working assumption, we propose the following pseudo-likelihood function for clustered multivariate competing risks data:

$$L(\boldsymbol{\beta}, \Lambda) = \prod_{i=1}^{n}\prod_{j=1}^{n_i}\prod_{h=1}^{H}\left[\prod_{k=1}^{K} F'_{kh}(X_{ijh}|\mathbf{Z}_{ijh}, \boldsymbol{\beta}, \boldsymbol{\Lambda})^{I(\Delta_{ijh}=k)}\left\{1 - \sum_{k=1}^{K} F_{kh}(X_{ijh}|\mathbf{Z}_{ijh}, \boldsymbol{\beta}, \boldsymbol{\Lambda})\right\}^{I(\Delta_{ijh}=0)}\right], \tag{2}$$

where $f'(x) = df(x)/dx$ for a function f.

2.2 | Algorithm

To obtain the nonparametric maximum pseudo-likelihood estimator (NPMPLE), we treat Λ_{kh} as a right continuous step function with jump size $\Lambda_{kh}\{t\}$ at time $t \in [0, \tau]$. Then, maximizing (2) with respect to $\boldsymbol{\beta}$ and $\Lambda_{kh}\{X_{ij}\}$ for $\Delta_{ijh} = k$, $k = 1, \ldots, K$, $h = 1, \ldots, H$ can be implemented for each outcome type through optimization algorithms as described in Zeng and Lin.[17] For time-independent covariates, we can use the explicit algorithm of Mao and Lin[14] and apply it to each outcome type. To elaborate the algorithm, let $t_{kh1} < \cdots < t_{khq_{kh}}$ be the distinct failure times of cause k within outcome type h, where q_{kh} is the number of cause k events within outcome type h. Denote $d_{khr} = \Lambda_{kh}\{t_{khr}\}$ for $r = 1, \ldots, q_{kh}, k = 1, \ldots, K$, and $h = 1, \ldots, H$. Let $t_{h1} < \cdots < t_{hq}$ be the distinct failure times regardless of causes with $t_{h,q+1} = \infty$ and let $\delta_{h1}, \ldots, \delta_{hq}$ and $d_{h1}, \ldots, d_{hq}$ be the corresponding causes and jump sizes, respectively, for outcome type h.

Define $\Lambda_{khr} = \sum_{a=1}^{r} I(\delta_{ha} = k)d_{ha}$ and $\Lambda_{(khr)} = \sum_{a=1}^{r} d_{kha}$. Denote $H_{kh} = \log(G'_{kh}) - G_{kh}$ and $\mathbf{Z}_{(khr)}$ the covariate vector for the subject having the kth cause of failure of outcome type h at time t_{khr}. Once we take the derivative of the pseudo-log-likelihood of (2) with respect to $d_{kh,r}$ and set that derivative to zero, we obtain the following relationship:

$$\begin{aligned} d_{kh,r+1}^{-1} = d_{kh,r}^{-1} &+ e^{\boldsymbol{\beta}_{kh}^T\mathbf{Z}_{(khr)}}H'_{kh}(e^{\boldsymbol{\beta}_{kh}^T\mathbf{Z}_{(khr)}}\Lambda_{(khr)}) \\ &- \sum_{t_{khr}<t_a<t_{kh,r+1}}\ \sum_{t_{ha}\le X_{ijh}<t_{h,a+1}} I(C_{ijh} < T_{ijh})\frac{\exp\{\boldsymbol{\beta}_{kh}^T\mathbf{Z}_{ijh} + H_{kh}(e^{\boldsymbol{\beta}_{kh}^T\mathbf{Z}_{ijh}}\Lambda_{kha})\}}{\sum_{k=1}^{K}\exp\{-G_{kh}(e^{\boldsymbol{\beta}_{kh}^T\mathbf{Z}_{ijh}}\Lambda_{kha})\} - K + 1}. \end{aligned} \tag{3}$$

Let $\alpha_{kh} = d_{kh1}$, $\boldsymbol{\alpha} = (\alpha_{11}, \ldots, \alpha_{KH})^T$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$. Define the pseudo-log-likelihood function $\ell(\boldsymbol{\beta}, \Lambda) = \log L(\boldsymbol{\beta}, \Lambda)$. Then, $\widehat{\Lambda}(\boldsymbol{\theta}) = \text{argmax}_{\Lambda}\ell(\boldsymbol{\theta}, \Lambda)$ can be obtained by using Algorithm (3) for each outcome. The pseudo-profile log-likelihood $\ell_p(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}, \widehat{\Lambda}(\boldsymbol{\theta}))$ can be computed as well. The Newton-Raphson algorithm can be used to obtain the NPMPLE of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$. We set the initial values of $\boldsymbol{\beta}$ and α_{kh} to $\mathbf{0}$ and $1/q_{kh}$, respectively. The proposed algorithm worked proficiently in our simulation studies in Section 3.

2.3 | Covariance estimator

Because the proposed pseudo-likelihood is not the true likelihood, one cannot use the inverse matrix of the information matrix for $\boldsymbol{\beta}$ and nonzero d_{kr}'s to estimate the covariance matrix of their estimators. Similarly, one cannot use the

negative second derivative of the pseudo-profile log-likelihood function to estimate the covariance matrix of $\widehat{\boldsymbol{\theta}}$. Here, we propose a sandwich variance estimator to estimate the covariance matrix of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Lambda}}$. To obtain the sandwich variance estimator, we treat the score equation based on the first derivative of the log-pseudo-likelihood function as the estimating equation similarly to GEE. Thus, it is robust to true correlation structure of outcomes and clusters misspecification.[18] Let the first and the second derivative of $\ell(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$ be $\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ and $\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$, respectively. We also define the first and the second derivative of the pseudo-profile log-likelihood, $\ell_p(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ as $\mathbf{U}_p(\boldsymbol{\theta})$ and $\mathbf{I}_p(\boldsymbol{\theta})$, respectively. Let $\text{Cov}(A)$ be the covariance matrix of $\mathbf{A}$. To estimate the covariance matrix of the parameter estimates, we propose the following sandwich variance estimators for the pseudo-log-likelihood and the pseudo-profile log-likelihood:

$$\{\mathbf{I}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})\}^{-1}\widehat{\text{Cov}}\{\mathbf{U}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\}\{\mathbf{I}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})\}^{-1},\ \{\mathbf{I}_p(\widehat{\boldsymbol{\theta}})\}^{-1}\widehat{\text{Cov}}\{\mathbf{U}_p(\boldsymbol{\theta}_0)\}\{\mathbf{I}_p(\widehat{\boldsymbol{\theta}})\}^{-1},$$

respectively, where $\boldsymbol{\beta}_0$, $\boldsymbol{\Lambda}_0$, and $\boldsymbol{\theta}_0$ are the true parameters of $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\theta}$, respectively. The first derivatives $\mathbf{U}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)$ and $\mathbf{U}_p(\boldsymbol{\theta}_0)$ are the sum of n independent mean zero terms. To estimate $\text{Cov}\{\mathbf{U}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\}$ and $\text{Cov}\{\mathbf{U}_p(\boldsymbol{\theta}_0)\}$, define $\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\Lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{h=1}^{H}\mathbf{U}_{ijh}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ and $\mathbf{U}_p(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{h=1}^{H}\mathbf{U}_{ijh,p}(\boldsymbol{\theta})$. Write $\mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\Lambda}) = \sum_{j=1}^{n_i}\sum_{h=1}^{H}\mathbf{U}_{ijh}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ and $\mathbf{U}_{i,p}(\boldsymbol{\theta}) = \sum_{j=1}^{n_i}\sum_{h=1}^{H}\mathbf{U}_{ijh,p}(\boldsymbol{\theta})$. Then, $\widehat{\text{Cov}}\{\mathbf{U}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\}$ and $\widehat{\text{Cov}}\{\mathbf{U}_p(\boldsymbol{\theta}_0)\}$ can be obtained by $\sum_{i=1}^{n}\mathbf{U}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})\mathbf{U}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})^T$ and $\sum_{i=1}^{n}\mathbf{U}_{i,p}(\widehat{\boldsymbol{\theta}})\mathbf{U}_{i,p}(\widehat{\boldsymbol{\theta}})^T$, respectively.

2.4 | Asymptotic property

In this section, we study the asymptotic properties of the proposed estimator. We impose the following regularity conditions:

Condition 1. The true value of the p-dimensioned $\boldsymbol{\beta}$, lies in the interior of a compact subset of the Euclidean space $\mathbb{R}^p$. The true value of $\boldsymbol{\Lambda}_{kh}$, denoted as $\boldsymbol{\Lambda}_{kh0}$, is continuously differentiable with $\boldsymbol{\Lambda}'_{kh0} > 0$ on $[0, \tau]$ for some constant $\tau > 0$.

Condition 2. The components of $\mathbf{Z}(.)$ are uniformly bounded and have bound total variation with probability 1. If $\boldsymbol{\beta}_{kh}^{\intercal}\mathbf{Z}(t) = d(t)$ almost surely for some constant function d for all $t \in [0, \tau]$, then $\boldsymbol{\beta}_{kh} = 0$ and $d(t) = 0$ for $k = 1, \ldots, K$, and $h = 1, \ldots, H$.

Condition 3. There is a constant δ_0 such that $\Pr(T_h \geq \tau|\mathbf{Z}) \geq \delta_0 > 0$ and $\Pr(C_h \geq \tau|\mathbf{Z}) = \Pr(C_h = \tau|\mathbf{Z}) \geq \delta_0 > 0$ with probability 1 for $h = 1, \ldots, H$.

Condition 4. The function G_{kh} is four-times differentiable with $G_{kh}(0) = 0$ and $G'_{kh}(x) > 0$. And for any $c_0 > 0$, $\limsup_{x\to\infty}\left[\{G_{kh}(c_0 x)\}^{-1}\log\left\{x\sup_{y\leq x}G'_{kh}(y)\right\}\right] = 0$.

Conditions 1 to 4 are similar to Conditions 1 to 4 of Mao and Lin.[14] We extend Conditions 1 to 4 of Mao and Lin[14] to our multivariate competing risks setting. Conditions 1 to 3 are standard conditions for right-censored data analysis. Condition 4 is satisfied by both the Box-Cox and logarithmic transformations (see Mao and Lin[14] for more details).

Theorem 1 (Consistency). *Under Conditions 1 to 4, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Lambda}}_{kh}(k = 1, \ldots, K, h = 1, \ldots, H)$ are strongly consistent. In other words,*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \sup_{t\in[0,\tau]}\sum_{k=1}^{K}\sum_{h=1}^{H}|\widehat{\boldsymbol{\Lambda}}_{kh}(t) - \boldsymbol{\Lambda}_{kh0}(t)| \to 0$$

almost surely, where $\|.\|$ denotes the Euclidean norm.

Theorem 2 (Asymptotic normality). *Under the regularity conditions, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}_0)$ converges weakly to a zero-mean Gaussian process.*

Theorem 2 implies $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $\sqrt{n}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}_0)$ converge to a multivariate zero-mean normal distribution and a multivariate zero-mean Gaussian process on $[0, \tau]^{\otimes KH}$: the KH-produce space of $[0, \tau]$. Confidence intervals for β_{kh} and $\Lambda_{kh}(t)$ can be obtained using the normal approximation with the sandwich variance estimator. Theorem 1 can be shown similarly to Mao and Lin[14] due to the independence working assumption. Thus, we omit its proof. The proof of Theorem 2 is provided in the Appendix.

3 | SIMULATIONS

In this section, we assess the performance of the proposed methods by extensive Monte Carlo (MC) simulation studies. Consider $H = 2$ outcome types, each with $K = 2$ competing risks. For each outcome type, we generate clustered competing risks data using the positive stable frailty model:[13]

$$F_{1h}(t|\omega, \mathbf{Z}) = 1 - \{1 - p(1 - e^{-t})\}^{\omega \exp(\boldsymbol{\gamma}_h^{\intercal}\mathbf{Z})}, \quad F_{2h}(t|\omega, \mathbf{Z}) = (1-p)^{\omega \exp(2)}(1 - e^{-t}),$$

where $0 < p < 1$ and the frailty ω follows the positive stable distribution with parameter $\alpha = 1$ for independent data, and 0.5 for clustered data. Due to the shared ω among outcome types and subjects within a cluster, all causes and subjects within a cluster are correlated for $\alpha = 0.5$. We generate n clusters, $n = 200, 400, 800$, and each cluster has $L = 2$ subjects. Two cluster-level covariates $\mathbf{Z} = (Z_1, Z_2)^{\intercal}$ are considered, where Z_1 and Z_2 follow uniform $[0, 1]$ and the Bernoulli distribution with success probability 0.5, respectively. Because of the Laplace transformations of the standard positive stable distribution, the true parameters are rescaled according to $\boldsymbol{\beta}_{kh} = \alpha\boldsymbol{\gamma}_{kh}$.[13] For $\alpha = 0.5$, we set $\boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = (1, 0)^{\intercal}$ for cause 1 of both outcome types. For $\alpha = 1$, we set $\boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = (0.5, 0)^{\intercal}$ for cause 1 of both outcome types. Thus, $\boldsymbol{\beta}_{11} = \boldsymbol{\beta}_{12} = (0.5, 0)^{\intercal}$ regardless of α values. The model for cause 2 does not depend on any covariates, so the true parameter vector for cause 2 is $\boldsymbol{\beta}_{21} = \boldsymbol{\beta}_{22} = (0, 0)^{T}$ for both outcome types. We choose $p = 0.2$ to ensure $F_{1h}(\infty|w, \mathbf{Z}) + F_{2h}(\infty|w, \mathbf{Z}) < 1$ regardless of the values of ω and Z. To guarantee that both cause 1 and 2 follow the subdistribution hazards model, we generate a latent event cause 3 with the probability $1 - F_{1h}(\infty|w, \mathbf{Z}) - F_{2h}(\infty|w, \mathbf{Z})$, and set the event times to infinity so that all observations with cause 3 are censored. We generate censoring time for each cluster from a positive stable frailty proportional hazard model,[19]

$$\lambda_c(t|\omega, \mathbf{Z}) = \omega_c \lambda_{0c}(t) \exp(\boldsymbol{\gamma}_c^{\intercal}\mathbf{Z}),$$

where $\boldsymbol{\gamma_c} = (3, 2)^{\intercal}$ and $\lambda_{0c} = 10^{-4}$ for both outcome types. The cluster frailty ω_c follows the positive stable distribution with $\alpha_c = 0.5$ and 1 for clustered and independent data, respectively. For clustered data ($\alpha = 0.5$), the overall probabilities for cause 1, cause 2, and censoring are approximately 37%, 40%, 23%, respectively for each outcome type. For independent data ($\alpha = 1$), the overall probabilities for cause 1, cause 2, and censoring are approximately 21%, 16%, and 63%, respectively for outcome types 1 and 2.

For each simulation, we report the average parameter estimates, average standard errors (SE), and coverage rates of the 95% confidence interval from 500 MC simulations. For the proposed method, we explore two modeling structures: treating the data as clustered and independent. The coverage rate and SE for clustered modeling and independent modeling are denoted as coverage$_c$ and SE$_c$, and coverage$_I$ and SE$_I$, respectively.

In the first scenario, we consider clustered data with $\alpha = 0.5$ and compare the proposed method to the proportional subdistribution hazards (PSH) model of Zhou et al[8] when the models for both causes are correctly specified. We use R package "crrSC"[20] to fit Zhou et al.[8] Because Zhou et al[8] was designed for analyzing one cause of a single outcome type, we fit four separate models of Zhou et al[8] for outcome types 1 and 2. Zhou et al[8] is limited to the covariate-independent censoring. We average over the 500 MC simulations for the mean parameter estimates.

To specify the models for each cause, we use the proposed semiparametric transformation model (1) to jointly model outcome types 1 and 2 using the form of the proportional subdistribution hazard model,[7] $G_{kh}(x) = x$, for cause 1 in both outcome types, and proportional odds model,[21] $G_{kh}(x) = \log(1 + x)$, for cause 2 in both outcome types. We report the simulation results in Table 1. While the estimates of Zhou et al[8] are biased due to covariate-dependent censoring, the estimates from the proposed method are approximately unbiased. In addition, the coverage rates for the 95% confidence intervals of the proposed method estimator after accounting for clusters are closer to 0.95 than those for the proposed method treating the data as independent data and that of Zhou et al.[8]

In addition, we examine the performance of the proposed model on independent data ($\alpha = 1$) while assuming the clustered structure of data. We compare the proposed model to fitting four separate PSH models and report these results in Table 1. The estimates of the proposed models treating the data as clustered and independent data are approximately unbiased and their coverage rates are around 0.95 for all parameters. In particular, the coverage rates of the proposed methods treating the data as clustered and independent data are close. The estimates of the PSH model are biased. The standard errors produced by the proposed models are only slightly larger than those by the PSH model.

Next, we examine the robustness of the proposed model under misspecified models for cause 1 in both outcome types. Specifically, we fit the proposed method using the proportional odds model for cause 1 and proportional subdistribution

TABLE 1 Comparing the estimated coefficients, standard error (SE) and coverage rates of the 95% confidence interval of the proposed method with Zhou et al[8] and PSH[7] for clustered ($\alpha = 0.5$) and independent data ($\alpha = 1$), respectively, from 500 iterations

			Outcome type 1				Outcome type 2			
			β_{111}	β_{112}	β_{211}	β_{212}	β_{121}	β_{122}	β_{221}	β_{222}
	n	**True value**	**0.5**	**0**	**0**	**0**	**0.5**	**0**	**0**	**0**
$\alpha = 0.5$	200	Proposed	0.507	−0.003	0.043	−0.013	0.493	0.004	0.013	−0.026
		Coverage_C	(0.933)	(0.919)	(0.943)	(0.943)	(0.939)	(0.933)	(0.917)	(0.955)
		SE_C	0.345	0.197	0.491	0.280	0.347	0.198	0.493	0.281
		Coverage_I	(0.855)	(0.846)	(0.879)	(0.909)	(0.865)	(0.873)	(0.871)	(0.917)
		SE_I	0.277	0.158	0.421	0.241	0.279	0.159	0.423	0.241
		Zhou et al	0.461	−0.020	−0.119	−0.096	0.451	−0.015	−0.132	−0.109
		Coverage	(0.947)	(0.931)	(0.945)	(0.935)	(0.949)	(0.947)	(0.923)	(0.945)
		SE	0.353	0.202	0.431	0.249	0.353	0.202	0.431	0.249
	400	Proposed	0.504	0.003	−0.008	−0.009	0.503	0.004	0.005	−0.017
		Coverage_C	(0.936)	(0.944)	(0.942)	(0.946)	(0.930)	(0.952)	(0.942)	(0.952)
		SE_C	0.245	0.140	0.348	0.199	0.245	0.140	0.348	0.199
		Coverage_I	(0.862)	(0.856)	(0.908)	(0.906)	(0.872)	(0.870)	(0.894)	(0.902)
		SE_I	0.196	0.112	0.298	0.170	0.196	0.112	0.298	0.170
		Zhou et al	0.475	−0.015	−0.138	−0.093	0.473	−0.015	−0.128	−0.101
		Coverage	(0.936)	(0.928)	(0.924)	(0.902)	(0.932)	(0.954)	(0.906)	(0.914)
		SE	0.248	0.142	0.304	0.175	0.248	0.142	0.305	0.175
	800	Proposed	0.501	0.005	−0.005	−0.005	0.495	0.006	−0.004	−0.009
		Coverage_C	(0.918)	(0.944)	(0.936)	(0.940)	(0.926)	(0.942)	(0.930)	(0.966)
		SE_C	0.173	0.099	0.245	0.140	0.173	0.099	0.246	0.140
		Coverage_I	(0.842)	(0.872)	(0.882)	(0.892)	(0.850)	(0.886)	(0.882)	(0.940)
		SE_I	0.138	0.079	0.210	0.120	0.138	0.079	0.210	0.120
		Zhou et al	0.473	−0.014	−0.137	−0.090	0.466	−0.013	−0.133	−0.094
		Coverage	(0.916)	(0.948)	(0.898)	(0.882)	(0.922)	(0.952)	(0.872)	(0.904)
		SE	0.175	0.100	0.214	0.124	0.175	0.100	0.215	0.124
$\alpha = 1$	200	Proposed	0.484	−0.004	0.10	−0.001	0.509	−0.010	0.028	−0.004
		Coverage_C	(0.948)	(0.962)	(0.944)	(0.944)	(0.950)	(0.946)	(0.950)	(0.950)
		SE_C	0.378	0.215	0.474	0.270	0.381	0.216	0.474	0.269
		Coverage_I	(0.946)	(0.968)	(0.948)	(0.946)	(0.956)	(0.952)	(0.946)	(0.952)
		SE_I	0.381	0.216	0.477	0.271	0.384	0.217	0.476	0.270
		PSH	0.447	−0.027	−0.043	−0.032	0.473	−0.033	−0.026	−0.036
		Coverage	(0.930)	(0.960)	(0.942)	(0.942)	(0.952)	(0.950)	(0.952)	(0.952)
		SE	0.380	0.216	0.432	0.247	0.383	0.217	0.430	0.245
	400	Proposed	0.490	−0.009	0.014	0.006	0.514	−0.007	0.022	0.008
		Coverage_C	(0.934)	(0.962)	(0.948)	(0.956)	(0.946)	(0.950)	(0.936)	(0.930)
		SE_C	0.267	0.152	0.335	0.190	0.268	0.152	0.335	0.190
		Coverage_I	(0.938)	(0.962)	(0.950)	(0.954)	(0.946)	(0.954)	(0.944)	(0.932)
		SE_I	0.268	0.152	0.336	0.190	0.269	0.153	0.336	0.190
		PSH	0.453	−0.033	−0.038	−0.025	0.477	−0.030	−0.032	−0.025
		Coverage	(0.928)	(0.968)	(0.948)	(0.944)	(0.944)	(0.946)	(0.942)	(0.930)
		SE	0.267	0.152	0.304	0.173	0.268	0.153	0.304	0.173
	800	Proposed	0.483	−0.008	0.017	−0.005	0.510	−0.003	0.014	0.000
		Coverage_C	(0.934)	(0.960)	(0.932)	(0.944)	(0.956)	(0.954)	(0.962)	(0.926)
		SE_C	0.189	0.108	0.236	0.134	0.189	0.107	0.237	0.134
		Coverage_I	(0.932)	(0.954)	(0.942)	(0.934)	(0.952)	(0.934)	(0.944)	(0.922)
		SE_I	0.190	0.108	0.237	0.134	0.190	0.107	0.237	0.134
		PSH	0.446	−0.031	−0.035	−0.036	0.473	−0.026	−0.040	−0.032
		Coverage	(0.936)	(0.954)	(0.944)	(0.934)	(0.954)	(0.934)	(0.944)	(0.926)
		SE	0.189	0.108	0.214	0.122	0.189	0.107	0.215	0.122

Note: For the proposed model, we report results from treating the data as clustered (Coverage_C and SE_C) and independent (Coverage_I and SE_I). The proposed model while treating the data as clustered and independent share the same estimated coefficients.

TABLE 2 Parameter estimate, standard error (SE), and coverage rate of the 95% confidence interval for cause 2 based on the proposed model or Zhou et al with misspecified model for cause 1 for both outcome types

		Outcome type 1		Outcome type 2	
		β_{211}	β_{212}	β_{221}	β_{222}
n	**True value**	**0**	**0**	**0**	**0**
200	Proposed	0.041	−0.021	0.030	−0.026
	Coverage_C	(0.913)	(0.932)	(0.907)	(0.952)
	SE_C	0.419	0.240	0.420	0.241
	Coverage_I	(0.873)	(0.901)	(0.861)	(0.911)
	SE_I	0.359	0.206	0.361	0.206
	Zhou et al	−0.122	−0.097	−0.135	−0.107
	Coverage	(0.944)	(0.932)	(0.924)	(0.946)
	SE	0.431	0.250	0.431	0.249
400	Proposed	0.006	−0.014	0.016	−0.020
	Coverage_C	(0.944)	(0.954)	(0.946)	(0.952)
	SE_C	0.298	0.171	0.299	0.171
	Coverage_I	(0.904)	(0.910)	(0.892)	(0.912)
	SE_I	0.255	0.146	0.256	0.146
	Zhou et al	−0.138	−0.093	−0.128	−0.101
	Coverage	(0.924)	(0.902)	(0.906)	(0.914)
	SE	0.304	0.175	0.305	0.175
800	Proposed	0.011	−0.010	0.012	−0.013
	Coverage_C	(0.942)	(0.940)	(0.934)	(0.970)
	SE_C	0.210	0.121	0.211	0.121
	Coverage_I	(0.870)	(0.902)	(0.880)	(0.940)
	SE_I	0.180	0.103	0.180	0.103
	Zhou et al	−0.137	−0.090	−0.133	−0.094
	Coverage	(0.898)	(0.882)	(0.872)	(0.904)
	SE	0.214	0.124	0.215	0.124

Note: We report results from treating the data as clustered (Coverage_C and SE_C) and independent (Coverage_I and SE_I).

model for cause 2 for both outcome types under the same settings of the first scenario ($\alpha = 0.5$). Thus, cause 1 is not correctly specified. Since cause 2 is not covariate-dependent, the models for cause 2 are still correctly specified. We report the parameter estimates, standard errors, and coverage rates of the correctly specified cause, namely, β_{211} and β_{212} for cause 2 of outcome type 1, and β_{221} and β_{222} for cause 2 of outcome type 2. As shown in Table 2, the bias of the parameter estimates of the proposed method for cause 2 after accounting for clustered data is small for both outcomes, and their coverage rates are approximately around 0.95. We also report the results from Zhou et al.[8] Their standard errors are in general higher than the proposed model's and the coverage rates depart from 0.95 especially as the sample size increases. These results suggest that the proposed methods are robust against model misspecification under this limited simulation setting.

We conduct additional simulations when both causes follow the PSH model with nonzero parameters. For each outcome type, we generate clustered competing risks data using the following PSH model with positive stable frailty:[13]

$$F_{1h}(t|\omega,\mathbf{Z}) = 1 - \{1 - p(1 - e^{-t})\}^{\omega \exp(\boldsymbol{\gamma}_{1h}^{\mathrm{T}}\mathbf{Z})}, \quad F_{2h}(t|\omega,\mathbf{Z}) = 1 - \{1 - p(1 - e^{-t})\}^{\omega \exp(\boldsymbol{\gamma}_{2h}^{\mathrm{T}}\mathbf{Z})}, \quad 0 < p \le 1.$$

TABLE 3 Comparing the estimated coefficients, standard error (SE) and coverage rates of the 95% confidence interval of the proposed method with Zhou et al[8] and PSH[7] for clustered ($\alpha = 0.5$) and independent data ($\alpha = 1$), respectively, from 500 iterations

			Outcome type 1				Outcome type 2			
			β_{111}	β_{112}	β_{211}	β_{211}	β_{121}	β_{122}	β_{221}	β_{222}
n		**True value**	**0.300**	**0.050**	**0.150**	**0.100**	**0.200**	**−0.100**	**0.050**	**−0.100**
$\alpha = 0.5$	250	Proposed	0.322	0.005	0.106	0.049	0.151	−0.036	0.063	−0.159
		Coverage_C	(0.936)	(0.954)	(0.952)	(0.934)	(0.952)	(0.932)	(0.944)	(0.936)
		SE_C	1.470	1.465	1.489	1.478	1.495	1.485	1.510	1.503
		Coverage_I	(0.892)	(0.914)	(0.890)	(0.872)	(0.876)	(0.860)	(0.878)	(0.884)
		SE_I	1.203	1.199	1.218	1.210	1.221	1.213	1.237	1.228
		Zhou et al	0.100	−0.144	−0.121	−0.107	−0.068	−0.185	−0.156	−0.310
		Coverage	(0.932)	(0.956)	(0.952)	(0.930)	(0.936)	(0.928)	(0.936)	(0.922)
		SE	1.472	1.471	1.491	1.484	1.493	1.490	1.511	1.511
	500	Proposed	0.318	0.054	0.086	0.046	0.215	−0.129	0.041	−0.177
		Coverage_C	(0.960)	(0.952)	(0.964)	(0.944)	(0.952)	(0.936)	(0.950)	(0.952)
		SE_C	1.041	1.033	1.055	1.046	1.056	1.049	1.073	1.063
		Coverage_I	(0.880)	(0.894)	(0.894)	(0.872)	(0.884)	(0.878)	(0.884)	(0.884)
		SE_I	0.849	0.844	0.860	0.854	0.862	0.855	0.873	0.865
		Zhou et al	0.093	−0.204	−0.138	−0.107	−0.004	−0.274	−0.225	−0.325
		Coverage	(0.960)	(0.950)	(0.960)	(0.942)	(0.944)	(0.940)	(0.944)	(0.944)
		SE	1.042	1.037	1.057	1.053	1.055	1.055	1.074	1.069
	1000	Proposed	0.290	0.051	0.061	0.079	0.170	−0.090	0.040	−0.149
		Coverage_C	(0.950)	(0.944)	(0.958)	(0.944)	(0.958)	(0.938)	(0.928)	(0.954)
		SE_C	0.733	0.727	0.744	0.738	0.746	0.741	0.759	0.752
		Coverage_I	(0.906)	(0.888)	(0.900)	(0.880)	(0.892)	(0.882)	(0.876)	(0.896)
		SE_I	0.598	0.594	0.606	0.601	0.606	0.602	0.616	0.610
		Zhou et al	0.064	−0.144	−0.169	−0.072	−0.049	−0.232	−0.222	−0.295
		Coverage	(0.940)	(0.934)	(0.946)	(0.946)	(0.938)	(0.932)	(0.916)	(0.938)
		SE	0.735	0.731	0.747	0.743	0.747	0.745	0.761	0.757
$\alpha = 1$	250	Proposed	0.360	0.071	0.163	0.183	0.167	−0.060	−0.062	−0.080
		Coverage_C	(0.950)	(0.932)	(0.944)	(0.936)	(0.934)	(0.950)	(0.958)	(0.944)
		SE_C	1.472	1.462	1.497	1.492	1.502	1.483	1.522	1.510
		Coverage_I	(0.954)	(0.928)	(0.948)	(0.950)	(0.940)	(0.948)	(0.954)	(0.952)
		SE_I	1.481	1.473	1.508	1.496	1.514	1.492	1.528	1.514
		PSH	0.165	−0.060	−0.034	0.047	−0.020	−0.189	−0.256	−0.205
		Coverage	(0.952)	(0.936)	(0.946)	(0.952)	(0.942)	(0.948)	(0.956)	(0.946)
		SE	1.478	1.472	1.507	1.498	1.513	1.494	1.525	1.517
	500	Proposed	0.288	0.052	0.153	0.119	0.190	−0.108	−0.031	−0.097
		Coverage_C	(0.950)	(0.932)	(0.958)	(0.954)	(0.944)	(0.942)	(0.956)	(0.948)
		SE_C	1.042	1.038	1.058	1.053	1.063	1.052	1.076	1.068
		Coverage_I	(0.950)	(0.934)	(0.958)	(0.954)	(0.942)	(0.946)	(0.954)	(0.942)
		SE_I	1.045	1.040	1.062	1.053	1.068	1.055	1.079	1.069
		PSH	0.091	−0.077	−0.044	−0.013	0.007	−0.233	−0.219	−0.224
		Coverage	(0.950)	(0.932)	(0.964)	(0.952)	(0.948)	(0.944)	(0.946)	(0.928)
		SE	1.042	1.040	1.061	1.053	1.067	1.056	1.076	1.071
	1000	Proposed	0.318	0.056	0.141	0.099	0.191	−0.130	0.011	−0.122
		Coverage_C	(0.958)	(0.954)	(0.954)	(0.938)	(0.954)	(0.952)	(0.948)	(0.942)
		SE_C	0.737	0.734	0.748	0.744	0.752	0.746	0.762	0.756
		Coverage_I	(0.954)	(0.952)	(0.954)	(0.930)	(0.956)	(0.954)	(0.950)	(0.944)
		SE_I	0.739	0.735	0.749	0.743	0.754	0.747	0.762	0.756
		PSH	0.122	−0.073	−0.058	−0.034	0.007	−0.254	−0.175	−0.248
		Coverage	(0.942)	(0.944)	(0.938)	(0.922)	(0.950)	(0.954)	(0.942)	(0.944)
		SE	0.737	0.735	0.748	0.743	0.753	0.747	0.760	0.757

Note: For the proposed model, we report results from treating the data as clustered (Coverage_C and SE_C) and independent (Coverage_I and SE_I). The proposed model while treating the data as clustered and independent share the same estimated coefficients.

TABLE 4 Sensitivity analysis for the estimated coefficients, standard error (SE) and coverage rates of the 95% confidence interval for clustered ($\alpha = 0.5$) from 500 iterations

			Outcome type 1		Outcome type 2	
			β_{111}	β_{112}	β_{121}	β_{122}
	n	**True value**	**0.3**	**0.05**	**0.2**	**−0.1**
$\alpha = 0.5$	250	Proposed	0.283	−0.014	0.115	−0.058
		Coverage_C	(0.930)	(0.950)	(0.948)	(0.928)
		SE_C	1.443	1.438	1.469	1.460
		Zhou et al	0.100	−0.144	−0.068	−0.185
		Coverage	(0.932)	(0.956)	(0.936)	(0.928)
		SE	1.472	1.471	1.493	1.490
	500	Proposed	0.280	0.076	0.180	−0.152
		Coverage_C	(0.960)	(0.946)	(0.948)	(0.936)
		SE_C	1.021	1.013	1.037	1.030
		Zhou et al	0.093	−0.204	−0.004	−0.274
		Coverage	(0.960)	(0.950)	(0.944)	(0.940)
		SE	1.042	1.037	1.055	1.055
	1000	Proposed	0.255	0.042	0.137	−0.112
		Coverage_C	(0.940)	(0.932)	(0.954)	(0.938)
		SE_C	0.713	0.707	0.730	0.725
		Zhou et al	0.064	−0.144	−0.049	−0.232
		Coverage	(0.940)	(0.934)	(0.938)	(0.932)
		SE	0.735	0.731	0.747	0.745

Note: For the proposed model, we report results from treating the data as clustered (Coverage_C and SE_C).

All causes and subjects within a cluster are correlated for $\alpha = 0.5$ as previous. We generate n clusters, $n = 250, 500, 1000$, and each cluster has $L = 4$ subjects. Two cluster-level covariates $\mathbf{Z} = (Z_1, Z_2)^\intercal$ are considered, where Z_1 and Z_2 follow uniforms $[0.1, 0.3]$ and $[0, 0.2]$, respectively. We set $\boldsymbol{\beta}_{11} = (0.3, 0.05)^\intercal$ and $\boldsymbol{\beta}_{12} = (0.2, -0.1)^\intercal$ for cause 1 of outcome type 1 and outcome type 2, respectively. For cause 2, we set $\boldsymbol{\beta}_{21} = (0.15, 0.1)^\intercal$ and $\boldsymbol{\beta}_{22} = (0.05, -0.1)^\intercal$ for outcome type 1 and outcome type 2, respectively. Following Logan et al,[13] we set $\boldsymbol{\gamma} = \boldsymbol{\beta}/\alpha$. Event times are generated in the same manner as in Table 1. We generate censoring time from the proportional hazards model with positive stable frailty with $\boldsymbol{\gamma}_c = (6, 4)^\intercal$ and $\lambda_{0c} = 0.1$ for both outcome types. For clustered data ($\alpha = 0.5$), the overall probabilities for cause 1, cause 2, and censoring are approximately 15%, 15%, 70%, respectively for each outcome type. For independent data ($\alpha = 1$), the overall probabilities for cause 1, cause 2, and censoring are approximately 14%, 14%, and 72% for both outcome types. Table 3 provides the results. While the estimates of the proposed method are approximately unbiased, those from Zhou et al are biased. The coverage rates of the propose method for clustered data get closer to 95% as the sample sizes grow. The coverage rates for the proposed method assuming independent data are well below 95%. Under the same data generation setting, we examine the robustness of the proposed model, where we use the PSH model for cause 1, but the proportional odds model for cause 2. Thus, the model for cause 2 is misspecified. Table 4 shows the results for cause 1. While the estimates of the proposed method are approximately unbiased, Zhou et al suffers from large biases. The coverage rates of the proposed method are closer to 95% compared with those of Zhou et al. Last, under the same setting with $\alpha = 0.5$ for Table 4, we estimate $F_1(t|Z_1 = 0.2, Z_2 = 0.1)$ and $F_2(t|Z_1 = 0.2, Z_2 = 0.1)$ at two time points. We choose 0.125 and 0.25 quantiles of all event times as the two time points t_1 and t_2, respectively, based on an MC simulation study with 10^5 replicates. Table 5 summarizes the bias. Although the model for cause 2 is misspecified, the estimates of the cumulative incidence for both cases are approximately unbiased for the proposed method, which shows the robustness of the proposed method. However, Zhou et al results in larger biases compared with the proposed method.

TABLE 5 Bias for marginal cumulative incidence rates at two time points with $\alpha = 0.5$ with 500 iterations

	Cause 1				Cause 2			
	$F_1(t_1\|z)$		$F_1(t_2\|z)$		$F_2(t_1\|z)$		$F_2(t_2\|z)$	
n	Proposed	Zhou et al	Proposed	Zhou et al	Proposed	Zhou et al	Proposed	Zhou et al
250	0.007	−0.014	0.008	−0.027	0.007	0.006	0.008	−0.015
500	0.003	0.016	0.003	0.028	0.002	−0.010	0.003	−0.020
1000	0.001	−0.016	0.001	−0.028	0.001	−0.011	0.001	−0.022

TABLE 6 Patient characteristics of Ballen et al[3] data

		UCB	8/8 MUD	7/8 MMUD
Conditioning regimen	MA	175	189	62
	NMA/RIC	340	573	186
Year of transplant	2008-2009	250	524	185
	2010-2011	265	238	63
Disease	AML	355	613	199
	ALL	160	149	49
ATG	No	404	500	120
	Yes	111	262	128
	Total	515	762	248

4 | DATA ANALYSIS

To illustrate the proposed method, we apply the proposed method to the dataset from Ballen et al[3] to jointly analyze two infection outcome types: bacterial and fungal infections. Ballen et al[3] explores the effect of alternative graft sources among patients without a matched sibling donor to received potentially curative HCT. The analysis for our application includes 1525 acute myeloid leukemia (AML) or acute lymphoid leukemia (ALL) patients ≥16 years old in the first or second complete remission without missing bacterial or fungal infection status. All patients received a transplant with a single or double unrelated umbilical cord blood (UCB) transplantation, a matched unrelated donor (MUD), or a single antigen/allele mismatched unrelated donor (MMUD) between 2008 and 2011. The study includes patients who received either myeloablative (MA) or non-myeloablative/reduced intensity conditioning (NMA/RIC) regimens. We consider 5 variables: donor source, conditioning regimen intensity, year of transplant, disease type, and antithymocyte globulin (ATG) use (yes vs no). The detailed frequencies of the covariates are provided in Table 6. There were 1289 patients with an bacterial infection and 437 patients with a fungal infection. Three hundred and eighty three patients experienced both infections. There were 906 deaths, where 106 deaths were without a bacterial infection, 593 deaths were without a fungal infection, and 73 deaths were without any infections. One of the primary objectives of Ballan et al[3] is to compare the incidences of bacterial and fungal infections after HCT among different donor types. They concluded that bacterial and fungal infections are more common among patients who received UCB or MMUD than MUD, and bacterial infections are more common among patients who received UCB than MMUD. The dataset includes patients from 149 transplant centers with significant transplant center effects for bacterial (P-value = 0.0005) and fungal (P-value < 0.0001) infections according to the score test of homogeneity.[22] The original study fitted the proportional cause-specific hazards model and did not consider the cluster effects in their analysis. We use the marginal proportional hazard model[23] to examine if the censoring distribution depends on covariates. For both outcome types, the censoring distribution depends on some covariates. Specifically, the censoring distribution depends on donor source and year of transplant for bacterial infection outcome type, and on donor source, conditioning regimen, and year of transplant for fungal infection outcome type at a significance level 0.05.

We analyze the data using the proposed methods under the proportional subdistribution hazards model for both infection types and death without infections. As a comparison, we also include the analysis results from the proposed method treating data as independent data and the PSH model of Zhou et al.[8] The proposed model simultaneously evaluates

TABLE 7 Results from multivariate semiparametric transformation model with clusters: Subdistribution hazard ratio (SHR) of bacterial and fungal infections

		Bacterial infection		Fungal infection	
	Variable	**SHR (95% CI)**	***P*-value**	**SHR (95% CI)**	***P*-value**
Proposed$_C$	*Donor source*		0.004*		0.096*
	UCB	1		1	
	MMUD	0.815 (0.665-0.998)	0.048	0.858 (0.579-1.273)	0.448
	MUD	0.777 (0.669-0.902)	0.001	0.744 (0.560-0.988)	0.041
	Conditioning regimen				
	MA	1		1	
	NMA/RIC	1.392 (1.227-1.579)	<0.001	0.967 (0.807-1.159)	0.719
	Year of transplant				
	2008-2009	1		1	
	2010-2011	0.893 (0.787-1.012)	0.077	0.977 (0.799-1.193)	0.816
	Disease				
	AML	1		1	
	ALL	1.054 (0.901-1.233)	0.514	1.038 (0.838-1.286)	0.732
	ATG, yes	0.849 (0.746-0.967)	0.014	1.052 (0.843-1.313)	0.656
Proposed$_I$	*Donor source*		<0.001*		0.022*
	UCB	1		1	
	MMUD	0.815 (0.683-0.972)	0.023	0.858 (0.642-1.148)	0.303
	MUD	0.777 (0.683-0.883)	<0.001	0.744 (0.603-0.919)	0.006
	Conditioning regimen				
	MA	1		1	
	NMA/RIC	1.392 (1.224-1.584)	<0.001	0.967 (0.783-1.195)	0.758
	Year of transplant				
	2008-2009	1		1	
	2010-2011	0.893 (0.794-1.004)	0.058	0.977 (0.794-1.200)	0.821
	Disease				
	AML	1		1	
	ALL	1.054 (0.924-1.202)	0.436	1.038 (0.838-1.286)	0.732
	ATG, yes	0.849 (0.751-0.961)	0.010	1.052 (0.855-1.294)	0.634
Zhou et al[8]	*Donor source*		0.003*		0.193*
	UCB	1		1	
	MMUD	0.803 (0.655-0.983)	0.007	0.889 (0.603-1.311)	0.552
	MUD	0.768 (0.660-0.895)	0.001	0.783 (0.590-1.038)	0.090
	Conditioning regimen				
	MA	1		1	
	NMA/RIC	1.395 (1.225-1.589)	<0.001	1.029 (0.854-1.241)	0.761
	Year of transplant				
	2008-2009	1		1	
	2010-2011	0.913 (0.804-1.037)	0.162	0.946 (0.775-1.153)	0.581
	Disease				
	AML	1		1	
	ALL	1.037 (0.880-1.221)	0.666	1.040 (0.824-1.312)	0.741
	ATG, yes	0.846 (0.744-0.961)	0.010	1.028 (0.829-1.274)	0.802

*Overall *P*-values.

TABLE 8 Results from multivariate semiparametric transformation model with clusters: Subdistribution hazard ratio of death without an infection for each outcome type

		Death without bacterial infection		Death without fungal infection	
	Variable	**SHR (95% CI)**	***P*-value**	**SHR (95% CI)**	***P*-value**
Proposed$_C$	*Donor source*		0.109*		0.002*
	UCB	1		1	
	MMUD	2.365 (1.047-5.339)	0.038	0.907 (0.609-1.349)	0.629
	MUD	1.346 (0.742-2.443)	0.328	0.645 (0.488-0.851)	0.002
	Conditioning regimen				
	MA	1		1	
	NMA/RIC	0.415 (0.228-0.756)	0.004	1.023 (0.821-1.275)	0.838
	Year of transplant				
	2008-2009	1		1	
	2010-2011	1.147 (0.687-1.916)	0.600	0.957 (0.780-1.173)	0.670
	Disease				
	AML	1		1	
	ALL	1.267 (0.684-2.349)	0.452	1.020 (0.793-1.311)	0.877
	ATG, yes	1.068 (0.558-2.046)	0.842	0.781 (0.590-1.035)	0.086
Proposed$_I$	*Donor source*		0.101*		0.001*
	UCB	1		1	
	MMUD	2.365 (1.068-5.237)	0.034	0.907 (0.655-1.255)	0.555
	MUD	1.346 (0.708-2.560)	0.364	0.645 (0.505-0.823)	<0.001
	Conditioning regimen				
	MA	1		1	
	NMA/RIC	0.415 (0.233-0.740)	0.003	1.023 (0.802-1.305)	0.854
	Year of transplant				
	2008-2009	1		1	
	2010-2011	1.147 (0.668-1.971)	0.619	0.957 (0.760-1.204)	0.705
	Disease				
	AML	1		1	
	ALL	1.267 (0.692-2.320)	0.443	1.020 (0.797-1.306)	0.875
	ATG, yes	1.068 (0.591-1.930)	0.827	0.781 (0.609-1.002)	0.052
Zhou et al[8]	*Donor source*		0.317*		0.005*
	UCB	1		1	
	MMUD	1.699 (0.727-3.969)	0.221	0.931 (0.617-1.404)	0.733
	MUD	1.040 (0.531-2.037)	0.909	0.665 (0.496-0.891)	0.006
	Conditioning regimen				
	MA	1		1	
	NMA/RIC	0.572 (0.303-1.081)	0.086	1.093 (0.858-1.393)	0.471
	Year of transplant				
	2008-2009	1		1	
	2010-2011	1.188 (0.676-2.088)	0.549	0.958 (0.779-1.178)	0.682
	Disease				
	AML	1		1	
	ALL	1.458 (0.706-3.009)	0.308	1.022 (0.773-1.352)	0.877
	ATG, yes	1.153 (0.607-2.189)	0.664	0.747 (0.559-0.998)	0.048

*Overall *P*-values.

covariate effects on not only bacterial and fungal infections, but also on their competing risks (death without bacterial infection and death without fungal infection). We report the analysis results for two infections in Table 7 and those for death without each of the two infections in Table 8. Despite of covariate-dependent censoring, the parameter estimates and P-values of the proposed method for clustered data (Proposed$_C$) and Zhou et al[8] are similar. This is likely due to relatively small magnitude of parameter estimates for the censoring distribution. Thus, they make little impact on difference in estimates between the proposed method and Zhou et al.[8] The 95% confidence intervals of the proposed method treating the data as independent data (Proposed$_I$) are slightly narrower than those of the proposed method treating the data as clustered data (Proposed$_C$). As a result, while the overall P-value for donor type of Proposed$_I$ for fungal infection is less than 0.05 (P-value = 0.022), that of Proposed$_C$ is greater than 0.05 (P-value = 0.096).

This reanalysis takes into account the significant cluster effects that were ignored in the original analysis. Moreover, the multivariate analysis also allows us to compare covariate effects across the two infections. For example, the P-value from comparing UCB and MMUD for bacterial infection is 0.048. Although it is less than 0.05, it could be a false positive given that we have conducted multiple comparisons. Let $\beta_{\text{mmud,bac}}$ and $\beta_{\text{mmud,fun}}$ be the parameters for MMUD of bacterial infection and fungal infection, respectively. We investigate whether $(\beta_{\text{mmud,bac}}, \beta_{\text{mmud,fun}})^\intercal = (0, 0)^\intercal$ and obtain P-value = 0.135. This suggests the finding on MMUD when compared with UCB in bacterial infection could be a false alarm. The multivariate outcome model reveals the covariate effects on the CIF of death without bacterial/fungal infection as in Table 8. In contrast with the analysis results for infections in Table 7, the impact of donor source on the CIF of death without infections is quite different between death without bacterial infection and death without fungal infection. In addition, the direction of the subdistribution hazards ratios for conditioning regimen intensity in the bacterial infection in Table 7 (1.392) and in death without bacterial infection in Table 8 (0.415) is opposite. Thus, compared with patients who received MAC, those who received NMA/RIC are less likely to suffer from deaths prior to bacterial infection, but more likely to experience bacterial infection. Higher death rates for those receiving MAC are somewhat expected given that MAC is more toxic and difficult to tolerate than NMA/RIC.[24] On the other hand, MAC and NMA/RIC make little impact on experiencing fungal infection and death without fungal infection.

5 | DISCUSSION

In contrast with the widely used Zhou et al,[8] the proposed semiparametric transformation model does not rely on the correct specification of the censoring distribution and allows for comparison of covariate effects across multivariate outcomes. In addition, it is more flexible than Zhou et al[8] due to using transformation of cumulative subdistribution hazards. In HCT studies, infections prior to death are clinically well studied, but death without infections is largely ignored in the analysis.[3,25] We have evaluated covariates effects on the CIF of both infections prior to death and death without infections by fitting the multivariate model. It has revealed the benefit of using MAC over NMA/RIC for bacterial infection, but using MAC may lead to more deaths without bacterial infection due to its toxicity. The proposed marginal model has been developed under the independence working assumption to handle clusters and multiple outcome types. Thus, it is robust in that misspecified models in one outcome type do not affect the covariate effects of the other outcome types.

Bellach et al[15] proposed another semiparametric transformation model for independent competing risks data. In contrast with Mao and Lin,[14] their model fits each cause in a separate model. Thus, a joint inference across all causes is infeasible. Like Zhou et al,[8] their model requires estimating the censoring distribution, and their asymptotics is limited to covariate-independent censoring. Nonetheless, because they fit each cause separately, misspecified models for one cause do not affect the inference on the other causes. On the other hand, misspecified models for some causes may yield incorrect inference on all causes in the proposed method and Mao and Lin[14] although the limited simulation studies of Mao and Lin[14] and Section 3 showed robustness against model misspecification. Therefore, extending Bellach et al[15] to clustered competing risks data and covariate-dependent censoring would be an interesting research problem in the future, where covariate-dependent censoring can be modeled using the marginal proportional hazards model.[23] Robustness against misspecified models for censoring also needs to be investigated.

The proposed method estimates the marginal effects on the CIF. Developing methods with random effects to evaluate conditional effects on the CIF for multivariate competing risks data would be worthy of investigation. In this case, one could consider two types of random effects: random effects to account for clusters and correlation among multivariate outcomes. Compared with marginal models, random effects models may better reveal the correlation structure within a cluster or between outcomes. When data have multiple types of clusters such as both matched pairs and transplant centers, multiple random effects can be included in the model. This would add more flexibility compared with the marginal model.

The proposed model includes the PSH model and the proportional odds model as a special case. However, choosing an appropriate transformation could be challenging in practice. In this case, one can treat η from the following Box-Cox transformation or the logarithmic transformation as an additional covariate:

$$\frac{(1+x)^{\eta}-1}{\eta} \quad \text{or} \quad \frac{1}{\eta}\log(1+\eta x), \quad \eta \geq 0.$$

However, this makes the interpretation of parameter estimates difficult depending on the estimate for η because of its nontrivial transformation. Nonetheless, it may result in a more accurate CIF prediction compared with when using PSH model or the proportional odds model.

Although this article studied the semiparametric transformation model for the subdistribution hazards model, one can develop a semiparametric transformation model for the cause-specific hazards model. More specifically, consider the following cause-specific hazard function:

$$\Lambda_{*kh}(t|\mathbf{Z}_{ijh}) = G_{kh}\left\{\int_0^t \exp(\boldsymbol{\beta}_{*kh}^T\mathbf{Z}_{ijh})d\Lambda_{*kh}(u)\right\},$$

where $\boldsymbol{\beta}_{*kh} = (\beta^*_{*kh1}, \ldots, \beta_{*khp})^T$ is a parameter vector, G_{kh} is a known increasing function, and $\Lambda_{*kh}(t)$ is an arbitrary increasing function. Then, the pseudo-likelihood function for clustered multivariate competing risks data is

$$\prod_{i=1}^{n}\prod_{j=1}^{n_i}\prod_{h=1}^{H}\left\{\prod_{k=1}^{K}\Lambda'_{*kh}(X_{ijh}|\mathbf{Z}_{ijh})^{I(\Delta_{ijh}=k)}\right\}\exp\left\{-\sum_{i=1}^{K}\Lambda_{*kh}(X_{ijh}|\mathbf{Z}_{ijh})\right\}.$$

Studying this model would be a worthy future project. When selecting a competing risk model for data analysis, one can compare the likelihood value of the semiparametric transformation models for the subdistribution hazards and the cause-specific hazards functions and select a model with a higher likelihood. However, in practice, we recommend choosing a model depending on what effects (ie, the cause-specific effects or direct effects on CIF) investigators are interested in studying.

ACKNOWLEDGEMENTS

Yizeng He, Soyoung Kim, and Kwang Woo Ahn were supported by the Advancing a Healthier Wisconsin Endowment (Project # 5520569) and the US National Cancer Institute (U24CA076518). Lu was supported by the US National Heart, Lung, and Blood Institute (R01HL149875). The authors would like to thank the Associate Editor and the two anonymous reviewers for their insightful comments which significantly improved the article.

DATA AVAILABILITY STATEMENT

The code for the simulation studies is available from the corresponding author upon reasonable request.

ORCID

Soyoung Kim https://orcid.org/0000-0003-1404-0575
Lu Mao https://orcid.org/0000-0002-8626-9822
Kwang Woo Ahn https://orcid.org/0000-0003-4567-8037

REFERENCES

1. Cullen K. Mass health examinations in the Busselton population, 1966 to 1970. *Med J Aust*. 1972;2(13):714-718.
2. Knuiman M, Divitini M, Olynyk J, Cullen D, Bartholomew H. Serum ferritin and cardiovascular disease: a 17-year follow-up study in Busselton, Western Australia. *Am J Epidemiol*. 2003;158(2):144-149.
3. Ballen K, Woo Ahn K, Chen M, et al. Infection rates among acute leukemia patients receiving alternative donor hematopoietic cell transplantation. *Biol Blood Marrow Transpl*. 2016;22(9):1636-1645.
4. Wohlfahrt J, Andersen PK, Melbye M. Multivariate competing risks. *Stat Med*. 1999;18(9):1023-1030.
5. Flowers CR, Costa LJ, Pasquini MC, et al. Efficacy of pharmacokinetics-directed Busulfan, cyclophosphamide, and etoposide conditioning and autologous stem cell transplantation for lymphoma: comparison of a multicenter phase II study and CIBMTR outcomes. *Biol Blood Marrow Transpl*. 2016;22(7):1197-1205.
6. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978;34(4):541.

7. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509.
8. Zhou B, Fine J, Latouche A, Labopin M. Competing risks regression for clustered data. *Biostatistics*. 2012;13(3):371-383.
9. He P, Eriksson F, Scheike TH, Zhang MJ. A proportional hazards regression model for the subdistribution with covariates-adjusted censoring weight for competing risks data. *Scand J Stat*. 2016;43(1):103-122.
10. Scheike TH, Zhang MJ, Gerds TA. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*. 2008;95(1):205-220.
11. Scheike TH, Sun Y, Zhang MJ, Jensen TK. A semiparametric random effects model for multivariate competing risks data. *Biometrika*. 2010;97(1):133-145.
12. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005;61(1):223-229.
13. Logan BR, Zhang MJ, Klein JP. Marginal models for clustered time-to-event data with competing risks using pseudovalues. *Biometrics*. 2011;67(1):1-7.
14. Mao L, Lin DY. Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *J Royal Stat Soc Ser B (Stat Methodol)*. 2017;79(2):573-587.
15. Bellach A, Kosorok MR, Rüschendorf L, Fine JP. Weighted NPMLE for the subdistribution of a competing risk. *J Am Stat Assoc*. 2019;114(525):259-270.
16. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.
17. Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *J Royal Stat Soc Ser B (Stat Methodol)*. 2007;69(4):507-564.
18. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44(4):1049-1060.
19. Martinussen T, Pipper C. Estimation in the positive stable shared frailty cox proportional hazards model. *Lifetime Data Anal*. 2005;11(1):99-115. doi:10.1007/s10985-004-5642-4
20. Zhou B, Latouche A. crrSC: competing risks regression for stratified and clustered data; 2013. R package version 1.1.
21. Bennett S. Analysis of survival data by the proportional odds model. *Stat Med*. 1983;2(2):273-277.
22. Commenges D, Andersen PK. Score test of homogeneity for survival data. *Lifetime Data Anal*. 1995;1(2):145-156. discussion 157-9.
23. Lee EW, Wei L, Amato DA, Leurgans S. Cox-type regression analysis for large numbers of small groups of correlated failure time observations; 1992:237-247; Springer.
24. Atilla E, Atilla PA, Demirer T. A review of myeloablative vs reduced intensity/non-myeloablative regimens in allogeneic hematopoietic stem cell transplantations. *Balkan Med J*. 2017;34(1):1-9.
25. Papanicolaou GA, Ustun C, Young JAH, et al. Bloodstream infection due to vancomycin-resistant Enterococcus is associated with increased mortality after hematopoietic cell transplantation for acute leukemia and myelodysplastic syndrome: a multicenter, retrospective cohort study. *Clin Infect Dis*. 2019;69(10):1771-1779.
26. Spiekerman CF, Lin DY. Marginal regression models for multivariate failure time data. *J Am Stat Assoc*. 1998;93(443):1164-1175.

How to cite this article: He Y, Kim S, Mao L, Ahn KW. Marginal semiparametric transformation models for clustered multivariate competing risks data. *Statistics in Medicine*. 2022;41(26):5349-5364. doi: 10.1002/sim.9573

APPENDIX

To prove Theorem 2, let BV_1 be the space of functions on $[0, \tau]$ that are uniformly bounded by 1 and with total variation bounded by 1. Define $V = \{\boldsymbol{v} \in \mathbb{R}^p : \|\boldsymbol{v}\| \leq 1\}$ and $W = BV_1^{\otimes KH}$, which is a KH product space of BV_1. Let $\boldsymbol{v} \in V$ and $\boldsymbol{w} = (m_{11}, \ldots, m_{KH}) \in W$. The score operator for Λ_{kh} for $m_{kh} \in BV_1$ is

$$
\begin{aligned}
i_{2kh}\{\boldsymbol{\beta}, \boldsymbol{\Lambda}\}[m_{kh}] = & \int_0^\tau \left[H'_{kh} \left\{ \int_0^t e^{\boldsymbol{\beta}_{kh}^\intercal \mathbf{Z}(s)} d\Lambda_{kh}(s) \right\} \int_0^t e^{\boldsymbol{\beta}_{kh}^\intercal \mathbf{Z}(s)} m_{kh}(s) d\Lambda_{kh}(s) + m_{kh}(t) \right] dN_{kh}(t) \\
& - \int_0^\tau \left\{ \tilde{\Psi}_{kh}(t; \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) \int_0^t m_{kh}(s) e^{\boldsymbol{\beta}_{kh}^\intercal \mathbf{Z}(s)} d\Lambda_{kh}(s) \right\} dN_0(t),
\end{aligned}
$$

where $\tilde{\Psi}_{kh}(t; \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) = S^{-1}(t; \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) \exp\left[H_{kh} \left\{ \int_0^t e^{\boldsymbol{\beta}_{kh}^\intercal \mathbf{Z}(s)} d\Lambda_{kh}(s) \right\} \right]$. And the score functions for $\boldsymbol{\beta}$ is $\boldsymbol{i}_1 = (\boldsymbol{i}_{111}^\intercal, \ldots, \boldsymbol{i}_{1KH}^\intercal)^\intercal$, where

$$
\begin{aligned}
\boldsymbol{i}_{1kh}(\boldsymbol{\beta}, \boldsymbol{\Lambda}) = & \int_0^\tau \left[H'_{kh} \left\{ \int_0^t e^{\boldsymbol{\beta}_{kh}^\intercal \mathbf{Z}(s)} d\Lambda_{kh}(s) \right\} \int_0^t \mathbf{Z}(s) e^{\boldsymbol{\beta}_{kh}^\intercal \mathbf{Z}(s)} d\Lambda_{kh}(s) \right] dN_{kh}(t) + \int_0^\tau \mathbf{Z}(t) dN_{kh}(t) \\
& - \int_0^\tau \left\{ \tilde{\Psi}_{kh}(t; \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) \int_0^t \mathbf{Z}(s) e^{\boldsymbol{\beta}_{kh}^\intercal \mathbf{Z}(s)} d\Lambda_{kh}(s) \right\} dN_0(t).
\end{aligned}
$$

For a sufficiently small $\delta > 0$, consider the following class of functions

$$\mathcal{F} = \left\{ \boldsymbol{i}_i(\boldsymbol{\beta}, \boldsymbol{\Lambda}), i_{2kh}(\boldsymbol{\beta}, \boldsymbol{\Lambda})[m_{kh}] : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| + \sup_{t\in[0,\tau]} \sum_{k=1}^{K}\sum_{h=1}^{H} |\boldsymbol{\Lambda}_{kh}(t) - \boldsymbol{\Lambda}_{kh0}(t)| < \delta, \right.$$
$$\left. m_{kh} \in BV_1,\ k = 1, \dots, K, h = 1, \dots, H \right\}.$$

Let $\mathbb{P}_n$ denote the empirical measure, P the underlying probability measure, and $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ the empirical process. Because L is fixed, by a similar argument to the proof of Theorem 3 of Spiekerman and Lin[26] using the uniform metric, we can show $\mathbb{G}_n$ indexed by $\mathcal{F}$ is asymptotically Gaussian. By the consistency of $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})$, the continuity of the score functions in the parameters, and the dominated convergence theorem,

$$\mathbb{G}_n \left\{ \boldsymbol{v}^{\intercal} \boldsymbol{i}_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) + \sum_{k=1}^{K}\sum_{h=1}^{H} i_{2kh}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})[m_{kh}] \right\} = \mathbb{G}_n \left\{ \boldsymbol{v}^{\intercal} \boldsymbol{i}_1(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0) + \sum_{k=1}^{K}\sum_{h=1}^{H} i_{2kh}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)[m_{kh}] \right\} + o_p(1) \tag{A1}$$

uniformly in $(\boldsymbol{v}, \boldsymbol{w})$. Furthermore, by Mao and Lin,[14] we have

$$\sqrt{n} \left\{ \boldsymbol{v}^{\intercal}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \sum_{k=1}^{K}\sum_{h=1}^{H} \int_0^{\tau} m_{kh} d(\widehat{\boldsymbol{\Lambda}}_{kh} - \boldsymbol{\Lambda}_{kh0}) \right\} = -\mathbb{G}_n \left\{ \tilde{\boldsymbol{v}}^{\intercal} \boldsymbol{i}_1(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0) + \sum_{k=1}^{K}\sum_{h=1}^{H} i_{2kh}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda})[\tilde{m}_{kh}] \right\} + o_p(1),$$

uniformly in $(\boldsymbol{v}, \boldsymbol{w})$ for some $\tilde{\boldsymbol{v}}$ and $\tilde{m}_{kh}$. A detailed expression for $\tilde{\boldsymbol{v}}$ and $\tilde{m}_{kh}$ can be found in Mao and Lin.[14] Therefore, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}_0)$ is asymptotically Gaussian.