Semiparametric modeling and estimation of the terminal behavior of recurrent marker processes before failure events

Kwun Chuen Gary Chan

Department of Biostatistics and Department of Health Services,

University of Washington, Seattle, Washington 98105, U.S.A.

and

Mei-Cheng Wang

Department of Biostatistics, Johns Hopkins University,

Baltimore, Maryland 21205, U.S.A.

December 8, 2015

Author's Footnote:

Kwun Chuen Gary Chan is Associate Professor, Department of Biostatistics and Department of Health Services, University of Washington, 1959 NE Pacific Street, Seattle WA 98195. (email: kcgchan@uw.edu). Mei-Cheng Wang is Professor, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore MD 21205. (email: mcwang@jhsph.edu). This research was partially supported by the National Institutes of Health grants P01 CA 098252, R01 AI 089341 and R01 HL 122212.

Abstract

Recurrent event processes with marker measurements are mostly and largely studied with forward time models starting from an initial event. Interestingly, the processes could exhibit important terminal behavior during a time period before occurrence of the failure event. A natural and direct way to study recurrent events prior to a failure event is to align the processes using the failure event as the time origin and to examine the terminal behavior by a backward time model. This paper studies regression models for backward recurrent marker processes by counting time backward from the failure event. A three-level semiparametric regression model is proposed for jointly modeling the time to a failure event, the backward recurrent event process, and the marker observed at the time of each backward recurrent event. The first level is a proportional hazards model for the failure time, the second level is a proportional rate model for the recurrent events occurring before the failure event, and the third level is a proportional mean model for the marker given the occurrence of a recurrent event backward in time. By jointly modeling the three components, estimating equations can be constructed for marked counting processes to estimate the target parameters in the three-level regression models. Large sample properties of the proposed estimators are studied and established. The proposed models and methods are illustrated by a community-based AIDS clinical trial to examine the terminal behavior of frequencies and severities of opportunistic infections among HIV infected individuals in the last six months of life.

KEYWORDS: Marked Counting Process; Partial Likelihood; Recurrent Event Process; Semiparametric models.

1. INTRODUCTION

In prospective follow-up studies, recurrent events with marker data are often collected where markers are observed at the occurrence of recurrent events. Examples include CD4 counts, viral load measurements, severity scores at recurrent opportunistic infections and medical costs at repeated hospital visits. Recurrent marker processes often exhibit a certain terminal behavior before failure events such as deaths. Lunney et al. (2003) studied the functional decline at the last year of life and concluded that heterogeneity in the decline is often observed in clinical settings. They suggested that a better understanding of functional decline before death will improve organization and delivery of palliative care. The understanding of marker behavior prior to the failure event could identify vulnerable subpopulations that will help prioritizing services or treatments in resources-limited settings. Using the human immunodeficiency virus (HIV) infection as an illustrating example, it is scientifically and clinically interesting to understand the terminal behavior of the frequency and severity of opportunistic infections before death. Evidence suggests that HIV-infected patients experienced higher frequency of AIDS-defining events before death, where the frequencies could vary with gender, risk behaviors or geographic location (Chan et al., 1995). Taking Alzheimer's Disease as another example, cognitive decline across the spectrum of numerous symptoms is commonly characterized by changes in biomarkers before diagnosis of the disease. Interestingly, backward-in-time performance of biomarkers for Alzheimer's Disease is widely recognized as an important research topic for signaling the occurrence of disease, though the data analyses conducted in the research area typically overlooked the effect of censoring from follow-up (Hall et al., 2000; Wilson et al. 2007, among other relevant papers). For end-stage renal diseases, Usvyat et al. (2013) studied the pattern of interdialytic weight gain, systolic blood pressure, serum albumin and C-reactive protein levels before death, and showed that race was associated with some biomarkers but not others. Based on the results, they made a clinical recommendation on how patients with end-stage renal diseases should be monitored.

When interests focus on the behavior of processes counting time forward from an initial event, Nelson (1988), Pepe and Cai (1993), Lawless and Nadeau (1995), Cook and Lawless (1997), Lin et al. (2000), Wang, Qin and Chiang (2001), Liu, Wolfe and Huang (2004), Zeng and Lin (2007), Liu, Schaubel and Kalbfleisch (2012) among others studied recurrent event processes, Lin (2000) studied medical cost processes, Pawitan and Self (1993) studied CD4 count processes, and Huang and Wang (2004), Ye, Kalbfleisch and Schaubel (2007), Huang, Qin and Wang (2010), Zhao, Zhou and Sun (2011) and Kalbfleisch et al. (2013) considered joint analysis of recurrent and terminal events. Such forward processes, however, cannot conveniently model the terminal behavior of the processes. In fact, most of the forward models studied in the literature implicitly assume that the recurrent marker processes show no change in pattern before failure events. An exception is Liu, Wolfe and Kalbfleisch (2007), who considered a joint model of repeated monthly medical costs and failure time which allows costs to increase uniformly for a period of time before death.

By aligning the time origins to the failure events, models on backward processes are natural and direct ways to study the terminal behavior of stochastic processes. Although the notion of backward process is implicitly employed in the medical publications listed above, the adopted analytical approaches are usually ad hoc and based on only uncensored data. For example, Chan et al. (1995) and Lunney et al. (2003) employed complete-case analyses that used only uncensored observations to study risk factors associated with the number of infections before death. A similar analytical strategy was also used by Hall et al. (2000) to study the association between biomarkers and Alzheimer's disease incidence, where uncensored observations are treated as cases and all the censored observations are treated as controls. Such analytical approaches are common in biomedical data applications and would typically result in biased estimates for population parameters when failure time and recurrent marker process are correlated. Specifically, if we treat right censored subjects as missing, the complete-case subsample is biased toward parameter values for a subpopulation with shorter failure times. To study the terminal behavior of dynamic processes, Chan and Wang (2010) introduced the backward recurrent event process that counts time backward from the failure event. In that paper they compared forward and backward models, and considered one sample estimation problems. By aligning the origins of the processes to failure events, the terminal behavior of stochastic processes could be naturally and directly studied by backward processes.

In this paper, we propose a three-level joint model of failure time and backward recurrent marker processes. We directly model the recurrent marker processes before failure events by aligning the origins of the processes at failure events, and study the terminal behavior of the processes by counting time backward from the failure events. One of the analytical challenges is to deal with censoring, as time origins of backward processes are only partially observed due to right censoring. Another challenge arises because we want to model the backward processes flexibly by semiparametric models which contain multiple nuisance functions. By jointly modeling the failure time, recurrent events and a marker measurement, estimating equations are proposed for follow-up data collected subject to right censoring.

The manuscript is organized as follows. A three-level semiparametric model is proposed in Section 2 to jointly model the hazard rate of failure, the rate of occurrence of recurrent events before death and the average level of marker measurement observed at each recurrence. Semiparametric inference for the proposed model is developed in Section 3: Estimation and inference of finite-dimensional target parameters are established in Section 3.1, estimation of infinite-dimensional functional parameters is studied in Section 3.2, and a diagnostic test for the proposed model is discussed in Section 3.3. Section 4 contains numerical results, which includes a simulation study and an analysis of a data set from the Terry Beirn Community Programs for Clinical Research on AIDS studies. Section 5 provides several concluding remarks. Proofs of the theoretical results are given in the Appendix.

2. A THREE-LEVEL SEMIPARAMETRIC MODEL

In this section, a semiparametric model is proposed for the backward recurrent marker processes. Let $(M(t), Q(t))$ be a bivariate process in the time interval $[0, \tau_1]$, where τ_1 is a pre-specified time corresponding to the maximal follow-up time, $M(t)$ is the cumulative number of recurrent events from the time origin to time t, and $Q(t)$ is a marker process which is potentially observable conditioning on the occurrence of an event, *i.e.* $dM(t) = 1$, $t \in [0, \tau_1]$. Suppose the marker data are observed only when recurrent events occur. Here, $M(\cdot)$ and $Q(\cdot)$ could be arbitrarily correlated. We define $V(t) = \int_0^t Q(s)dM(s)$ as a recurrent marker process. This process is well-defined since $Q(s)$ is defined given $dM(s) = 1$, or equivalently, given the occurrence of a recurrent event at time s.

In the Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA) study, the initial event is treatment randomization and the failure event is death, $M(t)$ is the number of opportunistic infections within t time units after randomization, $Q(t)$ is the severity score associated with an infection at time t and $V(t)$ is the severity-weighted number of opportunistic infections up to time t after randomization (Neaton et al., 1994). $V(t)$ is a measure of both the frequency and the severity of opportunistic infections. In the absence of censoring, recurrent marker processes are terminated by a failure event such as death. We call such an event a theoretical terminal event to distinguish them from an observed terminal event in the presence of incomplete follow-up. Let T be the time to the failure event, C be the censoring time and $Y = \min(T, C)$ be the observed terminal event. Often in the recurrent events literature, an observed terminal event is called a censoring event which may cause confusion, and we distinguish the observed terminal event from the potential censoring event in this paper. Let $\Delta = I(T \leq C)$ be an indicator that a failure event is observed, and X be a q-vector covariate. Extensions to time-varying covariates will be discussed in Section 5.

To study the terminal behavior of recurrent marker processes, we define the following backward processes: $M^B(u) = M(T) - M(T-u)^-$, $Q^B(u) = Q(T-u)$ and $V^B(u) = \int_0^u Q^B(v)dM^B(v) = V(T) - V(T-u)^-$, where the superscript $^-$ represents the left-hand limit. With slight abuse of notations, the superscript $^{\mathrm{T}}$ represents the transpose of a matrix or a column vector. Conditioning on covariate X, the outcomes of interests $(T, M^B(\cdot), Q^B(\cdot))$ are allowed to be correlated. A conditional independent censoring condition is assumed for estimation, in which $(T, M^B(\cdot), Q^B(\cdot))$ is conditionally independent of the censoring time C given covariate X. Note that the backward processes are defined relative to the failure event occurring at T, not the observed terminal event occurring at Y. In the CPCRA study, $M^B(u)$ is the number of opportunistic infections within u time units before death, $Q^B(u)$ is the severity score associated with an infection at time u before death and $V^B(u)$ is the severity-weighted number of opportunistic infections within u time units before death. The time origins for the backward processes, $u = 0$, are the failure events of interest. The mean function of backward processes has a direct interpretation as the average pattern of processes before failure events. The forward and backward processes are designed to study different scientific questions about the processes. For example, forward medical cost processes are more relevant for studying the total cost of care for cancer patients, but backward costs are more relevant for comparing and evaluating palliative care settings.

There are certain considerations for constructing regression models of backward processes. First of all, the failure time T, the backward recurrent event process $M^B(u)$ and the backward marker process $Q^B(u)$ are dependent in general and a regression model should allow such a dependence. Also, there is an order of defining the processes. A backward process can be viewed as a marked process defined at the failure event, which extends the notion of a marked variable discussed in Huang and Louis (1998), and a marked variable can be considered as a generalization of a cause of death in competing risk (Prentice et al., 1978; Sun, Gilbert and McKeague, 2009). In addition, a backward marker measurement is observed only when a recurrent event occurs. Note that parameters for marginal models of marked processes are not identified when we have limited study period, as discussed in Huang (2002) and Huang and Wang (2003). One way to overcome this identifiability problem is to jointly model both the failure time and the backward processes.

Following the above considerations, we propose the following three-level regression model for jointly modeling T, $M^B(u)$ and $Q^B(u)$. Let $h(t; x)$ be the hazard function at time t conditioning on a covariate value x, and $\lambda(u; x, t)$ be the backward rate of recurrent event at u time units before the failure event given a covariate value x and the occurrence of failure event at time t after the initial event, that is $\lambda(u; x, t)du = E(dM^B(u)|X = x, T = t)$. Conditioned on an occurrence of recurrent event at backward time u before the failure event, let $\mu(u; x, t)$ represents the mean of marker, $Q^B(u)$, given covariate value x and failure time t, that is $\mu(u; x, t) = E(Q^B(u)|X = x, T = t, M^B(du) = 1)$. In the CPCRA example, $\lambda(u; x, t)$ is the average number of opportunistic infections *per unit time* at backward time u before death given covariates $X = x$ and that death occurs at $T = t$, $\mu(u; x, t)$ is the mean severity score if an opportunistic infection occurs at backward time u before death. As the terminal behavior of processes occur within a rather short period of time before failure events, relevant scientific questions center on a short period τ_0 before death, where τ_0 is a pre-specified time period of interest which is usually much short than the maximum follow-up period τ_1. For example, Chan et al. (1995) studied the frequency of opportunistic infections within the last six months of life, and Chan and Wang (2010) studied the medical cost for cancer patients within the last year of life. While τ_0 is typically specified beforehand indicating a time period of scientific interest, it is ideal to pick τ_0 which is not too small and not too large. If τ_0 is small, only a small amount of backward data can be used and it will lead to a reduction of estimation efficiency. If τ_0 is large, as subjects with failure time less than τ_0 are not included in the model population, it implies that a good proportion of subjects will be excluded from the model population and this will subsequently affect the biomedical or public health interpretation of the analysis results.

We consider the following three-level model:

Level 1. Proportional hazard model of T conditioning on $X = x$,

$$h(t; x) = h_0(t) \exp(\xi_0^{\mathrm{T}} x)$$

Level 2. Proportional rate model of $M^B(u)$ conditioning on $(X = x, T = t)$,

$$\lambda(u; x, t) = \lambda_0(u) \exp(f_0(t) + \alpha_0^{\mathrm{T}} x)$$

Level 3. Proportional mean model of $Q^B(u)$ conditioning on $dM^B(u) = 1$ and $(X = x, T = t)$,

$$\mu(u; x, t) = \mu_0(u) \exp(g_0(t) + \beta_0^{\mathrm{T}} x)$$

where $t > 0$, $u \in [0, \tau_0]$, and $h_0(t)$, $f_0(t)$, $g_0(t)$, $\lambda_0(u)$ and $\mu_0(u)$ are unspecified functions. Since the models at levels 2 or 3 each contains two unspecified functions, we normalize $\lambda_0(u)$ and $\mu_0(u)$ for identifiability purposes. In particular, we assume $\int_0^{\tau_0} \lambda_0(u)du = 1$ and $\int_0^{\tau_0} \lambda_0(u)\mu_0(u)du = 1$. The parameters of main interests are ξ_0, α_0 and β_0, which have interpretations as the log relative hazards of the failure event, the log relative rate of the recurrent events before the failure event, and the log relative mean of the backward markers at an recurrent event before the failure event. The proposed three level models recognize the ordering of observation by sequential conditioning, and also allow the processes to be correlated with failure time. Moreover, the model parameters are estimable from follow-up data subject to incomplete follow-up, as will be discussed in section 3.

Note that the models at levels 2 and 3 together imply an integrated model for the recurrent marker process $V^B(u)$,

$$\nu(u; x, t) = \nu_0(t) \exp(l_0(t) + \gamma_0^{\mathrm{T}} x)$$

where $\nu(u; x, t)du = \mu(u; x, t) \times \lambda(u; x, t)du = E(dV^B(u)|X = x, T = t)$ is called the backward generalized rate of a recurrent marker process, $\nu_0(u) = \lambda_0(u) \times \mu_0(u)$, $l_0(t) = f_0(t) + g_0(t)$ and $\gamma_0 = \alpha_0 + \beta_0$. The parameter γ_0 has the interpretation of log relative mean of the backward recurrent marker process *per unit time.* In the CPCRA example, α_0 is log rate ratio of the number of opportunistic infections *per unit time* before death, β_0 is log mean ratio of the severity score of an opportunistic infection before death and $\gamma_0 = \alpha_0 + \beta_0$ is log mean ratio of the severity-weighted frequency of opportunistic infection *per unit time* before death.

In levels 2 and 3 models, we only specify the rate and the mean of the processes, but not the full distribution. While a hazard function completely specifies a survival distributions, the specification of the full distribution of marked point processes is very difficult in general (Cox and Isham, 1980). Although a Poisson process assumption and a parametric assumption for the mark distribution given the history of the processes can specify the full likelihood, such assumptions are usually considered very restrictive in the recurrent event literature (Lin et al. 2001; Cai et al., 2010; among others). In contrast with those fully parameterized models, our semiparametric rate and mean models possess some desirable features and therefore provide more flexibility in modeling and data analysis. Furthermore, while shared frailty models are viable alternatives, they often require full parametric specification which is unappealing in the analysis of recurrent events. As a comparison, through the target parameters and multiple nonparametric functional parameters, our joint models in levels 1 to 3 handle the dependence structure of the three outcome components flexibly, which is known to be notoriously difficult in the recurrent event literature.

3. ESTIMATION AND INFERENCE

3.1 Finite-dimensional target parameters

Let $\mathcal{N}_i(t) = I(Y_i \le t, \Delta_i = 1), i = 1, \ldots, n$, be counting processes of the observed failure events and $\mathcal{N}_i^*(t) = \mathcal{N}_i(t) - \int_0^t I(Y_i \ge s) \exp(\xi_0^{\mathrm{T}} X_i) dH_0(s)$ be martingale residual processes, where $H_0(t) = \int_0^t h_0(s)ds$. For notational convenience, we further define the following: For any real-valued q-dimensional vector c, let $S^{(k)}(t; c) = n^{-1} \sum_{i=1}^n I(Y_i \ge t) X_i^{\otimes k} \exp(c^{\mathrm{T}} X_i)$, $k = 0, 1, 2$ where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^{\mathrm{T}}$, and $s^{(k)}(t; c) = E[I(Y \ge t) X^{\otimes k} \exp(c^{\mathrm{T}} X)]$. Also, let $\bar{X}(t; c) = S^{(1)}(t; c)/S^{(0)}(t; c)$ and $\bar{x}(t; c) = s^{(1)}(t; c)/s^{(0)}(t; c)$.

It is well known that the log relative hazard parameter ξ_0 in the level 1 model can be estimated by maximizing the partial likelihood (Cox 1972), and is equivalent to solving the partial score equation $U_1(\tau_1; \xi) = 0$, where

$$U_1(t; \xi) = \frac{1}{n} \sum_{i=1}^n \int_0^t \left(X_i - \bar{X}(s; \xi)\right) d\mathcal{N}_i(t) = 0 \ .$$

A difficulty for estimating α_0 and β_0 in levels 2 and 3 models is that there are four unspecified nuisance functions, $\lambda_0(u), \mu_0(u), f_0(t), g_0(t)$, and there are two different time scales, the forward time scale t and the backward time scale u. For proportional hazards model with two time scales, Efron (2002) argued that the nuisance functions in two time scales cannot be eliminated simultaneously in the estimation of finite-dimensional parameters. Under the proposed three-level model, we develop the following method to eliminate the nuisance functions in both time scales and to estimate the target parameters α_0 and β_0. The method naturally extends the partial likelihood methodology to marked counting processes, with a major difference that the estimators for the levels 2 and 3 models cannot be derived from a profile likelihood approach similar to the partial likelihood for the level 1 model, since the rate and mean models do not specify the full distribution of the backward processes.

Define a bivariate process to record information of $M_i^B(u)$, $u \in [0, \tau_0]$, for a subject with an uncensored event in $[\tau_0, t]$: $\mathcal{M}_i(t, u) = M_i^B(u) I(\tau_0 \leq Y_i \leq t, \Delta_i = 1)$. For $t \in [\tau_0, \tau_1]$ and $u \in [0, \tau_0]$, under the conditional independent censoring condition, observe that

$$\begin{aligned} & E(\mathcal{M}_i(dt, du)|Y_i \geq t, X_i) \\ = \;& E(\mathcal{N}_i(dt)|T_i \geq t, C_i \geq t, X_i) \times E(\mathcal{M}_i(dt, du)|\mathcal{N}_i(dt) = 1, C_i \geq t, X_i) \\ = \;& E(\mathcal{N}_i(dt)|T_i \geq t, X_i) \times E(M_i^B(du)|\mathcal{N}_i(dt) = 1, X_i) \\ = \;& \lambda_0(u) h_0(t) \exp(f_0(t) + (\alpha_0 + \xi_0) X_i)\, dt du. \end{aligned} \tag{1}$$

Consider the reparametrization $\theta = \alpha + \xi$. For $t \in [\tau_0, \tau_1]$ and $u \in [0, \tau_0]$, (1) implies $E(\mathcal{M}_i^*(t, u; B_0^M, \theta_0)) = 0$ where

$$\mathcal{M}_i^*(t, u; B_0^M, \theta_0) = \mathcal{M}_i(t, u) - \int_{v=0}^{u} \int_{s=\tau_0}^{t} I(Y_i \geq s) \lambda_0(v) \exp(\theta_0 X_i)\, dB_0^M(s)\, dv \ , \tag{2}$$

$\theta_0 = \alpha_0 + \xi_0$ and $B_0^M(t) = \int_{\tau_0}^{t} \exp(f_0(s))\, dH_0(s)$. Furthermore, $\mathcal{M}_i^*(t, \tau_0; B_0^M, \theta_0) = \mathcal{M}_i(t, \tau_0) - \int_{s=\tau_0}^{t} I(Y_i \geq s) \exp(\theta_0 X_i)\, dB_0^M(s)$ since $\int_0^{\tau_0} \lambda_0(u) du = 1$. This motivates us to consider the following set of estimating equations:

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau_0} \int_{\tau_0}^{\tau_1} \mathcal{M}_i^*(dt, du; B^M, \theta) = 0 \tag{3}$$

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau_0} \int_{\tau_0}^{\tau_1} X_i \mathcal{M}_i^*(dt, du; B^M, \theta) = 0 \ . \tag{4}$$

For a given θ, the solution to (3) is

$$\hat{B}^M(t; \theta) = \sum_{i=1}^{n} \int_0^{\tau_0} \int_{\tau_0}^{t} \frac{\mathcal{M}_i(ds, du)}{\sum_{j=1}^{n} \int_0^{\tau_0} I(Y_i \geq s) \lambda_0(v) e^{\theta X_j}\, dv} = \sum_{i=1}^{n} \int_0^{\tau_0} \int_{\tau_0}^{t} \frac{\mathcal{M}_i(ds, du)}{S^{(0)}(s; \theta)} \ ,$$

where the last equality holds because $\int_0^{\tau_0} \lambda_0(u)du = 1$. Replacing B^M with $\hat{B}^M$, (4) becomes $U_2(\tau_1;\theta) = 0$ where

$$\begin{aligned} U_2(t;\theta) &= \frac{1}{n}\sum_{i=1}^n \int_0^{\tau_0}\int_{\tau_0}^t \left(X_i - \bar{X}(s;\theta)\right)\mathcal{M}_i(ds,dv) \\ &= \frac{1}{n}\sum_{i=1}^n \int_{\tau_0}^{\tau_1} \left(X_i - \bar{X}(s;\theta)\right) M_i^B(\tau_0)\mathcal{N}_i(ds) \ . \end{aligned}$$

Note that $U_2(t;\theta)$ only involves the target parameter θ, but not the nuisance functions $\lambda_0(u)$ and $B_0^M(t)$. Denote the solution of $U_2(\tau_1;\theta) = 0$ by $\hat{\theta}$, α_0 can be estimated by $\hat{\alpha} = \hat{\theta} - \hat{\xi}$.

To estimate β_0, we consider a marked counting process $\mathcal{V}_i(t,u) = V_i^B(u)I(\tau_0 \le Y_i \le t, \Delta_i = 1)$. Define $\phi_0 = \alpha_0 + \beta_0 + \xi_0$, $B_0^V(t) = \int_{\tau_0}^t \exp(f_0(s) + g_0(s))dH_0(s)$. Following similar arguments as above, we have $E(\mathcal{V}_i^*(t,u;B_0^V,\phi_0)) = 0$ where

$$\mathcal{V}_i^*(t,u;B_0^V,\phi_0) = \mathcal{V}_i(t,u) - \int_{v=0}^u \int_{s=\tau_0}^t I(Y_i \ge s)\lambda_0(v)\mu_0(v)\exp(\phi_0 X_i)\, dB_0^V(s)\, dv \ .$$

Furthermore, we can estimate ϕ_0 by solving the estimating equation $U_3(\tau_1;\phi) = 0$, where

$$\begin{aligned} U_3(t;\phi) &= \frac{1}{n}\sum_{i=1}^n \int_0^{\tau_0}\int_{\tau_0}^t \left(X_i - \bar{X}(s;\phi)\right)\mathcal{V}_i(ds,du) \\ &= \frac{1}{n}\sum_{i=1}^n \int_{\tau_0}^t \left(X_i - \bar{X}(s;\phi)\right) V_i^B(\tau_0)\mathcal{N}_i(ds) \ . \end{aligned}$$

Note that $U_3(t;\phi)$ only involves the target parameter ϕ, but not other nuisance functions. Denote the solution of $U_3(t;\phi) = 0$ by $\hat{\phi}$, β_0 can be estimated by $\hat{\beta} = \hat{\phi} - \hat{\alpha} - \hat{\xi} = \hat{\phi} - \hat{\theta}$.

To study the asymptotic results, we introduce the following notations: $\hat{\psi} = (\hat{\xi}^{\mathrm{T}}, \hat{\theta}^{\mathrm{T}}, \hat{\phi}^{\mathrm{T}})^{\mathrm{T}}$, $\psi_0 = (\xi_0^{\mathrm{T}}, \theta_0^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}$, $\eta_{i1} = \int_0^{\tau_1}(X_i - \bar{x}(t;\xi_0))\mathcal{N}_i^*(dt)$, $\eta_{i2} = \int_0^{\tau_0}\int_{\tau_0}^{\tau_1}(X_i - \bar{x}(t;\theta_0))\mathcal{M}_i^*(dt,du)$ and $\eta_{i3} = \int_0^{\tau_0}\int_{\tau_0}^{\tau_1}(X_i - \bar{x}(t;\phi_0))\mathcal{V}_i^*(dt,du)$, $\eta_i = (\eta_{i1}^{\mathrm{T}}, \eta_{i2}^{\mathrm{T}}, \eta_{i3}^{\mathrm{T}})^{\mathrm{T}}$, $\Sigma = E(\eta_1\eta_1^{\mathrm{T}})$, $A_1 = E(\int_0^{\tau_1}(X - \bar{x}(t;\xi_0))^{\otimes 2} I(Y \ge t)e^{\xi_0 X}dH_0(t))$, $A_2 = E(\int_{\tau_0}^{\tau_1}(X - \bar{x}(t;\theta_0))^{\otimes 2} I(Y \ge t)e^{\theta_0 X}dB_0^M(t))$, $A_3 = E(\int_{\tau_0}^{\tau_1}(X - \bar{x}(t;\phi_0))^{\otimes 2} I(Y \ge t)e^{\phi_0 X}dB_0^V(t))$, and $A = A_1 \oplus A_2 \oplus A_3$ where $\oplus$ denotes direct sum of matrices.

The following theorem indicates that the root-n scaled and centered versions of the parameter estimators of the three level models will jointly converge in distribution to a zero-mean $3 \times q$-variate normal distribution.

Theorem 1. *Under the regularity conditions stated in the appendix,* $\hat{\psi} \overset{a.s.}{\to} \psi_0$ *and* $\sqrt{n}(\hat{\psi} - \psi_0) \overset{d}{\to} N(0, V)$ *where* $V = A^{-1}\Sigma(A^{-1})^{\mathrm{T}}$.

The proof is given in Appendix A.1. As apparent in the definition of A and Σ, the estimation the asymptotic variance matrix V requires consistent estimation of certain cumulative versions of the baseline functions $H_0(t)$, $B_0^M(t)$ and $B_0^V(t)$, for which we discuss next.

3.2 Functional parameters

It is well known that $H_0(t)$ can be estimated by the Breslow (1974) estimator,

$$\hat{H}(t; \hat{\xi}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{N}_i(ds)}{S^{(0)}(s; \hat{\xi})} .$$

We could also estimate $B_0^M(t)$ and $B_0^V(t)$ by Breslow-type estimators

$$\hat{B}^M(t; \hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_0}\int_{\tau_0}^t \frac{\mathcal{M}_i(ds, du)}{S^{(0)}(s; \hat{\theta})}$$

and

$$\hat{B}^V(t; \hat{\phi}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_0}\int_{\tau_0}^t \frac{\mathcal{V}_i(ds, du)}{S^{(0)}(s; \hat{\phi})}$$

respectively. Note that the estimates $\hat{B}^M$ and $\hat{B}^V$ can be used for the estimation of the asymptotic covariance matrix of the finite-dimensional parameter, as well as for the model diagnostic procedure given in Section 3.3. A remark on the types of functional parameters considered in the literature is given in Section 5. By plugging in unknown parameters $(\xi_0, \theta_0, \phi_0, H_0(t), B_0^M(t), B_0^V(t))$ in the definition of A and Σ by their estimates, we obtain $\hat{A}$ and $\hat{\Sigma}$. Therefore, we can estimate the asymptotic variance matrix V by $\hat{V} = \hat{A}^{-1}\hat{\Sigma}(\hat{A}^{-1})^{\mathrm{T}}$, which is shown to be a consistent estimator of V in Appendix A.2.

Furthermore, we can estimate $L_0(u) = \int_0^u \lambda_0(v)\, dv$ and $N_0(u) = \int_0^u \nu_0(v)\, dv = \int_0^u \lambda_0(v)\mu_0(v)\, dv$ by

$$\hat{L}_0(u) = \sum_{i=1}^{n}\int_{\tau_0}^{\tau_1} \frac{\mathcal{M}_i(ds, u)}{\sum_{j=1}^n I(Y_j \geq s)e^{\hat{\theta}X_j}} \Bigg/ \sum_{i=1}^{n}\int_{\tau_0}^{\tau_1} \frac{\mathcal{M}_i(ds, \tau_0)}{\sum_{j=1}^n I(Y_j \geq s)e^{\hat{\theta}X_j}}$$

and

$$\hat{N}_0(u) = \sum_{i=1}^{n}\int_{\tau_0}^{\tau_1} \frac{\mathcal{V}_i(ds, u)}{\sum_{j} I(Y_j \geq s)e^{\hat{\phi}X_j}} \Bigg/ \sum_{i=1}^{n}\int_{\tau_0}^{\tau_1} \frac{\mathcal{V}_i(ds, \tau_0)}{\sum_{j=1}^n I(Y_j \geq s)e^{\hat{\phi}X_j}} .$$

To identify the limiting processes of the functional estimators, we define the following notations: $b_0^M(t;\theta_0) = \int_{\tau_0}^t \bar{x}(s;\theta_0)dB_0^M(s)$, $b_0^V(t;\phi_0) = \int_{\tau_0}^t \bar{x}(s;\theta_0)dB_0^V(s)$,

$$\Pi_i^M(t) = \int_{\tau_0}^t \frac{\mathcal{M}_i^*(ds,\tau_0)}{s^{(0)}(s;\theta_0)} - b_0^M(t;\theta_0)A_2^{-1}\eta_{2i} \ ,$$

$$\Pi_i^V(t) = \int_{\tau_0}^t \frac{\mathcal{V}_i^*(ds,\tau_0)}{s^{(0)}(s;\phi_0)} - b_0^V(t;\phi_0)A_3^{-1}\eta_{3i} \ ,$$

$$\Lambda_i^M(u) = \int_{\tau_0}^{\tau_1} \frac{\mathcal{M}_i^*(ds,u)}{s^{(0)}(s;\theta_0)} - L_0(u)b_0^M(\tau_1;\theta_0)A_2^{-1}\eta_{2i} \ ,$$

$$\Lambda_i^V(u) = \int_{\tau_0}^{\tau_1} \frac{\mathcal{V}_i^*(ds,u)}{s^{(0)}(s;\phi_0)} - N_0(u)b_0^V(\tau_1;\phi_0)A_3^{-1}\eta_{3i} \ ,$$

$$\Gamma_i^M(u) = \frac{\Lambda_i^M(u) - L_0(u)\Lambda_i^M(\tau_0)}{B_0^M(\tau_1)} \ ,$$

and

$$\Gamma_i^V(u) = \frac{\Lambda_i^V(u) - N_0(u)\Lambda_i^V(\tau_0)}{B_0^V(\tau_1)} \ .$$

The following theorem summarizes the large sample properties of the functional estimators.

Theorem 2. *Under the regularity conditions stated in the appendix,*

1. $\sqrt{n}(\hat{B}^M(t;\hat{\theta}) - B_0^M(t))$ *converges to a mean-zero Gaussian process with a covariance function* $\rho^M(t_1,t_2) = E(\Pi_1^M(t_1)\Pi_1^M(t_2))$ *for* $t_1,t_2 \in [\tau_0,\tau_1]$.

2. $\sqrt{n}(\hat{B}^V(t;\hat{\phi}) - B_0^V(t))$ *converges to a mean-zero Gaussian process with a covariance function* $\rho^V(t_1,t_2) = E(\Pi_1^V(t_1)\Pi_1^V(t_2))$ *for* $t_1,t_2 \in [\tau_0,\tau_1]$.

3. $\sqrt{n}(\hat{L}(u) - L_0(u))$ *converges to a mean-zero Gaussian process with a covariance function* $\varrho^M(u_1,u_2) = E(\Gamma_1^M(u_1)\Gamma_1^M(u_2))$ *for* $u_1,u_2 \in [0,\tau_0]$.

4. $\sqrt{n}(\hat{N}(u) - N_0(u))$ *converges to a mean-zero Gaussian process with a covariance function* $\varrho^V(u_1,u_2) = E(\Gamma_1^V(u_1)\Gamma_1^V(u_2))$ *for* $u_1,u_2 \in [0,\tau_0]$.

Inference for the functional parameters can be conducted by the following resampling method. For simplicity, we only discuss the construction of simultaneous confidence bands for $L_0(u), u \in [0,\tau_0]$ in detail and similar procedures can be applied to construct confidence bands for $N_0(u)$, $B_0^M(t)$ and $B_0^V(t)$. Let $\hat{\Gamma}_i^M(u)$ be an estimator of $\Gamma_i^M(u)$ by plugging in unknown quantities by consistent estimators defined earlier. The limiting process of $\sqrt{n}(\hat{L}(u)-L_0(u))$ can be approximated by $\tilde{\varrho}^M(u) = n^{-1/2}\sum_{i=1}^n \hat{\Gamma}_i^M(u)G_i$ where $G=(G_1,\ldots,G_n)$ are independent standard normal variables which are generated independent of the observed data. Conditioning on the data, the only randomness in $\tilde{\varrho}^M$ comes from $(G_1,\ldots,G_n)$, and we can show as in Appendix A.4 in Lin et al. (2000) that $\tilde{\varrho}^M(u)$ and $\sqrt{n}(\hat{L}(u)-L_0(u))$ have the same limiting process. By simulating many realizations of G, we can obtain $c_{\alpha/2}$ which is the $(1-\alpha)\times 100$ percentile of $\sup_{u\in[0,\tau_0]}|\tilde{\varrho}^M(u)/\sqrt{\hat{\varrho}^M(u,u)}|$, where $\hat{\varrho}^M(u,u) = n^{-1}\sum_{i=1}^n(\hat{\Gamma}_i^M(u))^2$. Since $L_0(u)$ is positive, we can construct confidence bands using the logarithmic transformation, and the values are given by $\hat{L}(u)\exp\{\pm n^{-1/2}c_{\alpha/2}\sqrt{\hat{\varrho}^M(u,u)}/\hat{L}(u)\}, u\in[0,\tau_0]$. Pointwise confidence intervals can be constructed by replacing $c_{\alpha/2}$ with the standard normal critical value $z_{\alpha/2}$.

3.3 Goodness-of-fit test

When the proposed model is true, residuals $\mathcal{N}^*$, $\mathcal{M}^*$ and $\mathcal{V}^*$ are unbiased. Therefore, residual-based diagnostic test statistics can be constructed to examine potential model violation. We focus on testing against violation of the exponential link functions, since the parameter interpretation and the validity of reparametrizations θ and ϕ depends critically on the exponential link. For the proportional hazards model, Lin et al. (1993) considered a goodness-of-fit test statistic based on the following process:

$$\mathcal{T}_1(x) = n^{-1}\sum_{i=1}^n I(\hat{\xi}^{\mathrm{T}}X_i \le x)\hat{\mathcal{N}}_i^*(\tau_1)\ .$$

They proposed a test based on $\sup_x|\mathcal{T}_1(x)|$ and showed that the test is consistent against incorrect link functions in the form of $g(\xi^{*T}X)$, where g is not the exponential function. Inspired by this idea, we consider two additional processes for the recurrent events and the marker processes:

$$\mathcal{T}_2(y) = n^{-1}\sum_{i=1}^n I(\hat{\theta}^{\mathrm{T}}X_i \le y)\hat{\mathcal{M}}_i^*(\tau_1,\tau_0)$$

and

$$\mathcal{T}_3(z) = n^{-1}\sum_{i=1}^{n} I(\hat{\phi}^{\mathrm{T}}X_i \le z)\hat{\mathcal{V}}_i^*(\tau_1,\tau_0)\ .$$

Let $a = (x, y, z)$ and $\mathcal{T}(a) = (\mathcal{T}_1(x), \mathcal{T}_2(y), \mathcal{T}_3(z))^{\mathrm{T}}$. A test statistic is given by $\mathcal{S} = \sup_a |\mathcal{T}(a)|$. We use the following simulation method to approximate the null distribution of $\mathcal{S}$. We show in Appendix A.4 that $\sqrt{n}\mathcal{T}(a)$ converges to a mean-zero Gaussian process under the null hypothesis that the model is correctly specified, and $\sqrt{n}\mathcal{T}(a) = n^{-1/2}\sum_{i=1}^{n}\mathcal{T}_i^{\dagger}(a) + o_p(1)$. The limiting distribution of $\sqrt{n}\mathcal{T}(a)$ can be approximated by $n^{-1/2}\sum_{i=1}^{n}\hat{\mathcal{T}}_i^{\dagger}(a)G_i$, where $\hat{\mathcal{T}}_i^{\dagger}(a)$ replaces unknown functions in $\mathcal{T}^{\dagger}(a)$ by their consistent estimators given in Appendix A.4 and $G = (G_1, \ldots, G_n)$ are independent normal variables generated independent of the data. By generating many realizations of G, the null distribution of $\sqrt{n}\mathcal{S}$ is approximated by the distribution of $\sup_a |n^{-1/2}\sum_{i=1}^{n}\hat{\mathcal{T}}_i^{\dagger}(a)G_i|$. Conditioning on observed data, the only randomness comes from G and we can use the arguments in Appendix A.5 of Lin et al. (2000) to show that the limiting processes of $\sqrt{n}\mathcal{T}(a)$ and $n^{-1/2}\sum_{i=1}^{n}\hat{\mathcal{T}}_i^{\dagger}(a)G_i$ are the same.

4. NUMERICAL STUDIES

We studied the finite sample performance of the proposed estimators by Monte Carlo simulations. Data were generated 5000 times in each simulation, and each simulated data set consisted of 100, 200 or 400 observations. We simulated the data as follows. Two covariates were included in the model: a Bernoulli(0.5) covariate, and a standard Gaussian covariate. Failure time was generated by $T = 0.75 + T'$ where T' is Weibull distributed with shape parameter 2 and scale parameter $1/\exp(0.5 \times \xi_0^{\mathrm{T}}X)$, where $\xi_0 = (0.5, 1)$. Censoring time was generated from a uniform $(0, 5)$ distribution independent of the failure time. Given T and X, the recurrent event processes at time u before death were generated from a non-homogeneous Poisson process with rate $2(1 - \exp(-u/2))\exp(-\log T + \alpha_0^{\mathrm{T}}X)$ where $\alpha_0 = (0.75, 0.25)$. Given T, X and occurrence of recurrent event at time u before death, the associated marks were generated from a gamma distribution with shape parameter 1 and scale parameter $\exp(-0.5u - 0.5\log T + \beta_0^{\mathrm{T}}X)$ where $\beta_0 = (-0.5, -0.25)$ and $u \in [0, 1]$.

The frequency of recurrent events and marker values were negatively correlated with survival time.

The simulation results are summarized in Table 1. In each case, the estimators for ξ_0, α_0, β_0 had small biases, the variance estimates were close to the sampling variation of the parameter estimates and the empirical coverage of the 95% confidence intervals based on the proposed sandwich variance estimate was close to the nominal value. We conducted additional simulations to study the performance of the proposed model diagnostic tests for the exponential link function. To approximate the null distributions, the resampling procedure described in Section 3.3 was performed 1000 times for each simulation data set. Under exponential link functions, the tests at 5% significance level had an empirical rejection proportion of 5.3% and 5.1% for $n = 200$ and $n = 500$, respectively. Under a misspecified linear link function, the simulated powers were 73.1% and 96.3% for $n = 200$ and $n = 500$, respectively. Under a logarithmic link function, the empirical powers were 82.2% and 99.2% for $n = 200$ and $n = 500$, respectively.

[Table 1 about here.]

The proposed methods were also applied to analyze the recurrent marker data collected from the ddC/ddI trials of the Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA) study. The CPCRA study is a randomized trial comparing didanosine (ddI) and zalcitabine (ddC) as treatments for HIV-infected patients. Recurrent opportunistic infections are common to HIV-infected individuals because of their compromised immune system. The trial collected the time from randomization to death or censoring, along with the occurrence of recurrent opportunistic diseases and the severity of each infection. Each individual was observed to experience between 0 and 5 recurrent infections, and a severity score is assessed by ten physicians at each event (Neaton et al, 1994). Table 2 shows the results from the proposed regression model with ddI/ddC treatment arm, age, gender and race (African American vs. non African American) as explanatory variables. The diagnostic tests for exponential link functions gave p-values of 0.43, 0.82 and 0.96 for models with

$\tau_0 = 4, 5, 6$ months. Results show that there is a significant gender and race effect on the frequency or severity of opportunistic infections up to six months before death. In particular, female had 76% less opportunistic infections in the last six months of life than male after controlling for other explanatory variables and failure time (95% C.I.: 93% less to 10% less). It is estimated that African Americans had 55% less opportunistic infections in the last six months of life compare to non-African Americans (95% C.I.: 75% less to 18% less). In contrast, using a forward proportional rate model as in Lin et al. (2000) and Cai et al. (2010), it is estimated that females had 23% less opportunistic infections (95% C.I.: 82% less to 220% more) than males and African Americans had 36% less opportunistic infections (95% C.I.: 60% less to 4% more) than non-African Americans. The estimated baseline backward cumulative functions $\hat{L}_0(u)$ and $\hat{N}_0(u)$ are given in Figure 1, which shows that the frequency and severity of opportunistic infections had a sharp increase between two to four weeks prior to death.

[Table 2 about here.]

[Figure 1 about here.]

5. CONCLUDING REMARKS

In this paper, we proposed a semiparametric regression model for studying the terminal behavior of recurrent marker processes. The model consists of three levels, the first level is a proportional hazards model for a failure time, the second level is a proportional rate model for a backward recurrent event process and the third level is a proportional mean model for a backward marker given a recurrent event occurrence before death. The combined model involves three-level modeling in which the backward process models are constructed with failure time included as a covariate, and the model for backward marker measurement conditions on both the failure time and the occurrence of a recurrent event. This three-level conditional model takes into account of the order of data observation and dependence is allowed among the three outcome variables. For this complicated problem, we proposed a conceptually sophisticated yet analytically simple procedure for estimation and inference. Specifically, by carefully constructing the flexible model structure in Section 3, we end up with multiple cancellation of nuisance functions in levels 1 to 3 models from the proposed estimating equations, which we consider a novel procedure for estimation. Alternative models can be developed in the future, for example, using subject-specific latent variables to model the dependence among the failure time, the backward recurrent event process and the backward marker process.

A referee indicated that Liu et al. (2007) proposed a turn-back-time method to negate the impact of recurrent events immediately prior to death. In particular, they modeled longitudinal costs over discrete times from an initial event, and assumed that there is a uniform increase in costs for the last b months, which has a similar flavor as the approach of Wulfsohn and Tsiatis (1997). While we agree that the model of Liu et al. is novel and interesting, it is important to point out that our model and analytical approach are, in fact, quite different from the one in Liu et al. (2007), whose main interest was to model the association of the risk factors with monthly medical costs. The approach of Liu et al. did not consider modeling recurrent events (such as repeated hospitalizations) which in fact have an important role to determine the medical cost. Also, the regression parameters for covariates are assumed to be the same for the initial period and the final period before death. Since only monthly costs are modeled, their approach does not serve to answer questions such as whether a certain treatment is associated with more frequent hospitalizations. Our model focused on random recurrent events where markers are observed only at recurrent events but not on regular intervals. Another important distinction is that they artificially censor the cost information before failure or censoring events in order to focus on estimation of the initial periods. Censoring of cost information before failure events is also considered by Huang and Wang (2003). Our focus is entirely in the opposite direction: instead of discarding information from the last period before the events, we only use those information to answer important scientific questions relating to this period.

Similar to the difficulty encountered in the existing literature of flexible semiparametric joint modeling (Lin and Ying, 2001, among others), the de-convolution of mixed functions is generally hard and tedious, if not impossible, and in this paper we only estimate mixed functional parameters such as $B_0^M(t)$ and $B_0^V(t)$. While non-mixed cumulative functions can be estimated when failure time is assumed to be independent of the processes given covariates (Lin et al, 2000, Cai et al., 2010), our model is more general and can handle informative survival and recurrent events which is known to be notoriously difficult in the recurrent event literature. Estimation of cumulative mixed function is a tradeoff for a more general model, but the functional estimates can still be interpreted as cumulative mark-specific hazard functions (Huang and Louis, 1998), and are essential for the estimation of the asymptotic covariance matrix of the finite-dimensional parameter (as shown in Appendix A.2), and for the model diagnostic procedure given in Section 3.3.

An associate editor suggested us to consider an extension to time-varying covariates, which requires careful thoughts as there are two time scales in the models (forward and backward times). Let $X(t)$ be a vector of time-varying covariates and $\mathcal{X}(t)$ the history of the time-varying process $X(\cdot)$ up to t. Suppose we consider the following models:

Level 1**.** Proportional hazard model of T conditioning on $\mathcal{X}(t)$, depends only on $X(t) = x(t)$,

$$h(t;x) = h_0(t)\exp(\xi_0^{\mathrm{T}} x(t))$$

Level $2^{\dagger}$**.** Proportional rate model of $M^B(u)$ conditioning on $(T = t, \mathcal{X}(t))$, depends only on $(T = t, X(t-u) = x(t-u))$,

$$\lambda(u; x(t-u), t) = \lambda_0(u)\exp(f_0(t) + \alpha_0^{\mathrm{T}} x(t-u))$$

Level $3^{\dagger}$**.** Proportional mean model of $Q^B(u)$ conditioning on $T = t, \mathcal{X}(t)$ and $dM^B(u) = 1$ only depends on $(T = t, X(t-u) = x(t-u))$:

$$\mu(u; x(t-u), t) = \mu_0(u)\exp(g_0(t) + \beta_0^{\mathrm{T}} x(t-u)) \ .$$

For $t \geq \tau_0$ and $u \in [0, \tau_0]$, we observe analogous to (1) that

$$E(\mathcal{M}_i(dt, du)|Y_i \geq t, \mathcal{X}_i(t)) = \lambda_0(u)h_0(t)\exp(f_0(t) + \alpha_0 X_i(t-u) + \xi_0 X_i(t))\, dtdu \ ,$$

and define

$$\mathcal{M}_i^\dagger(t,u;B_0^M,\alpha_0,\xi_0) = \mathcal{M}_i(t,u) - \int_{v=0}^{u}\int_{s=\tau_0}^{t} I(Y_i \geq s)\lambda_0(v)\exp\left(\alpha_0 X_i(s-v) + \xi_0 X_i(s)\right) dB_0^M(s)\, dv\ ,$$

then $E(\mathcal{M}_i^\dagger(t,u;B_0^M,\theta_0)) = 0$.

Following previous arguments, we derive

$$\hat{B}^M(t;\alpha,\xi) = \sum_{i=1}^{n}\int_0^{\tau_0}\int_{\tau_0}^{t} \frac{\mathcal{M}_i(ds,du)}{\sum_{j=1}^{n}\int_0^{\tau_0} I(Y_i \geq s)\lambda_0(v)e^{\alpha^{\mathrm{T}}X_j(s-v)+\xi^{\mathrm{T}}X_j(s)}\, dv}\ .$$

Note that $\lambda_0(v)$ is **not** eliminated through integration as in the case of time-independent covariates. In fact, it is typical that one of the nuisance parameters cannot be eliminated under a proportional hazard model with two time scales (Efron, 2002). As discussed in Efron (2002), we can use a cubic regression spline for approximating $\lambda_0(u)$ for $u \in [0,\tau_0]$. Since we assume that $\int_0^{\tau_0}\lambda_0(u)du = 1$, $\lambda_0(u)$ is analogous to a density function on $[0,\tau_0]$, and we can use the log-spline density approximation as in Kooperberg and Stone (1991): $\lambda_0(u) = \exp(\kappa_0^{\mathrm{T}}B(u) - c(\kappa_0))$ where $B(u) = (B_1(u),\ldots,B_p(u))^{\mathrm{T}}$ are spline basis functions, κ_0 is a p-dimensional vector of unknown coefficients and $c(\kappa_0) = \log(\int \exp(\kappa_0^{\mathrm{T}}B(u))du)$ is a normalizing constant. Let $X_i^\dagger(t,u) = (B(u), X_i(t-u))^{\mathrm{T}}$, a system of estimating equations can be constructed as

$$\int_0^{\tau_0}\int_{\tau_0}^{\tau_1} (X_i^\dagger(s,u) - \bar{X}_i^\dagger(s))\mathcal{M}_i(dt,du) = 0$$

where

$$\bar{X}_i^\dagger(t) = \frac{\sum_{j=1}^{N} I(Y_j \geq t)\int_0^{\tau_0} X_j(t,u)\exp(\kappa^{\mathrm{T}}B(u) + \alpha^{\mathrm{T}}X_j(t-u) + \hat{\xi}^{\mathrm{T}}X_j(t))\, du}{\sum_{j=1}^{N} I(Y_j \geq t)\int_0^{\tau_0}\exp(\kappa^{\mathrm{T}}B(u) + \alpha^{\mathrm{T}}X_j(t-u) + \hat{\xi}^{\mathrm{T}}X_j(t))\, du}\ .$$

The above estimating equation is $p+q$ dimensional with $p+q$ unknown parameters. A similar estimating equation can be derived for estimating $\gamma_0 = \alpha_0 + \beta_0$. Alternatively, when the level 2 and 3 models are formulated as

Level 2^***.** Proportional rate model of $M^B(u)$ conditioning on $(T = t, \mathcal{X}(t))$, depends only on $(T = t, X(t) = x(t))$,

$$\lambda(u; x(t), t) = \lambda_0(u)\exp(f_0(t) + \alpha_0^{\mathrm{T}}x(t))$$

Level 3***.** Proportional mean model of $Q^B(u)$ conditioning on $T = t, \mathcal{X}(t)$ and $dM^B(u) = 1$ only depends on $(T = t, X(t) = x(t))$:

$$\mu(u; x(t), t) = \mu_0(u) \exp(g_0(t) + \beta_0^{\mathrm{T}} x(t)) \ .$$

Then,

$$E(\mathcal{M}_i(dt, du) | Y_i \geq t, \mathcal{X}_i(t)) = \lambda_0(u) h_0(t) \exp(f_0(t) + (\alpha_0 + \xi_0) X_i(t)) \, dt du \ ,$$

and

$$\begin{aligned} \hat{B}^M(t; \alpha, \xi) &= \sum_{i=1}^n \int_0^{\tau_0} \int_{\tau_0}^t \frac{\mathcal{M}_i(ds, du)}{\sum_{j=1}^n \int_0^{\tau_0} I(Y_i \geq s) \lambda_0(v) e^{(\alpha+\xi) X_j(s)} \, dv} \\ &= \sum_{i=1}^n \int_0^{\tau_0} \int_{\tau_0}^t \frac{\mathcal{M}_i(ds, du)}{\sum_{j=1}^n I(Y_i \geq s) \lambda_0(v) e^{(\alpha+\xi) X_j(s)}} \ . \end{aligned}$$

In this case, $\lambda_0(u)$ can be eliminated from the estimation of finite-dimensional parameters, and the method proposed for time-independent covariates can be directly extended. For this model, time-varying covariates are assumed observable at the failure event, but we can use time-varying covariates which are defined as summary measures of the end-of-life history of certain processes. For example, with a time-varying treatment status $Z(t)$, $X(t)$ can be defined as the fraction of years in which an individual takes the treatment between time $t - \tau_0$ and t, that is $X(t) = \int_{t-\tau_0}^t Z(t) dt / \tau_0$. Using this definition, $X(t)$ is well defined for individuals with $Y \geq \tau_0$, which are individuals being included in the estimation of level 2 and 3 models. We can also change the domain of integration from $[0, \tau_1]$ to $[\tau_0, \tau_1]$ for proportional hazards model under this definition of $X(t)$.

In this paper we study the terminal behavior of recurrent marker processes before failure events. In applications, it could be clinically more relevant to consider the complete history of recurrent events for evaluating treatments on disease progression. Nevertheless, studying the full history of recurrent marker processes is not straightforward, particularly when it is complicated by the presence of terminal behavior before failure event of interest. In this paper we focus on modeling the terminal behavior of processes as it is often an important scientific question. A promising future research direction is to jointly model the forward and backward processes to study the complete history of processes, for which one has to pay extra attention to model consistency in joint modeling.

APPENDIX.

A.1 Asymptotic properties of finite-dimensional parameters

We assume the following regularity conditions:

(a) Parameters $(\xi, \theta, \phi) \in \Xi \times \Theta \times \Phi$, where Ξ, Θ and Φ are compact subspaces of $\mathbb{R}^q$.

(b) $P(C \geq \tau_1) > 0$.

(c) The support for $M^B(\tau_0), V^B(\tau_0)$ and X are bounded.

(d) The matrix A is positive definite.

These conditions are analogous to those of Andersen and Gill (1982) for the proportional hazards model and Lin et al. (2000) for the proportional rate model for recurrent events.

We first prove the consistency of the proposed estimators. It follows from Andersen and Gill (1982) that the maximum partial likelihood estimator $\hat{\xi}$ is a consistent estimator of ξ_0. To show the consistency of $\hat{\theta}$, we first consider

$$L_2(\theta) = \frac{1}{n}\left[\sum_{i=1}^{n}\int_{\tau_0}^{\tau_1}(\theta-\theta_0)^{\mathrm{T}}X_i\mathcal{M}_i(dt,\tau_0) - \int_{\tau_0}^{\tau_1}\log\left\{\frac{S^{(0)}(t;\theta)}{S^{(0)}(t;\theta_0)}\right\}\mathcal{M}(dt,\tau_0)\right].$$

where $\mathcal{M}(t,u) = n^{-1}\times\sum_{i=1}^{n}\mathcal{M}_i(t,u)$. Since $\mathcal{M}(t,\tau_0)$ and $S^{(0)}(t;\theta)$ have bounded variations, it follows from strong law of large numbers that $L_2(\theta)$ converges almost surely to

$$\mathcal{L}_2(\theta) = E\left[\int_{\tau_0}^{\tau_1}(\theta-\theta_0)^{\mathrm{T}}X_1\mathcal{M}_1(dt,\tau_0) - \int_{\tau_0}^{\tau_1}\log\left\{\frac{s^{(0)}(t;\theta)}{s^{(0)}(t;\theta_0)}\right\}\mathcal{M}_1(dt,\tau_0)\right]$$

for every θ. Note that

$$\frac{\partial^2 L_2(\theta)}{\partial\theta^2} = -\frac{1}{n}\sum_{i=1}^{n}\int_{\tau_0}^{\tau_1}[X_i-\bar{X}(t;\theta)]^{\otimes 2}I(Y_i\geq t)\exp(\theta^{\mathrm{T}}X_i)\frac{\mathcal{M}_i(dt,\tau_0)}{S^{(0)}(t;\theta)}$$

which is negative semidefinite and $\partial L_2/\partial\theta$ is $U_2(\tau_1;\theta)$. Also, it can be seen that $\partial\mathcal{L}_2(\theta_0)/\partial\theta = 0$ and $\partial^2\mathcal{L}_2(\theta_0)/\partial\theta^2 = -E(\eta_{21}\eta_{21}^{\mathrm{T}})$ under the model assumptions. Since $L_2(\theta)$ and $\mathcal{L}_2(\theta)$ are both concave functions and the parameter space is compact, it follows that $\sup_{\theta\in\Theta}|L_2(\theta) - \mathcal{L}_2(\theta)| \overset{a.s.}{\to} 0$ (Rockafellar, 1970). Note that $\hat{\theta}$ is the unique maximizer for L_2 and θ_0 is the unique maximizer of $\mathcal{L}_2$. Therefore, following the arguments in Appendix A.1 of Lin et al. (2000), we can show that $\hat{\theta}$ converges almost surely to θ_0. Strong consistency of $\hat{\phi}$ can be proven using similar arguments.

We will then prove the weak convergence results. Let $U^*(t;\psi) = (U_1(t;\xi)^{\mathrm{T}}, U_2(t,\theta)^{\mathrm{T}}, U_3(t,\phi)^{\mathrm{T}})^{\mathrm{T}}$. It can be seen that

$$U^*(t;\psi_0) = \bar{\mathcal{D}}_X^*(t) - \int_0^t \bar{X}^*(s;\psi_0)d\bar{\mathcal{D}}^*(s)$$

where $\bar{\mathcal{D}}^*(t) = n^{-1}\times(\sum\mathcal{N}_i^*(t)^{\mathrm{T}}, \sum\mathcal{M}_i^*(t,\tau_0)^{\mathrm{T}}, \sum\mathcal{V}_i^*(t,\tau_0)^{\mathrm{T}})^{\mathrm{T}}$ and $\bar{\mathcal{D}}_X^*(t) = n^{-1}\times(\sum X_i\mathcal{N}_i^*(t)^{\mathrm{T}}, \sum X_i\mathcal{M}_i^*(t,\tau_0)^{\mathrm{T}}, \sum X_i\mathcal{V}_i^*(t,\tau_0)^{\mathrm{T}})^{\mathrm{T}}$. Since $\mathcal{N}_i^*(t)$, $\mathcal{M}_i^*(t,\tau_0)$ and $\mathcal{V}_i^*(t,\tau_0)$ can be expressed as sums and products of monotone functions in t with bounded second moments, which implies that $\bar{\mathcal{D}}^*(t)$ and $\bar{\mathcal{D}}_X^*(t)$ are Donsker (van der Vaart and Wellner (1997), p. 215). Then it follows from functional central limit theorem (Pollard (1990), p.53; van der Vaart and Wellner (1997), p. 211) that $(\sqrt{n}\bar{\mathcal{D}}(t), \sqrt{n}\bar{\mathcal{D}}_X(t))$ converges weakly to zero-mean Gaussian processes. Also, $\sqrt{n}\{S^{(0)}(t;c) - s^{(0)}(t;c)\}$ and $\sqrt{n}\{S^{(1)}(t;c) - s^{(1)}(t;c)\}$ converges weakly to zero-mean Gaussian processes as shown in Lemma 5.1 in Tsiatis (1981). Using Lemma 3 of Gill (1989), one can establish that the mapping $U^*(\cdot,\phi)$ is compactly differentiable with respect to the supremum norm, and by the functional delta method, one can conclude that

$$\sqrt{n}U^*(\tau;\psi_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_i + o_p(1)$$

where $\eta_i = (\eta_{i1}^{\mathrm{T}}, \eta_{i2}^{\mathrm{T}}, \eta_{i3}^{\mathrm{T}})^{\mathrm{T}}$, $\eta_{i1} = \int_0^{\tau_1}(X_i - \bar{x}(t;\xi_0))\mathcal{N}_i^*(dt)$, $\eta_{i2} = \int_0^{\tau_0}\int_{\tau_0}^{\tau_1}(X_i - \bar{x}(t;\theta_0))\mathcal{M}_i^*(dt,du)$ and $\eta_{i3} = \int_0^{\tau_0}\int_{\tau_0}^{\tau_1}(X_i - \bar{x}(t;\phi_0))\mathcal{V}_i^*(dt,du)$. It follows that the limiting variance of $\sqrt{n}U^*(\tau_1;\psi_0)$ is Σ.

By Taylor series expansion,

$$\sqrt{n}(\hat{\psi} - \psi_0) = \sqrt{n}\hat{A}^{-1}(\psi^*)U^*(\tau_1;\psi_0)$$

where $\hat{A}(\psi) = -\partial U^*(\tau_1; \psi)/\partial\psi$ and ψ^* is on the line segment joining $\hat{\psi}$ and ψ_0. Let $A(\psi) = E[-\partial U^*(\tau_1; \psi)/\partial\psi]$. By similar arguments as in the proof of consistency, one can show that $\sup_{\psi \in \Psi} |\hat{A}(\psi) - A(\psi)| \stackrel{a.s.}{\to} 0$ where $\Psi = \Xi \times \Theta \times \Phi$. Together with the fact that $\hat{\psi}$ is strongly consistent for ψ_0, $A(\psi)$ is continuous at ψ_0 and $A(\psi_0) = A$, we have $\hat{A}(\psi^*)$ converges almost surely to A. Therefore,

$$\sqrt{n}(\hat{\psi} - \psi_0) = \sqrt{n} A^{-1} U^*(\tau_1; \psi_0) + o_p(1)$$

and $\sqrt{n}(\hat{\psi} - \psi_0)$ converges in distribution to a multivariate normal distribution with mean 0 and variance $V = A^{-1} \Sigma A^{-1}$.

A.2 Consistency of $\hat{B}^M(t; \hat{\theta})$, $\hat{B}^V(t; \hat{\phi})$ and $\hat{V}$

Since $\{V(\tau_0)N(t), t \in [\tau_0, \tau_1]\}$ is a Donsker class, it is also Glivenko-Cantelli. Combined with the Glivenko-Cantelli property of $S^{(0)}(t; \theta)$, we have

$$\hat{B}^M(t; \theta) \stackrel{a.s.}{\to} \int_{\tau_0}^{t} \left\{ \frac{s^{(0)}(u; \theta_0)}{s^{(0)}(u; \theta)} \right\} dB_0^M(u)$$

uniformly for $t \in [\tau_0, \tau_1]$ and θ in a neighborhood of θ_0. Also the derivative of $\hat{B}^M(t; \theta)$ with respect to θ is uniformly bounded for all large n and for θ in a neighborhood of θ_0, strong consistency of $\hat{\theta}$ implies $\hat{B}^M(t; \hat{\theta})$ converges almost surely to $B_0^M(t)$ uniformly in $t \in [\tau_0, \tau_1]$. Similarly, $\hat{B}^V(t; \hat{\phi})$ converges almost surely to $B_0^V(t)$ uniformly in $t \in [\tau_0, \tau_1]$. These results, together with the almost sure convergence of $\hat{\phi}$, $\bar{X}(t; \hat{\xi})$, $\bar{X}(t; \hat{\theta})$ and $\bar{X}(t; \hat{\phi})$ to ϕ_0, $\bar{x}(t, \xi_0)$, $\bar{x}(t, \theta_0)$ and $\bar{x}(t, \phi_0)$ respectively, implies that $n^{-1} \sum_{i=1}^{n} ||\hat{\eta}_i - \eta_i||_F^2 \stackrel{a.s.}{\to} 0$, where $|| \cdot ||_F$ is the matrix Frobenius norm. In addition, $n^{-1} \sum_{i=1}^{n} \eta_i \eta_i^{\mathrm{T}} \stackrel{a.s.}{\to} \Sigma$ by the strong law of large numbers. Therefore, $\hat{\Sigma} \stackrel{a.s.}{\to} \Sigma$. Furthermore, $\hat{\psi}$ and $\hat{A}(\psi_0)$ converges almost surely to ψ_0 and A and this implies the almost sure convergence of $\hat{A}$ to A. Therefore, $\hat{V} = \hat{A}^{-1} \hat{\Sigma} (\hat{A}^{-1})^{\mathrm{T}} \stackrel{a.s.}{\to} A^{-1} \Sigma (A^{-1})^{\mathrm{T}} = V$.

A.3 Weak convergence of functional parameters

Consider the decomposition

$$\sqrt{n}[\hat{B}^M(t;\hat{\theta}) - B_0^M(t)] = \sqrt{n}[\hat{B}^M(t;\theta_0) - B_0^M(t)] + \sqrt{n}[\hat{B}^M(t;\hat{\theta}) - \hat{B}^M(t;\theta_0)] \ . \qquad (5)$$

The first term on the right-hand side of (5) is

$$n^{-1/2}\sum_{i=1}^n \int_0^{\tau_0}\int_{\tau_0}^t \frac{\mathcal{M}_i^*(ds,du)}{S^{(0)}(s;\theta_0)} = n^{-1/2}\sum_{i=1}^n \int_0^{\tau_0}\int_{\tau_0}^t \frac{\mathcal{M}_i^*(ds,du)}{s^{(0)}(s;\theta_0)} + o_p(1) \ ,$$

which can be shown by arguments as in Appendix A.1. Furthermore, Taylor series expansion shows that the second term on the right-hand side of equation (5) equals $-\sqrt{n}b^M(t;\theta^*)(\hat{\theta} - \theta_0)$, and $b^M(t;\theta)$ converges uniformly to

$$b_0^M(t;\theta) = \int_{\tau_0}^t \bar{x}(u;\theta)dB_0^M(u) \ .$$

Combine with the results we established earlier that $\sqrt{n}(\hat{\theta} - \theta_0) = n^{-1/2}\sum_{i=1}^n \eta_{i2} + o_p(1)$, the second term on the right-hand side of (5) equals

$$-n^{-1/2}\sum_{i=1}^n b_0^M(t;\theta_0)A_2^{-1}\eta_{i2} + o_p(1) \ .$$

Hence, $\sqrt{n}[\hat{B}^M(t;\hat{\theta}) - B_0^M(t)] = n^{-1/2}\sum_{i=1}^n \Pi_i^M(t) + o_p(1)$. The desired result follows from the functional central limit theorem.

To show the weak convergence of $\sqrt{n}(\hat{L}(u) - L_0(u))$, we first define

$$\hat{\mathcal{B}}^M(u;\hat{\theta}) = \frac{1}{n}\sum_{i=1}^n \int_0^u \int_{\tau_0}^{\tau_1} \frac{\mathcal{M}_i(ds,dv)}{S^{(0)}(s;\hat{\theta})} \ .$$

Since $\mathcal{M}(t,u)$ is increasing in t and u, the collection $\{\mathcal{M}(t,u), t \in [\tau_0,\tau_1], u \in [0,\tau_0]\}$ is Donsker and is also Glivenko-Cantelli. Using arguments as in Appendix A.2, we can show that $\hat{\mathcal{B}}^M(u;\hat{\theta})$ converges almost surely to $L_0(u)B_0^M(\tau_1)$ uniformly in $u \in [0,\tau_0]$. Note that $\hat{\mathcal{B}}^M(\tau_0,\theta) = \hat{B}^M(\tau_1;\theta)$, and $\hat{\mathcal{B}}^M(\tau_0,\hat{\theta})$ converges almost surely to $L_0(\tau_0)B_0^M(\tau_1) = B_0^M(\tau_1)$. Therefore, $\hat{L}_0(u) = \hat{\mathcal{B}}^M(u;\hat{\theta})/\hat{\mathcal{B}}^M(\tau_0;\hat{\theta})$ converges almost surely to $L_0(u)$ uniformly for $u \in [0,\tau_0]$. Using similar arguments, we can show that $\hat{N}_0(u)$ converges almost surely to $N_0(u)$ uniformly for $u \in [0,\tau_0]$. Using similar arguments as above, we can show that $\sqrt{n}(\hat{\mathcal{B}}^M(u;\hat{\theta}) - L_0(u)B_0^M(\tau_1)) = n^{-1/2}\sum_{i=1}^n \Lambda_i^M(u) + o_p(1)$ where

$$\Lambda_i^M(u) = \int_0^u \int_{\tau_0}^{\tau_1} \frac{\mathcal{M}_i^*(ds,dv)}{s^{(0)}(s;\theta)} - L_0(u)b_0^M(\tau_1;\theta_0)A_2^{-1}\eta_{2i} \ .$$

By functional delta method, we can show that

$$\sqrt{n}(\hat{L}(u) - L_0(u)) = n^{-1/2} \sum_{i=1}^{n} \left[\frac{\Lambda_i^M(u) - L_0(u)\Lambda_i^M(\tau_0)}{B_0^M(\tau_1)} \right] + o_p(1) \ .$$

A.4 The null distribution of $\mathcal{S}$

Let $a = (x, y, z) \in \mathbb{R}^3$, $\mathcal{T}(a) = (\mathcal{T}_1(x), \mathcal{T}_2(y), \mathcal{T}_3(z))^{\mathrm{T}}$. By arguments as in Appendix A.3 and in the Appendices of Lin and Ying (1993) and Lin et al. (2000), we can show that under the null hypothesis that the model assumption is correct, $\sqrt{n}\mathcal{T}(a) = n^{-1/2} \sum_{i=1}^{n} \mathcal{T}_i^*(a) + o_p(1)$, where $\mathcal{T}_i^*(a) = (\mathcal{T}_{1i}^*(x), \mathcal{T}_{2i}^*(y), \mathcal{T}_{3i}^*(z))^{\mathrm{T}}$,

$$\mathcal{T}_{1i}^*(x) = \int_0^{\tau_1} \left[I(\xi_0^{\mathrm{T}} X \le x) - \frac{s_r(t, x; \xi_0)}{s^{(0)}(t; \xi_0)} - b_r(x; \xi_0) A_1^{-1} \{X_i - \bar{x}(t; \xi_0)\} \right] \mathcal{N}_i^*(dt) \ ,$$

$$\mathcal{T}_{2i}^*(y) = \int_{\tau_0}^{\tau_1} \left[I(\theta_0^{\mathrm{T}} X \le y) - \frac{s_r(t, y; \theta_0)}{s^{(0)}(t; \theta_0)} - b_r^M(y; \theta_0) A_2^{-1} \{X_i - \bar{x}(t; \theta_0)\} \right] \mathcal{M}_i^*(dt, \tau_0) \ ,$$

$$\mathcal{T}_{31}^*(z) = \int_{\tau_0}^{\tau_1} \left[I(\phi_0^{\mathrm{T}} X \le z) - \frac{s_r(t, z; \phi_0)}{s^{(0)}(t; \phi_0)} - b_r^V(z; \phi_0) A_3^{-1} \{X_i - \bar{x}(t; \phi_0)\} \right] \mathcal{V}_i^*(dt, \tau_0) \ ,$$

$$s_r(t, x, c) = E[I(Y \ge t, c^{\mathrm{T}} X \le x) \exp(c^{\mathrm{T}} X)] \ ,$$

$$b_r(x; \xi_0) = E\left[\int_0^{\tau_1} I(Y \ge t, \xi_0^{\mathrm{T}} X \le x) \exp(\xi_0^{\mathrm{T}} X)(X - \bar{x}(t; \xi_0)) dH_0(t) \right] \ ,$$

$$b_r^M(y; \theta_0) = E\left[\int_{\tau_0}^{\tau_1} I(Y \ge t, \theta_0^{\mathrm{T}} X \le y) \exp(\theta_0^{\mathrm{T}} X)(X - \bar{x}(t; \theta_0)) dB_0^M(t) \right] \ ,$$

$$b_r^V(z; \phi_0) = E\left[\int_{\tau_0}^{\tau_1} I(Y \ge t, \phi_0^{\mathrm{T}} X \le z) \exp(\phi_0^{\mathrm{T}} X)(X - \bar{x}(t; \theta_0)) dB_0^V(t) \right] \ .$$

Consistent estimators of these functions are given by

$$S_r(t, x, c) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \ge t, c^{\mathrm{T}} X_i \le x) \exp(c^{\mathrm{T}} X_i) \ ,$$

$$B_r(x; \hat{\xi}_0) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau_1} I(\hat{\xi}^{\mathrm{T}} X_i \le x)(X_i - \bar{X}(s; \hat{\xi})) \mathcal{N}_i(ds) \ ,$$

$$B_r^M(y; \hat{\theta}_0) = \frac{1}{n} \sum_{i=1}^{n} \int_{\tau_0}^{\tau_1} I(\hat{\theta}^{\mathrm{T}} X_i \le y)(X_i - \bar{X}(s; \hat{\theta})) \mathcal{M}_i(ds, \tau_0) \ ,$$

$$B_r^V(z; \hat{\phi}_0) = \frac{1}{n} \sum_{i=1}^{n} \int_{\tau_0}^{\tau_1} I(\hat{\phi}^{\mathrm{T}} X_i \le z)(X_i - \bar{X}(s; \hat{\phi})) \mathcal{V}_i(ds, \tau_0) \ .$$

By the strong consistency of $\hat{\xi}$, $\hat{\theta}$, $\hat{\phi}$ and the uniform strong law of large numbers, $S_r(u, x, c)$, $B_r(x; \hat{\xi})$,$B_r^M(y; \hat{\theta})$,$B_r^V(z; \hat{\phi})$ converge almost surely to $s_r(u, x, c), b_r(x; \xi_0), b_r^M(y; \theta_0), b_r^V(z; \phi_0)$.

REFERENCES

Andersen, P. K., and Gill, R. D. (1982), "Cox's regression model for counting processes: a large sample study," *The Annals of Statistics*, 10(4), 1100–1120.

Breslow, N. (1972), "Contribution to the discussion of the paper by DR Cox," *Journal of the Royal Statistical Society, Series B*, 34(2), 216–217.

Cai, J., Zeng, D., and Pan, W. (2010), "Semiparametric proportional means model for marker data contingent on recurrent event," *Lifetime data analysis*, 16(2), 250–270.

Chan, I. S., Neaton, J. D., Saravolatz, L. D., Crane, L. R., and Osterberger, J. (1995), "Frequencies of opportunistic diseases prior to death among HIV-infected persons," *AIDS*, 9, 1145–1152.

Chan, K. C. G., and Wang, M.-C. (2010), "Backward estimation of stochastic processes with failure events as time origins," *The Annals of Applied Statistics*, 4(3), 1602–1620.

Cook, R., and Lawless, J. (1997), "Marginal analysis of recurrent events and a terminating event," *Statistics in Medicine*, 16(8), 911–924.

Cox, D. R. (1972), "Regression models and life-tables," *Journal of the Royal Statistical Society. Series B.*, 34(2), 187–220.

Cox, D. R., and Isham, V. (1980), *"Point processes"* CRC Press: London.

Efron, B. (2002), "The two-way proportional hazards model," *Journal of the Royal Statistical Society: Series B*, 64(4), 899–909.

Gill, R. D. (1989), "Non-and semi-parametric maximum likelihood estimators and the von mises method (part 1)," *Scandinavian Journal of Statistics*, 16(2), 97–128.

Hall, C. B., Lipton, R. B., Sliwinski, M., and Stewart, W. F. (2000), "A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease," *Statistics in medicine*, 19(11-12), 1555–1566.

Huang, C.-Y., Qin, J., and Wang, M.-C. (2010), "Semiparametric Analysis for Recurrent Event Data with Time-Dependent Covariates and Informative Censoring," *Biometrics*, 66(1), 39–49.

Huang, C.-Y., and Wang, M.-C. (2004), "Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data," *Journal of the American Statistical Association*, 99(468), 1153–1165.

Huang, Y. (2002), "Calibration regression of censored lifetime medical cost," *Journal of the American Statistical Association*, 97(457), 318–327.

Huang, Y., and Louis, T. A. (1998), "Nonparametric estimation of the joint distribution of survival time and mark variables," *Biometrika*, 85(4), 785–798.

Huang, Y., and Wang, M.-C. (2003), "Frequency of Recurrent Events at Failure Time," *Journal of the American Statistical Association*, 98(463), 663–670.

Kalbfleisch, J. D., Schaubel, D. E., Ye, Y., and Gong, Q. (2013), "An Estimating Function Approach to the Analysis of Recurrent and Terminal Events," *Biometrics*, 69(2), 366–374.

Kooperberg, C., and Stone, C. J. (1991), "A study of logspline density estimation," *Computational Statistics & Data Analysis*, 12(3), 327–347.

Lawless, J., and Nadeau, C. (1995), "Some simple robust methods for the analysis of recurrent events," *Technometrics*, 37(2), 158–168.

Lin, D. Y. (2000), "Proportional means regression for censored medical costs," *Biometrics*, 56(3), 775–778.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000), "Semiparametric regression for the mean and rate functions of recurrent events," *Journal of the Royal Statistical Society: Series B*, 62(4), 711–730.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993), "Checking the Cox model with cumulative sums of martingale-based residuals," *Biometrika*, 80(3), 557–572.

Lin, D. Y., and Ying, Z. (2001), "Semiparametric and nonparametric regression analysis of longitudinal data," *Journal of the American Statistical Association*, 96(453), 103–126.

Liu, D., Schaubel, D. E., and Kalbfleisch, J. D. (2012), "Computationally efficient marginal models for clustered recurrent event data," *Biometrics*, 68(2), 637–647.

Liu, L., Wolfe, R. A., and Huang, X. (2004), "Shared frailty models for recurrent events and a terminal event," *Biometrics*, 60(3), 747–756.

Liu, L., Wolfe, R. A., and Kalbfleisch, J. D. (2007), "A shared random effects model for censored medical costs and mortality," *Statistics in medicine*, 26(1), 139–155.

Lunney, J. R., Lynn, J., Foley, D. J., Lipson, S., and Guralnik, J. M. (2003), "Patterns of functional decline at the end of life," *Journal of the American Medical Association*, 289(18), 2387–2392.

Neaton, J., Wentworth, D., Rhame, F., Hogan, C., Abrams, D., and Deyton, L. (1994), "Considerations in choice of a clinical endpoint for AIDS clinical trials. Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA).," *Statistics in medicine*, 13(19-20), 2107–2125.

Nelson, W. (1988), "Graphical analysis of system repair data," *Journal of Quality Technology*, 20(1), 24–35.

Pawitan, Y., and Self, S. (1993), "Modeling disease marker processes in AIDS," *Journal of the American Statistical Association*, 88(423), 719–726.

Pepe, M., and Cai, J. (1993), "Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates," *Journal of the American Statistical Association*, 88(423), 811–820.

Pollard, D. (1990), *"Empirical processes: Theory and applications"* Institute of Mathematical Statistics: Hayward, CA.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V., and Breslow, N. (1978), "The analysis of failure times in the presence of competing risks," *Biometrics*, 34, 541–554.

Rockafellar, R. (1970), *"Convex analysis"* Princeton: Princeton NJ.

Sun, Y., Gilbert, P. B., and McKeague, I. W. (2009), "Proportional hazards models with continuous marks," *The Annals of statistics*, 37(1), 394–426.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990), "Martingale-based residuals for survival models," *Biometrika*, 77(1), 147–160.

Tsiatis, A. A. (1981), "A large sample study of Cox's regression model," *The Annals of Statistics*, 9(1), 93–108.

Usvyat, L. A., Barth, C., Bayh, I., Etter, M., von Gersdorff, G. D., Grassmann, A., Guinsburg, A. M., Lam, M., Marcelli, D., Marelli, C. et al. (2013), "Interdialytic weight gain, systolic blood pressure, serum albumin, and C-reactive protein levels change in chronic dialysis patients prior to death," *Kidney International*, 84, 149–157.

Van der Vaart, A. W., and Wellner, J. A. (1996), *"Weak Convergence"* Springer: New York NY.

Wang, M.-C., Qin, J., and Chiang, C. (2001), "Analyzing recurrent event data with informative censoring," *Journal of the American Statistical Association*, 96(455), 1057–1065.

Wilson, R. S., Beck, T. L., Bienias, J. L., and Bennett, D. A. (2007), "Terminal cognitive decline: accelerated loss of cognition in the last years of life," *Psychosomatic Medicine*, 69(2), 131–137.

Wulfsohn, M. S., and Tsiatis, A. A. (1997), “A joint model for survival and longitudinal data measured with error,” *Biometrics*, 53(1), 330.

Ye, Y., Kalbfleisch, J. D., and Schaubel, D. E. (2007), “Semiparametric analysis of correlated recurrent and terminal events,” *Biometrics*, 63(1), 78–87.

Zeng, D., and Lin, D. Y. (2007), “Semiparametric transformation models with random effects for recurrent events,” *Journal of the American Statistical Association*, 102(477), 167–180.

Zhao, X., Zhou, J., and Sun, L. (2011), “Semiparametric Transformation Models with Time-Varying Coefficients for Recurrent and Terminal Events,” *Biometrics*, 67(2), 404–414.

Figure 1: Estimates of baseline backward cumulative functions and 95% confidence bands for CPCRA data. (a) $\hat{L}_0(u)$, (b) $\hat{N}_0(u)$.

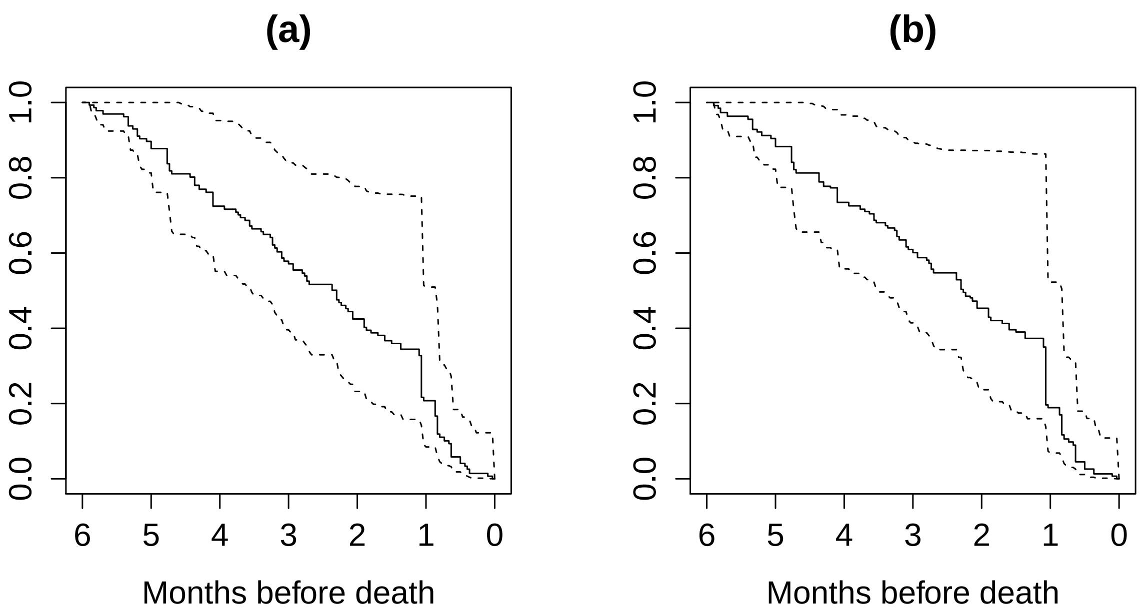

Table 1: *Summary of simulation study: Bias, standard error and coverage probabilities.*

(a) Sample size=100

	Truth	Bias	SSE	SEE	CP(%)
ξ_0	0.50	0.015	0.279	0.274	94.6
	1.00	0.033	0.297	0.287	92.8
α_0	0.75	0.013	0.245	0.233	93.7
	0.25	0.017	0.282	0.268	92.9
β_0	-0.5	0.011	0.216	0.199	91.7
	-0.25	0.001	0.237	0.216	9.13
$L_0(0.5)$	0.27	0.002	0.053	0.048	9.17
$N_0(0.5)$	0.316	0.001	0.078	0.069	90.6
$B_0^M(1.5)$	0.835	0.005	0.266	0.258	94.4
$B_0^V(1.5)$	0.539	0.006	0.199	0.189	93.5

(b) Sample size=200

	Truth	Bias	SSE	SEE	CP(%)
ξ_0	0.50	0.004	0.190	0.189	94.9
	1.00	0.015	0.204	0.198	93.2
α_0	0.75	0.002	0.166	0.161	94.2
	0.25	0.008	0.189	0.184	93.8
β_0	-0.5	0.002	0.149	0.143	93.7
	-0.25	0.003	0.162	0.153	92.6
$L_0(0.5)$	0.27	0.001	0.036	0.035	93.4
$N_0(0.5)$	0.316	0.001	0.054	0.051	93.6
$B_0^M(1.5)$	0.835	0.001	0.186	0.182	94.6
$B_0^V(1.5)$	0.539	0.001	0.139	0.136	94.3

(c) Sample size=400

	Truth	Bias	SSE	SEE	CP(%)
ξ_0	0.50	0.004	0.134	0.132	94.7
	1.00	0.004	0.139	0.139	94.7
α_0	0/75	0.003	0.114	0.113	95.2
	0.25	0.006	0.133	0.129	94.0
β_0	-0.5	0.004	0.103	0.101	94.4
	-0.25	0.003	0.111	0.108	93.8
$L_0(0.5)$	0.27	0.001	0.026	0.025	94.2
$N_0(0.5)$	0.316	0.001	0.039	0.037	93.6
$B_0^M(1.5)$ β_0	0.835	0.002	0.132	0.128	94.3
$M_0^V(1.5)$	0.539	0.002	0.100	0.096	93.6

Bias and SSE are the sampling bias and sampling standard error respectively. SEE is the sample average of the standard error estimator, and CP is the empirical coverage probability of the 95% confidence intervals.

Table 2: *Regression Analysis of the CPCRA data with four explanatory variables: Trt(1=ddC,0=ddI), Age, Sex(1=female, 0=male), Race(1=African American,0=others)*

		$\tau_0 = 4$ months		$\tau_0 = 5$ months		$\tau_0 = 6$ months	
		Est	95%C.I.	Est	95%C.I.	Est	95%C.I.
ξ_0	Trt	-0.21	(-0.49,0.08)				
	Age	0.01	(-0.01,0.03)				
	Sex	0.18	(-0.32,0.68)				
	Race	0.05	(-0.30,0.40)				
α_0	Trt	-0.03	(-0.48,0.42)	0.02	(-0.40,0.44)	0.09	(-0.36,0.54)
	Age	0.01	(-0.02,0.03)	0.01	(-0.02,0.03)	0.01	(-0.02,0.04)
	Sex	-0.76	(-1.82,0.31)	-1.40	(-2.71,-0.76)	-1.42	(-2.73,-0.10)
	Race	-1.19	(-1.96,-0.44)	-0.76	(-1.35,-0.18)	-0.80	(-1.39,-0.20)
β_0	Trt	-0.07	(-0.24,0.11)	-0.02	(-0.17,0.13)	-0.01	(-0.16,0.14)
	Age	0.00	(-0.01,0.01)	0.00	(-0.01,0.01)	0.00	(-0.01,0.01)
	Sex	-0.58	(-1.43,0.27)	-0.85	(-2.28,0.58)	-0.85	(-2.29,0.58)
	Race	0.09	(-0.14,0.33)	0.14	(-0.02,0.29)	0.07	(-0.16,0.31)
γ_0	Trt	-0.10	(-0.56,0.37)	-0.03	(-0.45,0.44)	0.08	(-0.38,0.55)
	Age	0.01	(-0.02,0.04)	0.01	(-0.02,0.03)	0.01	(-0.02,0.03)
	Sex	-1.34	(-2.73,0.05)	-2.25	(-4.21,-0.29)	-2.26	(-4.24,-0.29)
	Race	-1.09	(-1.09,-0.28)	-0.63	(-1.23,-0.03)	-0.72	(-1.36,-0.09)