Statistica Sinica **27** (2017), 1059-1077
doi:https://doi.org/10.5705/ss.202015.0347

SEMIPARAMETRIC REGRESSION ANALYSIS OF RECURRENT GAP TIMES IN THE PRESENCE OF COMPETING RISKS

Chia-Hui Huang, Yi-Hau Chen and Ya-Wen Chuang

National Taipei University, Academia Sinica and Taichung Veterans General Hospital

Abstract: When a disease progression goes through several stages marked by a nonterminal, recurrent event such as relapse, or a terminal event such as death, which terminates the progression, researchers can be concerned with the duration or gap times between successive events (stages) and wish to study the covariates effects on the gap times. How previous events or gap times affect the current gap time can be also of interest. We propose a unifying framework for joint regression analysis of gap times between successive events. The proposed mixture modeling framework consists of a logistic regression for predicting the path of transition (to a nonterminal or terminal event) at each stage, and a proportional hazards model for predicting the gap times for transition to the nonterminal and terminal events at each stage; these components of the model are conditional on the past event history and stage-specific covariates. In particular, when the number of stages is fixed at one or two, the proposed framework can be applied to the analysis of conventional competing risks or semicompeting risks data. We develop a semiparametric maximum likelihood inference procedure for the proposed models. For which the large sample theory follows directly from martingale theory. Explicit expressions for the information matrix are derived, which facilitate direct variance estimation and convenient computation. Simulation results reveal the proposal's worth, and applications to two clinical studies illustrate its utility.

Key words and phrases: Competing risks, martingale processes, mixture model, multiple events, recurrent data.

1. Introduction

In event time studies, individuals can experience multiple events. With competing risks data, subjects are assumed to be susceptible to several dependent events and only the first occurring event time is observable (Tsiatis (1998)). There are also studies in which individuals can experience a nonterminal event before reaching the terminal event. Data of this type are called semicompeting

risks data (Fine, Jiang and Chappel (2001)). Here the nonterminal event might recur a number of times in a subject. Joint analysis of such multiple events is often of interest, and it imposes substantial challenges owing to the need to account for dependence among multiple events.

Various modeling strategies have been proposed for joint analysis of competing or semicompeting risks. Extending the model in Fine (1999) for a single event, Chang et al. (2007) and Lu and Peng (2008) consider the mixture model framework for competing risks. In this, a parametric or semiparametric model is specified for each of the event time distributions conditional on event type, and a multinomial model for the event type; these models can depend on covariates. An alternative framework consists of semiparametric transformation models for marginal regression models, and a copula model for the joint distribution of the events (Peng and Fine (2007); Hsieh, Wang and Ding (2008); Chen (2010, 2012)). In the analysis of a recurrent event together with a dependent terminal event, random effects models have been applied when the event times measured from a common origin (Liu, Wolfe and Huang (2004); Ye, Kalbfleisch and Scaubel (2007); Rondeau et al. (2007); Zeng and Lin (2009)) or the gap times between consecutive recurrent events (Huang and Liu (2007)) are of interest. In such models, dependence among event times or gap times among events in the same subject are implicitly introduced via the random effects.

A more general framework for analysis of multiple events can be formulated as multi-state models. As elaborated in Andersen and Keiding (2002), the occurrence of an event in a subject can be viewed as the subject's transition from one state to another in a multi-state model. Andersen, Abildstrom and Rosthøj (2002) proposed analysis of competing risks under the framework of multi-state models. Xu, Kalbfleisch and Tai (2010) develop an illness-death model with shared frailty to address positive dependence between event times for semicompeting risks data. Hu and Tsodikov (2014) utilize the illness-death model and proportional hazards assumption to study the joint distribution of the nonterminal and terminal event times in semicompeting risks. Such illness-death, or more general, multi-state models explicitly characterize dynamics of disease progression and hence can be useful in clinical studies.

In this work we propose a unifying modeling and analysis framework for multi-event settings. We consider the mixture regression model for the illness-death process as in Chang et al. (2007) and Lu and Peng (2008), and extend this model to allow for the building-block illness-death process to repeatedly evolve through multiple stages. The mixture regression model considered consists of a

logistic regression for predicting the transition to a nonterminal "relapse" state or to a terminal "death" state at each stage, and a proportional hazards model for predicting the gap or duration times for transition to the "relapse" and "death" states at each stage, both components conditional on the past event history and stage-specific covariates. When some Markov assumption is reasonable, such models can incorporate one or a few prior gap times as predictors. When observation is only to the occurrence of the first event, the proposed models reduce to those considered in Chang et al. (2007) and Lu and Peng (2008). Our proposal can also be applied when the nonterminal event is non-recurrent, the conventional setting of semicompeting risks (Fine, Jiang and Chappel (2001)).

We develop semiparametric maximum likelihood inference procedure for the proposed models, where the score functions are expressed as martingales, so that the large sample theory follows from martingale theory. Explicit expressions for the information matrix are derived, that facilitate direct variance estimation and convenient computation. In addition to the covariate effects on the gap times between events, the proposed analysis allows for direct assessments of impacts from previous events or gap times on the current gap time, that differs from such existing methods as Huang and Liu (2007).

This article is structured as follows. In Section 2, we introduce the data structure considered, and the proposed models. The maximum likelihood estimation procedure and its asymptotic properties are in Sections 3 and 4, respectively. The results of simulation studies, and applications to clinical studies are reported in Sections 5 and 6. The final section provides some discussions and conclusions.

2. Model and Data

Suppose that disease progression can be classified into two states, R and D, in which D is terminal and R is a nonterminal state. It is assumed that the evolution of the disease can reach state R repeatedly before entering D. Figure 1 illustrates a cancer progression, where R denotes cancer relapse and D denotes death. Stage k ($k = 1, 2, 3, \cdots$) of the progression is the period between the $(k-1)$th and kth relapse.

Let ζ_k ($k = 1, 2, 3, \cdots$) be the "path" indicator at stage k, which equals 1 if a subject follows the path from state 0 or R to R, and equals 0 if a subject moves directly to D. For subjects with $\zeta_k = 1$, let T_k^R be the kth gap time from state $0 \to R$ (if $k = 1$) or $R \to R$ (if $k > 1$). For subjects with $\zeta_k = 0$, let T_k^D be the kth gap time from state $0 \to D$ (if $k = 1$) or $R \to D$ (if $k > 1$). Let

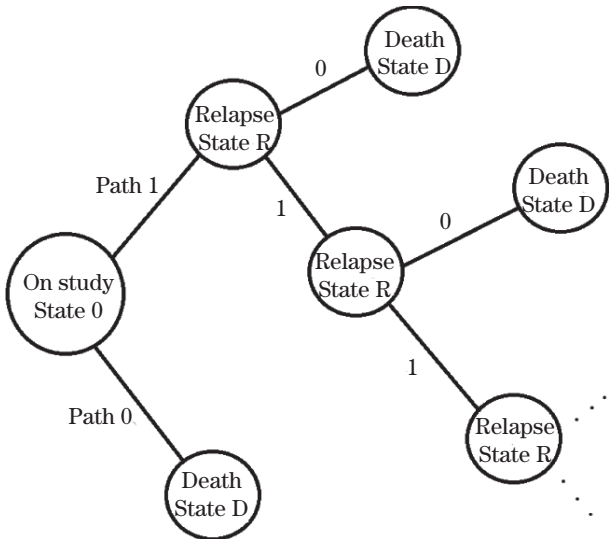

Figure 1. The illness-death model.

$S_l^* = \sum_{k=1}^{l} T_k$ be the cumulative duration time from the initial study time up to the lth event, where $T_k = T_k^R$ when $\zeta_k = 1$ and $T_k = T_k^D$ when $\zeta_k = 0$. Let $\zeta_0 := 1$, and for $k = 1, 2, \cdots$, $\mathcal{H}_k$ be the event time information prior to stage k when $\zeta_{k-1} = 1$. Let X_k $(k = 1, 2, \cdots)$ be a set of stage-specific covariates. For simplicity we assume X_k is constant within stage k.

To model the probability that $\zeta_k = 1$, we assume that, conditional on $\zeta_{k-1} = 1$, $\mathcal{H}_k$, and X_k, ζ_k follows a logistic regression model (Farewell (1982); Lu and Peng (2008)) with covariates W_k and parameter β_p :

$$P(\zeta_k = 1 \mid \mathcal{H}_k, X_k, \zeta_{k-1} = 1) = p_k(\beta_p) = \frac{\exp(\beta_p' W_k)}{1 + \exp(\beta_p' W_k)}, \ \zeta_0 := 1, \ k \geq 1, \quad (2.1)$$

where W_k is a vector of functions of $\mathcal{H}_k$ and X_k that are relevant for predicting the path status at stage k. In practice, W_k may include some subset of X_k and the duration time T_{k-1}^R at the immediate previous stage; such a choice is based on a Markov-type assumption that the path status at stage k depends on the previous event times only through the duration time in the previous stage.

Given the path status ζ_k at stage k, $\mathcal{H}_k$ and X_k $(k = 1, 2, \cdots)$, the hazard functions for the duration times T_k^R and T_k^D at stage k are modeled by proportional hazards models:

$$\lim_{\Delta \to 0} \Delta^{-1} P(t \leq T_k^R < t + \Delta \mid T_k^R \geq t, \zeta_k = 1, \mathcal{H}_k, X_k) = \exp(\theta_r' Q_r) \lambda_r(t),$$
$$\lim_{\Delta \to 0} \Delta^{-1} P(t \leq T_k^D < t + \Delta \mid T_k^D \geq t, \zeta_k = 0, \mathcal{H}_k, X_k) = \exp(\theta_d' Q_d) \lambda_d(t),$$

where λ_r and λ_d are nonnegative, unknown baseline hazard functions for the recurrent and terminal events, respectively, and Q_r and Q_d are vectors of functions of $\mathcal{H}_k$ and X_k that are relevant for predicting the duration times T_k^R and T_k^D. For example, we may consider the specific models given as

$$\lim_{\Delta\to 0} \Delta^{-1} P(t \le T_k^R < t+\Delta \mid T_k^R \ge t, \zeta_k = 1, \mathcal{H}_k, X_k) = \exp(\beta_r' Z_k + \alpha_r T_{k-1}^R)\lambda_r(t), \tag{2.2}$$

$$\lim_{\Delta\to 0} \Delta^{-1} P(t \le T_k^D < t+\Delta \mid T_k^D \ge t, \zeta_k = 0, \mathcal{H}_k, X_k) = \exp(\beta_d' Z_k + \alpha_d T_{k-1}^R)\lambda_d(t), \tag{2.3}$$

where $T_0^R := 0$, Z_k is some subset of X_k. In this example, the regression parameters β_r, β_d assess the covariate effects on gap times for the recurrent and terminal events, and α_r and α_d are the association parameters measuring the effect of immediate previous gap time on the current gap times for relapse and death, respectively. Models (2.2) and (2.3) are based on the Markov assumption that, given the immediate previous duration time T_{k-1}^R, the current duration times T_k^R and T_k^D are independent of all earlier duration times $\{T_1^R, \ldots, T_{k-2}^R\}$. We focus on the specifics of models (2.2) and (2.3).

We further allow the existence of an external censoring time V^* that is independent of the recurrent gap times and the terminal event time given $\{X_k : k \ge 1\}$. Let $(V_i^*, X_{ik}, \zeta_{ik})$, $i = 1, \cdots, n$, be n independent and identically distributed replicates of (V^*, X_k, ζ_k), with (T_{ik}^R, T_{ik}^D), $i = 1, \cdots, n$, independent and identically distributed replicates from models (2.2)-(2.3). Let τ denote the maximum follow-up time in the study and $V_i = \min(V_i^*, \tau)$. The observed recurrent data consist of $\{\left(S_{ik}, \delta_{ik}^R, \delta_{ik}^D, X_{ik}\right), i = 1, 2, \cdots, n; k = 1, 2, \cdots\}$, where $S_{ik} = \min(S_{ik}^*, V_i)$, and $\delta_{ik}^R = \zeta_{ik} I(S_{ik}^* \le V_i)$, $\delta_{ik}^D = (1-\zeta_{ik}) I(S_{ik}^* \le V_i)$ indicating whether the kth duration times for relapse and death are uncensored or not.

The path indicator ζ_{ik} is not directly observable, but $\delta_{ik}^R = 1$ implies $\zeta_{ik} = 1$, and $\delta_{ik}^D = 1$ implies $\zeta_{ik} = 0$.

Remark 1. Our modeling framework is similar to that in Fine (1999) and Chang et al. (2007) as developed for non-recurrent events. Although such a type of model conditions on the path information, it is a natural extension of the semiparametric transformation models to the competing risks setting, and is useful in the study of covariate-specific probability of failure from a given cause over time. Here the regression coefficients β_r and β_d are the effects of the covariate for patients who would subsequently experience the recurrence event and the death event, respectively. Similarly, the baseline hazards $\lambda_r(t)$ and $\lambda_d(t)$ are referred to the

two groups of patients.

Remark 2. Although the proposed analysis builds upon the hazards models in (2.2) and (2.3), the cumulative incidence function (CIF) for relapse and death events can be simply found with (2.1). To obtain the CIF for the relapse and death events at stage k, $P(T_k \le t, \zeta_k = 1 \mid X_k, T_{k-1})$ and $P(T_k \le t, \zeta_k = 0 \mid X_k, T_{k-1})$, we note that

$$\begin{aligned} &P(T_k \le t, \zeta_k = 1 \mid X_k, T_{k-1}) \\ &= P(\zeta_k = 1 \mid \zeta_{k-1} = 1, X_k) P(T_k \le t \mid \zeta_k = 1, X_k, T_{k-1}) \\ &= \frac{\exp(\beta_p' W_k)}{1 + \exp(\beta_p' W_k)} \Big[1 - \exp\{-\exp(\alpha_r T_{k-1} + \beta_r' Z_k)\Lambda_r(t)\}\Big], \end{aligned}$$

and analogously for $P(T_k \le t, \zeta_k = 0 \mid X_k, T_{k-1})$. Therefore, the CIF can be estimated directly by evaluating the probability of $(\zeta_k = 1)$, the proportion of subjects who experience relapse, and the distribution of the failure time given the event type through its corresponding cumulative hazard function. A practical example with CIF estimation is provided in Section 6.2.

3. Maximum Likelihood Method

In the following, since gap times are considered, the time index t is reset to zero at the occurrence of each "relapse" event (nonterminal event). Based on (2.1), (2.2) and (2.3), we derive the likelihood and the score using the counting process approach. Let $G_{ik} = S_{ik} - S_{i,k-1}$ be the gap times subject to censoring with $G_{i0} := 0$, and $Y_{ik}(t) = I(G_{ik} \ge t)$ the "at risk" process. Let $N_{ik}^R(t) = \delta_{ik}^R I(G_{ik} \le t)$ and $N_{ik}^D(t) = \delta_{ik}^D I(G_{ik} \le t)$ be the counting processes for the transitions to states R and D, respectively. When death or censoring occurs in the kth stage and there is no way to observe a following event in subject i, we have $G_{il} := 0$ and $Y_{il}(\cdot) := 0$ for $l > k$ by definition.

In our formulation the population at each stage is considered as a mixture of two groups of subjects following different paths. The path status for a subject at some stage is unknown until the "relapse" or "death" event occurs. All subjects are subject to independent censoring, and the stage-specific path status for a subject is unknown if censoring occurs. According to (2.1), (2.2) and (2.3), the cause-specific hazard function for $N_{ik}^R(t)$ is

$$Y_{ik}(t)\Theta_{ik}(t-;\Omega)\exp(\beta_r' Z_{ik} + \alpha_r G_{i,k-1})\lambda_r(t), \Theta_{ik}(t;\Omega) = \frac{p_{ik}(\beta_p)\exp[-\mu_{ik}(\alpha_r)\eta_{ik}(\beta_r)\Lambda_r(t)]}{p_{ik}(\beta_p)\exp[-\mu_{ik}(\alpha_r)\eta_{ik}(\beta_r)\Lambda_r(t)] + \bar{p}_{ik}(\beta_p)\exp[-\mu_{ik}(\alpha_d)\eta_{ik}(\beta_d)\Lambda_d(t)]},$$

where $\Omega = (\boldsymbol{\alpha}, \boldsymbol{\beta}, d\boldsymbol{\Lambda})$, $\boldsymbol{\alpha} = (\alpha_r, \alpha_d)$, $\boldsymbol{\beta} = (\beta_p, \beta_r, \beta_d)$, and $d\boldsymbol{\Lambda} = (d\Lambda_r, d\Lambda_d)$. Here $p_{ik}(\beta_p)$ is the conditional probability of $(\zeta_{ik} = 1)$ given in (2.1) with W_k replaced by W_{ik}, $\bar{p}_{ik}(\beta_p) = 1 - p_{ik}(\beta_p)$, $\eta_{ik}(\beta) = \exp(\beta' Z_{ik})$, $\mu_{ik}(\alpha) = \exp(\alpha G_{i,k-1})$, and λ_r and λ_d are, respectively, the derivatives of Λ_r and Λ_d. The Θ_{ik} are interpreted as the conditional probability of observing $N^R_{ik}(t) = 1$ at time t conditional on past event information and that the subject i is at risk at time t. Similarly, the cause-specific hazard function for $N^D_{ik}(t)$ is $Y_{ik}(t)\bar{\Theta}_{ik}(t-;\Omega)\exp(\beta_d' Z_{ik} + \alpha_d G_{i,k-1})\lambda_d(t)$, and $\bar{\Theta}_{ik} = 1 - \Theta_{ik}$ is interpreted analogously to Θ_{ik}.

Let $dM^R_{ik}(t) = dN^R_{ik}(t) - Y_{ik}(t)\Theta_{ik}(t-;\Omega)\exp(\beta_r' Z_{ik} + \alpha_r G_{i,k-1})d\Lambda_r(t)$, $dM^D_{ik}(t) = dN^D_{ik}(t) - Y_{ik}(t)\bar{\Theta}_{ik}(t-;\Omega)\exp(\beta_d' Z_{ik} + \alpha_d G_{i,k-1})d\Lambda_d(t)$. Let K be a sufficiently large number denoting the maximum number of the observed events among subjects. Then the loglikelihood function for the observed data is given by

$$\begin{aligned}
\ell(\Omega) = &\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \big[\log\Theta_{ik}(t-;\Omega) + \beta_r' Z_{ik} + \alpha_r G_{i,k-1} + \log d\Lambda_r(t)\big] dN^R_{ik}(t) \\
&- \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau Y_{ik}(t)\Theta_{ik}(t-;\Omega)\exp(\beta_r' Z_{ik} + \alpha_r G_{i,k-1})d\Lambda_r(t) \\
&+ \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \big[\log\bar{\Theta}_{ik}(t-;\Omega) + \beta_d' Z_{ik} + \alpha_d G_{i,k-1} + \log d\Lambda_d(t)\big] dN^D_{ik}(t) \\
&- \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau Y_{ik}(t)\bar{\Theta}_{ik}(t-;\Omega)\exp(\beta_d' Z_{ik} + \alpha_d G_{i,k-1})d\Lambda_d(t). \qquad (3.1)
\end{aligned}$$

Remark 3. In (2.1), (2.2) and (2.3), we have assumed common baseline hazards and regression parameters $\Omega = (\boldsymbol{\alpha}, \boldsymbol{\beta}, d\boldsymbol{\Lambda})$ across stages. The models can be modified to allow for stage-specific parameters. A practical issue regarding such modeling is that estimation instability can arise if only sparse data are available for certain gap times.

Remark 4. When only the data from the first stage are available, the loglikelihood function (3.1) reduces to

$$\begin{aligned}
&\sum_{i=1}^{n}\int_0^\tau \big[\log\Theta_i(t-;\Omega) + \beta_r' Z_i + \log d\Lambda_r(t)\big] dN^R_i(t) \\
&\quad + \big[\log\bar{\Theta}_i(t-;\Omega) + \beta_d' Z_i + \log d\Lambda_d(t)\big] dN^D_i(t) \\
&\quad - \sum_{i=1}^{n}\int_0^\tau Y_i(t)\Big[\Theta_i(t-;\Omega)\exp(\beta_r' Z_i)d\Lambda_r(t) + \bar{\Theta}_i(t-;\Omega)\exp(\beta_d' Z_i)d\Lambda_d(t)\Big].
\end{aligned}$$

This loglikelihood function is equivalent to that in Chang et al. (2007), as developed for competing risks data. As pointed out in Chang et al. (2007), (2.1), (2.2) and (2.3) are not identifiable when $\beta_p = \beta_r = \beta_d = 0$.

Remark 5. Fixing the number of stages as two, and setting $p_2(\beta_p) = 0$ and hence the non-terminal event can occur at most once in a subject, the proposed models can be applied to the "semicompeting risks" data. In this case, (3.1) is

$$\begin{aligned}
&\sum_{i=1}^{n} \int_0^\tau \Big\{ \big[\log \Theta_{i1}(t-;\Omega) + \beta_r' Z_{i1} + \log d\Lambda_r(t) \big] dN_{i1}^R(t) \\
&\quad - Y_{i1}(t) \Theta_{i1}(t-;\Omega) \exp(\beta_r' Z_{i1}) d\Lambda_r(t) \Big\} \\
&\quad + \sum_{i=1}^{n} \int_0^\tau \Big\{ \big[\log \bar{\Theta}_{i1}(t-;\Omega) + \beta_d' Z_{i1} + \log d\Lambda_d(t) \big] dN_{i1}^D(t) \\
&\quad - Y_{i1}(t) \bar{\Theta}_{i1}(t-;\Omega) \exp(\beta_d' Z_{i1}) d\Lambda_d(t) \Big\} \\
&\quad + \sum_{i=1}^{n} \int_0^\tau \Big\{ \big[\beta_d' Z_{i2} + \alpha_d G_{i1} + \log d\Lambda_d(t) \big] dN_{i2}^D(t) \\
&\quad - Y_{i2}(t) \exp(\beta_d' Z_{i2} + \alpha_d G_{i1}) d\Lambda_d(t) \Big\},
\end{aligned}$$

where the first two terms correspond to the first-occurring event, either a relapse, death or censoring event, and the last term correspond to the death or censoring event that is observed subsequent to the relapse event observed at stage 1.

Remark 6. The proposed models (2.2) and (2.3) for gap times condition on the past event history. This avoids the complication of "induced dependent censoring" for gap time analysis (Wang and Wells (1998); Lin, Sun and Ying (1999)) due to the dependence of the current and previous gap times.

We propose nonparametric maximum likelihood estimation for the regression parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and the baseline hazard functions $\boldsymbol{\Lambda}$. As in Zeng and Lin (2006), Zeng and Lin (2007), and Chen (2009), the baseline hazard functions are treated as non-decreasing step functions with jumps only at the observed event times, and $d\boldsymbol{\Lambda}$ denotes the collection of jump sizes of $\boldsymbol{\Lambda}$. It is assumed that the counting processes $N_{ik}^R(t)$ and $N_{ik}^D(t)$ cannot jump simultaneously. We give explicit expressions for the score and information matrix for $(\boldsymbol{\alpha}, \boldsymbol{\beta}, d\boldsymbol{\Lambda})$ in the Supplementary Material. In the computation, the jump sizes $d\Lambda_{r^*}$ and $d\Lambda_{d^*}$ and the regression parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are solved simultaneously from the corresponding score equations. As in Zeng and Lin (2006, 2007), and (Chen (2009, 2010)), we implement the computation via the Matlab function *fminunc* to solve

the system of nonlinear estimating equations. Since we have explicit expressions for the information (Hessain) matrix, such implementation is quite stable and fast.

4. Asymptotic Results

Let $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, and $d\widehat{\boldsymbol{\Lambda}}$ be the nonparametric maximum likelihood estimators of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, d\boldsymbol{\Lambda})$, $\widehat{\Omega} = (\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, d\widehat{\boldsymbol{\Lambda}})$, and $\Omega^0 = (\boldsymbol{\alpha}^0, \boldsymbol{\beta}^0, d\boldsymbol{\Lambda}^0)$ the true value of Ω. As can be seen from the Supplement Material, the estimating functions for Ω and the cumulative hazard functions $\Lambda_r(t)$ and $\Lambda_d(t)$ at Ω^0 are all martingales. Hence we can apply the martingale central limit theorem to establish the large sample properties of $\widehat{\Omega}$. The following regularity conditions are required, and the proofs of the theorems are sketched in the Supplement Material.

(A1) The true Λ_r and Λ_d are strictly increasing and differentiable; $\boldsymbol{\alpha}^0$ and $\boldsymbol{\beta}^0$ are in the interior of a compact parameter space.

(A2) With probability one, $P(S_k \geq \tau | \mathcal{H}_k, X_k) > 0$, and $P(\delta_k^R = \delta_k^D = 0, S_k = \tau \,|\mathcal{H}_k, X_k) > 0,\ k = 1, 2.$

(A3) The covariates X_k $(k \geq 1)$ are bounded in $[0, \tau]$, and linearly independent.

(A4) The information matrix $\mathcal{I}$ is positive definite and $n^{-1}\mathcal{I}$ converges in probability to a deterministic and positive definite matrix $\mathcal{I}^0$.

Theorem 1. *If* (A1)-(A4) *hold, then* $\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}$ *converges to* $\boldsymbol{\alpha}^0, \boldsymbol{\beta}^0$ *with probability one, and* $\widehat{\boldsymbol{\Lambda}}_r$, $\widehat{\boldsymbol{\Lambda}}_d$ *converge to* $\boldsymbol{\Lambda}_r^0$, $\boldsymbol{\Lambda}_d^0$ *uniformly in the interval* $[0, \tau]$ *with probability one, where* $\widehat{\boldsymbol{\Lambda}}_s(t) = \int_0^t d\widehat{\boldsymbol{\Lambda}}_s(u)$ *for* $s \in \{r, d\}$, $t \in [0, \tau]$.

Following Chen (2010, 2012), and Hu and Tsodikov (2014), consider the linear functional

$$\sqrt{n}\left\{\int_0^\tau \boldsymbol{a}'\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\right) + \boldsymbol{b}'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right) + \boldsymbol{\gamma}(t)'\left(d\widehat{\boldsymbol{\Lambda}}(t) - d\boldsymbol{\Lambda}^0(t)\right)\right\}, \tag{4.1}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$ are vectors and $\boldsymbol{\gamma}(t)$ is a function with bounded total variation in $[0, \tau]$. Let $\boldsymbol{\Gamma}$ be the vector consisting of the values of $\boldsymbol{\gamma}(t)$ evaluated at the observed event times corresponding to the jumps $\{d\boldsymbol{\Lambda}\}$. Let $\mathcal{E}' = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\Gamma}')$.

Theorem 2. *If* (A1)-(A4) *hold,* $\sqrt{n}\{\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0, \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0, \widehat{\boldsymbol{\Lambda}}(\cdot) - \boldsymbol{\Lambda}^0(\cdot)\}$ *converges weakly to a zero-mean Gaussian process, the linear functional* (4.1) *is asymptotically normal with mean zero and variance-covariance matrix* $\mathcal{E}'(\mathcal{I}^0)^{-1}\mathcal{E}$, *and the information matrix* $\mathcal{I}^0$ *can be consistently estimated by* $n^{-1}\widehat{\mathcal{I}}$.

The explicit expression for the negative Hessian matrix of the loglikelihood function is provided in the Supplement Material.

5. Simulation Studies

Simulation studies were performed to demonstrate the finite sample performances of the proposed inference procedure. First we considered covariates $X = (X_1, X_2)$ constant over stages, with X_1 a bernoulli trial with success probability 0.5, X_2 a standard normal distribution truncated at ± 3. We took $W_k = (1, X_1)$, and $\beta_p = (0, 0)$ or $(-0.5, 1)$ in (2.1) so that the probabilities of transition to the "relapse" state and to the "death" state were equal. We generated $\{(T_k^R, T_k^D) : k = 1, 2, \cdots\}$ from (2.2)-(2.3) with $Z_k = (X_1, X_2)$, $\beta_r = (-0.3, 0.7)$, $\beta_d = (-0.5, 0.5)$, and $(\alpha_r, \alpha_d) = (0, 0)$ or $(0.1, -0.1)$. The censoring variable V^* was uniform distribution on $[0, 20]$ and the maximum follow-up time $\tau = 10$.

The proposed method was applied to the setting of competing risks data: observation stopped when a "relapse", "death", or censoring event was first observed. Here the baseline function $\lambda_r(t) = 0.05t$ and $\lambda_d(t) = 0.1$, led to roughly 30% of the events being "relapse" and 22% of events being "death" for the two sets of β_p. The simulation results are shown in Table 1 for the sample sizes $n = 150$ and 300. For each parameter, the bias, simulation standard deviation (SD) and mean of the estimated standard standard errors (SE) are summarized to illustrate the finite sample properties of the nonparametric maximum likelihood estimate. We see that the proposed estimates of regression parameters $\boldsymbol{\beta}$ are nearly unbiased and the estimated standard errors found by the observed information matrix are close to the simulation standard deviations. The estimation for $\boldsymbol{\Lambda}$ under the smaller sample size of $n = 150$ yields a relatively larger bias; however, that bias was negligible when $n = 300$. The 95% confidence intervals give coverage probabilities at the desired level, and the results improve with increased sample size. Some of the coverage rates being lower than 90% may be due to the low event rates. To check this, we did a simulation study with data generated by the same values of regression coefficients as in Table 1, but the baseline hazard rates $\lambda_r(t)$ and $\lambda_d(t)$ were enhanced to be $(0.25t, 0.5)$. The results are tabulated in the Supplement Material, showing that the coverage rates are closer to the nominal level of 95%.

Under the same setups we applied our method to the semicompeting risks setting with $\alpha_d = 0$. The simulation results are reported in Table 2. Here, for estimation of β_d and Λ_d, there is extra information from subjects whose first-

Table 1. Simulation results for the competing risks data, $K = 1$, where the statistics are based on the 1,000 replications, and sample sizes $n = 150$ or 300 in each replication. Here $\lambda_r(t) = 0.05t$, $\lambda_d(t) = 0.1$, and about 30% (22%) of the first-occurring event was "relapse" ("death").

		Scenario 1: $\beta_p = (0, 0)$				Scenario 2: $\beta_p = (-0.5, 1)$			
n	Parameter	Bias	SD	SE	CP (%)	Bias	SD	SE	CP (%)
150	$\beta_p : 1$	0.057	0.372	0.381	95.4	0.077	0.382	0.386	94.9
	$\beta_p : X_1$	0.023	0.537	0.512	94.6	0.023	0.577	0.535	92.6
	$\beta_r : X_1 = -0.3$	−0.010	0.446	0.407	92.4	0.018	0.460	0.422	92.2
	$\beta_r : X_2 = 0.7$	0.039	0.173	0.156	94.0	0.046	0.166	0.156	93.5
	$\beta_d : X_1 = -0.5$	−0.041	0.481	0.454	94.1	−0.007	0.527	0.488	93.3
	$\beta_d : X_2 = 0.5$	0.036	0.165	0.160	95.2	0.035	0.165	0.156	95.3
	$\Lambda_r(\tau/4) = 0.156$	−0.005	0.062	0.059	88.5	−0.005	0.071	0.066	86.3
	$\Lambda_r(\tau/2) = 0.625$	0.002	0.218	0.208	90.1	−0.006	0.250	0.237	89.0
	$\Lambda_r(3\tau/4) = 1.406$	0.041	0.565	0.538	91.0	0.071	0.658	0.622	89.7
	$\Lambda_d(\tau/4) = 0.25$	0.023	0.116	0.104	93.6	0.012	0.103	0.090	92.3
	$\Lambda_d(\tau/2) = 0.5$	0.059	0.234	0.201	93.9	0.043	0.204	0.172	95.1
	$\Lambda_d(3\tau/4) = 0.75$	0.114	0.387	0.324	95.9	0.100	0.359	0.281	95.1
300	$\beta_p : 1$	0.034	0.276	0.259	93.1	0.022	0.265	0.259	95.2
	$\beta_p : X_1$	0.031	0.382	0.356	94.2	0.032	0.394	0.372	93.6
	$\beta_r : X_1 = -0.3$	−0.027	0.305	0.283	93.2	−0.018	0.302	0.292	94.0
	$\beta_r : X_2 = 0.7$	0.026	0.112	0.106	92.7	0.025	0.113	0.106	94.2
	$\beta_d : X_1 = -0.5$	−0.016	0.324	0.304	94.6	0.037	0.346	0.326	94.1
	$\beta_d : X_2 = 0.5$	0.007	0.107	0.107	95.7	0.017	0.112	0.104	94.5
	$\Lambda_r(\tau/4) = 0.156$	−0.002	0.043	0.042	90.8	−0.002	0.047	0.046	91.6
	$\Lambda_r(\tau/2) = 0.625$	0.001	0.146	0.141	92.0	0.009	0.167	0.163	92.5
	$\Lambda_r(3\tau/4) = 1.406$	0.042	0.375	0.358	92.1	0.067	0.432	0.410	93.0
	$\Lambda_d(\tau/4) = 0.25$	0.011	0.069	0.067	95.2	0.001	0.061	0.058	93.9
	$\Lambda_d(\tau/2) = 0.5$	0.032	0.135	0.126	95.4	0.012	0.111	0.106	95.1
	$\Lambda_d(3\tau/4) = 0.75$	0.063	0.223	0.196	95.9	0.027	0.169	0.160	95.3

occurring event is "relapse", so the standard deviations are expected to be smaller than the comparable competing risks setting considered in Table 1. Results in Table 2 suggest that the proposed inference method still performs well in this setting.

The last setting considered had the nonterminal event recurrent and the number of occurrences of this event in each subject ranging from zero to three. The setups were similar to those in the previous simulations, but we set the covariates $W_k = (1, X_1, T_{k-1}^R)$ as (2.1) for the transition probability at stage k to depend on the previous gap time. The baseline functions were $\lambda_r(t) = 0.05t$ and $\lambda_d(t) = 0.1$, so that at stages 1 to 3 the proportions of "relapse" among the events observed at those stages were 30%, 8% and 2%, respectively, and the proportions

Table 2. Simulation results for semicompeting risks data, $K = 2$, $\lambda_r(t) = 0.05t$ and $\lambda_d(t) = 0.1$. About 30% (22%) of the first-occurring event is "relapse" ("death") and 11% of subjects have both.

		Scenario 1: $\beta_p = (0, 0)$				Scenario 2: $\beta_p = (-0.5, 1)$			
n	Parameter	Bias	SD	SE	CP (%)	Bias	SD	SE	CP (%)
150	$\alpha_d = 0$	−0.021	0.091	0.088	95.1	−0.022	0.096	0.089	94.8
	$\beta_p : 1$	0.054	0.363	0.403	96.1	0.075	0.369	0.409	96.7
	$\beta_p : X_1$	0.012	0.514	0.481	93.6	0.040	0.524	0.504	95.1
	$\beta_r : X_1 = -0.3$	−0.014	0.461	0.397	90.7	0.003	0.456	0.416	91.7
	$\beta_r : X_2 = 0.7$	0.028	0.170	0.157	93.2	0.038	0.173	0.156	93.5
	$\beta_d : X_1 = -0.5$	−0.045	0.350	0.341	93.9	−0.006	0.358	0.343	94.6
	$\beta_d : X_2 = 0.5$	0.027	0.132	0.140	96.0	0.021	0.135	0.137	95.9
	$\Lambda_r(\tau/4) = 0.156$	−0.004	0.062	0.060	90.5	−0.003	0.072	0.067	86.1
	$\Lambda_r(\tau/2) = 0.625$	0.005	0.230	0.211	91.3	0.005	0.258	0.242	89.2
	$\Lambda_r(3\tau/4) = 1.406$	0.092	0.625	0.575	91.1	0.081	0.673	0.633	90.8
	$\Lambda_d(\tau/4) = 0.25$	0.022	0.089	0.097	96.2	0.015	0.089	0.087	95.2
	$\Lambda_d(\tau/2) = 0.5$	0.051	0.180	0.188	95.8	0.038	0.168	0.165	95.8
	$\Lambda_d(3\tau/4) = 0.75$	0.098	0.305	0.296	97.0	0.080	0.291	0.263	96.4
300	$\alpha_d = 0$	−0.012	0.060	0.060	95.9	−0.006	0.061	0.060	95.6
	$\beta_p : 1$	0.043	0.257	0.286	95.4	0.015	0.260	0.276	96.9
	$\beta_p : X_1$	−0.006	0.359	0.336	93.4	0.019	0.353	0.349	95.2
	$\beta_r : X_1 = -0.3$	−0.008	0.285	0.277	94.9	−0.013	0.310	0.288	92.9
	$\beta_r : X_2 = 0.7$	0.025	0.110	0.110	94.9	0.017	0.112	0.106	94.0
	$\beta_d : X_1 = -0.5$	−0.003	0.230	0.231	96.1	0.002	0.229	0.229	94.5
	$\beta_d : X_2 = 0.5$	0.010	0.087	0.098	97.3	0.008	0.087	0.093	96.2
	$\Lambda_r(\tau/4) = 0.156$	−0.004	0.042	0.043	92.5	0.000	0.047	0.047	93.3
	$\Lambda_r(\tau/2) = 0.625$	−0.001	0.141	0.147	93.3	0.010	0.169	0.166	94.1
	$\Lambda_r(3\tau/4) = 1.406$	0.038	0.382	0.379	95.3	0.058	0.438	0.418	93.6
	$\Lambda_d(\tau/4) = 0.25$	0.008	0.058	0.065	95.6	0.004	0.052	0.055	94.9
	$\Lambda_d(\tau/2) = 0.5$	0.017	0.109	0.123	96.6	0.011	0.095	0.103	96.0
	$\Lambda_d(3\tau/4) = 0.75$	0.034	0.170	0.189	96.9	0.026	0.149	0.156	96.3

of "death" in the events observed at the three stages were 22%, 5% and 1%. Table 3 summarizes the results for the four cases with sample size $n = 300$. We see these the proposed estimators work well and, as expected, the bias and the standard deviations of (β_r, β_d) and (Λ_r, Λ_d) are smaller in this setting with recurrent events, compared to the settings without recurrence.

6. Data Applications

The proposed analysis is applied to two clinical trial studies. The first one is the non-gram-positive organisms study for kidney patients taking the peritoneal dialysis (PD) treatment, where data from 575 subjects were collected at Taichung

Table 3. Simulation results for recurrent event data with $K = 3$, $\lambda_r(t) = 0.05t$, and $\lambda_d(t) = 0.1$. The proportions of "relapse" events are $30\%, 8\%, 2\%$ from stages 1 to 3; the proportions of "death" events are $22\%, 5\%, 1\%$ from stages 1 to 3.

	Scenario 1: $(\alpha_r, \alpha_d) = (0, 0)$; $\beta_p = (0, 0, 0.1)$				Scenario 2: $(\alpha_r, \alpha_d) = (0, 0)$; $\beta_p = (-0.5, 1, 0.1)$			
Parameter	Bias	SD	SE	CP (%)	Bias	SD	SE	CP (%)
α_r	0.026	0.089	0.093	93.7	0.022	0.085	0.088	95.0
α_d	−0.031	0.104	0.117	95.7	−0.024	0.105	0.116	95.7
$\beta_p : 1$	0.036	0.235	0.238	95.6	0.033	0.233	0.241	95.5
$\beta_p : X_1$	0.009	0.333	0.319	94.1	0.043	0.354	0.332	94.2
$\beta_p :$ *previous gap*	−0.034	0.132	0.139	92.2	−0.028	0.133	0.140	93.6
$\beta_r : X_1 = -0.3$	−0.025	0.253	0.250	94.7	−0.026	0.259	0.255	94.2
$\beta_r : X_2 = 0.7$	0.021	0.096	0.093	94.5	0.024	0.091	0.092	94.8
$\beta_d : X_1 = -0.5$	−0.003	0.277	0.268	95.6	0.034	0.325	0.290	93.6
$\beta_d : X_2 = 0.5$	0.015	0.096	0.095	95.1	0.006	0.092	0.093	95.8
$\Lambda_r(\tau/4) = 0.156$	−0.006	0.037	0.037	91.3	−0.004	0.040	0.041	92.6
$\Lambda_r(\tau/2) = 0.625$	0.000	0.127	0.129	93.4	0.006	0.141	0.147	94.3
$\Lambda_r(3\tau/4) = 1.406$	0.059	0.359	0.335	95.4	0.054	0.360	0.370	95.1
$\Lambda_d(\tau/4) = 0.25$	0.009	0.063	0.061	94.1	0.009	0.055	0.056	95.7
$\Lambda_d(\tau/2) = 0.5$	0.031	0.118	0.116	95.5	0.025	0.105	0.102	95.2
$\Lambda_d(3\tau/4) = 0.75$	0.052	0.188	0.176	96.1	0.040	0.162	0.152	95.9
	Scenario 3: $(\alpha_r, \alpha_d) = (0.1, -0.1)$; $\beta_p = (0, 0, 0.1)$				Scenario 4: $(\alpha_r, \alpha_d) = (0.1, -0.1)$; $\beta_p = (-0.5, 1, 0.1)$			
α_r	0.009	0.082	0.081	94.6	0.002	0.070	0.078	95.5
α_d	−0.008	0.115	0.122	97.4	−0.008	0.120	0.125	96.1
$\beta_p : 1$	0.019	0.242	0.238	94.5	0.038	0.235	0.242	96.1
$\beta_p : X_1$	0.030	0.309	0.313	96.2	0.025	0.337	0.330	94.8
$\beta_p :$ *previous gap*	0.007	0.133	0.130	93.4	0.016	0.133	0.136	95.3
$\beta_r : X_1 = -0.3$	−0.026	0.253	0.243	94.0	−0.005	0.265	0.251	93.2
$\beta_r : X_2 = 0.7$	0.015	0.098	0.091	94.5	0.018	0.095	0.090	93.2
$\beta_d : X_1 = -0.5$	0.013	0.276	0.266	95.3	0.048	0.321	0.290	93.7
$\beta_d : X_2 = 0.5$	0.013	0.097	0.097	95.4	0.010	0.099	0.095	94.4
$\Lambda_r(\tau/4) = 0.156$	−0.001	0.038	0.037	92.9	−0.004	0.042	0.040	91.2
$\Lambda_r(\tau/2) = 0.625$	0.009	0.130	0.130	94.1	0.001	0.150	0.145	92.6
$\Lambda_r(3\tau/4) = 1.406$	0.048	0.355	0.333	93.6	0.022	0.375	0.362	93.3
$\Lambda_d(\tau/4) = 0.25$	0.004	0.062	0.060	94.0	0.004	0.055	0.055	95.1
$\Lambda_d(\tau/2) = 0.5$	0.016	0.121	0.114	95.5	0.013	0.108	0.100	94.9
$\Lambda_d(3\tau/4) = 0.75$	0.036	0.190	0.174	96.2	0.027	0.165	0.151	94.5

Veterans General Hospital in Taiwan from 1996 to 2011 (Chen, Chuang and Shen (2015)). The other study, with a sample size of 1977, focused on chronic myeloid leukemia and was conducted by the European Society for Blood and Marrow Transplantation (EBMT).

Table 4. Analysis of peritoneal dialysis study. *: p-value is less than 5%.

	Path indicator		Relapse state		Death state	
Variable	Estimated	SE	Estimated	SE	Estimated	SE
Intercept	−0.200	0.081*	-	-	-	-
Previous gap	0.047	0.293	0.639	0.247*	1.379	0.244*
Diabetes	-	-	0.228	0.246	0.291	0.147*
Cardiovascular	-	-	−0.061	0.215	0.134	0.163

Table 5. Analysis of EBMT data. *: p-value is less than 5%.

	Path indicator		Relapse state		Death state	
Variable	Estimated	SE	Estimated	SE	Estimated	SE
Intercept	−0.038	0.269	-	-	-	-
Previous gap	-	-	-	-	−0.460	0.108*
Medium	−0.742	0.218*	1.101	0.213*	0.307	0.093*
High	−0.491	0.295	1.471	0.338*	1.073	0.129*

6.1. Analysis of PD data

In the PD study, the terminal event was death or taking the alternative treatment, hemodialysis. The nonterminal event was infection caused by the non-gram-positive organisms. A total of 265 (46%) subjects experienced the terminal event and 126 (22%) subjects had infection as the first-occurring event. The maximum number of observed stages was 6. We applied the proposed method to these data with covariates that included diabetes, cardiovascular disease, and previous gap time (rescaled to have a range between 0 and 1). From Table 4, subjects were more likely to reach the death state than the relapse state at each stage during the process of PD treatment. For the gap times to the nonterminal and terminal events, the parameters α_r and α_d in models (2.2) and (2.3) were estimated to be positive. Thus the previous gap time significantly accelerated the subsequent gap time regardless of whether the subsequent event was a nonterminal or terminal event.

The results in Table 4 reveal that diabetes could increase the risks for both non-gram-positive infection and death in each stage of the PD process, although only its effects on death attained statistically significance. The estimated cumulative hazard functions for the occurrence of infection and death for subjects with/without diabetes are shown in Figure 2, together with the associated pointwise 95% confidence intervals. The right panel of Figure 2 suggests that subjects with diabetes are associated with higher risks (cumulative hazards) of the death event than subjects without diabetes. A similar pattern is also observed for the risks of infection between diabetes and non-diabetes patients as shown in the left

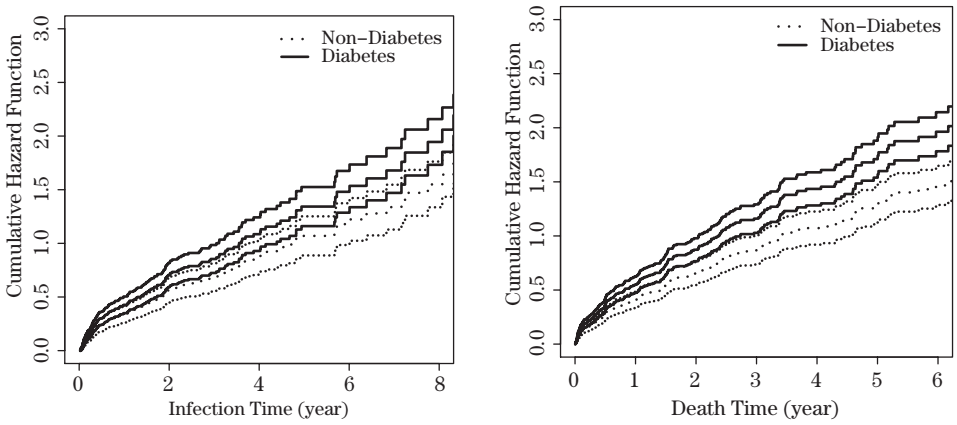

Figure 2. The estimated cumulative hazard functions for recurrent infection and terminal event time with 95% confidence intervals.

panel of Figure 2, although the difference between the two groups of patients is insignificant over time. The cardiovascular disease status does not affect the occurrences of infection and death at each stage of the PD process.

To access the adequacy of the proposed mixture model, the martingale residual plot based on the estimated martingale residuals $\{(\hat{M}_{ik}^R(\tau), \hat{M}_{ik}^D(\tau))\}$ is depicted in the Supplementary Material for both the nonterminal and the terminal events. The martingale residuals show no systematic patterns and hence suggest that the specified models are adequate.

6.2. Analysis of EMBT data

In the EBMT data, about 23% (35%) of subjects had relapse (death) as the first-occurring event, and 9% of subjects experienced both events. We considered semicompeting risks analysis, where relapse is the nonterminal event and death is the terminal event, and three gratwohl score groups, the "low risk" (the reference), "medium risk" and "high risk" groups, were included as the covariates in each state. The time to relapse, if observed, was also included as a covariate to access its effects on the time to death after relapse. Since this is semicompeting risks data, we only fit the path indicator model (2.1) to the data at stage 1 and the probability of relapse at the second stage was set to zero (see Remark 3).

The estimates for the regression coefficients of the path indicator model (2.1), and the regression coefficients of gap time models (2.2) and (2.3) are reported in Table 5. We see that, compared to the low and high risk groups, the medium risk group has a significantly lower chance to experience relapse than to experience death as the first-occurring event. In analysis of time to relapse, we see that

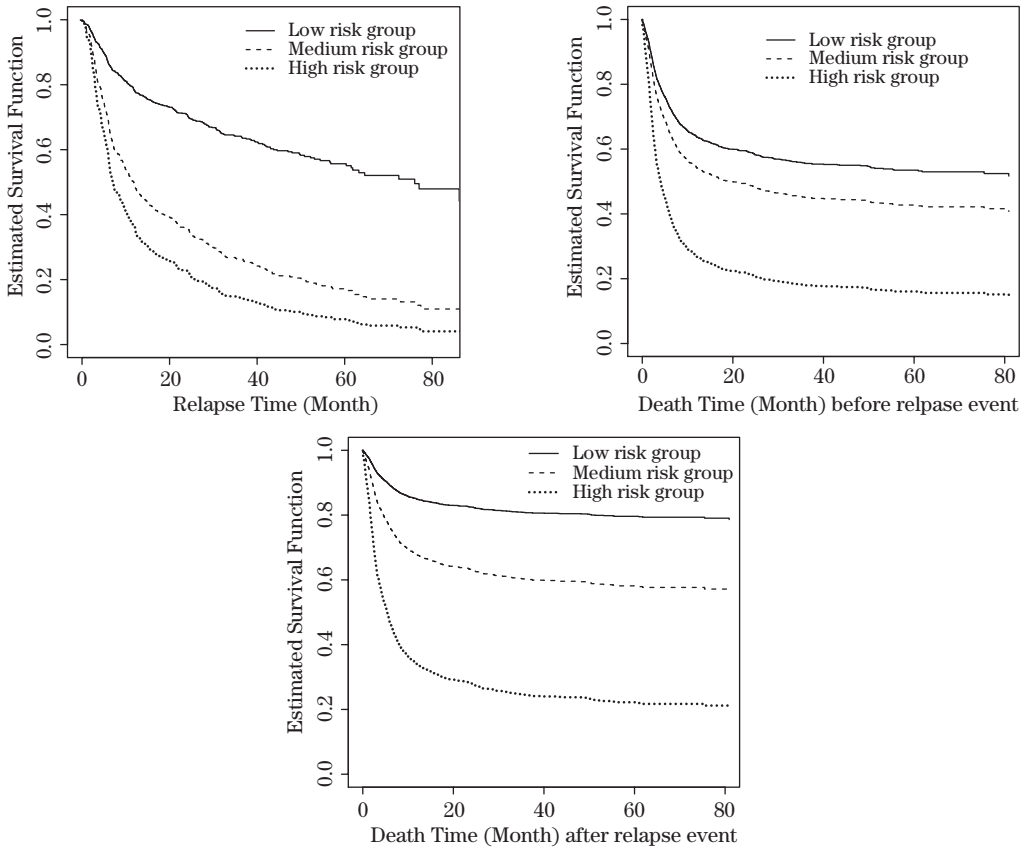

Figure 3. The estimated survival curves for the relapse and death times.

medium and high risk groups have significantly higher rates of relapse than the low risk group, and the same phenomenon is observed for time to death. The estimate of α_d in model (2.3) is negative and significant, implying significant decelerating effects on death by the relapse time. In the top-left panel of Figure 3, we see the estimated survival functions for relapse as the first-occurring event (stage 1) among the three risk groups, and the top-right panel shows the corresponding estimated survival functions for death. In the bottom panel of Figure 3, the estimated survival functions for death after relapse (stage 2) are shown for the three risk groups with the previous duration time to relapse given by the respective group-specific sample medians. We see that the medium and high risk groups have significantly lower survival rates for the relapse event. The survival function for death at stage 1 in the medium risk group is close to that in the low risk group. The survival functions for death after relapse (stage 2), after adjusting for the previous duration time to relapse, seem to exhibit quite

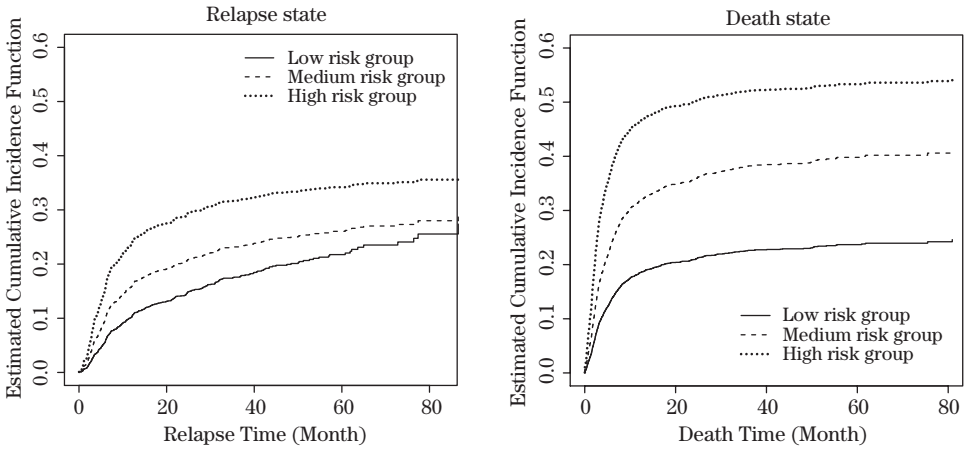

Figure 4. The estimated cumulative incidence functions for relapse (left) and death state (right) at stage 1.

different patterns than the survival functions for death at stage 1.

The estimated CIF of "relapse" and "death" events at stage 1 are depicted in Figure 4 for subjects in the high, medium, and low risk groups. In the left panel, we see that the CIF curves for the relapse event have no significant difference among the three groups of patients. On the other hand, high risk group tends to have higher cumulative incidence rates of death than the medium group, and the medium group also have higher cumulative incidence rates than the low risk group. The model checking of the proposed model is provided in the Supplementary Material, showing the models specified do not severely deviate from the observed data.

7. Conclusions

Our proposal facilitates joint analysis of recurrent events and dependent competing risks, with the analysis focus on gap times. In addition to the covariate effects on the gap times between events, the proposed analysis also examines effects from previous event or gap times on the current gap time directly through event time regression models, rather than indirectly through unobserved frailty or random effects. One limitation for this approach may be that a specific association pattern for the relationship between the previous and the current gap times needs to be assumed in the analysis. However, our simulation results suggest that the analysis is robust to moderate model misspecification (see the Supplementary Material).

Although we restricted our attention to the case where there is one nonter-

minal, possibly recurrent event and one terminal event, the proposed framework can be readily extended to allow for multiple terminal and nonterminal events by generalizing the logistic model component in the mixture regression modeling to a multinomial logistic regression model for multiple transition paths. The event time regression models for predicting gap times can also be made even more flexible by generalizing the semiparametric proportional hazards models to the semiparametric transformation models (Chen (2010, 2012)), which include the proportional hazards and proportional odds models as special cases.

Supplementary Materials

Supplementary material include the loglikelihood function (3.1), score equations of parameters, information matrix, the proof of Theorems 1 and 2 and additional results from Sections 5, 6 and 7.

Acknowledgment

The authors are grateful for the helpful comments from an associate editor and two referees. This reseach was supported by Taiwan Ministry of Science and Techonology Grant 102-2118-M-305-004.

References

Andersen, P.K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research* **11**, 91-115.

Andersen, P.K., Abildstrom, S.Z. and Rosthøj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research* **11**, 203-215.

Chang, I.S., Hsiung, C.A., Wen, C.C., Wu, Y.J. and Yang, C.C. (2007). Non-parametric maximum-likelihood estimation in a semiparametric mixture model for competing-risks data. *Scand. J. Statist.* **34**, 870-895.

Chen, C.M., Chuang, Y.W. and Shen, P.S. (2015). Two-stage estimation for multivariate recurrent event data with a dependent terminal event. *Biometrical Journal* **57**, 215-233.

Chen, Y.H. (2009). Weighted Breslow-type and maximum likelihood estimation in semiparametric transformation models. *Biometrika* **96**, 591-600.

Chen, Y.H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *J. Roy. Statist. Soc. B* **72**, 235-251.

Chen, Y.H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression model. *Lifetime Data Anal.* **18**, 36-57.

Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041-1046.

Fine, J.P. (1999). Analyzing competing risks data with transformation models. *J. Roy. Statist. Soc. B* **61**, 817-830.

Fine, J.P., Jiang, H. and Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907-919.

Hsieh, J.J., Wang, W.J. and Ding, A.A. (2008). Regression analysis based on semicompeting risks data. *J. Roy. Statist. Soc. B* **70**, 3-20.

Hu, C. and Tsodikov, A. (2014). Semiparametric regression analysis for time-to-event marked endpoints in cancer studies. *Biostatistics* **15**, 513-525.

Huang, X. and Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics* **63**, 389-397.

Lin, D.Y., Sun, W. and Ying, Z. (1999). Nonparametric estimation of gap time distributions for serial events with censored data. *Biometrika* **86**, 59-70.

Liu, L., Wolfe, R.A. and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747-756.

Lu, W. and Peng, L. (2008). Separametric analysis of mixture regression models with competing risks data. *Lifetime Data Anal.* **14**, 231-252.

Peng, L. and Fine, J.P. (2007). Regression modeling of semicompeting risks data. *Biometrics* **63**, 96-108.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. and Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* **8**, 708-721.

Tsiatis, A.A. (1998) Competing risks. *Encyclopedia of Biostatistics* **1**, 824-834.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Wang, W.J. and Wells, M. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85**, 561-572.

Xu, J., Kalbfleisch, J.D. and Tai, B. (2010). Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics* **66**, 716-725.

Ye, Y., Kalbfleisch, J.D. and Scaubel, D.E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63**, 78-87.

Zeng, D. and Lin, D.Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627-640.

Zeng, D. and Lin, D.Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. Roy. Statist. Soc. B* **69**, 507-564.

Zeng, D. and Lin, D.Y. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* **65**, 746-752.

Department of Statistics, National Taipei University, Taipei, Taiwan.

E-mail: chuang2342@mail.ntpu.edu.tw

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

E-mail: yhchen@stat.sinica.edu.tw

Division of Nephrology, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan.

E-mail: colaladr@yahoo.com.tw

(Received October 2015; accepted May 2016)