Published in final edited form as:

Lifetime Data Anal. 2017 October ; 23(4): 605–625. doi:10.1007/s10985-016-9378-8.

A new approach to regression analysis of censored competing-risks data

Yuxue Jin and

Quantitative Marketing, Google, New York 10011, U.S.A

Tze Leung Lai

Department of Statistics, Stanford University, California 94305, U.S.A

Abstract

An approximate likelihood approach is developed for regression analysis of censored competing-risks data. This approach models directly the cumulative incidence function, instead of the cause-specific hazard function, in terms of explanatory covariates under a proportional subdistribution hazards assumption. It uses a self-consistent iterative procedure to maximize an approximate semiparametric likelihood function, leading to an asymptotically normal and efficient estimator of the vector of regression parameters. Simulation studies demonstrate its advantages over previous methods.

Keywords

Asymptotic efficiency; Cumulative incidence function; Empirical process theory; Hazard function of subdistribution; Martingale central limit theorem; Semiparametric likelihood; Volterra equation

1 Introduction

Competing-risks data are inherent in cancer studies for which the response to treatment can be classified in terms of failure from disease processes and non-disease-related causes. For example, after treatments for leukemia and lymphoma, patients may either relapse or die in remission. In quality-of-life studies, the risks of the different failure types need to be considered for cost- effectiveness analyses. Although Cox's proportional hazards model has been routinely used to analyze competing-risks data with covariates, what it models is the *cause-specific hazard rate*

$$\lambda_j(t; Z) = \lim_{h \rightarrow 0} h^{-1} P\{t \leq Y \leq t+h, J=j | Y \geq t, Z\}, \quad (1)$$

where Z is a vector of covariates, Y is the failure time and J is the cause of failure, with the causes labeled $1, 2, \dots, j^*$. This approach treats each cause separately as the only cause for the observed failures that can be attributed to it, and treats the observed failures due to other causes as censoring outcomes. Viewing competing risks as multiple endpoints each of which precludes the occurrence of the others, Fine and Gray (1999) proposed to consider, instead of (1), the hazard function of the cumulative incidence due to cause j (Gray, 1988):

$$\tilde{\lambda}_j(t; Z) = -\frac{d}{dt} \log(1 - F_j(t; Z)), \quad (2)$$

where $F_j(t; Z)$ is the *cumulative incidence function*

$$F_j(t; Z) = P\{Y \leq t, J = j | Z\}. \quad (3)$$

Note that $1 - F_j(t; Z) = 1 - P\{Y < t, J = j | Z\} = P\{Y > t | Z\} + P\{Y < t, J \neq j | Z\}$.

In the absence of covariates, the cumulative incidence function can be estimated by the Aalen–Johansen estimator based on a finite-state Markov chain model of competing risks; see the survey by Klein and Shu (2002). Although the proportional hazards model $\lambda_j(t; Z) = \lambda_j^0(t) e^{\beta^T Z}$ can be readily extended to the hazard functions $\tilde{\lambda}_j$ of the competing risks, regression analysis of $\tilde{\lambda}_j$ under the model $\tilde{\lambda}_j(t; Z) = \tilde{\lambda}_j^0(t) \exp(\beta^T Z)$ becomes much more difficult. Three approaches to this problem have been developed. Fine and Gray (1999) used the following modification of risk sets to extend partial likelihood to competing-risks data. The risk set at time t includes the subjects that are naturally at risk and those who have failed prior to time t due to other competing risks and are not censored at time t . A subject who has failed prior to time t due to other competing risks is weighted by the probability that it is not censored at time t given that it has not been censored at the failure time. The recent review by Beyersmann and Scheike (2013) lists pertinent references on the method. Scheike et al (2008) used weighted counting processes as responses in a binomial regression model for cumulative incidence functions. Both methods assume a common censoring distribution and conditions under which it can be estimated consistently. Their performance is highly sensitive to how well the censoring distribution can be estimated. Klein and Andersen (2005) circumvented this issue by using jackknife pseudo-values to model the cumulative incidence functions directly. These pseudo-values are used in a generalized estimating equation to obtain estimates of model parameters. Binder et al (2014) and Logan and Wang (2013) review subsequent development of the pseudo-value approach. Although Graw et al (2009) have shown this approach to be consistent and asymptotically normal, the method is not asymptotically efficient as there is loss of information in these pseudo-values.

To obtain an efficient estimator without first estimating the censoring distribution, we develop a self-consistent approximation to the semiparametric likelihood function for regression analysis of the cumulative incidence function on covariates under the assumption of proportional subdistribution hazards. The approximate semiparametric likelihood function is maximized by an iterative procedure that consists of two steps at each iteration. The first step estimates the baseline hazard functions of the competing risks by using Nelson–Aalen-type estimates that assume the regression parameters to be known, and the second step updates the regression parameter estimates by maximizing the log-likelihood function with the baseline hazards fixed at the first-step estimates. As will be shown in Section 2, a difficulty with carrying out this maximization is the constraint

$$1 - \sum_{j=1}^{j^*} F_j(t-; Z) \geq 0, \quad (4)$$

the left-hand side of which appears in the denominator of the jump $\Lambda_{ij}(t)$ of the cumulative cause-specific hazard of subject i associated with cause j . Thus, the semiparametric partial likelihood involves terms that have the left-hand side of (4) in the denominator, making the iteration procedure unstable and analytically intractable. To circumvent this difficulty, we modify the log-likelihood function by introducing a smooth weight function that sets to 0 terms for which the left-hand side of (4) falls below a threshold depending on n , and that increases smoothly to 1 after reaching a larger threshold. Details are given in Section 2 for the more general setting of time-varying covariates.

An asymptotic theory of the proposed estimator is developed in Section 3, where we establish its asymptotic normality and efficiency, derive consistent estimates of its asymptotic covariance matrix, and justify χ^2 -approximations to the associated likelihood ratio tests. An illustrative application to bone marrow transplantation data and simulation studies comparing the proposed estimator with previous methods are given in Section 4. Further discussion of the novelty of our approach and its extensions is given in Section 5.

2 Regression modeling of competing-risks data

To fix the ideas, consider the case of $j^* = 2$ competing risks and assume the proportional hazards model for the cumulative incidence function (3), i.e.,

$$F_j(t; Z) = 1 - \exp \left\{ - \int_0^t \exp(\beta_j^\top Z(s)) dA_j(s) \right\}, \quad j=1, 2, \quad (5)$$

with time-varying exogenous covariate vector $Z(\cdot)$ of dimension p and unknown parameters β_j and $A_j(\cdot)$. The data consist of $(T_i, \delta_i, \delta_i J_i, Z_i)$, $1 \leq i \leq n$, where $T_i = \min(Y_i, C_i)$, $\delta_i = I_{\{Y_i \leq C_i\}}$, C_i is the censoring time, Y_i is the failure time and J_i is the cause of failure of subject i . Letting $\Lambda_{ij}(t)$ denote the cause-specific cumulative hazard, at time $t = T_i$ of subject i associated with cause j , note that

$$d\Lambda_{ij}(t) = \frac{(1 - F_j(t-; Z_i)) \exp(\beta_j^\top Z_i(t)) dA_j(t)}{1 - F_1(t-; Z_i) - F_2(t-; Z_i)}, \quad (6)$$

where $dA_j(t) = a_j(t)dt$ if A_j is differentiable at t , with derivative a_j , and $dA_j(t) = \Delta A_j(t)$ if A_j has a jump at t , with jump size $\Delta A_j(t)$. In the sequel we assume that F_1 and F_2 are continuous so that $1 - F_1(t; Z_i) - F_2(t; Z_i) = \exp(-\Lambda_{1i}(t) - \Lambda_{2i}(t))$, and use β_j to denote the actual but unknown parameter value, $B_0 = (\beta_1, \beta_2)$, and b_j to denote the corresponding variable in the

likelihood function, $B = (b_1, b_2)$. Under the assumption that C_j is conditionally independent of (Y_j, J_j) given Z_j , the likelihood function can be expressed as

$$\prod_{i=1}^n (\Delta \Lambda_{i1}(T_i))^{I_{\{\delta_i=1, J_i=1\}}} (\Delta \Lambda_{i2}(T_i))^{I_{\{\delta_i=1, J_i=2\}}} \exp(-\Lambda_{i1}(T_i) - \Lambda_{i2}(T_i)), \quad (7)$$

in which Λ_{i1} and Λ_{i2} are regarded as functions of b_1, b_2 and will be denoted below by $\Lambda_{i1}(\cdot; b_1, b_2)$ and $\Lambda_{i2}(\cdot; b_1, b_2)$. Although the assumption of continuity of F_j precludes the occurrence of jumps in Λ_{ij} , we use Λ_{ij} in (7) because we approximate $d\Lambda_{ij}(T_i)$ by the jump size $\hat{\Lambda}_{ij}(T_i)$ of $\hat{\Lambda}_{ij}$ that jumps at T_i . Thus, although F_j is assumed to be continuous, its estimate $\hat{F}_j(\cdot; Z_j, B)$ is piecewise constant and only jumps at uncensored type- j event times. Define $\hat{F}_j(t; Z_j, B)$ via (5) by replacing β_j with b_j and $A_j(s)$ with $\hat{A}_j(s; B)$, which is defined below in (8). Similarly define $\hat{\Lambda}_{ij}(t; B)$ via (6) by replacing β_j with b_j , $A_j(t)$ with $\hat{A}_j(t; B)$ and $F_j(t; Z_j)$ with $\hat{F}_j(t; Z_j, B)$. The idea of replacing $a_j(t)dt$ by the jump size $\hat{A}_j(t)$ is also used by Zeng and Lin (2007, p. 515) in their general framework of NPMLE (nonparametric maximum likelihood estimation) which in the case $j^* = 1$ (Cox regression model) “reduces to the maximum partial likelihood estimator of β and the Breslow estimator of A ” and which will be discussed in Section 5, where the difference between their approach and ours is explained.

Let $\hat{F}_+ = \hat{F}_1 + \hat{F}_2$. To avoid giving unduly large weight to subject i when its denominator $1 - \hat{F}_+(t; Z_i, B)$ is too small, we follow Lai and Ying (1991, 1994) and introduce a weight function $p_n(1 - \hat{F}_+(t; Z_i, B))$ to smoothly truncate the observed failures at which $1 - \hat{F}_+$ falls below some threshold. Specifically, we multiply $\log(\hat{\Lambda}_{ij}(T_i))$ by $p_n(1 - \hat{F}_+(t; Z_i, B))$, which automatically enforces the constraint $F_1 + F_2 = 1$ in the modified log-likelihood. The function p_n is a nondecreasing, twice continuously differentiable function on $[0, 1]$ such that $p_n(u) = 0$ for $u \leq \varepsilon_n$ and $p_n(u) = 1$ for $u \geq 2\varepsilon_n$, with $\varepsilon_n > 0$ such that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Details of its choice are given in the first paragraph of Section 4.1.

We use an iterative algorithm to maximize the modified log-likelihood function; each iteration of which consists of the following two steps:

Step 1—At the k th iteration, given the current estimates $B^{(k-1)} = (\hat{b}_1, \hat{b}_2)$ of the regression parameters, estimate A_j by $\hat{A}_j(t; B^{(k-1)})$, where

$$\hat{A}_j(t; B) = \int_0^t \frac{\sum_{i=1}^n p_n(1 - \hat{F}_+(s; Z_i, B))}{\hat{S}_j(s; B)} dN_{ij}(s), \quad (8)$$

is the Breslow-type estimator of $A_j(t)$, $N_{ij}(t) = I_{\{T_i \leq t, \delta_i=1, J_i=j\}}$ and

$$\hat{S}_j(s; B) = \sum_{i=1}^n I_{\{T_i \geq s\}} p_n(1 - \hat{F}_+(s-; Z_i, B)) \frac{\{1 - \hat{F}_j(s-; Z_i, B)\} \exp(b_j^\top Z_i(s))}{1 - \hat{F}_+(s-; Z_i, B)}, \quad (9)$$

in which $\hat{F}_j(\cdot; Z_i, B)$ is a step function whose jumps are related to those of the step function $\hat{A}_j(\cdot; B)$ via (5) with $\beta_j = \hat{b}_j$. Note that the step functions $\hat{A}_j$ and $\hat{F}_j$ can be defined recursively at the ordered jumps $T_{(1)} < T_{(2)} < \dots$, beginning with $T_{(1)}$. This is similar to the forward recursive formula in Appendix A.2. in Zeng and Lin (2007).

Step 2—In the likelihood function (7), replace Λ_{ij} defined in (6) by $\hat{\Lambda}_{ij}$, which is the corresponding function of (b_1, b_2) with $\hat{A}_j(\cdot; B^{(k-1)})$ and $\hat{F}_j(\cdot; Z_i, B)$ replacing $A_j(\cdot)$ and $F_j(\cdot; Z_i)$. Note that $\hat{F}_j(\cdot; Z_i, B)$ varies in (b_1, b_2) only through $\exp(\beta_j^\top Z_i(\cdot))$ while $\hat{A}_j(\cdot; B^{(k-1)})$ is fixed at $B^{(k-1)}$. We update the estimates of the regression parameters by maximizing over b_1, b_2 the modified log-likelihood

$$\begin{aligned} \mathcal{L}(b_1, b_2) &= \sum_{i=1}^n \sum_{j=1}^2 \left[p_n(1 - \hat{F}_+(T_i-; Z_i, B^{(k-1)})) I_{\{\delta_i=1, j_i=j\}} \log(\Delta \hat{\Lambda}_{ij}(T_i; b_1, b_2)) \right. \\ &\quad - \sum_{i': T_{i'} \leq T_i} p_n(1 - \hat{F}_+(T_{i'}-; Z_i, B^{(k-1)})) \Delta \hat{\Lambda}_{ij}(T_{i'}; b_1, b_2) \left. \right] \end{aligned} \quad (10)$$

The iterative procedure stops when the change in the modified log-likelihood function (10) or in the updated $(\hat{b}_1, \hat{b}_2)$ falls below a prespecified threshold. Note that $\hat{\Lambda}_{ij}(T_i; b_1, b_2)$ and therefore $\hat{\Lambda}_{ij}(T_i; b_1, b_2)$ are also smooth functions of b_1 and b_2 that appear in $\exp(b_j^\top Z_i(\cdot))$. Taking partial derivatives with respect to the components of b_1 and b_2 to maximize (10), we can compute the maximizer by solving the estimating equation

$$U_j(b_1, b_2) = 0, \text{ for } j=1, 2, \quad (11)$$

where $U_j(b_1, b_2) = (\partial / \partial b_j) \mathcal{L}(b_1, b_2)$ and $\partial / \partial b_j$ denotes the gradient vector of partial derivatives with respect to the components of b_j . The partial derivative treats $p_n(\cdot)$ and $\hat{A}_j(\cdot)$ in the modified log-likelihood as free of (b_1, b_2) , in view of the iterative scheme (10), and only $\hat{\Lambda}_{ij}(\cdot; b_1, b_2)$ as varying with (b_1, b_2) through $(1 - \hat{F}_j(\cdot; Z_i, b_1, b_2)) / (1 - \hat{F}_+(\cdot; Z_i, b_1, b_2))$ and $\exp(b_j^\top Z_i(\cdot))$.

3 Asymptotic theory

3.1 Assumptions, notations, and representation of the score statistic

Assume the continuity of A_j on $[0, \tau)$ and a given compact set K that is known to contain $(\beta_1^\top, \beta_2^\top)^\top$ in its interior, where $\tau = \sup\{t : P(T_i < t) < 1\}$, recalling that $T_i = \min(Y_i, C_i)$ are i.i.d. In addition, assume that (Z_i, C_i) are i.i.d. random vectors, and that for $j = 1, 2$, F_j has a continuously differentiable density function f_j on $[0, \tau)$ satisfying

$$\int_0^{\tau-\rho} \sup_{s \leq t \leq s+\rho} |f_j'(t)| ds < \infty \text{ for some } \rho > 0. \quad (12)$$

This regularity condition, which is similar to those assumed by Lai and Ying (1991, 1992, 1994), is used to analyze U_j/b_K and thereby establish the asymptotic linearity of the score function $(U_1^\top(B), U_2^\top(B))^\top$ in a small neighborhood of B_0 . Moreover, analogous to Ying and Lai's approach, the non-decreasing, twice continuously differentiable function p_n is chosen to satisfy the following conditions:

$$(C1) \quad p_n(u) = 0 \text{ for } u \leq \varepsilon_n, p_n(u) = 1 \text{ for } u \geq 2\varepsilon_n, \text{ with } 0 < \varepsilon_n \rightarrow 0.$$

$$(C2) \quad \max_{\varepsilon_n \leq u \leq 2\varepsilon_n} \{n^{-a}|p_n'(u)| + n^{-2a}|p_n''(u)|\} = O(1), \text{ for some } 0 < a < 1/6.$$

Let $U(B) = (U_1^\top(B), U_2^\top(B))^\top$. Differentiating the right hand side of (10), in which we treat $p_n(\cdot)$ and $\hat{A}_j(\cdot)$ as free of (b_1, b_2) , with respect to b_j gives the representation

$$U_j(B) = \sum_{i=1}^n \left\{ \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i, B)) \hat{\eta}_{ij}(t; B) (dN_{ij}(t) - d\hat{\Lambda}_{ij}(t; B)) \right. \\ \left. + \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i, B)) \hat{\phi}_{ij}(t; B) (dN_{i\bar{j}}(t) - d\hat{\Lambda}_{i\bar{j}}(t; B)) \right\}, \quad (13)$$

in which $j \in \{1, 2\}$. The $\hat{\eta}_{ij}$ and $\hat{\phi}_{ij}$ are given explicitly by

$$\hat{\eta}_{ij}(t; B) = Z_i(t) - \int_0^t Z_i(s) \exp(b_j^\top Z_i(s)) d\hat{A}_j(s; B) + \frac{(1 - \hat{F}_j(t-; Z_i, B)) \int_0^t Z_i(s) \exp(b_j^\top Z_i(s)) d\hat{A}_j(s; B)}{1 - \hat{F}_1(t-; Z_i, B) - \hat{F}_2(t-; Z_i, B)}, \quad (14)$$

$$\hat{\phi}_{ij}(t; B) = \frac{(1 - \hat{F}_j(t-; Z_i, B)) \int_0^t Z_i(s) \exp(b_j^\top Z_i(s)) d\hat{A}_j(s; B)}{1 - \hat{F}_1(t-; Z_i, B) - \hat{F}_2(t-; Z_i, B)}. \quad (15)$$

Recalling that the (Z_i, C_i) are assumed to be i.i.d., we shall use (T, δ, Z) to denote an independent copy of (T_i, δ_i, Z_i) .

3.2 Uniform law of large numbers for U_j and U_j/b_k

A key step in our development of the semiparametric likelihood theory is the uniform convergence of $n^{-1}U_j(b_1, b_2)$ and its partial derivatives $n^{-1}U_j/b_k$. This result involves the solution $(\tilde{A}_j(t; B), \tilde{F}_j(t; Z, B), j = 1, 2)$ of the system of integral equations (in t , for fixed B):

$$\begin{aligned}\tilde{A}_j(t; B) &= \int_0^t \frac{E\{p_n(1-\tilde{F}_+(s; Z, B))dN_j(s)\}}{E\{I_{\{T \geq s\}} \pi_n(1-\tilde{F}_+(s; Z, B))(1-\tilde{F}_j(s; Z, B))\exp(b_j^\top Z(s))\}} \\ \tilde{F}_j(t; Z, B) &= 1 - \exp\left\{-\int_0^t \exp(b_j^\top Z(s))d\tilde{A}_j(s; B)\right\} \quad j=1, 2,\end{aligned}\quad (16)$$

where $\tilde{F}_+ = \tilde{F}_1 + \tilde{F}_2$ and $\pi_n(u) = p_n(u)/u$. In view of the smoothness assumptions on F_j and p_n and conditions (C1)–(C2), the solution of (16) exists and is unique by the Picard-Lindelöf theorem, and is continuously differentiable in the components of B (Coddington and Levinson, 1955). Let

$$\tilde{A}_{ij}(t; Z, B) = \int_0^t I_{\{T \geq s\}} \pi_n(1-\tilde{F}_+(s; Z, B)) \left\{1-\tilde{F}_j(s; Z, B)\right\} e^{b_j^\top Z(s)} d\tilde{A}_j(s; B), \quad (17)$$

and define $\eta_{ij}(t; B)$ and $\phi_{ij}(t; B)$ by (14) and (15) with $\tilde{A}_j, \tilde{F}_1$ and $\tilde{F}_2$ replacing $\hat{A}_j, \hat{F}_1$ and $\hat{F}_2$. Define

$$\begin{aligned}\tilde{U}_j(B) = E \left\{ \int_0^\tau p_n(1-\tilde{F}_+(t; Z, B)) \eta_j(t; Z, B) \left(dN_j(t) - d\tilde{A}_j(t; Z, B) \right) + \int_0^\tau p_n(1-\tilde{F}_+(t; Z, B)) \phi_j(t; Z, B) \left(dN_j(t) - d\tilde{A}_j(t; Z, B) \right) \right\}.\end{aligned}\quad (18)$$

In the Appendix we prove the following lemma that enables us to derive the asymptotic linearity of $U_j(B)$ via $\tilde{U}_j(B)$.

Lemma 1—For $j, k = 1, 2$,

$$\sup_{\{b_1, b_2\} \in K} \left\{ |n^{-1}U_j(b_1, b_2) - \tilde{U}_j(b_1, b_2)| + |(\partial/\partial b_k)U_j(b_1, b_2)/n - (\partial/\partial b_k)\tilde{U}_j(b_1, b_2)| \right\} \xrightarrow{P} 0. \quad (19)$$

Let $\tilde{U}(B) = (\tilde{U}_1^\top(B), \tilde{U}_2^\top(B))^\top$ and define

$$\Sigma = \begin{pmatrix} (\partial/\partial b_1)\tilde{U}_1(\beta_1, \beta_2) & (\partial/\partial b_2)\tilde{U}_1(\beta_1, \beta_2) \\ (\partial/\partial b_1)\tilde{U}_2(\beta_1, \beta_2) & (\partial/\partial b_2)\tilde{U}_2(\beta_1, \beta_2) \end{pmatrix}. \quad (20)$$

Since $\sup_{(b_1, b_2) \in K} |n^{-1}U_j(b_1, b_2) - \tilde{U}_j(b_1, b_2)| \xrightarrow{P} 0$ by Lemma 1, it follows that $\hat{B}$ converges to B_0 in probability under the assumption

$$\tilde{U}(b_1, b_2) \neq 0 \text{ for } (b_1, b_2) \in K - \{(\beta_1, \beta_2)\}. \quad (21)$$

Since $U_j(\hat{B}) = 0$ for $j = 1, 2$, it follows from Lemma 1 and (20) that

$$\begin{aligned} -n^{-1/2}U(B_0) &= n^{-1/2} \left\{ U(\hat{B}) - U(B_0) \right\} \\ &= \left\{ n^{-1} \int_0^1 \dot{U}(B_0 + t(\hat{B} - B_0)) dt \right\} \sqrt{n} \begin{pmatrix} \hat{\beta}_1^\top - \beta_1^\top & \hat{\beta}_2^\top - \beta_2^\top \end{pmatrix}^\top \\ &= (\Sigma + o_p(1)) \sqrt{n} \begin{pmatrix} \hat{\beta}_1^\top - \beta_1^\top & \hat{\beta}_2^\top - \beta_2^\top \end{pmatrix}^\top. \end{aligned} \quad (22)$$

in which we use $\dot{U}$ to denote the Jacobian matrix of partial derivative of U with respect to the components of (b_1, b_2) and apply the mean value theorem for vector-valued functions (Lang, 1968).

3.3 Volterra equation and weak convergence to Gaussian martingale

Let $M_{ij} = N_{ij} - A_{ij}$. Then $\{M_{ij}(t), t \geq 0\}$ is a martingale. Replacing B by B_0 , $\hat{F}_j$ by F_j and $\hat{A}_j$ by A_j in (14) and (15) gives the population counterparts $\eta_{ij}(t)$ and $\phi_{ij}(t)$ of $\hat{\eta}_{ij}(t; B_0)$ and $\hat{\phi}_{ij}(t; B_0)$. Denoting $\hat{A}_j(\cdot; B_0)$ and $\hat{F}_j(\cdot; Z, B_0)$ simply by $\hat{A}_j(\cdot)$ and $\hat{F}_j(\cdot; Z)$, define $\mathcal{A}_n(t) = n^{1/2}(\hat{A}_1(t) - A_1(t), \hat{A}_2(t) - A_2(t))^\top$,

$$\begin{aligned} S_j(t) &= E \left[I_{\{T \geq t\}} e^{\beta_j^\top Z(t)} \{1 - F_j(t; Z)\} \pi_n(1 - F_+(t; Z)) \right], \\ W_n(t) &= n^{-1/2} \left(\int_0^t \sum_{i=1}^n p_n(1 - F_+(s, Z_i)) (dM_{i1}(s)/S_1(s), dM_{i2}(s)/S_2(s)) \right)^\top. \end{aligned}$$

Letting $\pi_n^*(t; Z) = p_n(1 - F_+(t; Z))/[1 - F_+(t; Z)]^2$, define the 2×2 matrix $d\mathcal{W}_n(t)$ whose (i, j) th element is

$$E \left\{ I_{\{T \geq t\}} e^{2\beta_j^\top Z(t)} F_j(t; Z) (1 - F_j(t; Z)) \pi_n^*(t; Z) \right\} \frac{dA_j(t)}{S_j(t)} \text{ if } i=j, \\ E \left\{ I_{\{T \geq t\}} e^{\beta_1^\top Z(t) + \beta_2^\top Z(t)} (1 - F_1(t; Z)) (1 - F_2(t; Z)) \pi_n^*(t; Z) \right\} \frac{dA_i(t)}{S_i(t)} \text{ if } i \neq j. \quad (23)$$

In the Appendix, we prove the following.

Lemma 2— $\mathcal{A}_n(t)$ satisfies the Volterra equation

$$\mathcal{A}_n(t) = W_n(t) - \int_0^t d\Psi_n(s) \mathcal{A}_n(s-) + o_p(1), t < \tau, \quad (24)$$

and is given by

$$\mathcal{A}_n(t) = \int_0^t \exp \{ -(\Psi_n(t) - \Psi_n(s)) \} dW_n(s) + o_p(1), \quad (25)$$

in which the exponential function of a matrix is defined by an infinite series of matrix products and the $o_p(1)$ term refers to a remainder that converges to 0 in probability as $n \rightarrow \infty$. Hence

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t p_n(1 - (F_1 + F_2)(s; Z_i)) [(\eta_{i1}(s), \phi_{i2}(s)) dM_{i1}(s), (\eta_{i2}(s), \phi_{i1}(s)) dM_{i2}(s)], \mathcal{A}_n^\top(t) \right), \quad (26)$$

0 $\leq t < \tau$, converges weakly to a Gaussian martingale in the Skorohod space $D^k[0, \tau]$ as $n \rightarrow \infty$, where $k = 6$ denotes the dimensionality of the vector (26).

3.4 Asymptotic normality of $n^{-1/2}U_f(B_0)$ and $\hat{\beta}_j$

In the Appendix we make use of (13), (24) and Lemma 2 to prove the asymptotic normality of the score function at the true regression parameter values, analogous Cox regression in proportional hazards models.

Lemma 3—

$$n^{-1/2} (U_1^\top(B_0), U_2^\top(B_0))^\top \xrightarrow{\mathcal{D}} N(0, V), \quad (27)$$

in which $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and V is the limiting covariance matrix.

Note that the second and third equalities in (22) are basically the asymptotic linearity property of $U(B)$ mentioned at the beginning of Section 3.1. Combining (22) with Lemma 3 yields the following.

Theorem 1—Suppose F_1 and F_2 have continuously differentiable density functions f_1 and f_2 on $[0, \tau]$ satisfying (12), and that (21) holds and Σ is nonsingular, where Σ is defined in (20). Then

$$\sqrt{n}(\hat{\beta}_1^\top - \beta_1^\top, \hat{\beta}_2^\top - \beta_2^\top)^\top \xrightarrow{\mathcal{D}} N(0, \Sigma^{-1} V (\Sigma^{-1})^\top). \quad (28)$$

3.5 Asymptotic efficiency and large-sample inference

In the Appendix we show that $\Sigma = -V$, where V is the covariance matrix of the limiting, normal distribution in Lemma 3, in the proof of the following theorem on the asymptotic efficiency of $\hat{B}$ in the sense of attaining the information bound for the asymptotic variance of a regular estimator.

Theorem 2—Under the assumptions of Theorem 1, and with $\mathcal{L}$ defined by (10),

$$2 \{ \mathcal{L}(\hat{B}) - \mathcal{L}(B_0) \} \xrightarrow{\mathcal{D}} \chi_{2p}^2. \quad (29)$$

Moreover, for any regular estimator $(\hat{\beta}_1^\top, \hat{\beta}_2^\top)$ of $(\beta_1^\top, \beta_2^\top)$, the limiting distribution of $\sqrt{n}(\hat{\beta}_1^\top - \beta_1^\top, \hat{\beta}_2^\top - \beta_2^\top)^\top$ is the convolution of $N(0, -\Sigma^{-1})$ and another distribution as $n \rightarrow \infty$.

The $p \times 2$ parameter matrix $B = (b_1, b_2)$ takes value in a $2p$ -dimensional linear space Ω , which explains why the limiting χ^2 -distribution in (29) has $2p$ degrees of freedom. In view of (29), $\{B: 2(\mathcal{L}(\hat{B}) - \mathcal{L}(B)) < \chi_{2p, 1-\alpha}^2\}$ is an approximate $(1 - \alpha)$ -level confidence ellipsoid for B . Moreover, to test the null hypotheses $H_0: B \in \Omega_0$ at level α , where Ω_0 is a $(2p - m)$ -dimensional submanifold of Ω , we can use the approximate likelihood ratio test that rejects H_0 if $2\{\mathcal{L}(\hat{B}) - \mathcal{L}(\hat{B}_0)\} \geq \chi_{m, 1-\alpha}^2$, where $\hat{B}_0$ is the approximate maximum likelihood estimator of B under the constraint that it belongs to Ω_0 . The derivation of the limiting χ_m^2 -distribution is given in the last paragraph of the Appendix. Computation of $\hat{B}_0$ uses the same iterative algorithm as that in Section 2, but the maximization in Step 2 of each iteration is under the constraint $B \in \Omega_0$.

4 Numerical studies and discussion

4.1 Comparative simulation study of estimators

This section describes a simulation study of the iterative maximum likelihood method in Section 2, denoted by ML. For comparison, we also consider Fine and Gray's method, denoted by FG, which is implemented by using the R package *cmprsk*, and Klein and Andersen's method, denoted by KA, which is implemented by using the R functions in Klein et al (2008), assuming the working independence model for the jackknife pseudo-values in the generalized estimating equations. We study three scenarios involving a univariate time-invariant covariate Z and the cumulative incidence functions

$$F_j(t; Z) = 1 - \{1 - \rho_j(1 - e^{-t})\}^{\exp(Z\beta_j)}, \quad j=1, 2. \quad (30)$$

To simulate a failure time and a failure cause, we first simulate the failure time T according to the total failure probability $F_+(t; Z) = F_1(t; Z) + F_2(t; Z)$, and then assign probability $(dF_j/dF_+)(T; Z)$ for the j th cause of failure. We also use the following weight function. Letting $p(u) = 0$ for $u = 0$ and $p(u) = 1$ for $u = 1$, define $p_n(t) = p(2c^{-1}n^\gamma t - 1)$. In particular, we use $p(u) = 6u^5 - 15u^4 + 10u^3$, $0 < u < 1$, and $c = 0.1$ and $\gamma = 0.16$ in the numerical examples in Sections 4.1 and 4.2.

In the three scenarios, Z is Bernoulli with success probability $1/2$. The censoring times are uniformly distributed on an interval (B_L, B_U) . The model parameters, sample size n and censoring rates r are listed in Table 1, where each result is based on 10,000 simulations. Table 1 shows that in Scenarios I and II, ML has smaller variances and mean squared errors for the regression parameter estimates than FG and KA, whose pseudo-values incur some loss of information. In Scenario 4 and FG and ML have comparable mean squared errors, which are markedly smaller than those of KA when $n = 100$.

4.2 Application to a bone marrow transplantation study

We apply our approach to a bone marrow transplantation study that compares the competing risks of death in remission and relapse after transplant using three different types of donors. The data set consists of 415 patients whose transplants were performed at Stanford Hospital and Clinics between 1986 and 1997. Among them, 156 patients were treated for acute lymphoblastic leukemia, 130 for acute myelogenous leukemia, and 130 for chronic myelogenous leukemia. Before transplants, 219 patients were at an early stage of disease (defined as in the first remission or in the chronic phase) and 197 were at an advanced stage. The median age of patients at transplant was 27 years. Patients were followed up to five years and there were 93 patients who relapsed and 139 deaths in remission. The regression covariates are time-invariant and include the patient's age at transplant (after centering at the sample mean) and indicator variables for the disease and donor-type types shown in Table 2. The disease types are acute (A) and chronic (C); "acute" consists of acute lymphoblastic and acute myelogenous leukemia, "chronic" refers to chronic myelogenous leukemia. The donor types are HLA-identical sibling (SIBL) with 303 cases, partially mismatched related (PMR)

with 25 cases, and HLA-matched unrelated (UR) with 87 cases. This explains the interaction covariates PMR-A, UR-A, etc, in Table 2. The reference group in the assumed proportional hazards model is SIBL-A with age 26 years at transplant and at an early stage of the disease. The results in Table 2 show significant effects of stage and the UR-C class membership on relapse, and of age and the UR-C class membership on death in remission over the reference group described at the end of Table 2.

A diagnostic check of the proportionality assumption (5) in the cumulative incidence functions is provided by the plot of the residuals $\tilde{T}_{ij} = \tilde{\Lambda}_{ij}(T_j)$, where $\tilde{\Lambda}_{ij}$ is the Nelson-Aalen estimate of the cause-specific cumulative hazard $\Lambda_{ij}(s)$, which under (5) is given by

$$\Lambda_{ij}(t) = \int_0^t \frac{1 - F_j(s; Z_i)}{1 - F_1(s; Z_i) - F_2(s; Z_i)} \exp(\beta_j^\top Z_i(s)) dA_j(s)$$

for subject i and cause $j = 1$ (relapse) or 2 (death in remission). Noting that in the uncensored case the random variables $\tilde{T}_{ij}, i = 1, \dots, n$, constitute a sample of n independent exponential variables with common mean 1 when $\tilde{\Lambda}_{ij}(s)$ is replaced by the true cause-specific hazard $\Lambda_{ij}(s)$, Kay (1977) proposed to use the Nelson-Aalen plot of these residuals to check if they are approximately exponentially distributed; see also Andersen et al (1993, pp. 448 and 555). Figure 1 gives the Nelson-Aalen plots of the residuals of relapse and death in remission at advance stage of disease. The line with unit slope, which corresponds to the cumulative function of exponential random variables, lies within the 95% pointwise confidence bands. This is also true for patients at an early stage and also for the plots stratified by the other covariates, which are not shown here.

Although we have focused on the application of our approach to the analysis of these data, we have also examined whether the conclusions would change when other methods are applied. Table 2 includes the results for these methods, and the p -values in the table show that FG, KL, and Cox regression also give significant effects of the covariates that are significant in the ML analysis.

5 Discussion

Zeng and Lin (2007, 2010) have developed a general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data, establishing its asymptotic normality and efficiency. Their theory is about directly maximizing the semiparametric likelihood function that contains the nonparametric component to have jumps at uncensored data, and is particularly effective for transformation models. They rely exclusively on modern empirical process theory and point out that “the counting process martingale theory, which has been the workhorse behind the development of survival analysis in the last quarter of a century, plays no role” in this development. In contrast, our approach has relied on both martingale theory in Sections 3.1 and 3.2 and modern empirical process theory in Section 3.3.

Unlike the applications considered by Zeng and Lin, for regression analysis of censored competing-risks data involving cumulative incidence functions, directly maximizing the likelihood function is analytically intractable and computationally unstable because of the potentially vanishing denominator in (6). The modified likelihood approach in Section 2 circumvents this difficulty by using the weights $p_n(1 - \hat{F}_+(s-; Z_i, B))$ in (8) and (9). Instead of directly maximizing the likelihood function, it is a score function approach that involves estimating equations. The weights reduce to 1 when the $\hat{F}_+(s-; Z_i, B)$ are not too close to 1, and therefore the modified likelihood is a smooth approximation of the likelihood function.

This method to handle small risk set sizes in regression analysis of censored survival data dates back to Lai and Ying (1991, 1992, 1994) and involves exponential inequalities of empirical processes developed by Lai and Ying (1988). It leads to asymptotically efficient estimators which have efficiency gains over previous $\sqrt{n}$ -consistent estimators that are reviewed in the second paragraph of Section 1 and implemented in the comparative simulation study of Section 4.1. To ensure convergence of the iterative procedure in Section 2 to solve the estimating equations (11), one can start with a $\sqrt{n}$ -consistent preliminary estimator, e.g., one of the estimators in the second paragraph of Section 1. The competing-risks regression model considered herein is easier to analyze than those considered by Lai and Ying (1991, 1992, 1994) in the following aspect. Because the covariate effects on the cumulative incidence functions are of the proportional-hazards type, the estimators $\hat{A}_j$ and $\hat{S}_j$ in (8) and (9) amount to exponential tilting (reweighting) of the point masses of uncensored observations that define the counting processes $N_{j\cdot}$ yielding smooth objective functions of the regression parameters in the modified likelihood (10) or the score function (11). In contrast, the covariate effects of the accelerated failure time models in Lai and Ying (1991, 1992, 1994) are additive and result in estimating functions that have jump discontinuities in the regression parameters, requiring more elaborate exponential bounds from empirical process theory to derive asymptotic linearity results of the type in (22). On the other hand, the estimators (8) and (9) are defined recursively over the uncensored event times, from the smallest to the largest one, and this aspect is not encountered in the analysis of Lai and Ying. The novelty of our method to address this lies in the limiting (or population) versions, defined by the system of integral equations (16), of the recursions based on the sample data. Zucker (2005) also faces a similar problem in extending Cox regression to the case where the covariates are subject to error. His pseudo-partial likelihood approach leads to estimators that involve $\int_0^t \dots \sum_{i=1}^n dN_i(s)$; see Zucker (2005, p. 1274). By showing that these random functions of t and b form a uniformly bounded and equicontinuous family, he applies the Arzela-Ascoli theorem and the functional strong law of large numbers to show the existence of a unique consistent root to the pseudo-partial likelihood equation. This argument is subsequently used by Zucker et al (2008) and Gorfine et al (2009) for semiparametric shared frailty models in prospective and case-control survival studies, and by Martinussen et al (2011) in proving consistency of these parameter estimates in the Aalen additive gamma model. Because the recursions (8) and (9) are technically more difficult and involve weights $p_n(\cdot)$ to dampen the effects of small denominators, we need a more powerful functional-analytic approach that identifies (8) and $\hat{F}_j$ (defined by replacing A_j in (5) by $\hat{A}_j$) as sample versions of the system of integral equations (16) so that we can use the exponential

inequalities for empirical processes and their stochastic integrals in Lai and Ying (1988, 1991) in the associated perturbation analysis of (16).

Geskus (2011) has extended Fine and Gray's method for regression analysis of cumulative incidence functions to left truncated and right censored (LTRC) data. Since the approach of Lai and Ying (1992, 1994) also works for LTRC data, we can likewise extend the approximate likelihood approach developed herein to LTRC data.

Acknowledgments

We thank the reviewers for their careful reading of the paper and valuable comments that led to its substantial improvement. The research of the second author was supported by National Science Foundation grant DMS 1407828 and National Cancer Institute grant 1 P30 CA124435.

References

- Andersen, PK., Borgan, Ø., Gill, RD., Keiding, N. Springer Series in Statistics. Springer-Verlag; New York: 1993. Statistical Models Based on Counting Processes.
- Bellman, RE. Perturbation techniques in mathematics, engineering and physics. Holt, Rinehart & Winston; New York: 1964.
- Beyersmann, J., Scheike, TH. Classical regression models for competing risks. In: Klein, JP, van Houwelingen, HC, Ibrahim, JG., Scheike, TH., editors. Handbook of Survival Analysis, Handbooks of Modern Statistical Methods. Chapman and Hall/CRC; Boca Raton: 2013. p. 157-177.
- Binder N, Gerds TA, Andersen PK. Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime Data Anal.* 2014; 20(2):303–315. DOI: 10.1007/s10985-013-9247-7 [PubMed: 23430270]
- Coddington, EA., Levinson, N. Theory of Ordinary Differential Equations. McGraw-Hill Book Co; New York-Toronto-London: 1955.
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Amer Statist Assoc.* 1999; 94(446):496–509. DOI: 10.2307/2670170
- Geskus RB. Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring. *Biometrics.* 2011; 67(1):39–49. DOI: 10.1111/j.1541-0420.2010.01420.x [PubMed: 20377575]
- Gorfine M, Zucker DM, Hsu L. Case-control survival analysis with a general semiparametric shared frailty model—a pseudo full likelihood approach. *Annals of statistics.* 2009; 37(3):1489. [PubMed: 19684881]
- Graw F, Gerds TA, Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal.* 2009; 15(2):241–255. DOI: 10.1007/s10985-008-9107-z [PubMed: 19051013]
- Gray RJ. A class of K -sample tests for comparing the cumulative incidence of a competing risk. *Ann Statist.* 1988; 16(3):1141–1154. DOI: 10.1214/aos/1176350951
- Kay R. Proportional hazard regression models and the analysis of censored survival data. *J Roy Statist Soc Ser C Appl Statist.* 1977; 26(3):227–237. [jstor.org/stable/2346962](https://doi.org/10.2307/2346962).
- Khalil, H. Nonlinear Systems. 2. Prentice Hall; Upper Saddle River, NJ: 2002.
- Klein J, Shu Y. Multi-state models for bone marrow transplantation studies. *Statist Methods Med Res.* 2002; 11(2):117–139. DOI: 10.1191/0962280202sm277ra
- Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics.* 2005; 61(1):223–229. DOI: 10.1111/j.0006-341X.2005.031209.x [PubMed: 15737097]
- Klein JP, Gerster M, Andersen PK, Tarima S, Perme MP. SAS and R functions to compute pseudo-values for censored data regression. *Comp Methods Programs Biomed.* 2008; 89(3):289–300. DOI: 10.1016/j.cmpb.2007.11.017

- Lai TL, Ying Z. Stochastic integrals of empirical-type processes with applications to censored regression. *J Multivariate Anal.* 1988; 27(2):334–358.
- Lai TL, Ying Z. Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *Ann Statist.* 1991; 19(3):1370–1402.
- Lai TL, Ying Z. Asymptotically efficient estimation in censored and truncated regression models. *Statist Sinica.* 1992; 2(1):17–46.
- Lai TL, Ying Z. A missing information principle and M -estimators in regression analysis with censored and truncated data. *Ann Statist.* 1994; 22(3):1222–1255.
- Lang, S. *Analysis I.* Addison-Wesley Publishing Co; Boston: 1968.
- Logan, BR., Wang, T. Pseudo-value regression models. In: Klein, JP, van Houwelingen, HC, Ibrahim, JG., Scheike, TH., editors. *Handbook of Survival Analysis, Handbooks of Modern Statistical Methods.* Chapman and Hall/CRC; Boca Raton: 2013. p. 199–219.
- Martinussen T, Scheike TH, Zucker DM. The aalen additive gamma frailty hazards model. *Biometrika.* 2011; asr049.
- Scheike TH, Zhang MJ, Gerds TA. Predicting cumulative incidence probability by direct binomial regression. *Biometrika.* 2008; 95(1):205–220. DOI: 10.1093/biomet/asm096
- van der Vaart, AW. *Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics.* Vol. 3. Cambridge University Press; Cambridge: 1998.
- Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *J R Stat Soc Ser B Stat Methodol.* 2007; 69(4):507–564. with discussion and a reply by the authors. DOI: 10.1111/j.1369-7412.2007.00606.x
- Zeng D, Lin DY. A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statist Sinica.* 2010; 20(2):871–910.
- Zucker DM. A pseudo-partial likelihood method for semiparametric survival regression with covariate errors. *Journal of the American Statistical Association.* 2005; 100(472):1264–1277.
- Zucker DM, Gorfine M, Hsu L. Pseudo-full likelihood estimation for prospective survival analysis with a general semiparametric shared frailty model: Asymptotic theory. *Journal of Statistical Planning and Inference.* 2008; 138(7):1998–2016.

Appendix

Proof. of Lemma 1

To begin with, we have applied at the beginning of Section 3.2 the theory of integral equations to show the existence of $(\tilde{A}_j, \tilde{F}_j)$, $j = 1, 2$. We now use arguments involving exponential inequalities of empirical processes to relate quantities like $\hat{A}_j(t; B)$ and $\hat{F}_j(s-; Z_i, B)$ to their population versions, similar to those in Lai & Ying (1991, Appendix A; 1994, Section 4.2) to show that there exists $\rho > 0$ such that

$$\sup_{\|b_1 - \beta_1\| + \|b_2 - \beta_2\| \leq \rho, 0 < t < \tau} \left\{ \left| \hat{A}_j(t; B) - \tilde{A}_j(t; B) \right| + \|D\hat{A}_j(t; B) - D\tilde{A}_j(t; B)\| \right\} = O_p(n^{-1/2}) \quad \text{for } j=1, 2,$$

(31)

where Dg denotes the Jacobian matrix of the partial derivatives of g with respect to the components of B and $\|C\| = \max |c_{ij}|$ is the max-norm of a matrix $C = (c_{ij})$. A key tool in these arguments is provided by the aforementioned empirical process theory that enables us to approximate the averages of random functions by their expectations. Another important

tool is to introduce the smooth weight function p_n that satisfies (C1) and (C2) to circumvent difficulties when the denominator $1 - \hat{F}_+(s-; Z_i, B)$ becomes too small. The smoothness of $\hat{S}_j$ and $\hat{A}_j$ makes the arguments considerably simpler than those in Lai and Ying (1991, 1994) since we can differentiate them with respect to the components of B and use assumptions (12) and (C2) to bound the derivatives. Before analyzing the partial derivatives in $D\hat{A}_j - D\tilde{A}_j$, we first prove the result for $\hat{A}_j - \tilde{A}_j$ in (31) and the corresponding result for $\hat{F}_j - \tilde{F}_j$. Recalling that $\pi_n(u) = p_n(u)u$, our basic idea is to identify $\hat{A}_j(t; B)$ in (8) and $\hat{F}_j(t-; Z, B)$ as sample versions of (16) that replaces $E\{p_n(\cdot \cdot \cdot) dN_j(s)\}$ in (16) by

$$n^{-1} \sum_{i=1}^n p_n(1 - \hat{F}_+(s-; Z_i, B)) dN_{ij}(s),$$

and $E\{I_{\{T \geq s\}} \cdots \exp(b_j^T Z(s))\}$ in (16) by $n^{-1} \hat{S}_j(s; B)$ defined in (9), in which $\hat{F}_j(t-; Z_i, B)$ corresponds to the second equation in (16) with $\tilde{A}_j$ replaced by $\hat{A}_j$. Perturbation analysis (Bellman, 1964; Khalil, 2002) of the system of integral equations (16) together with exponential inequalities of empirical processes (Lai and Ying, 1988, 1991) can then be used to show uniform convergence of these sample versions of (16) to their population counterparts $\tilde{A}_j$ and $\tilde{F}_j$ at rate $O_p(n^{-1/2})$. Making use of this and the exponential inequalities for empirical processes again, we can replace sample means in the integrands and integrators in the stochastic integral representation (35) of $n^{-1/2} U_j(B)$, thereby proving the uniform convergence of $n^{-1} U_j(b_1, b_2) - \tilde{U}_j(b_1, b_2)$ in (19). Since $U_j(B)$ and $\tilde{U}_j(B)$ are smooth functions of the components of B , $D\hat{A}_j - D\tilde{A}_j$ and $D\hat{F}_j - D\tilde{F}_j$ can be analyzed similarly, thereby proving (31), which can then be used to prove the corresponding result in (/ b_n) $U_j(b_1, b_2)/n$ in (19).

Proof. of Lemma 2

Let $\hat{p}_n(t-; Z_j) = p_n(1 - \hat{F}_+(t-; Z_j, B_0))$ and define $p_n(t-; Z_j)$ similarly with $\hat{F}_+$ replaced by F_+ . Note that

$$\begin{aligned} \sqrt{n}(d\hat{A}_j(t) - dA_j(t)) &= \frac{1}{n^{-1} \hat{S}_j(t)} n^{-1/2} \sum_{i=1}^n [\hat{p}_n(t-; Z_i)(dN_{ij}(t) - dA_{ij}(t)) \\ &- I_{\{T_i \geq t\}} e^{\beta_j^T Z_i(t)} \left(\frac{\hat{p}_n(t-; Z_i)(1 - \hat{F}_j(t-; Z_i))}{1 - \hat{F}_i(t-; Z_i) - \hat{F}_2(t-; Z_i)} - \frac{\hat{p}_n(t-; Z_i)(1 - F_j(t; Z_i))}{1 - F_1(t; Z_i) - F_2(t; Z_i)} \right) dA_j(t)]. \end{aligned} \quad (32)$$

To simplify the arguments and highlights the main ideas, we focus on the case where Z_i does not vary with time, for which $1 - F_j(t; Z_i) = \exp(-e^{\beta_j^T Z_i} A_j(t))$, thus expressing $F_j(\cdot; Z_i)$ as a smooth increasing function of $A_j(\cdot)$. In this case, the right-hand side of (32) has two terms, the first of which is

$$n^{-1/2} \sum_{i=1}^n \frac{p_n(t-; Z_i)}{S_j(t)} dM_{ij}(t) + n^{-1/2} \sum_{i=1}^n \left\{ \frac{\hat{p}_n(t-; Z_i)}{n^{-1} \hat{S}_j(t)} - \frac{p_n(t-; Z_i)}{S_j(t)} \right\} dM_{ij}(t). \quad (33)$$

The second summand in (33) can be written as $n^{-1/2} \sum_{i=1}^n (\varepsilon_{n,ij}(t)/S_j(t)) dM_{ij}(t)$, in which $\varepsilon_{n,ij}(t)$ is a predictable process such that $\max_{t < \tau, 1 \leq i \leq n} |\varepsilon_{n,ij}(t)|/p_n(t-; Z_i) \xrightarrow{P} 0$. Similarly, the second term on the right-hand side of (32) can be expressed as

$$\left\{ n^{1/2} (\hat{A}_j(t) - A_j(t)) \psi_{jj}(t) + n^{1/2} (\hat{A}_{\bar{j}}(t) - A_{\bar{j}}(t)) \psi_{\bar{j}j}(t) + \delta_{n,j}(t) \right\} dA_j(t)/S_j(t),$$

in which $\bar{j} = j$, $\delta_{n,j}(t)$ is a predictable process such that $\int_0^\tau (|\delta_{n,j}(t)|/S_j(t)) dA_j(t) \xrightarrow{P} 0$, and $\psi_{jj}(t) dA_j(t)/S_j(t)$ and $\psi_{\bar{j}j}(t) dA_j(t)/S_j(t)$ are the (j, j) th and $(\bar{j}, j)$ th elements of $d\Psi_n(t)$, respectively. This follows from applying the first-order Taylor expansion to the summands and then the law of large numbers to the sum in the second term on the right-hand side of (32).

Expressing $1 - F_j(\cdot; Z_j)$ as a function of $A_j(\cdot)$ as noted at the beginning of last paragraph, we can take the derivative of the quotient in (32) with respect to A_j and also with respect to $A_{\bar{j}}$ because the denominator involves both F_j and $F_{\bar{j}}$ explaining the formula in (23) in which the expectation is a consequence of the law of large numbers. Replacing the $o_p(1)$ term in (24) by

$$\int_0^t \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varepsilon_{n,i1}(s) \frac{dM_{i1}(s)}{S_1(s)}, \varepsilon_{n,i2}(s) \frac{dM_{i2}(s)}{S_2(s)} \right)^\top + \int_0^t \left(\delta_{n,1}(s) \frac{dA_1(s)}{S_1(s)}, \delta_{n,2}(s) \frac{dA_2(s)}{S_2(s)} \right)^\top,$$

the solution (25) of (24) follows from Theorem II.6.3 in Andersen et al (1993).

For time-varying Z_j , we can modify the preceding argument. Although the basic idea is the same, the technical details are more complicated as the Taylor expansion has to be carried out directly instead of by differentiation with respect to A_j ; see (34) below for example. Since $dN_{j1}(t) dN_{j2}(t) = 0$, $\langle M_{j1}, M_{j2} \rangle = 0$, i.e., M_{j1} and M_{j2} are orthogonal martingales, and therefore we can apply the martingale central limit theorem (Andersen et al, 1993) and (25) to conclude that (26) converges weakly to a Gaussian martingale in $D^k[0, \tau]$.

Proof. of Lemma 3

Since $\hat{F}_j$ is defined by replacing A_j in (5) by $\hat{A}_j$, it follows that

$$\hat{F}_j(t; Z) - F_j(t; Z) = (1 - F_j(t; Z)) \left[1 - \exp \left\{ - \int_0^t e^{\beta_j^\top Z(s)} d(\hat{A}_j(s) - A_j(s)) \right\} \right]. \quad (34)$$

We next combine this with (24) and apply Taylor's theorem and the law of large numbers for the i.i.d. (Z_i, C_i) to decompose $n^{1/2}U_j(B_0)$ into six martingales summands plus a $o_p(1)$ remainder.

From (13), (24) and (5), it follows that $n^{1/2}U_j(B_0)$ can be written as

$$\begin{aligned}
 & \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i)) \eta_{ij}(t) dM_{ij}(t) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i)) \phi_{ij}(t) dM_{i\bar{j}}(t) \\
 & - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i)) \left[\frac{1 - \hat{F}_j(t-; Z_i)}{1 - \hat{F}_1(t-; Z_i) - \hat{F}_2(t-; Z_i)} - \frac{1 - F_j(t; Z_i)}{1 - F_1(t; Z_i) - F_2(t; Z_i)} \right] \times \hat{\eta}_{ij}(t) e^{\beta_j^\top Z_i(t)} d\hat{A}_j(t) \\
 & - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i)) \left[\frac{1 - \hat{F}_{\bar{j}}(t-; Z_i)}{1 - \hat{F}_1(t-; Z_i) - \hat{F}_2(t-; Z_i)} - \frac{1 - F_{\bar{j}}(t; Z_i)}{1 - F_1(t; Z_i) - F_2(t; Z_i)} \right] \times \hat{\phi}_{ij}(t) e^{\beta_{\bar{j}}^\top Z_i(t)} d\hat{A}_{\bar{j}}(t) \\
 & - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i)) \left[\frac{\eta_{ij}(t) \{1 - F_j(t; Z_i)\}}{1 - F_1(t; Z_i) - F_2(t; Z_i)} e^{\beta_j^\top Z_i(t)} \right] d(\hat{A}_j(t) - A_j(t)) \\
 & - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{T_i} p_n(1 - \hat{F}_+(t-; Z_i)) \left[\frac{\phi_{ij}(t) \{1 - F_{\bar{j}}(t; Z_i)\}}{1 - F_1(t; Z_i) - F_2(t; Z_i)} e^{\beta_{\bar{j}}^\top Z_i(t)} \right] d(\hat{A}_{\bar{j}}(t) - A_{\bar{j}}(t)) \\
 & + o_p(1).
 \end{aligned}
 \tag{35}$$

Making use of (24), it can be shown similarly that the third summand in (35) can be expressed as

$$\begin{aligned}
 & \sqrt{n} \int_0^\tau \int_0^t E\{\eta_{ij}(t) I_{\{T_i \geq t\}} e^{\beta_j^\top (Z_i(t) + Z_i(s))} \left[\frac{F_{\bar{j}}(t; Z_i)(1 - F_j(t; Z_i))}{(1 - F_1(t; Z_i) - F_2(t; Z_i))^2} \right] \times p_n(1 - F_+(t; Z_i))\} d(\hat{A}_j(s) - A_j(s)) dA_j(t) \\
 & + \sqrt{n} \int_0^\tau \int_0^t E\{\eta_{ij}(t) I_{\{T_i \geq t\}} e^{\beta_j^\top Z_i(t) + \beta_{\bar{j}}^\top Z_i(s)} \left[\frac{(1 - F_1(t; Z_i))(1 - F_2(t; Z_i))}{(1 - F_1(t; Z_i) - F_2(t; Z_i))^2} \right] \times p_n(1 - F_+(t; Z_i))\} d(\hat{A}_{\bar{j}}(s) - A_{\bar{j}}(s)) dA_j(t) + o_p(1).
 \end{aligned}
 \tag{36}$$

Replacing $(j, \bar{j})$ by $(\bar{j}, j)$ and η_{ij} by ϕ_{ij} in (36), the fourth summand in (35) can be expressed similarly. The terms inside the fifth and the sixth summands in (35) can also be replaced by the corresponding expected values, again using the law of large numbers. Hence Lemma 3 follows from (35) and the weak convergence of (26) to a Gaussian martingale.

Proof. of Theorem 2

Although the nuisance parameters F_1 and F_2 in the semiparametric likelihood (7) are infinite-dimensional, we can use parametric submodels such that the information bound for estimating B in the parametric submodel is attained by $\hat{B}$, proving its asymptotic efficiency which is formulated in Theorem 2 in terms of the Hájek convolution theorem; see Andersen et al (1993, Sections VIII.1.3 and VIII.2.4) and van der Vaart (1998, Chapter 25). Moreover, standard likelihood theory for the least favorable parametric submodel shows that $V = -\Sigma$. In fact, if we impose the stronger assumptions that (a) the support $\mathcal{Z}$ of Z_i is bounded, (b) ess

$\sup C_j = \tau < \infty$, and (c) $\inf_{\mathcal{X} \times \mathcal{Z}} \{1 - F_1(\tau, z) - F_2(\tau, z)\} > 0$, then we do not need to introduce the weight function $p_n(\cdot)$ in (10), which then corresponds to the semiparametric log-likelihood function considered by Zeng and Lin (2007, 2010), whose asymptotic efficiency theory for semiparametric maximum likelihood estimators can be applied to this case. Removing these assumptions requires modification of the likelihood function to address the difficulties caused by small values of $1 - \hat{F}_+(T_j^-, Z_j, B)$. This is the idea, introduced by Lai and Ying (1991, 1992, 1994), behind the smooth weight function $p_n(1 - \hat{F}_+(t^-, Z_j, B))$ that is asymptotically negligible and can yield the same information bound as that under the additional restrictive assumptions (a), (b), and (c).

Fig. 1. Nelson-Aalen plots of the residuals of relapse (left) and death in remission (right) for patients at advanced stage (dotted lines). The dash-dotted lines are the 95% pointwise confidence bands; the solid line has unit slope.

Table 1

Bias, variance and mean squared error (MSE) of the estimate of β_1 (first row for each sample size) and of β_2 (second row) by the iterative maximum likelihood method (ML), Fine and Gray's method (FG) and Klein and Andersen's method (KA) in three scenarios with different censoring rates r . The numbers in the parentheses are the standard errors of the simulations.

n	Bias			Variance			MSE		
	ML	FG	KA	ML	FG	KA	ML	FG	KA
(a) Scenario I: $(\rho_1, \rho_2, \beta_1, \beta_2, B_L, B_U) = (0.25, 0.5, 1, -0.5, 0.045, 4), r = 37.1\%$									
100	-0.077 (0.0039)	0.039 (0.0042)	0.057 (0.0045)	0.151 (0.0032)	0.179 (0.0040)	0.202 (0.0044)	0.157 (0.0023)	0.180 (0.0032)	0.206 (0.0036)
	-0.068 (0.0037)	-0.016 (0.0038)	-0.024 (0.0041)	0.137 (0.0031)	0.142 (0.0030)	0.166 (0.0035)	0.142 (0.0026)	0.142 (0.0022)	0.166 (0.0026)
200	-0.045 (0.0027)	0.016 (0.0028)	0.023 (0.0030)	0.074 (0.0015)	0.078 (0.0016)	0.088 (0.0018)	0.076 (0.0011)	0.079 (0.0012)	0.089 (0.0013)
	-0.043 (0.0026)	-0.008 (0.0026)	-0.009 (0.0028)	0.067 (0.0014)	0.067 (0.0014)	0.077 (0.0016)	0.069 (0.0011)	0.067 (0.0010)	0.077 (0.0012)
500	-0.018 (0.0017)	0.002 (0.0018)	0.005 (0.0019)	0.030 (0.0006)	0.031 (0.0006)	0.035 (0.0007)	0.030 (0.0004)	0.031 (0.0005)	0.035 (0.0005)
	-0.016 (0.0016)	-0.002 (0.0016)	-0.003 (0.0017)	0.025 (0.0005)	0.026 (0.0005)	0.029 (0.0006)	0.025 (0.0004)	0.026 (0.0004)	0.029 (0.0004)
(b) Scenario II: $(\rho_1, \rho_2, \beta_1, \beta_2, B_L, B_U) = (0.2, 0.4, 1, -0.5, 0.02, 1), r = 74.6\%$									
200	0.029 (0.0046)	0.048 (0.0053)	0.059 (0.0050)	0.215 (0.0048)	0.283 (0.0255)	0.248 (0.0060)	0.215 (0.0038)	0.285 (0.0254)	0.252 (0.0051)
	-0.035 (0.0050)	-0.019 (0.0047)	-0.026 (0.0046)	0.247 (0.0484)	0.219 (0.0163)	0.215 (0.0047)	0.248 (0.0484)	0.220 (0.0161)	0.216 (0.0036)
500	0.007 (0.0028)	0.016 (0.0028)	0.022 (0.0029)	0.079 (0.0017)	0.080 (0.0017)	0.085 (0.0018)	0.079 (0.0012)	0.080 (0.0013)	0.085 (0.0013)
	-0.012 (0.0027)	-0.010 (0.0027)	-0.013 (0.0028)	0.070 (0.0014)	0.073 (0.0015)	0.078 (0.0016)	0.071 (0.0010)	0.073 (0.0011)	0.078 (0.0012)
(c) Scenario III: $(\rho_1, \rho_2, \beta_1, \beta_2, B_L, B_U) = (0.1, 0.1, 1, -0.5, 10, 20), r = 74.5\%$									
100	0.051 (0.0585)	0.107 (0.0965)	0.161 (0.1544)	0.342 (0.0732)	0.931 (0.7891)	2.383 (3.000)	0.345 (0.0573)	0.943 (0.7949)	2.41 (3.0263)
	-0.844 (0.4150)	-0.334 (0.2386)	-0.766 (0.4767)	17.2 (8.030)	5.69 (2.456)	22.7 (10.51)	17.9 (8.341)	5.80 (2.416)	23.3 (10.59)
200	0.034 (0.0281)	0.033 (0.0286)	0.036 (0.0287)	0.158 (0.0250)	0.164 (0.0258)	0.165 (0.0258)	0.160 (0.0200)	0.165 (0.0205)	0.166 (0.0206)
	-0.059 (0.0638)	-0.037 (0.0493)	-0.056 (0.0729)	0.814 (1.0008)	0.485 (0.3102)	1.064 (1.3930)	0.818 (1.0034)	0.487 (0.3086)	1.067 (1.3959)
500	0.011 (0.0109)	0.014 (0.0107)	0.016 (0.0108)	0.059 (0.0055)	0.058 (0.0053)	0.058 (0.0053)	0.059 (0.0041)	0.058 (0.0039)	0.058 (0.0039)
	-0.009 (0.0149)	-0.018 (0.0150)	-0.019 (0.0151)	0.111 (0.0103)	0.113 (0.0105)	0.114 (0.0106)	0.111 (0.0076)	0.114 (0.0078)	0.114 (0.0078)

Table 2

Estimated regression parameter vectors $\hat{\beta}_R$ and $\hat{\beta}_D$ and the p -values in the regression model for relapse (R) and death in remission (D)

		Covariate						
		PMR-A	UR-A	SIBL-C	PMR-C	UR-C	Stage	Age
$\hat{\beta}_R$		(a) Relapse						
	ML	-0.196	0.111	-0.457	-0.342	-33.961	0.914	-0.003
	FG	-0.120	0.387	-0.438	-0.357	-9.954	1.138	0.002
	KL	-0.154	0.308	-0.487	-0.755	-2.79e+13	1.144	0.001
	Cox	0.326	0.628	-0.544	0.225	-17.074	1.295	0.012
p -value	ML	0.555	0.831	0.656	0.801	0.000	0.000	0.429
	FG	0.799	0.162	0.211	0.726	0.001	0.000	0.823
	KL	0.756	0.271	0.229	0.642	0.000	0.000	0.955
	Cox	0.493	0.018	0.134	0.827	0.003	0.000	0.119
$\hat{\beta}_D$		(b) Death in Remission						
	ML	1.046	0.376	-0.252	0.870	1.005	0.071	0.027
	FG	1.063	0.719	-0.215	0.878	1.076	0.120	0.031
	KL	0.999	0.614	-0.327	0.836	1.026	0.020	0.027
	Cox	1.110	0.857	-0.264	0.846	1.020	0.239	0.033
p -value	ML	0.010	0.192	0.417	0.084	0.000	0.655	0.000
	FG	0.002	0.011	0.380	0.074	0.000	0.550	0.000
	KL	0.008	0.032	0.191	0.154	0.000	0.919	0.000
	Cox	0.003	0.002	0.290	0.077	0.000	0.219	0.000