Statistica Sinica **16**(2006), 1367-1385

MODELING THE SUBDISTRIBUTION OF A COMPETING RISK

Liuquan Sun[1], Jingxia Liu[2], Jianguo Sun[3] and Mei-Jie Zhang[2]

[1]*Chinese Academy of Sciences,* [2]*Medical College of Wisconsin and* [3]*University of Missouri*

Abstract: Competing risk failure time data occur frequently in medical studies, and a number of methods have been proposed for the analysis of these data. To assess covariate effects, a standard approach is to model the cause-specific hazard functions of different failure types. Recently, Fine and Gray (1999) proposed directly modeling the subdistribution of a competing risk with a Cox type model. In this paper, we consider a more flexible and general hazard model for the subdistribution. It is a combination of the additive model and the Cox model and allows one to perform a more detailed study of covariate effects. Inference procedures are developed for estimation of both parametric and nonparametric components of the model, and the asymptotic properties of the proposed estimators are established. Robust variance estimates along with some goodness-of-fit test procedures are also presented, and the prediction of the cumulative incidence function is discussed. The proposed methodology is applied to a set of competing risk data from a bone marrow transplant study.

Key words and phrases: Additive hazards model, competing risk, Cox model, estimating functions, prediction of cumulative incidence function.

1. Introduction

This paper discusses regression analysis of competing-risk failure time data that are inherent to medical research in which responses to treatments can be classified in terms of failure from disease processes or non-disease-related causes. To assess covariates effects, the standard approach for the analysis is to model the cause-specific hazard functions of different failure types using either parametric or semiparametric models (Cheng, Fine and Wei (1998), Shen and Cheng (1999) and Scheike and Zhang (2003)). However, these methods do not allow the analyst to directly assess the effect of a covariate on the cause-specific subdistribution (Fine and Gray (1999)). In the following, a flexible additive multiplicative hazard model is investigated for modeling the cause-specific subdistribution.

The proportional hazards model is perhaps the most commonly used regression model in survival analysis (Andersen, Borgan, Gill and Keiding (1993)).

However, for some situations, the model may only provide a rough summary of the effect of some covariates. An alternative is the additive hazards model (Aalen (1980), McKeague and Sasieni (1994) and Lin and Ying (1994)) that postulates a different relationship for the hazard and covariates than does the Cox model. In general, the subject matter seldom indicates clearly which of the models is to be preferred, and they may need to be used together to complement each other.

If the aim is to give a detailed description of covariate effects and to predict some survival probabilities, more flexible models are needed. To this end, for right-censored non-competing risk failure time data, several more general and flexible models have been proposed that combine the Cox and additive models. For example, Lin and Ying (1995) considered an additive-multiplicative hazard model, that is closely related to the proportional excess model (Andersen and Væth (1989) and Sasieni (1996)). Scheike and Zhang (2002) proposed an additive-multiplicative Cox-Aalen regression model in which some covariate effects work additively on the risk and other covariates have multiplicative effects. Martinussen and Scheike (2002) presented a flexible additive-multiplicative hazard model that allows both fixed and time-varying covariate effects.

For right-censored competing risk data, a lot of research has been focused on cause-specific hazard functions. In contrast, only limited work exists concerning the direct modeling of the cumulative incidence function. Fine and Gray (1999) studied the Cox model and developed some estimating equation-based inference procedures. In the following, we consider the modeling of the subdistribution using the flexible model given in Martinussen and Scheike (2002). In addition to developing some estimation procedures, robust variance estimates are presented and used for the prediction of cumulative incidence functions along with some goodness-of-fit test procedures. Note that the martingale formulas of variance estimators given in Martinussen and Scheike (2002) cannot be easily used for the prediction. More comments on this are given below.

The remainder of the paper is organized as follows. Section 2 introduces some notation and the regression model. Section 3 discusses inference about both the nonparametric and parametric components of the model. By using inverse probability of censoring weighting (IPCW) techniques (Robins and Rotnitzky (1992)), weighted score estimating equations are derived for estimation of the model and the asymptotic properties of the resulting estimators are established. Furthermore, some robust estimators of the standard errors are presented. In Section 4, an approach is proposed for predicting the cumulative incidence function, and some goodness-of-fit test procedures are discussed in Section 5 for model and

covariate selection. Section 6 applies the proposed methodology to a set of competing risk data arising from a bone marrow transplant study, and Section 7 concludes with some remarks.

2. Notation and Model

Consider a competing risk study with K different types of failures. Let T^0 and C denote the failure and censoring times, $\epsilon \in \{1, \ldots, K\}$ be the cause of failure, and $\boldsymbol{X}$ and $\boldsymbol{Z}$ be vectors of covariates of dimensions q and p, respectively. Assume that right-censored failure time data are observed and given by $\{T, \delta, \delta\epsilon, \boldsymbol{X}, \boldsymbol{Z}\}$, where $T = \min(T^0, C)$, $\delta = \mathcal{I}(T^0 \leq C)$ and $\mathcal{I}(\cdot)$ is the indicator function. In the following, we focus on modeling the cumulative incidence function for failure from Cause 1 $F_1(t; \boldsymbol{X}, \boldsymbol{Z}) = P\{T^0 \leq t, \epsilon = 1 | \boldsymbol{X}, \boldsymbol{Z}\}$ given the covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$. The inference about the cumulative incidence functions for other failure types can be performed similarly. Also, for simplicity of presentation, only time independent covariates will be considered, the inference procedures can be easily generalized to time-dependent covariate situations.

To estimate F_1, following Gray (1988) and Fine and Gray (1999), we consider the modeling of the hazard function of F_1 instead of the cause-specific hazard function. A major advantage of this is that one can directly estimate F_1 without simultaneously estimating subdistributions corresponding to other failure types. Specifically, define

$$\begin{aligned}\lambda_1(t; \boldsymbol{X}, \boldsymbol{Z}) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\{t \leq T^0 \leq t + \Delta t, \epsilon = 1 | T^0 \geq t \cup (T^0 \leq t \cap \epsilon \neq 1), \boldsymbol{X}, \boldsymbol{Z}\} \\ &= -\frac{d}{dt} \log\{1 - F_1(t; \boldsymbol{X}, \boldsymbol{Z})\}.\end{aligned}$$

Then $F_1(t; \boldsymbol{X}, \boldsymbol{Z}) = 1 - \exp\{-\Lambda_1(t; \boldsymbol{X}, \boldsymbol{Z})\}$, where $\Lambda_1(t; \boldsymbol{X}, \boldsymbol{Z}) = \int_0^t \lambda_1(u; \boldsymbol{X}, \boldsymbol{Z}) du$. One can think of λ_1 as the hazard function for the improper random variable $T^* = \mathcal{I}(\epsilon = 1) \times T + \{1 - \mathcal{I}(\epsilon = 1)\} \times \infty$.

In the following, it is assumed that λ_1 has the form

$$\lambda_1(t; \boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{\alpha}^T(t) \boldsymbol{X} + \lambda_{10}(t) \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}), \tag{1}$$

where $\boldsymbol{\alpha}(t)$ is an unknown q-vector of time-varying components representing the effects of covariates $\boldsymbol{X}$ on λ_1, $\boldsymbol{\beta}_0$ is a p-vector of unknown regression parameters denoting the effects of covariates $\boldsymbol{Z}$ on λ_1, and $\lambda_{10}(t)$ is an unspecified baseline hazard function. The above model assumes that covariates affect λ_1 in two ways: additive covariate effects described by an additive model; multiplicative covariate effects characterized by the Cox model. The additive model allows for time-varying covariate effects, while the Cox model allows only a common time-dependence through the baseline. Martinussen and Scheike (2002) considered

the above model for non-competing risk situations and, in this case, the model may be thought of as an excess hazards model where the background mortality is unknown and modeled by the additive model through the $\boldsymbol{X}$'s. The excess risk due to the covariates $\boldsymbol{Z}$ is described by the Cox model. Note that the additive model and the Cox model are two submodels of the flexible model above.

An alternative to (1) is to replace the second part by $\boldsymbol{\beta}_0^T \boldsymbol{Z}$. This is the model discussed in McKeague and Sasieni (1994). A potential problem with this model is that it can be sensitive to long follow-up time since the cumulative hazard function involves the function $\boldsymbol{\beta}_0^T \boldsymbol{Z}\, t$. In contrast, under (1), the corresponding form is $\exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \int_0^t \lambda_{10}(s)ds$. More comments on this alternative model are given below.

In the next section, for inference about (1), using IPCW techniques, we derive weighted score functions for both nonparametric and parametric components of the model. In the following, for simplicity, assume that C is independent of $T^0, \epsilon, \boldsymbol{X}$ and $\boldsymbol{Z}$. The methods developed can be generalized to allow for dependence between C and $(\boldsymbol{X}, \boldsymbol{Z})$ (e.g., Fine and Gray (1999)).

3. Estimation Procedures

Let $\{T_i^0, C_i, \epsilon_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\}$ $(i = 1, \ldots, n)$ be n independent replicates of $\{T^0, C, \epsilon, \boldsymbol{X}, \boldsymbol{Z}\}$. Then the observed data are $\{T_i, \delta_i, \delta_i\epsilon_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\}$. Define $N_i(t) = \mathcal{I}(T_i^0 \le t, \epsilon = 1)$ and $Y_i(t) = 1 - N_i(t-)$. Note that for the complete data (no censoring), individuals observed to fail remain in the risk set indefinitely as long as they have not failed from Cause 1. Define $r_i(t) = \mathcal{I}(C_i \ge T^0 \wedge t)$. If $r_i(t) = 1$, $N_i(t)$ and $Y_i(t)$ are computable up to time t; if $r_i(t) = 0$, individuals are observed up to time C_i and $N_i(t)$ and $Y_i(t)$ are not observable. However, $r_i(t)N_i(t)$ and $r_i(t)Y_i(t)$ are computable for $r_i(t) = 0$ or 1.

Let $G(t) = P\{C \ge t\}$ and $\widehat{G}(t)$ denote the Kaplan-Meier estimator of G based on the data $\{T_i, 1 - \delta_i, i = 1, \ldots, n\}$. The quantity $r_i(t)/G(T_i \wedge t)$ has expectation 1 conditional on $T_i^0, \epsilon_i, \boldsymbol{X}_i$ and $\boldsymbol{Z}_i$. Consider a time-dependent weight function of $R_i(t) = r_i(t)\widehat{G}(t)/\widehat{G}(T_i \wedge t)$. Define $dN_i^*(t) = R_i(t)dN_i(t)$ and $Y_i^*(t) = R_i(t)Y_i(t)$, and let $\boldsymbol{N}^* = (N_1^*, \ldots, N_n^*)^T$, $\boldsymbol{X}^* = (Y_1^*\boldsymbol{X}_1, \ldots, Y_n^*\boldsymbol{X}_n)^T$, $\boldsymbol{Z}^* = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)^T$, $\phi_i = \phi_i(\boldsymbol{\beta}) = Y_i^* \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$, $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\beta}) = (\phi_1, \ldots, \phi_n)^T$, $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\boldsymbol{\beta}) = \mathrm{diag}(\phi_i)$, $\boldsymbol{W} = \mathrm{diag}(w_i)$, and $\boldsymbol{V} = \mathrm{diag}(v_i)$, where $\boldsymbol{w} = (w_1, \ldots, w_n)^T$ and $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ are known weight functions. Note that here and in the following, we suppress dependence on time unless we wish to emphasize it.

To estimate the unknown parameters $\boldsymbol{\beta}_0$, $A(t) = \int_0^t \alpha(u)du$ and $\Lambda_{10}(t) = \int_0^t \lambda_{10}(u)du$, following Martinussen and Scheike (2002) and using IPCW tech-

niques, we propose to employ the following score functions

$$\begin{aligned}
&\int_0^\tau \boldsymbol{Z}^{*^T}\boldsymbol{\Phi}\boldsymbol{V}\{d\boldsymbol{N}^* - \boldsymbol{X}^* dA - \boldsymbol{\phi} d\Lambda_{10}\} = 0,\\
&\boldsymbol{X}^{*^T}\boldsymbol{W}\{d\boldsymbol{N}^* - \boldsymbol{X}^* dA - \boldsymbol{\phi} d\Lambda_{10}\} = 0,\\
&\boldsymbol{\phi}^T\boldsymbol{W}\{d\boldsymbol{N}^* - \boldsymbol{X}^* dA - \boldsymbol{\phi} d\Lambda_{10}\} = 0,
\end{aligned}$$

where τ is a prespecified constant such that $P\{T \geq \tau\} > 0$. For a given $\boldsymbol{\beta}$, solving the second and third score equations gives a weighted Aalen estimator

$$\widehat{A}(t;\boldsymbol{\beta}) = \int_0^t \left(\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}d\boldsymbol{N}^*(u) \tag{2}$$

for A, and a Breslow estimator

$$\widehat{\Lambda}_{10}(t;\boldsymbol{\beta}) = \int_0^t (\boldsymbol{\phi}^T\boldsymbol{W}\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T\boldsymbol{W}\boldsymbol{H}d\boldsymbol{N}^*(u) \tag{3}$$

for Λ_{10}, where $\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{X}^*(\boldsymbol{X}^{*^T}\boldsymbol{W}\boldsymbol{Q}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*^T}\boldsymbol{W}\boldsymbol{Q}$ and $\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{\phi}(\boldsymbol{\phi}^T\boldsymbol{W}\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T\boldsymbol{W}$. By inserting (2) and (3) in the first score equation for $\boldsymbol{\beta}_0$, we obtain

$$U(\boldsymbol{\beta};\tau) = \int_0^\tau \boldsymbol{Z}^{*^T}\boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{Q}\boldsymbol{H}d\boldsymbol{N}^*(t). \tag{4}$$

By letting $w_i = hv_i$ for some function h, $U(\boldsymbol{\beta};\tau)$ reduces to the Cox-like score equation

$$U(\boldsymbol{\beta};\tau) = \sum_{i=1}^n \int_0^\tau \left\{\boldsymbol{Z}_i - \frac{\sum_{j=1}^n w_j\phi_j\boldsymbol{Z}_j Y_j^* \exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)}{\sum_{j=1}^n w_j\phi_j Y_j^* \exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)}\right\} v_i\phi_i d\tilde{\boldsymbol{N}}_i(t) = 0,$$

where $d\tilde{\boldsymbol{N}} = \boldsymbol{H}d\boldsymbol{N}^*$. Let $\widehat{\boldsymbol{\beta}}$ denote the solution to (4). Once $\widehat{\boldsymbol{\beta}}$ is obtained, one can estimate the cumulative baseline hazard function Λ_{10} and the additive components $A(t)$ by the Breslow estimator $\widehat{\Lambda}_{10}(t) = \widehat{\Lambda}_{10}(t;\widehat{\boldsymbol{\beta}})$ and the weighted Aalen estimator $\widehat{A}(t) = \widehat{A}(t;\widehat{\boldsymbol{\beta}})$, respectively. If there is only a single cause of failure, (2)−(4) reduce to the score equations of Martinussen and Scheike (2002).

To investigate asymptotic properties of $\widehat{\boldsymbol{\beta}}$, $\widehat{\Lambda}_{10}(t)$ and $\widehat{A}(t)$, let $\widehat{\boldsymbol{\phi}}$, $\widehat{\boldsymbol{\Phi}}$, $\widehat{\boldsymbol{Q}}$ and $\widehat{\boldsymbol{H}}$ be defined as $\boldsymbol{\phi}$, $\boldsymbol{\Phi}$, $\boldsymbol{Q}$ and $\boldsymbol{H}$ with $\boldsymbol{\beta}$ replaced by $\widehat{\boldsymbol{\beta}}$. In the Appendix, it is shown that $n^{-1/2}U(\boldsymbol{\beta}_0;\tau)$ has an asymptotic normal distribution with mean zero and a covariance matrix that can be consistently estimated by

$$\widehat{\Sigma}_U = n^{-1}\sum_{i=1}^n \widehat{\Psi}_{1i}(\tau)\widehat{\Psi}_{1i}^T(\tau), \tag{5}$$

where

$$\Psi_{1i}(\tau) = \int_0^\tau \boldsymbol{D}_{1i}(t;\widehat{\boldsymbol{\beta}})R_i(t)d\widehat{M}_i(t) + \int_0^\tau \frac{\widehat{\boldsymbol{q}}_1(t)}{\widehat{\pi}(t)}d\widehat{M}_i^c(t),$$

$$\boldsymbol{D}_1(t;\boldsymbol{\beta}) = \big(\boldsymbol{D}_{11}(t;\boldsymbol{\beta}),\ldots,\boldsymbol{D}_{1n}(t;\boldsymbol{\beta})\big) = \boldsymbol{Z}^{*T}\widehat{\boldsymbol{\Phi}}\boldsymbol{V}\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{H}},$$

$$\widehat{\boldsymbol{q}}_1(t) = -n^{-1}\sum_{i=1}^n \int_0^t \boldsymbol{D}_{1i}(u;\widehat{\boldsymbol{\beta}})R_i(u)\mathcal{I}(u \geq t > T_i)d\widehat{M}_i(u),$$

$$\widehat{\pi}(t) = n^{-1}\sum_{i=1}^n \mathcal{I}(T_i \geq t),$$

$$\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(u)\big[\boldsymbol{X}_i^T(u)d\widehat{A}(u) + \exp\big(\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_i(u)\big)d\widehat{\Lambda}_{10}(u)\big],$$

$$\widehat{M}_i^c(t) = \mathcal{I}(T_i \leq t, \delta_i = 0) - \int_0^t \mathcal{I}(T_i \geq u)d\widehat{\Lambda}^c(u),$$

$$\widehat{\Lambda}^c(t) = \int_0^t \frac{\sum_{j=1}^n d\mathcal{I}(T_j \leq u, \delta_j = 0)}{\sum_{j=1}^n \mathcal{I}(T_j \geq u)}.$$

It thus follows that $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has an asymptotic normal distribution with mean zero and covariance matrix that can be consistently estimated by

$$\widehat{\Sigma}_{\boldsymbol{\beta}} = n^{-1}\mathcal{I}^{-1}(\widehat{\boldsymbol{\beta}})\sum_{i=1}^n \widehat{\Psi}_{1i}(\tau)\widehat{\Psi}_{1i}^T(\tau)\mathcal{I}^{-1}(\widehat{\boldsymbol{\beta}})^T\ , \tag{6}$$

where

$$\mathcal{I}(\boldsymbol{\beta}) = -\frac{1}{n}\frac{\partial U(\boldsymbol{\beta};\tau)}{\partial\boldsymbol{\beta}} = -\frac{1}{n}\big[\int_0^\tau \boldsymbol{Z}^{*T}(\partial\boldsymbol{\Phi}/\partial\boldsymbol{\beta})\big\{\{\boldsymbol{V}\boldsymbol{Q}\boldsymbol{H}d\boldsymbol{N}^*(t)\}\otimes\mathcal{I}_p\big\} + \int_0^\tau \boldsymbol{Z}^{*T}\boldsymbol{\Phi}\boldsymbol{H}^T\boldsymbol{V}(\partial\boldsymbol{Q}/\partial\boldsymbol{\beta})\big\{\boldsymbol{H}d\boldsymbol{N}^*(t)\otimes\mathcal{I}_p\big\}\big]\ .$$

Furthermore, we show in Appendix that $n^{1/2}\{\widehat{A}(t) - A(t)\}$ converges weakly to a zero-mean Gaussian process whose covariance function at (t_1, t_2) can be consistently estimated by

$$\widehat{\sigma}_A(t_1,t_2) = n^{-1}\sum_{i=1}^n \widehat{\Psi}_{2i}(t_1)\widehat{\Psi}_{2i}^T(t_2), \tag{7}$$

where

$$\widehat{\Psi}_{2i}(t) = \widehat{\boldsymbol{C}}_1(t)\mathcal{I}^{-1}(\widehat{\boldsymbol{\beta}})\widehat{\Psi}_{1i}(t) + \int_0^t \boldsymbol{D}_{2i}(u;\widehat{\boldsymbol{\beta}})R_i(u)d\widehat{M}_i(u) + \int_0^t \frac{\widehat{\boldsymbol{q}}_2(u,t)}{\widehat{\pi}(u)}d\widehat{M}_i^c(u),$$

$$\widehat{\boldsymbol{C}}_1(t) = -\int_0^t \big(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\big)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{\Phi}}\boldsymbol{Z}^*d\widehat{\Lambda}_{10}(u),$$

$$\boldsymbol{D}_2(t;\boldsymbol{\beta}) = \big(\boldsymbol{D}_{21}(t;\boldsymbol{\beta}),\ldots,\boldsymbol{D}_{2n}(t;\boldsymbol{\beta})\big) = n\big(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^{*}\big)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}},$$
$$\widehat{\boldsymbol{q}}_2(u,t) = -n^{-1}\sum_{i=1}^{n}\int_0^t \boldsymbol{D}_{2i}(s;\widehat{\boldsymbol{\beta}})R_i(s)\mathcal{I}(s \geq u > T_i)d\widehat{M}_i(s).$$

Similarly, $n^{1/2}\{\widehat{\Lambda}_{10}(t)-\Lambda_{10}(t)\}$ converges weakly to a zero-mean Gaussian process whose covariance function at (t_1,t_2) can be consistently estimated by

$$\widehat{\sigma}_{\Lambda_0}(t_1,t_2) = n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_{3i}(t_1)\widehat{\Psi}_{3i}(t_2), \tag{8}$$

where

$$\widehat{\Psi}_{3i}(t) = \widehat{\boldsymbol{C}}_2(t)\mathcal{I}^{-1}(\widehat{\boldsymbol{\beta}})\widehat{\Psi}_{1i}(t) + \int_0^t \boldsymbol{D}_{3i}(u;\widehat{\boldsymbol{\beta}})R_i(u)d\widehat{M}_i(u) + \int_0^t \frac{\widehat{\boldsymbol{q}}_3(u,t)}{\widehat{\pi}(u)}d\widehat{M}_i^c(u),$$
$$\widehat{\boldsymbol{C}}_2(t) = -\int_0^t (\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}})^{-1}\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{H}}\widehat{\boldsymbol{\Phi}}\boldsymbol{Z}^{*}d\widehat{\Lambda}_{10}(u),$$
$$\boldsymbol{D}_3(t;\boldsymbol{\beta}) = (\boldsymbol{D}_{31}(t;\boldsymbol{\beta}),\ldots,\boldsymbol{D}_{3n}(t;\boldsymbol{\beta})) = n(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}})^{-1}\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{H}},$$
$$\widehat{\boldsymbol{q}}_3(u,t) = -n^{-1}\sum_{i=1}^{n}\int_0^t \boldsymbol{D}_{3i}(s;\widehat{\boldsymbol{\beta}})R_i(s)\mathcal{I}(s \geq u > T_i)d\widehat{M}_i(s).$$

Note that the variance-covariance estimators given above are robust estimates and different from those given in Martinusseen and Scheike (2002). It would be difficult to use the latter estimators for the prediction of the cumulative incidence function and the model checking discussed in next sections.

4. Prediction of Cumulative Incidence Functions

One of the main goals in survival analysis is to predict certain survival probabilities for future subjects. To predict F_1 under (1) for a patient with a particular set of covariates $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Z} = \boldsymbol{z}$, one can first estimate the cumulative subdistribution hazard $\Lambda_1(t;\boldsymbol{x},\boldsymbol{z})$ by

$$\widehat{\Lambda}_1(t;\boldsymbol{x},\boldsymbol{z}) = \int_0^t \boldsymbol{x}^T(u)d\widehat{A}(u) + \int_0^t \exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{z}(u))d\widehat{\Lambda}_{10}(u)\,. \tag{9}$$

The predicted cumulative incidence is then given by $\widehat{F}_1(t;\boldsymbol{x},\boldsymbol{z}) = 1-\exp\{-\widehat{\Lambda}_1(t;\boldsymbol{x},\boldsymbol{z})\}$.

Furthermore, an estimate $\widehat{t}_p$ for the $100p$th percentile t_p of $F_1(t|\boldsymbol{x},\boldsymbol{z})$ can be obtained by solving the equation $\widehat{F}_1(t;\boldsymbol{x},\boldsymbol{z}) = 1-p$, where $0<p<1$ is such that $t_p < \tau$. Using the functional δ-method, one can show that for a known, monotone, absolutely continuous transformation g, $n^{1/2}\{g(\widehat{F}_1(t;\boldsymbol{x},\boldsymbol{z}))-g(F_1(t;\boldsymbol{x},\boldsymbol{z}))\}$

converges weakly to a zero-mean Gaussian process whose covariance function at (t_1, t_2) can be consistently estimated by

$$\widehat{\sigma}_F(t_1,t_2) = n^{-1}\dot{g}\big(\widehat{F}_1(t_1;\boldsymbol{x},\boldsymbol{z})\big)\dot{g}\big(\widehat{F}_1(t_2;\boldsymbol{x},\boldsymbol{z})\big)\big(1-\widehat{F}_1(t_1;\boldsymbol{x},\boldsymbol{z})\big) \times \big(1-\widehat{F}_1(t_2;\boldsymbol{x},\boldsymbol{z})\big)\sum_{i=1}^{n}\widehat{\Psi}_{4i}(t_1)\widehat{\Psi}_{4i}(t_2), \tag{10}$$

where $\dot{g}(t) = dg(t)/dt$ and

$$\widehat{\Psi}_{4i}(t) = \int_0^t \boldsymbol{x}^T(u)d\widehat{\Psi}_{2i}(u) + \int_0^t \exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{z}(u))d\widehat{\Psi}_{3i}(u) + \int_0^t \exp(\widehat{\boldsymbol{\beta}}^T\boldsymbol{z}(u))\boldsymbol{z}^T(u)d\widehat{\Lambda}_{10}(u)\mathcal{I}^{-1}(\widehat{\boldsymbol{\beta}})\widehat{\Psi}_{1i}(\tau).$$

The above transformation g is usually chosen to stabilize the variance and to ensure that pointwise and simultaneous confidence intervals for the probability $F_1(t;\boldsymbol{x},\boldsymbol{z})$ are bounded between 0 and 1. One commonly used choice is $g = \log(-\log)$.

Sometimes one is interested in constructing confidence bands for $A(t)$, $\Lambda_{10}(t)$, $F_1(t;\boldsymbol{x},\boldsymbol{z})$ or t_p. This may be analytically difficult since the limiting Gaussian processes for $n^{1/2}\{\widehat{A}(t) - A(t)\}$, $n^{1/2}\{\widehat{\Lambda}_{10}(t) - \Lambda_{10}(t)\}$ and $n^{1/2}\{g\big(\widehat{F}_1(t;\boldsymbol{x},\boldsymbol{z})\big) - g\big(F_1(t;\boldsymbol{x},\boldsymbol{z})\big)\}$ do not have independent increments. To this end, we propose to use the following simulation approach to approximate these limiting distributions as in Lin, Fleming and Wei (1994) and Scheike and Zhang (2003). Let $\{G_i; i = 1, ..., n\}$ be a simple random sample of size n from the standard normal distribution and independent of the observed data. Then one can construct the simultaneous confidence bands for $A(t)$, $\Lambda_{10}(t)$, $F_1(t;\boldsymbol{x},\boldsymbol{z})$ or t_p by replacing $\widehat{M}_i(t)$ and $\widehat{M}_i^c(t)$ with $G_i\widehat{M}_i(t)$ and $G_i\widehat{M}_i^c(t)$, respectively, and repeatedly generating normal random samples $\{G_i; i = 1, \ldots, n\}$ given the observed data. Note that since $\Lambda_{10}(t)$ is nonnegative, one may want to use the log transformation for the construction of its confidence bands.

5. Model Selection

This section considers the goodness of fit test of the model and the assessment of time-varying covariates. For these, we develop some asymptotically sound test procedures.

To evaluate the goodness of fit of the covariates included in the multiplicative part of the model, following Lin, Wei and Ying (1993) and Wei (1984), consider the cumulative score processes. The observed score process is given

by $n^{-1/2}U(\widehat{\boldsymbol{\beta}};t)$, and its asymptotic distribution is equivalent to the asymptotic distribution of

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\Big\{\widehat{\Psi}_{1i}(t)-\mathcal{I}(\widehat{\boldsymbol{\beta}},t)\mathcal{I}^{-1}(\widehat{\boldsymbol{\beta}},\tau)\widehat{\Psi}_{1i}(\tau)\Big\},$$

where $\mathcal{I}(\beta,t)$ is the minus of the derivative of $n^{-1}U(\beta;t)$. Note that if the multiplicative part of the model is appropriate, the components of the score process should behave as under the null. This suggests that we can use the following test statistics

$$\mathcal{F}_1=\sup_{0\le t\le\tau}\left|n^{-\frac{1}{2}}U_j(\widehat{\boldsymbol{\beta}};t)\right|,\quad (j=1,\ldots,p),$$

where $U_j(\boldsymbol{\beta};t)$ denotes the jth component of $U(\boldsymbol{\beta};t)$. The percentiles of this test statistic can be estimated empirically using a number of simulated processes as discussed in the previous section, or in Lin, Wei and Ying (1993).

Now consider testing if covariate j, included in the additive part of the model, is significant. For this, we suggest the test statistic

$$\mathcal{F}_2=\sup_{0\le t\le\tau}\left|\frac{\widehat{A}_j(t)}{\widehat{\sigma}^2_{A,j}(t)}\right|,$$

where $\widehat{A}_j$ is the jth component of $\widehat{A}$ and $\widehat{\sigma}^2_{A,j}(t)$ is the estimate of the variance of $\widehat{A}_j(t)$.

Sometimes one may also be interested in testing if an additive component has indeed a time-varying effect. To this end, we propose the test statistic

$$\mathcal{F}_3=\sup_{0\le t\le\tau}\left|\widehat{A}_j(t)-\frac{\widehat{A}_j(\tau)}{\tau}t\right|.$$

Note that $\mathcal{F}_2$ evaluates the departure of $\widehat{A}_j(t)$ from the null, while $\mathcal{F}_3$ measures the departure between $\widehat{A}_j(t)/t$ and the estimate of the constant effect under the null, $\widehat{A}_j(\tau)/\tau$. Also note that the asymptotic distribution of $n^{1/2}\{\widehat{A}(t)-A(t)\}$ is equivalent to the asymptotic distribution of $n^{-1/2}\sum_{i=1}^{n}\widehat{\Psi}_{2i}(t)$, where $\widehat{\Psi}_{2i}(t)$ is defined in (7). Then the percentiles of the above two test statistics can be simulated as before. The proposed tests are simple to implement and are omnibus. Additionally, one can plot the estimated cumulative regression function and use the plots to visually examine whether a covariate has a time-varying effect on the cumulative incidence function.

To fit (1), one needs to choose which covariates have time-varying effects and which covariates have constant effects on the hazard function. There are different ways to do this. One method is to begin by assuming that all covariates have

time-varying effects and to fit (1) with $Z = 0$. Then by using the $\mathcal{F}_3$ test given above, we can assign significant covariates to $\boldsymbol{X}$ and non-significant time-varying covariates to $\boldsymbol{Z}$. More details on this are given in the next section through the example. McKeague and Sasieni (1994) described some other approaches.

6. An Application

This section applies the proposed methodology to a set of cancer data from the International Bone Marrow Transplant Registry (IBMTR). IBMTR is a voluntary working group of over 300 transplant centers worldwide that contribute data on their allogeneic bone marrow transplants to the Statistical Center at the Medical College of Wisconsin. For this study, we analyzed the data arising from 408 myelodysplasia (MDS) patients treated with HLA-identical sibling bone marrow transplantation (BMT) from 1989 to 1997 and who had complete information of platelets at their transplantation (Sierra, Perez, Rozman, Carreras, Klein, Rizzo, Davies, Lazarus, Bredeson, Marks, Boogaerts, Goldman, Champlin, Keating, Weisdorf, de Witte and Horowitz (2002)). In the study two competing failure types − treatment related mortality (TRM), defined as death in complete remission, and relapse, were considered, with the focus on TRM and the three risk factors that have been suggested to be significantly associated with TRM. These are patient age (continuous, centered at a mean of 35, and ranging from 2 to 64), platelets before transplantation (binary variable, $\geq 100 \times 10^9/L$ or $< 100 \times 10^9/L$) and graft-versus-host disease (GVHD) prophylaxis (binary variable, T-cell depletion BMT or no T-cell depletion BMT).

In the analysis, to visually examine which risk factors had a time-varying effect, we first fitted the data to (1) with covariates $\boldsymbol{X}$ including age, platelet and GVHD prophylaxis and $Z = 0$ for all cases. Note that this submodel of (1) is additive. Figure 1 gives the estimated additive function $A(t)$ corresponding to each risk factor and the baseline, along with their confidence intervals and simultaneous confidence bands. It seems from Figure 1 that platelets and age had time-varying effects and GVHD prophylaxis had a constant effect on the cumulative incidence function of TRM. To verify these time-varying effects, we performed the $\mathcal{F}_3$ test given in the previous section by fitting an additive model first. The test gave p-values of 0.042, 0.002, 0.155 for risk factors of platelets, age and GVHD prophylaxis, respectively, confirming what was seen in Figure 1. Note that if we use the cause-specific hazards approach, only platelets would have a time-varying effect for TRM (Sierra et al. (2002)). Also in this case, since the cumulative incidence probability is a function of both the cause-specific hazards of TRM and relapse, unlike the proposed approach, it is hard to identify which covariates having time-varying effects on the cumulative incidence function.

To fit (1) given the above results, let $\boldsymbol{X}$ include platelet and age and $\boldsymbol{Z}$ be GVHD prophylaxis. That is, assume that platelet and age affected the hazard of TRM time-dependently and GVHD prophylaxis affected the hazard of TRM constantly. For the goodness-of-fit assessment of the proposed model, we first considered the $\mathcal{F}_1$ test for GVHD prophylaxis and obtained $p = 0.64$, indicating that GVHD prophylaxis could be included in the multiplicative part of the model. Second, we performed $\mathcal{F}_2$ and $\mathcal{F}_3$ tests for platelet and age. The $\mathcal{F}_3$ test gave p-values similar to those obtained before and confirmed that platelets and age had time-varying effects. The $\mathcal{F}_2$ test gave p-values of 0.003 and < 0.001 for platelets and age, respectively, which indicates that both risk factors had significant effect on the cumulative incidence probability of TRM. The above results suggest that (1) is an appropriate model for this data set. Fitting (1) yielded $\hat{\beta} = -0.54$ with the estimated standard error of $\hat{\sigma} = 0.26$, resulting in a p-value of 0.04 for testing $\beta = 0$. This indicates that GVHD prophylaxis had a significant effect on the cumulative incidence probability of TRM, and the patients who received T-cell depleted BMT had less chance of developing acute and chronic GVHD. Complication of acute and chronic GVHD is a main cause of death after BMT. Therefore, T-cell depleted BMT reduces the probability of TRM.

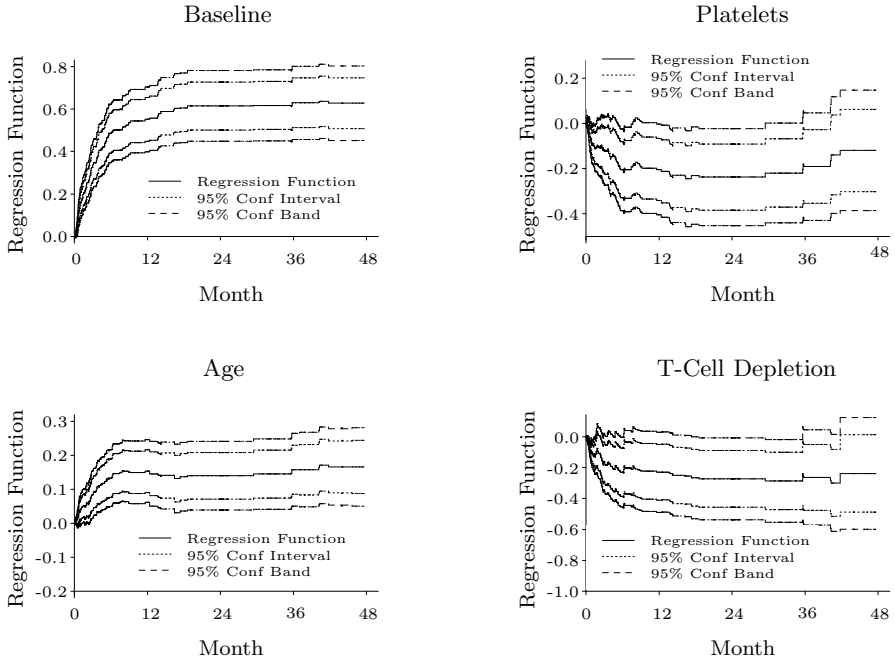

Figure 1. Regression function estimation for treatment related mortality (TRM).

To further examine the appropriateness of the model, we performed a subgroup analysis. Specifically, we considered four subgroups: patients with platelets $< 100\times10^9/L$ and who received a no T-cell depleted BMT; patients with platelets $\geq 100 \times 10^9/L$ and who received a no T-cell depleted BMT; old patients (age > 35) with platelets $< 100 \times 10^9/L$ and who received a no T-cell depleted BMT; younger patients (age ≤ 35) with platelets $< 100 \times 10^9/L$ and who received a no T-cell depleted BMT. The number of patients and mean patient ages for the four subgroups are $(N_1 = 254, \mu_1 = 36)$, $(N_2 = 100, \mu_2 = 32)$, $(N_3 = 141, \mu_3 = 47)$ and $(N_4 = 113, \mu_4 = 23)$, respectively. For each subgroup, we obtained the nonparametric estimate of the cumulative incidence function (CIF) of TRM given by $\int_0^t \widehat{S}(u^-)d\widehat{\Lambda}_1(u)$, and its 95% confidence bands, by using the approach given in Lin (1997). Here $\widehat{S}(u)$ is the Kaplan-Meier estimator of leukemia-free survival function and $\widehat{\Lambda}_1(u)$ is Nelson-Aalen's estimator for the cumulative cause-specific hazard function of TRM. They are displayed in Figure 2 together with the predicted CIF based on (1). It can be seen that all predicted CIF curves are almost identical to the nonparametric estimates and this suggests that (1) fits the data reasonably well.

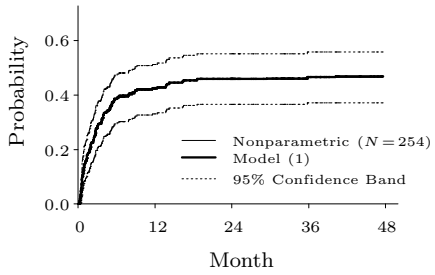

Figure 2.1. For subgroup patients with platelets $< 100 \times 10^9/L$ and no T-Cell Depleted BMT.

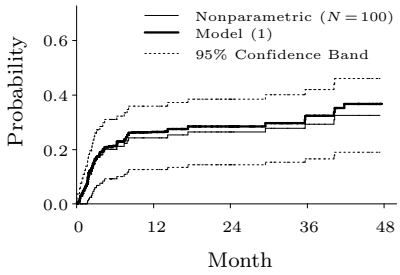

Figure 2.2. For subgroup patients with platelets $> 100 \times 10^9/L$ and no T-Cell Depleted BMT.

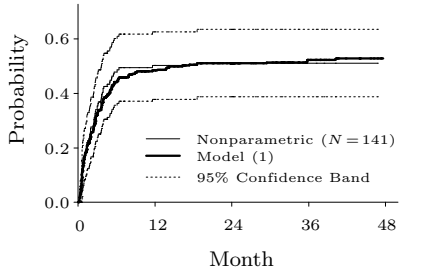

Figure 2.3. For subgroup patients with platelets $< 100 \times 10^9/L$; no T-Cell Depleted BMT and age > 35 years.

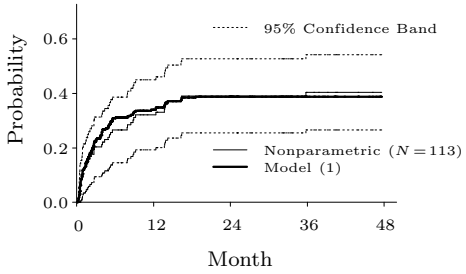

Figure 2.4. For Subgroup Patients with Platelets $< 100 \times 10^9/L$; no T-Cell Depleted BMT and age < 35 years.

Figure 2. Esitmated cumulative incidence function for TRM.

To give an idea about the robust variance estimate given in this paper, we compared the estimated standard errors of CIF by using the nonparametric approach based on the subgroup cases given in Lin (1997) and under (1) (for formulas, see Section 4). The estimates indicated that the two variance estimations are in a similar range, and the model-based method gives smaller variances as expected.

7. Concluding Remarks

This paper presented a flexible model for regression analysis of competing risk failure time data, and inference procedures were proposed for estimating its parametric and nonparametric components. The model includes the commonly used Cox model as a special case, it allows time-varying covariate effects, and provides a way to perform a more detailed analysis of covariate effects. In addition to the establishment of the asymptotic properties of the proposed estimates, robust variance estimation and some goodness-of-fit test procedures were provided that allow for prediction of survival probabilities and model checking.

Our model and proposed estimators are generalizations of those discussed in Martinusseen and Scheike (2002) for non-competing risk situations, though there are some differences between methods. One is that Martinusseen and Scheike (2002) derived the large-sample properties of their approach by using martingale central limit theory, and their variance estimators are martingale-based, which makes their use for prediction difficult. In contrast, the empirical process theory used here for the asymptotic properties and the development of robust variance estimates can be easily used for the prediction, as shown in Section 4.

To model the hazard function of the cumulative incidence functions, one could consider an alternative that replaces the Cox model in the second part of (1) by, for example, the proportional odds model. Inference procedures similar to those given above could be developed. Another alternative, mentioned before and suggested by a referee, is the partly parametric additive risk model proposed by McKeague and Sasieni (1994), which assumes that all covariates have additive effects. The model allows a different type of covariate effects compared to (1) for covariates with constant effects and is useful in many situations (Lin and Ying (1994) and McKeague and Sasieni (1994)). McKeague and Sasieni proposed some efficient estimators for the non-competing risks survival data and it would be useful to generalize their method to the situation considered here. However, this is beyond the scope of this paper.

Acknowledgements

The authors wish to thank the Editor, an associate editor and two referees for their helpful comments and suggestions, which greatly improved the paper. The

research of Drs. Liuquan Sun and Jianguo Sun was supported by grants from the National Natural Science Foundation of P.R. China and the National Institutes of Health, respectively, and Dr. Mei-Jie Zhang's research was supported by National Cancer Institute grant 2 R01 CA54706-10.

Appendix

In the following, for a given matrix B_n, $\mathcal{E}\{B_n\}$ denotes the limit of B_n, and $\approx$ means asymptotically equivalent. Let $\tilde{\phi}_i$ be ϕ_i with R_i replaced by $\tilde{R}_i$, where $\tilde{R}_i(t) = r_i(t)G(t)/G(T_i \wedge t)$. Define $dM_i = dN_i - Y_i[\boldsymbol{X}_i^T dA + \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}_i)d\Lambda_{10}]$, $dM_i^* = R_i dM_i$, $\boldsymbol{M}^* = (M_1^*, \ldots, M_n^*)^T$ and $\tilde{\boldsymbol{D}}_1 = (\tilde{\boldsymbol{D}}_{11}, \ldots, \tilde{\boldsymbol{D}}_{1n})$, where

$$\begin{aligned}\tilde{\boldsymbol{D}}_{1i} = {} & \boldsymbol{Z}_i\tilde{\phi}_i v_i - \mathcal{E}\{\boldsymbol{Z}^{*T}\boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{\phi}(\boldsymbol{\phi}^T\boldsymbol{W}\boldsymbol{\phi})^{-1}\}\tilde{\phi}_i w_i \\ & -\mathcal{E}\{\boldsymbol{Z}^{*T}\boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{Q}\boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}\boldsymbol{X}^*)^{-1}\}\tilde{R}_i Y_i \boldsymbol{X}_i w_i \\ & +\mathcal{E}\{\boldsymbol{Z}^{*T}\boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{Q}\boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{\phi}(\boldsymbol{\phi}^T\boldsymbol{W}\boldsymbol{\phi})^{-1}\}\tilde{\phi}_i w_i.\end{aligned}$$

Note that $\boldsymbol{Q}\boldsymbol{\phi} = 0, \boldsymbol{H}\boldsymbol{\phi} = \boldsymbol{\phi}$ and $\boldsymbol{H}\boldsymbol{X}^* = 0$. Then we have

$$\begin{aligned}n^{-\frac{1}{2}}U(\boldsymbol{\beta}_0;\tau) &= n^{-\frac{1}{2}}\int_0^\tau \boldsymbol{Z}^{*T}\boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{Q}\boldsymbol{H}d\boldsymbol{M}^*(t) \\ &\approx n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^\tau \tilde{\boldsymbol{D}}_{1i}(t)R_i(t)dM_i(t) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^\tau \tilde{\boldsymbol{D}}_{1i}(t)\tilde{R}_i(t)dM_i(t) \\ &\quad + n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^\tau \tilde{\boldsymbol{D}}_{1i}(t)\Big\{\frac{\widehat{G}(t)}{\widehat{G}(T_i \wedge t)} - \frac{G(t)}{G(T_i \wedge t)}\Big\}r_i(t)dM_i(t). \quad \text{(A.1)}\end{aligned}$$

Let $\Lambda^c(t)$ be the cumulative hazard function of the censoring distribution, and

$$M_i^c(t) = \mathcal{I}(T_i \le t, \delta_i = 0) - \int_0^t \mathcal{I}(T_i \ge u)d\Lambda^c(u).$$

Then we have

$$\frac{\widehat{G}(t)}{\widehat{G}(T_i \wedge t)} - \frac{G(t)}{G(T_i \wedge t)} \approx -\frac{G(t)\mathcal{I}(T_i < t)}{G(T_i \wedge t)}\sum_{j=1}^n \int_{T_i}^t \frac{dM_j^c(u)}{\sum_{k=1}^n \mathcal{I}(T_k \ge u)} \qquad \text{(A.2)}$$

(e.g., Gill (1980)). The second term in (A.1) can be approximated by

$$n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^\tau \frac{\boldsymbol{q}_1(t)}{\pi(t)}dM_i^c(t),$$

where $\boldsymbol{q}_1(t) = -\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \int_0^\tau \tilde{\boldsymbol{D}}_{1i}(u)\tilde{R}_i(u)\mathcal{I}(u \geq t > T_i)dM_i(u)$, and $\pi(t) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathcal{I}(T_i \geq t)$. Thus, we have $n^{-1/2}U(\boldsymbol{\beta}_0;\tau) \approx n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_{1i}(\tau)$, which is a sum of i.i.d. zero-mean terms, where

$$\boldsymbol{\xi}_{1i}(\tau) = \int_0^\tau \tilde{\boldsymbol{D}}_{1i}(t)\tilde{R}_i(t)dM_i(t) + \int_0^\tau \frac{\boldsymbol{q}_1(t)}{\pi(t)}dM_i^c(t).$$

Utilizing the Multivariate Central Limit Theorem, $n^{-1/2}U(\boldsymbol{\beta}_0;\tau)$ converges in distribution to a normal random variable with mean zero and variance matrix $\Sigma_U = E\{\boldsymbol{\xi}_{1i}(\tau)\boldsymbol{\xi}_{1i}^T(\tau)\}$, which can be estimated by (5). Now a Taylor series expansion yields (6).

For $\widehat{A}$, note that $\boldsymbol{\phi}(\boldsymbol{\beta}_0) = \boldsymbol{\phi}(\widehat{\boldsymbol{\beta}}) + \boldsymbol{\Phi}^*\boldsymbol{Z}^*(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}})$ and $\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{\phi}} = 0$, where $\boldsymbol{\Phi}^* = \operatorname{diag}(\boldsymbol{\phi}(\boldsymbol{\beta}^*))$ and $\boldsymbol{\beta}^*$ is on the line segment between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. It can be checked that

$$\begin{aligned}
&n^{\frac{1}{2}}(\widehat{A}(t) - A(t))\\
&= n^{\frac{1}{2}}\int_0^t \left(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\left\{\boldsymbol{X}^* dA(u) + \boldsymbol{\phi}(\boldsymbol{\beta}_0)d\Lambda_{10}(u) + d\boldsymbol{M}^*(u)\right\} - A(t)\\
&= n^{\frac{1}{2}}\int_0^t \left(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{\Phi}^*\boldsymbol{Z}^*(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}})d\Lambda_{10}(u)\\
&\quad + n^{\frac{1}{2}}\int_0^t \left(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}d\boldsymbol{M}^*(u)\\
&= \boldsymbol{C}_1^*(t)n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + n^{\frac{1}{2}}\int_0^t \left(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}d\boldsymbol{M}^*(u), \qquad \text{(A.3)}
\end{aligned}$$

where $\boldsymbol{C}_1^*(t) = -\int_0^t \left(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{\Phi}^*\boldsymbol{Z}^* d\Lambda_{10}(u)$. The first term in (A.3) can be approximated by

$$n^{-\frac{1}{2}}\sum_{i=1}^n \boldsymbol{C}_1(t)\tilde{\mathcal{I}}^{-1}(\boldsymbol{\beta}_0)\Psi_{1i}(\tau),$$

where $\boldsymbol{C}_1(t)$ and $\tilde{\mathcal{I}}(\boldsymbol{\beta}_0)$ are the limits of $C_1^*(t)$ and $\mathcal{I}(\boldsymbol{\beta}_0)$, respectively. Let $\tilde{\boldsymbol{D}}_2 = (\tilde{\boldsymbol{D}}_{21}, \ldots, \tilde{\boldsymbol{D}}_{2n})$, where

$$\begin{aligned}
\tilde{\boldsymbol{D}}_{2i} = &\int_0^t \mathcal{E}\left\{n\left(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\right)^{-1}\right\}\tilde{R}_i Y_i \boldsymbol{X}_i w_i\\
&-\mathcal{E}\left\{n\left(\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\widehat{\boldsymbol{\phi}}\left(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}}\right)^{-1}\right\}\tilde{\boldsymbol{\phi}}_i w_i.
\end{aligned}$$

Then the second term in (A.3) can be approximated by

$$\begin{aligned}
&n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^t \tilde{\boldsymbol{D}}_{2i}(u)\tilde{R}_i(u)dM_i(u)\\
&\quad + n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^t \tilde{\boldsymbol{D}}_{2i}(u)\left\{\frac{\widehat{G}(u)}{\widehat{G}(T_i \wedge u)} - \frac{G(u)}{G(T_i \wedge u)}\right\} r_i(u)dM_i(u). \qquad \text{(A.4)}
\end{aligned}$$

Using (A.2), the second term in (A.4) can be approximated by

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \frac{\boldsymbol{q}_2(u,t)}{\pi(u)}dM_i^c(u),$$

where $\boldsymbol{q}_2(u,t) = -\lim_{n\to\infty} n^{-1}\sum_{i=1}^n\int_0^t \boldsymbol{D}_{2i}(s)\tilde{R}_i(s)\mathcal{I}(s\geq u > T_i)dM_i(s)$. Therefore,

$$n^{\frac{1}{2}}(\widehat{A}(t) - A(t)) \approx n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\xi}_{2i}(t), \tag{A.5}$$

which is a sum of i.i.d. zero-mean terms for fixed t, where

$$\boldsymbol{\xi}_{2i}(t) = \boldsymbol{C}_1(t)\tilde{\mathcal{I}}^{-1}(\boldsymbol{\beta}_0)\Psi_{1i}(\tau) + \int_0^t \tilde{\boldsymbol{D}}_{2i}(u)\tilde{R}_i(u)dM_i(u) + \int_0^t \frac{\boldsymbol{q}_2(u,t)}{\pi(u)}dM_i^c(u).$$

By the Multivariate Central Limit Theorem, $n^{1/2}\big(\widehat{A}(t)-A(t)\big)$ converges in finite dimensional distributions to a zero-mean Gaussian process. Using empirical theory as in Lin, Wei, Yang and Ying (2000), we can show that $n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_{2i}(t)$ is tight. Thus, $n^{1/2}\big(\widehat{A}(t)-A(t)\big)$ converges weakly to a zero-mean Gaussian process whose covariance function at (t_1,t_2) can be consistently estimated by (7).

For Λ_{10}, let $\tilde{\boldsymbol{D}}_3 = (\tilde{\boldsymbol{D}}_{31},\ldots,\tilde{\boldsymbol{D}}_{3n})$, where

$$\begin{aligned}
&\tilde{\boldsymbol{D}}_{3i}\\
&= \mathcal{E}\big\{n\big(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}}\big)^{-1}\big\}\tilde{\boldsymbol{\phi}}_i w_i - \mathcal{E}\big\{n\big(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}}\big)^{-1}\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\boldsymbol{X}^*\big(X^{*^T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\big)^{-1}\tilde{R}_iY_i\boldsymbol{X}_i w_i\\
&\quad + \mathcal{E}\big\{n\big(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}}\big)^{-1}\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\boldsymbol{X}^*\big(\boldsymbol{X}^{*^T}\boldsymbol{W}\widehat{\boldsymbol{Q}}\boldsymbol{X}^*\big)^{-1}\boldsymbol{X}^{*^T}\boldsymbol{W}\widehat{\boldsymbol{\phi}}\big(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}}\big)^{-1}\big\}\tilde{\boldsymbol{\phi}}_i w_i.
\end{aligned}$$

Define $\boldsymbol{C}_2^*(t) = -\int_0^t(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}})^{-1}\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{H}}\boldsymbol{\Phi}^*\boldsymbol{Z}^*d\Lambda_{10}(u)$, and $\boldsymbol{C}_2(t)$ as the limit of $\boldsymbol{C}_2^*(t)$. Similar as before, it follows from $\widehat{\boldsymbol{H}}\boldsymbol{X}^* = 0$ and $\widehat{\boldsymbol{H}}\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}$ that

$$\begin{aligned}
&n^{\frac{1}{2}}(\widehat{\Lambda}_{10}(t) - \Lambda_{10}(t))\\
&= n^{\frac{1}{2}}\int_0^t(\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{\phi}})^{-1}\widehat{\boldsymbol{\phi}}^T\boldsymbol{W}\widehat{\boldsymbol{H}}\big\{\boldsymbol{X}^*dA(u) + \boldsymbol{\phi}(\boldsymbol{\beta}_0)d\Lambda_{10}(u) + d\boldsymbol{M}^*(u)\big\} - \Lambda_{10}(t)\\
&\approx \boldsymbol{C}_2(t)n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t\tilde{\boldsymbol{D}}_{3i}(u)R_i(u)dM_i(u)\\
&\approx n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{C}_2(t)\tilde{\mathcal{I}}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{\xi}_{1i}(\tau) + n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t\tilde{\boldsymbol{D}}_{3i}(u)\tilde{R}_i(u)dM_i(u)\\
&\quad + n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t\tilde{\boldsymbol{D}}_{3i}(u)\left\{\frac{\widehat{G}(u)}{\widehat{G}(T_i\wedge u)} - \frac{G(u)}{G(T_i\wedge u)}\right\}r_i(u)dM_i(u)\\
&\approx n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\xi}_{3i}(t),
\end{aligned}\tag{A.6}$$

which is a sum of i.i.d. zero-mean terms for fixed t, where

$$\boldsymbol{\xi}_{3i}(t) = \boldsymbol{C}_2(t)\tilde{\mathcal{I}}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{\xi}_{1i}(\tau) + \int_0^t \tilde{\boldsymbol{D}}_{3i}(u)\tilde{R}_i(u)dM_i(u) + \int_0^t \frac{\boldsymbol{q}_3(u,t)}{\pi(u)}dM_i^c(u),$$

and $\boldsymbol{q}_3(u,t) = -\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \int_0^t \tilde{\boldsymbol{D}}_{3i}(s)\tilde{R}_i(s)\mathcal{I}(s \geq u > T_i)dM_i(s)$. By the Multivariate Central Limit Theorem, $n^{1/2}\big(\widehat{\Lambda}_{10}(t) - \Lambda_{10}(t)\big)$ converges in finite dimensional distributions to a zero-mean Gaussian process. Using empirical theory as in Lin, Wei, Yang and Ying (2000), we can also show that $n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_{3i}(t)$ is tight. Thus, $n^{1/2}\big(\widehat{\Lambda}_{10}(t) - \Lambda(t)\big)$ converges weakly to a zero-mean Gaussian process whose covariance function at (t_1, t_2) can be consistently estimated by (8).

For $\widehat{F}_1(t; x.z)$, it follows from (A.5) and (A.6) that

$$\begin{aligned}
&n^{\frac{1}{2}}\big(\widehat{\Lambda}_1(t;\boldsymbol{x},\boldsymbol{z}) - \Lambda_1(t;\boldsymbol{x},\boldsymbol{z})\big) \\
&\quad = n^{\frac{1}{2}}\int_0^t \boldsymbol{x}^T(u)d\{\widehat{A}(u) - A(u)\} + \int_0^t \exp\big(\boldsymbol{\beta}_0^T\boldsymbol{z}(u)\big)d\{\widehat{\Lambda}_{10}(u) - \Lambda_{10}(u)\} \\
&\qquad + n^{\frac{1}{2}}\int_0^t \Big\{\exp\big(\widehat{\boldsymbol{\beta}}^T\boldsymbol{z}(u)\big) - \exp\big(\boldsymbol{\beta}_0^T\boldsymbol{z}(u)\big)\Big\}d\widehat{\Lambda}_{10}(u) \\
&\quad \approx n^{-\frac{1}{2}}\sum_{i=1}^n \boldsymbol{\xi}_{4i}(t),
\end{aligned}$$

which is a sum of i.i.d. zero-mean terms for fixed t, where

$$\begin{aligned}
\boldsymbol{\xi}_{4i}(t) = &\int_0^t \boldsymbol{x}^T(u)d\boldsymbol{\xi}_{2i}(u) + \int_0^t \exp(\boldsymbol{\beta}_0^T\boldsymbol{z}(u))d\boldsymbol{\xi}_{3i}(u) \\
&+ \int_0^t \exp\big(\boldsymbol{\beta}_0^T\boldsymbol{z}(u)\big)\boldsymbol{z}^T(u)d\Lambda_{10}(u)\tilde{\mathcal{I}}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{\xi}_{1i}(\tau).
\end{aligned}$$

As before, it can be checked that $n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_{4i}(t)$ is tight. With an application of the functional δ-method, $n^{1/2}\{g\big(\widehat{F}_1(t;\boldsymbol{x},\boldsymbol{z})\big) - g\big(F_1(t;\boldsymbol{x},\boldsymbol{z})\big)\}$ converges weakly to a zero-mean Gaussian process whose covariance function at (t_1, t_2) can be consistently estimated by (10).

References

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes. In *Lecture Notes in Statistics* 2*: Mathematical Statistics and Probability Theory* (Edited by W. Klonecki, A. Kozek and J. Rosinski), 1-25. Springer-Verlag, New York.

Andersen, P. K. and Væth, M. (1989). Simple parametric and nonparametric models for excess and relative mortality. *Biometrics* **45**, 523-535.

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

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

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.* **94**, 496-509.

Gill, R. (1980). *Censoring and Stochastic Integral*. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Statist.* **16**, 1141-1154.

Lin, D. Y. (1997) Non-parametric inference for cumulative incidence function in competing risks studies. *Statist. Medicine* **16**, 901-910.

Lin, D. Y., Fleming, T. R. and Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73-81.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

Lin, D. Y. and Ying, Z. (1995). Semiparametric analysis of general additive multiplicative hazard models for counting processes. *Ann. Statist.* **23**, 1712-1734.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate function of recurrent events. *J. R. Statist. Soc. Ser. B* **69**, 711-730.

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

Martinussen, T. and Scheike, T. H. (2002) A flexible additive multiplicative hazard model. *Biometrika* **89**, 283-298.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501-514.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology-Methodological Issues* (Edited by N. Jewell, K. Dietz and V. Farewell), 24-33. Birkhauser, Boston.

Sasieni, P. D. (1996). Proportional excess hazards. *Biometrika* **83**, 127-141.

Scheike, T. H. and Zhang, M. J. (2002). An additive-multiplicative Cox-Aalen regression model. *Scand. J. Statist.* **29**, 75-88.

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

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

Sierra, J., Perez, W. S., Rozman, C., Carreras, E., Klein, J. P., Rizzo, J. D., Davies, S. D., Lazarus, H. M., Bredeson, C. N., Marks, D. I., Boogaerts, M. A., Goldman, J., Champlin, R. E., Keating, A. Weisdorf, D. J., de Witte, T. M. and Horowitz, M. H. (2002). Bone marrow transplantation from HLA-identical siblings as treatment from myelodysplasia. *Blood* **100**, 1997-2004.

Wei, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *J. Amer. Statist. Assoc.* **79**, 649-652.

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, P.R.China.

E-mail: slq@amt.ac.cn

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

E-mail: liujingxia@yahoo.com

Department of Statistics, University of Missouri, 146, Middlebush Hall, Columbia, MO 65211, U.S.A.

E-mail: tsum@stat.missouri.edu

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

E-mail: meijie@mcw.edu

(Received September 2004; accepted May 2005)