CrossMark

A Joint Modeling Approach for Longitudinal Data with Informative Observation Times and a Terminal Event

Lianqiang Qu[1] · Liuquan Sun[2] · Xinyuan Song[3]

Abstract

In this article, we propose a new joint modeling approach for the analysis of longitudinal data with informative observation times and a dependent terminal event. We specify a semiparametric mixed effects model for the longitudinal process, a proportional rate frailty model for the observation process, and a proportional hazards frailty model for the terminal event. The association among the three related processes is modeled via two latent variables. Estimating equation approaches are developed for parameter estimation, and the asymptotic properties of the proposed estimators are established. The finite sample performance of the proposed estimators is examined through simulation studies, and an application to a medical cost study of chronic heart failure patients is illustrated.

Keywords Dependent terminal event · Estimating equations · Frailty · Informative observation · Longitudinal data · Mixed effects

✉ Liuquan Sun
  slq@amt.ac.cn

  Lianqiang Qu
  qulianq@mail.ccnu.edu.cn

  Xinyuan Song
  xysong@sta.cuhk.edu.hk

[1] School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, Hubei,
    People's Republic of China

[2] Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese
    Academy of Sciences, Beijing 100190, People's Republic of China

[3] Department of Statistics, The Chinese University of Hong Kong, Hong Kong,
    People's Republic of China

🕭 Springer

1 Introduction

Longitudinal data are frequently encountered in a wide variety of fields, such as medical follow-up studies and observational investigations. In many applications, observation times are usually irregular, and may be correlated with the longitudinal process. Various methods for analyzing longitudinal data with informative observation times have been developed [1,5,9,13–16,20,22]. For example, Lin et al. [9] considered a marginal regression model and proposed a class of inverse intensity-of-visit process-weighted estimators. Sun et al. [16] suggested a joint model for the longitudinal and observation processes via a shared latent variable. Liang et al. [5] proposed a class of joint models via two latent variables. All the above methods primarily analyze longitudinal data with informative observation times in the absence of a dependent terminal event.

In many situations, a dependent terminal event such as death, which precludes the follow-up, may exist. Moreover, Liu et al. [11] indicated that regarding the dependent terminal event as noninformative censoring may yield biased estimates in modeling the hospital visits and the longitudinal medical costs. It is also common that the longitudinal process may be correlated with both observation times and a terminal event. For example, Liu et al. [10] and Sun et al. [17], in analyzing data from a medical cost study, found that patients visiting the hospital more often tended to pay more for each visit, and these patients also had a higher mortality rate. That is, the longitudinal medical costs could be correlated with both hospital visits and death. Thus, there is clearly a need to develop suitable models for analyzing longitudinal data, which accounts for both informative observation times and a dependent terminal event simultaneously.

However, researches on the joint analysis of longitudinal data with informative observation times and a dependent terminal event are limited. Liu et al. [10] suggested a joint random effects model, where the random effects are assumed to be normally distributed. Sun et al. [17] proposed a joint modeling approach via two latent variables, where the dependence structure between two latent variables are left unspecified. Moreover, in the analysis of a bladder cancer data conducted by the Veterans Administration Cooperative Urological Research Group, Liang et al. [5] argued that the treatment assignment has a random effect on the tumor recurrence process. However, the existing methods for analyzing longitudinal data with informative observation times and a dependent terminal event did not consider the case where some covariates have random effects. This motivates us to develop the present study.

In this article, we propose a new joint model for analysis of longitudinal data with informative observation times and a dependent terminal event in the spirit of the works of Liang et al. [5], which allows for inference about the random effects of covariates. Specifically, a semiparametric mixed effects model is specified for the longitudinal process, a proportional rate frailty model is used for the observation process, and a proportional hazards frailty model is used for the terminal event. The association among the three related processes is modeled via two latent variables. The proposed joint model generalizes the approach of Liang et al. [5] by taking the terminal event into account. In addition, unlike Liang et al. [5], our model does not

assume that the observation process is a nonhomogeneous Poisson process. Thus, it is more comprehensive and flexible.

The rest of the article is organized as follows. Section 2 describes the joint model of the longitudinal process, the observation times, and the terminal event. In Sect. 3, estimating equation approaches are proposed for regression parameters of interest, and the asymptotic properties of the proposed estimators are established. Some simulation results for evaluating the proposed methods are reported in Sect. 4. An application to a medical cost study of chronic heart failure patients from the University of Virginia Health System is provided in Sect. 5, and some concluding remarks are given in Sect. 6. All proofs are relegated to the Appendix.

2 Model Specification

Consider a longitudinal study involving n independent subjects. For the ith subject, $i = 1, \ldots, n$, denote $Y_i(t)$ as the longitudinal process of interest at time t. Let $X_i(t)$ be the $p \times 1$ vector of external covariates as described in Kalbfleisch and Prentice [2]. Also, let C_i be the censoring time and D_i be the terminal event time such as death. In addition, define $T_i = \min(C_i, D_i)$ and $\Delta_i(t) = I(T_i \geq t)$, where $I(\cdot)$ is the indicator function. Let $\tilde{N}_i^R(t)$ be the counting process denoting the number of the observation times in the time interval $[0, t]$ and the observation process $N_i^R(t) = \tilde{N}_i^R(t \wedge T_i)$, where $a \wedge b$ is the minimum of a and b. Note that the longitudinal process $Y_i(t)$ is observed only at the jump points of $N_i^R(t)$. Let $\tilde{N}_i^D(t)$ denote the terminal process before or at time t, and $N_i^D(t) = \tilde{N}_i^D(t \wedge T_i)$. Also let υ_i be a nonnegative unobserved frailty. In what follows, we assume that given $X_i(t)$, the censoring time C_i is independent of $\{Y_i(\cdot), \ \tilde{N}_i^R(\cdot), \ \tilde{N}_i^D(\cdot), \ D_i, \ \upsilon_i\}$.

For the longitudinal process, following Liang et al. [5], we consider the following semiparametric mixed effects model for $Y_i(t)$:

$$Y_i(t) = \mu_0(t) + \gamma_0^T X_i(t) + u_i^T Z_i(t) + \epsilon_i(t), \tag{1}$$

where $\mu_0(t)$ is an unspecified smooth function, γ_0 is a vector of unknown regression parameters, $Z_i(t)$ is a q-dimensional subvector of $(1, X_i(t)^T)^T$, u_i is a q-dimensional vector of subject-specific random effects, and $\epsilon_i(t)$ is a zero-mean measurement error process and independent of u_i. For identifiability of model (1), the random effects u_i are assumed to have zero mean.

Following Ye et al. [19], we consider a (partial) marginal rate of the observation times given $X_i(t)$, $D_i = s$ and υ_i, which is defined as

$$d\Lambda_R(t|\upsilon_i, X_i(t)) = P\{d\tilde{N}_i^R(t) = 1|\upsilon_i, X_i(t), D_i = s\}, \ \ s \geq t.$$

For the analysis, we specify the observation process model as

$$d\Lambda_R(t|\upsilon_i, X_i(t)) = \upsilon_i \exp\{\beta_0^T X_i(t)\}d\Lambda_0^R(t), \tag{2}$$

where β_0 is an unspecified p-dimensional regression parameters, and $\Lambda_0^R(t)$ is cumulative baseline function. Note that $\mathrm{d}\Lambda_R(t|v_i, X_i(t))$ may depend on $X_i(t)$ and the frailty v_i, but does not depend on the terminal event time $D_i = s \geq t$. This implies that given covariates $X_i(t)$, v_i accounts for the correlation between the observation times and the terminal event. Also, it follows that $\mathrm{d}\Lambda_R(t|v_i, X_i(t)) = P\{\mathrm{d}\tilde{N}_i^R(t) = 1|v_i, X_i(t), D_i \geq t\}$ (e.g., [3]), which indicates that given $X_i(t)$ and v_i, $\mathrm{d}\Lambda_R(t|v_i, X_i(t))$ specifies the marginal rate of the observation times among those subjects surviving to time t. In addition, $P\{\mathrm{d}\tilde{N}_i^R(t) = 1|v_i, X_i(t), D_i = s\} = 0$, $s < t$, that is, the occurrence of additional observation times is precluded by the terminal event.

For the terminal event, we consider the following proportional hazards frailty model:

$$\mathrm{d}\Lambda_D(t|v_i, X_i(t)) = v_i \exp\{\alpha_0^T X_i(t)\}\mathrm{d}\Lambda_0^D(t), \tag{3}$$

where α_0 is an unknown p-dimensional regression parameters, and $\Lambda_0^D(t)$ is the cumulative baseline hazard function.

For models (1), (2), and (3), we assume that the association between the random effects u_i and v_i is formulized as $E(u_i|v_i) = \eta_0(v_i - 1)$, where η is a q-dimensional parameter. As in Ye et al. [19], we assume that given $X_i(t)$, the frailty v_i has a gamma distribution with mean 1 and variance θ, where $E(v_i|X_i(t)) = 1$ is fixed for identifiability of models (1), (2), and (3).

Remark 1 In the medical cost study of chronic heart failure patients in Sect. 5, a higher mortality rate is associated with a higher frequency of hospital visits. Thus, we assume that the observation process and the terminal event have a positive association through a common frailty variable v_i in the same fashion. However, the proposed method can be extended to allow for a negative association between the observation process and the terminal event. More details can be found in the third paragraph of Sect. 6.

Remark 2 Note that

$$E\{Y_i(t) - \mu_0(t) - \gamma_0^T X_i(t)|v_i, X_i(t)\} = E\{u_i^T|v_i, X_i(t)\}Z_i(t).$$

We followed the suggestion of Liang et al. [5] to assume the linear relationship between u_i and v_i for computational simplicity. In fact, the proposed method can be extended to the case that $E\{u_i|v_i, X_i(t)\} = g(v_i; \eta)$, where $g(v_i; \eta)$ is a q-dimensional vector with each component being a polynomial in v_i. A further discussion can be found in the second paragraph of Sect. 6.

Remark 3 Although model (1) that allows for inference about the random effects of covariates was considered by Liang et al. [5] and others (e.g., [16,22]), and models (2) and (3) that allow the common frailty to directly relate to the association between the observation and terminal event processes were investigated by Kalbfleisch et al. [3], a joint modeling for simultaneously accommodating all of the aforementioned features has not been considered in the literature. The proposed model fits this gap and provides a flexible framework to allow subject-specific observation process with continuous missing patterns as well as various types of associations between the longitudinal process and the observation times in the presence of a dependent terminal event.

3 Estimation Procedure

To handle the unobserved frailty υ_i we proceed with the estimation procedure by analyzing the average rates obtained by taking the conditional expectation of models (2) and (3) with respect to υ_i given $D_i \geq t$ and $X_i(t)$ ([3]). Note that

$$E[\upsilon_i|D_i \geq t, X_i(t)] = \frac{1}{1 + \theta \int_0^t \exp\{\alpha^T X_i(u)\} d\Lambda_0^D(u)}.$$

Then we have the following marginal rates:

$$d\Lambda_R(t|X_i(t)) = \psi_i(t)^{-1} \exp\{\beta_0^T X_i(t)\} d\Lambda_0^R(t), \tag{4}$$

and

$$d\Lambda_D(t|X_i(t)) = \psi_i(t)^{-1} \exp\{\alpha_0^T X_i(t)\} d\Lambda_0^D(t), \tag{5}$$

where

$$\psi_i(t) \equiv \psi_i(t; \alpha, \theta, \Lambda_0^D) = 1 + \theta \int_0^t \exp\{\alpha^T X_i(u)\} d\Lambda_0^D(u).$$

For the unobserved random effects u_i, using $E(u_i|\upsilon_i) = \eta_0(\upsilon_i - 1)$ and taking the conditional expectation of u_i given $D_i \geq t$ and $X_i(t)$, we obtain

$$E[u_i|D_i \geq t, X_i(t)] = \eta_0[(1 + \theta)/\psi_i(t) - 1]. \tag{6}$$

Let $\mathcal{A}_0(t) = \int_0^t \mu_0(s) d\Lambda_0^R(s)$, $\xi = (\alpha^T, \beta^T, \theta, \gamma^T, \eta^T)^T$, and $\xi_0 = (\alpha_0^T, \beta_0^T, \theta_0, \gamma_0^T, \eta_0^T)^T$ be the true value of ξ. Define

$$dM_i(t; \xi, \mathcal{A}) = Y_i(t)\psi_i(t) dN_i^R(t) - \Delta_i(t)\left\{\gamma^T X_i(t) + \eta^T B_i(t)\right\} \exp\{\beta^T X_i(t)\} d\Lambda_0^R(t)$$
$$- \Delta_i(t) \exp\{\beta^T X_i(t)\} d\mathcal{A}(t), \tag{7}$$

where

$$B_i(t) \equiv B_i(t; \alpha, \theta, \Lambda_0^D) = [(1 + \theta)/\psi_i(t) - 1]Z_i(t).$$

In view of (4) and (6), we have that $E\{M_i(t; \xi_0, \mathcal{A}_0)|X_i(t), \Delta_i(t)\} = 0$ under models (1) and (2). Thus, for given ξ, $d\Lambda_0^R(t)$, $\psi_i(t)$, and $B_i(t)$, a reasonable estimator of $\mathcal{A}_0(t)$ is the solution to the following estimating equation:

$$\sum_{i=1}^n \int_0^t \left[Y_i(u)\psi_i(u) dN_i^R(u) - \{\gamma^T X_i(u) + \eta^T B_i(u)\} \exp\{\beta^T X_i(u)\} d\Lambda_0^R(u)\right.$$
$$\left. - \Delta_i(u) \exp\{\beta^T X_i(u)\} d\mathcal{A}(u)\right] = 0, \quad 0 \leq t \leq \tau,$$

where τ is a constant such that $P\{T_i \geq \tau\} > 0$. Denote this estimator by $\hat{\mathcal{A}}_0(t; \alpha, \beta, \theta, \Lambda_0^D)$. Let $\hat{\alpha}$, $\hat{\beta}$, $\hat{\theta}$, $\hat{\Lambda}_0^R(t)$, $\hat{\psi}_i(t)$, and $\hat{B}_i(t)$ be the estimates of α, β, θ, $\Lambda_0^R(t)$, $\psi_i(t)$, and $B_i(t)$, respectively, which will be discussed later. In view of (7), applying the generalized estimating equation approach [4] and replacing $\mathcal{A}_0(t)$ with the above estimator, we specify the following estimating function for γ and η :

$$\sum_{i=1}^{n} \int_0^\tau \begin{pmatrix} X_i(t) - \bar{X}(t, \hat{\beta}) \\ \hat{B}_i(t) - \bar{\hat{B}}(t) \end{pmatrix} \Big[Y_i(t)\hat{\psi}_i(t)\mathrm{d}N_i^R(t)$$
$$- \{\gamma^T X_i(t) + \eta \hat{B}_i(t)\} \exp\{\hat{\beta}^T X_i(t)\}\mathrm{d}\hat{\Lambda}_0^R(t) \Big] = 0, \qquad (8)$$

where

$$\bar{X}(t; \beta) = \frac{\sum_{j=1}^{n} \Delta_j(t) X_j(t) \exp\{\beta^T X_j(t)\}}{\sum_{j=1}^{n} \Delta_j(t) \exp\{\beta^T X_j(t)\}},$$

and

$$\bar{\hat{B}}(t) = \frac{\sum_{j=1}^{n} \Delta_j(t) \hat{B}_j(t) \exp\{\hat{\beta}^T X_j(t)\}}{\sum_{j=1}^{n} \Delta_j(t) \exp\{\hat{\beta}^T X_j(t)\}}.$$

Denote the solution to (8) as $(\hat{\gamma}^T, \hat{\eta}^T)^T$, which has an explicit form:

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\eta} \end{pmatrix} = \left[\sum_{i=1}^{n} \int_0^\tau \begin{pmatrix} X_i(t) - \bar{X}(t, \hat{\beta}) \\ \hat{B}_i(t) - \bar{\hat{B}}(t) \end{pmatrix}^{\otimes 2} \Delta_i(t) \exp\{\hat{\beta}^T X_i(t)\}\mathrm{d}\hat{\Lambda}_0^R(t) \right]^{-1}$$
$$\times \left[\sum_{i=1}^{n} \int_0^\tau \begin{pmatrix} X_i(t) - \bar{X}(t, \hat{\beta}) \\ \hat{B}_i(t) - \bar{\hat{B}}(t) \end{pmatrix} Y_i(t)\hat{\psi}_i(t)\mathrm{d}N_i^R(t) \right], \qquad (9)$$

where $a^{\otimes 2} = aa^T$ for any vector a.

Now we consider the estimators $\hat{\alpha}$, $\hat{\beta}$, $\hat{\theta}$, $\hat{\Lambda}_0^R(t)$, and $\hat{\Lambda}_0^D(t)$. Under models (2) and (3), α, β, θ, $\Lambda_0^R(t)$, and $\Lambda_0^D(t)$ can be consistently estimated using a similar method to Kalbfleisch et al. [3]. Specifically, define

$$\mathrm{d}M_i^R(t) = \psi_i(t)\mathrm{d}N_i^R(t) - \Delta_i(t) \exp\{\beta^T X_i(t)\}\mathrm{d}\Lambda_0^R(t),$$

and

$$\mathrm{d}M_i^D(t) = \psi_i(t)\mathrm{d}N_i^D(t) - \Delta_i(t) \exp\{\alpha^T W_i(t)\}\mathrm{d}\Lambda_0^D(t).$$

It then follows from (4) and (5) that $M_i^R(t)$ and $M_i^D(t)$ are zero-mean stochastic processes. Thus, for given $\psi_i(t)$, we can use the following estimating equations to estimate α, β, $\Lambda_0^R(t)$ and $\Lambda_0^D(t)$:

$$\sum_{i=1}^{n} \int_0^\tau \left\{ X_i(t) - \bar{X}(t; \alpha) \right\} \psi_i(t) \mathrm{d} N_i^D(t) = 0,$$

$$\sum_{i=1}^{n} \int_0^\tau \left\{ X_i(t) - \bar{X}(t; \beta) \right\} \psi_i(t) \mathrm{d} N_i^R(t) = 0,$$

$$\sum_{i=1}^{n} \left[\psi_i(t) \mathrm{d} N_i^D(t) - \Delta_i(t) \exp\{\alpha^T X_i(t)\} \mathrm{d} \Lambda_0^D(t) \right] = 0, \quad 0 \le t \le \tau,$$

$$\sum_{i=1}^{n} \left[\psi_i(t) \mathrm{d} N_i^R(t) - \Delta_i(t) \exp\{\beta^T X_i(t)\} \mathrm{d} \Lambda_0^R(t) \right] = 0, \quad 0 \le t \le \tau.$$

However, the weight function $\psi_i(t)$, also includes unknown parameters θ, which must be estimated. For this, define

$$\omega_{1i}(t) = E[\tilde{N}_i^R(t)|X_i(t), D_i = t],$$
$$\omega_{2i}(t) = E[\tilde{N}_i^R(t)|X_i(t), D_i > t].$$

Under the assumed models, we have

$$\omega_{1i}(t) = (\theta + 1)\psi_i(t)^{-1} \int_0^t \exp\{\beta^T X_i(u)\} \mathrm{d} \Lambda_0^R(u),$$
$$\omega_{2i}(t) = \psi_i(t)^{-1} \int_0^t \exp\{\beta^T X_i(u)\} \mathrm{d} \Lambda_0^R(u).$$

Thus,

$$\theta + 1 = \frac{\omega_{1i}(t)}{\omega_{2i}(t)}. \tag{10}$$

In view of (10), as discussed in Kalbfleisch et al. [3], we specify the following estimating equation for θ:

$$\sum_{i=1}^{n} \int_0^\tau \left\{ N_i^R(t) - (\theta + 1)\omega_{2i}(t) Q(t) \right\} \mathrm{d} N_i^D(t) = 0,$$

where

$$Q(t) = \frac{\sum_{j=1}^{n} \omega_{2j}(t)^{-1} \Delta_j^*(t) N_j^R(t)}{\sum_{j=1}^{n} \Delta_j^*(t)},$$

with $\Delta_j^*(t) = \Delta_j(t)\{1 - N_j^D(t)\}$ being an indicator that subject j is at risk at t and dies after t.

Let $\rho = (\alpha^T, \beta^T, \theta, \Lambda_0^D, \Lambda_0^R)^T$. We can estimate ρ using the solutions to the equations $U(\rho) = (U_1^T, U_2^T, U_3, U_4, U_5)^T = 0$, where

$$U_1 = \sum_{i=1}^n \int_0^\tau \left\{ X_i(t) - \bar{X}(t; \alpha) \right\} \psi_i(t) \mathrm{d}N_i^D(t),$$

$$U_2 = \sum_{i=1}^n \int_0^\tau \left\{ X_i(t) - \bar{X}(t; \beta) \right\} \psi_i(t) \mathrm{d}N_i^R(t),$$

$$U_3 = \sum_{i=1}^n \int_0^\tau \left\{ N_i^R(t) - (\theta + 1) Q(t) \omega_{2i}(t) \right\} \mathrm{d}N_i^D(t) = 0,$$

$$U_4 = \sum_{i=1}^n \left[\psi_i(t) \mathrm{d}N_i^D(t) - \Delta_i(t) \exp\{\alpha^T X_i(t)\} \mathrm{d}\Lambda_0^D(t) \right], \quad 0 \le t \le \tau,$$

$$U_5 = \sum_{i=1}^n \left[\psi_i(t) \mathrm{d}N_i^R(t) - \Delta_i(t) \exp\{\beta^T X_i(t)\} \mathrm{d}\Lambda_0^R(t) \right], \quad 0 \le t \le \tau.$$

Let $\hat{\rho}$ denote the solutions to $U(\rho) = 0$. Note that the first terms of the estimating functions U_4 and U_5 represent two pure jump processes with jumps at observed event times. Thus, the solutions to $U_4 = 0$ and $U_5 = 0$ must be piecewise constant functions with jumps only at the observed terminal event times and the observation times (across all subjects), respectively, which yield the Aalen-Breslow-type estimators $\hat{\Lambda}_0^D(t)$ and $\hat{\Lambda}_0^R(t)$ [6]. Since estimation of each parameter depends on a subset of the other parameters, the solutions to the above estimating equations can be obtained through a recursive procedure (e.g., [3]). Thus, the estimators $\hat{\psi}_i(t)$ and $\hat{B}_i(t)$ can be obtained by replacing α, θ and $\Lambda_0^D(t)$ with $\hat{\alpha}, \hat{\theta}, \hat{\Lambda}_0^D(t)$ in $\psi_i(t; \alpha, \theta, \Lambda_0^D)$ and $B_i(t; \alpha, \theta, \Lambda_0^D)$, respectively. To summarize, we propose the following two-step estimation procedure:

Step 1 First obtain estimator $\hat{\rho}$ by solving the equations $U(\rho) = 0$. Then calculate the estimators $\hat{\psi}_i(t)$ and $\hat{B}_i(t)$ for $1 \le i \le n$.

Step 2 Plug $\hat{\eta}, \hat{\psi}_i(t)$ and $\hat{B}_i(t)$ into Eq. (9) to obtain the estimators $\hat{\gamma}$ and $\hat{\eta}$.

We use the criterion that the absolute differences of the consecutive iterations of parameter estimates is less than 10^{-3} to check convergence. The algorithm in Step 1 converges most times in general, but nonconvergence occurs occasionally depending on the setups. In the simulation studies, the percentage of nonconvergence is about 2% under different setups with sample size $n = 600$.

As discussed in Ye et al. [19] and Kalbfleisch et al. [3], under the regularity conditions (C1)-(C4) stated in the Appendix, $\hat{\rho}$ exists and is unique and consistent. Then using the uniform strong law of large numbers, one can show that $\hat{\gamma}$ and $\hat{\eta}$ are consistent. Let $\hat{\xi} = (\hat{\alpha}^T, \hat{\beta}^T, \hat{\theta}, \hat{\gamma}^T, \hat{\eta}^T)^T$. Thus, $\hat{\xi}$ is consistent to ξ_0. The asymptotic distribution of $\hat{\xi}$ is stated in the following theorem with the proof given in Appendix.

Theorem 1 *Under the regularity conditions (C1)-(C4) stated in Appendix, $n^{1/2}(\hat{\xi} - \xi_0)$ converges in distribution to a normal random vector with mean zero, and covariance matrix $\Gamma^{-1} \Sigma (\Gamma^T)^{-1}$, where Σ and Γ are given in Appendix.*

The asymptotic covariance matrix can be consistently estimated by the usual plug-in method. However, Σ has a complicated analytic form, and it may be unstable to estimate Σ when the plug-in method is used with a moderate sample size. Here, we propose to use the bootstrap method to estimate the covariance matrix of $\hat{\xi}$. In the following simulation studies with sample size $n = 600$, we find that the covariance estimation is fairly accurate when 100 bootstrap samples are used.

Remark 4 Note that the estimator $\hat{\theta}$ may not always be nonnegative. For this case, we propose to estimate θ by using $\hat{\theta}^* = \hat{\theta} I(\hat{\theta} \geq 0)$ with a nonnegative constraint. Based on the arguments that are similar to those in Lin and Ying [7] and Zeng, Chen, and Ibrahim [21], we can show that $\hat{\theta}^*$ possesses the same asymptotic normality as $\hat{\theta}$ does if $\theta_0 > 0$ and that $n^{1/2}\hat{\theta}^*$ converges in distribution to $\varpi^2 GI(G \geq 0)$ if $\theta_0 = 0$, where G denotes the standard normal distribution, and ϖ^2 is the asymptotic variance of $n^{1/2}\hat{\theta}$.

Finally, for comparison purposes, we summarize the differences between our proposed model and the model of Liu et al. [10] (denoted by LHO) as follows:

(1) For the longitudinal process, the LHO's method assumed a random effects model, which neither included an unspecified baseline function of time nor considered the case where some covariates have random effects. Instead, we propose a semiparametric mixed effects model, which includes an unspecified baseline function of time (intercept) and allows for inference about the random effects of time-varying covariates.

(2) For the observation process, the LHO's method considered a proportional intensity frailty model and assumed that conditional on the random effects, the observation process is independent of the terminal event time. We instead consider a proportional rate frailty model that specifies the marginal rate of the observation process given survival and assume that the observation process is not independent of the terminal event even conditional on the random effects.

(3) For the terminal event, both methods propose the proportional hazards frailty model. The LHO's method assumed that the terminal event depends on the longitudinal and the observation processes through two independent random effects, respectively. The proposed method, however, formulates the associations among the three related processes through two dependent random effects, and the dependence structure is assumed to have a linear or polynomial form.

(4) For the estimation procedures, the LHO's method conducted maximum likelihood estimation on the basis of the assumption that the two random effects are independently normally distributed. Thus, their estimation results are expected to be sensitive to departures from this assumption. Unlike their procedure, we use an estimating equation approach for parameter estimation while assuming that the frailty υ_i follows a gamma distribution. As demonstrated by the simulation studies in Sect. 4, the proposed method is robust to misspecification of the frailty distribution.

4 Simulation

We conducted simulation studies to examine the finite sample properties of the proposed estimators. In the study, two covariates $X_i = (X_{i1}, X_{i2})^T$ were considered, where X_{i1} follows a Bernoulli distribution with success probability 0.5, and X_{i2} follows a uniform distribution on $(0, 1)$. Set $Z_i = X_{i1}$. The frailty υ_i was generated from a gamma distribution with unit mean and variance $\theta_0 = 0, 0.5$ or 1. The censoring time was generated from a uniform distribution on $(c, 5)$, where c is chosen to yield about 30% censoring for the terminal event. Given the frailty υ_i and the covariates X_i, the terminal event time D_i was generated from model (3) with $\Lambda_0^D(t) = 0.5t$ and $\alpha_0 = (0.3, 0.5)^T$. The observation times were generated from a Poisson process with the intensity function $\lambda_R = 1.5\upsilon_i \exp\{-0.3X_{i1} + 0.7X_{i2}\}$. The average number of observations per subject was about 2 under the preceding settings. For given υ_i, $u_i = \eta_0(\upsilon_i - 1) + e_i$, where e_i is a normal distribution with mean zero and variance 0.5, and $\eta_0 = -1, 0$, or 1. The longitudinal response $Y_i(t)$ was generated from model (1) with $\mu_0(t) = 2t + 1$ and $\gamma_0 = (1, -0.5)^T$, where the measurement errors $\epsilon_i(t)$ are generated independently from a standard normal distribution for all t. The results presented below are based on 500 replications with sample size $n = 600$. The asymptotic covariance was estimated using the bootstrap method with 100 bootstrap samples, which were found to be adequate.

The simulation results for estimation of γ_0, θ_0, and η_0 are summarized in Table 1, which includes the bias (Bias) given by the sample means of the estimates minus the true values, the sample standard errors (SE), the sample mean of the standard error estimate (SEE), and the 95% empirical coverage probabilities (CP) based on the normal approximation. Table 1 shows that our proposed method performed well for the situations considered here. Specifically, the proposed estimators were practically unbiased, and the standard error estimators were very accurate based on the bootstrap method. The 95% empirical coverage probabilities were reasonable. Note that when $\theta_0 = 0$, η_0 is unidentifiable. However, based on our simulation results, the estimators of γ_0 and θ_0 are still performed well. In addition, the estimates of $\Lambda_0^D(t)$ and $\Lambda_0^R(t)$ are provided in Fig. 1, which indicates that the estimators are accurate.

For comparison, we also considered the method of Liang et al. [5] (denoted by LLY), who studied models (1) and (2) without the terminal event. Under the same setup as above, the comparison results are also reported in Table 1. The results indicate that the LLY's method may lead to biased estimates when the corresponding independent conditions are violated (i.e., $\theta_0 \neq 0$). Figure 2 illustrates how the biases of the LLY's method may arise. Notably, the SE's of the proposed method are greater than those of LLY's method. A major reason is that the proposed method involves the estimators of the parameters α_0 and $\Lambda_0^D(t)$ in model (3), and these estimators introduce additional uncertainty to the proposed estimation procedure.

We also conducted simulation studies to examine the performance of the proposed estimators when the gamma distribution was misspecified. We considered two scenarios for the frailty υ_i: (i) υ_i followed a log-normal distribution with unit mean and variance 0.5; (ii) υ_i was generated as one-tenth of a Poisson variable with mean 10. The other setups were the same as in Table 1 with $\gamma_0 = (1, -0.5)^T$, $\eta_0 = -1, 0$, or 1. The results are given in Table 2. The proposed estimators still performed reason-

Table 1 Simulation results for estimation of γ_0, θ_0, and η_0 when covariates are time-independent

θ	η		Our method				LLY's method	
			Bias	SE	SEE	CP	Bias	SE
0.5	−1	γ_1	0.0038	0.3609	0.3474	0.950	0.0583	0.1961
		γ_2	−0.0236	0.4339	0.4012	0.942	0.0922	0.2856
		η	−0.0448	1.9235	1.8994	0.964	0.8274	0.2301
		θ	−0.0014	0.0728	0.0750	0.956	–	–
	0	γ_1	0.0093	0.3473	0.3361	0.950	0.1216	0.1968
		γ_2	−0.0271	0.4133	0.3936	0.936	0.0650	0.2613
		η	−0.0465	1.9469	1.8996	0.948	−0.1035	0.2217
		θ	0.0056	0.0747	0.0750	0.944	–	–
	1	γ_1	0.0197	0.3369	0.3372	0.950	0.1694	0.2025
		γ_2	0.0241	0.4013	0.4018	0.942	0.0282	0.2946
		η	−0.0510	1.9603	1.9376	0.958	−1.0386	0.2347
		θ	0.0032	0.0719	0.0751	0.948	–	–
1	−1	γ_1	−0.0162	0.4436	0.4345	0.944	−0.0367	0.2219
		γ_2	−0.0108	0.5158	0.5187	0.952	0.1035	0.3293
		η	0.0301	1.2861	1.2529	0.952	0.8327	0.1718
		θ	0.0032	0.1096	0.1154	0.930	–	–
	0	γ_1	−0.0227	0.4120	0.4163	0.960	0.1233	0.2106
		γ_2	0.0010	0.5172	0.4947	0.938	0.0802	0.2937
		η	0.0367	1.2649	1.2912	0.956	−0.0843	0.1711
		θ	0.0013	0.1142	0.1157	0.956	–	–
	1	γ_1	−0.0056	0.4252	0.4170	0.946	0.2689	0.2547
		γ_2	−0.0134	0.5077	0.5107	0.936	0.0173	0.3183
		η	−0.0080	1.3113	1.3404	0.952	−0.9830	0.2147
		θ	−0.0026	0.1178	0.1153	0.932	–	–
0	*	γ_1	−0.0202	0.2653	0.2652	0.948	−0.0013	0.1651
		γ_2	0.0154	0.2887	0.2818	0.950	−0.0026	0.2667
		θ	0.0067	0.0389	0.0358	0.952	–	–

*When $\theta = 0$, η is not identifiable

ably well for the two scenarios considered, and the proposed method was robust to misspecification of the frailty distribution.

Furthermore, we conducted simulation studies for the setting with time-varying covariates. In the study, we took two time-dependent covariates as $X_{i1}(t) = \tilde{X}_{i1}t$ and $X_{i2}(t) = \tilde{X}_{i2}t$, where $\tilde{X}_{i1}$ and $\tilde{X}_{i2}$ were independently generated from a uniform distribution on (0, 1). Set $X_i(t) = (X_{i1}(t), X_{i2}(t))^T$ and $Z_i(t) = X_{i1}(t)$. The censoring time was generated from a uniform distribution on $(c, 2)$, where c is chosen to yield about 30% censoring for the terminal event. The other setups were the same as in Table 1, except that $\Lambda_0^D(t) = t$ and $\Lambda_0^R(t) = 5t$. The results are summarized in Table 3 with $n = 600$. It can be seen that the proposed method still performed satisfactorily in this case.
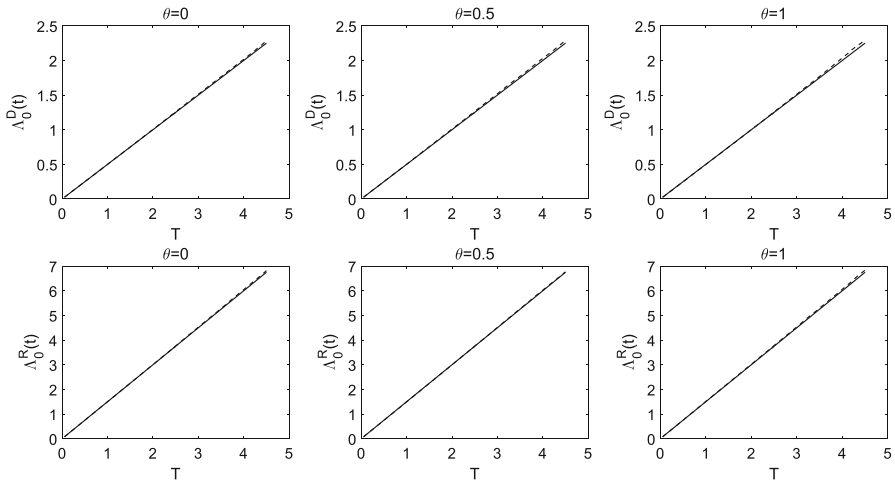
Fig. 1 The first row is for the estimates of $\Lambda_0^D(t)$ and the second row is for the estimates of $\Lambda_0^R(t)$. The dashed lines are the proposed estimators, and the solid lines are the true functions
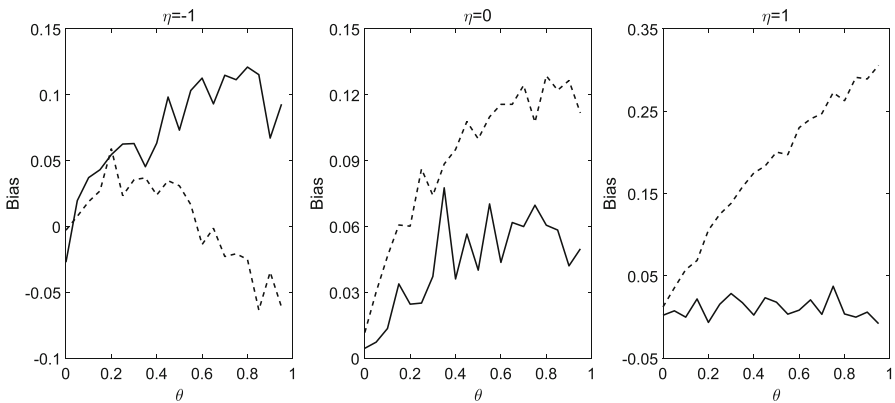

Fig. 2 Bias curves of LLY's method for the estimates of γ_1 and γ_2. The dashed lines are for γ_1, and the solid lines are for γ_2

5 An Application

In this section, we applied the proposed methods to the medical cost data of chronic heart failure patients that have been analyzed by Liu et al. [10], Sun et al. [17] and among others. These data were from the clinical data repository at the University of Virginia Health System, which included a total of 1475 patients aged 60–89 years who were first diagnosed with heart failure and treated in 2004. The follow-up ended with each patient's last hospital admission up to July 31, 2006, or death date, which was obtained from the Death Certificate Data at the Virginia Department of Vital Statistics. During follow-up, 297 patients (20%) died and others were censored. For each patient, three baseline covariates were measured: race, age, and gender. Preliminary studies implied that patients visiting the hospital more often tended to pay more for each

Table 2 Sensitivity analysis for the misspecification of the frailty

η		Log-normal		Poisson	
		γ_1	γ_2	γ_1	γ_2
−1	Bias	0.0296	0.0233	−0.0065	−0.0270
	SE	0.3462	0.3865	0.2256	0.3203
	SEE	0.3252	0.3695	0.2621	0.3017
	CP	0.928	0.928	0.946	0.936
0	Bias	0.0065	0.0187	−0.0028	−0.0093
	SE	0.3272	0.3680	0.2626	0.3094
	SEE	0.3146	0.3539	0.2597	0.2993
	CP	0.942	0.936	0.946	0.934
1	Bias	0.0156	0.0074	−0.0012	−0.0099
	SE	0.3193	0.3858	0.2651	0.3108
	SEE	0.3151	0.3655	0.2619	0.3029
	CP	0.948	0.924	0.946	0.946

Table 3 Simulation results for estimation of γ_0, θ_0, and η_0 when covariates are time-dependent

θ	η		Bias	SE	SEE	CP
0.5	−1	γ_1	0.0079	0.5101	0.5084	0.940
		γ_2	−0.0208	0.4665	0.4623	0.944
		η	0.0415	3.6579	3.6125	0.932
		θ	−0.0191	0.0704	0.0663	0.940
	0	γ_1	0.0086	0.5088	0.5022	0.952
		γ_2	−0.0222	0.4443	0.4434	0.956
		η	−0.0465	3.6516	3.5756	0.928
		θ	0.0056	0.0747	0.0750	0.940
	1	γ_1	−0.0207	0.4709	0.5033	0.962
		γ_2	−0.0371	0.4407	0.4459	0.940
		η	−0.0510	3.4737	3.5493	0.960
		θ	0.0032	0.0706	0.6990	0.950
0	*	γ_1	0.0135	0.3967	0.3881	0.940
		γ_2	−0.0103	0.3552	0.3777	0.956
		θ	0.0105	0.0278	0.0296	0.944

*When $\theta = 0$, η is not identifiable

visit, and these patients also had a higher mortality rate. That is, the medical cost (longitudinal process) may be strongly correlated with the hospital visits (observation times) and the death (terminal event). To show further how the longitudinal profiles and observational times profiles are associated with terminal events, we plotted the scatter diagrams of the frequency of hospital visits and the log sum of medical cost (until the observed survival time T_i) versus the observed survival time in Figs. 3 and 4, respectively. These plots show that the medical costs and hospital visits are positively correlated with the death. Given that gender had been shown to have no effect on the
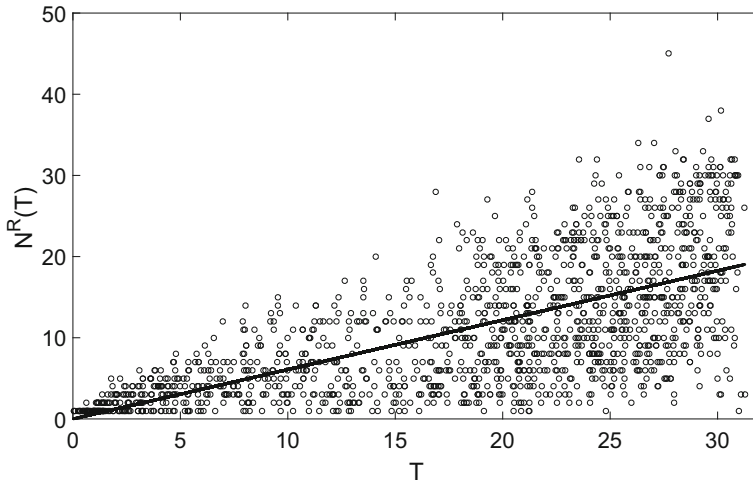
Fig. 3 Scatter plot of the frequency of hospital visits versus the observed survival time
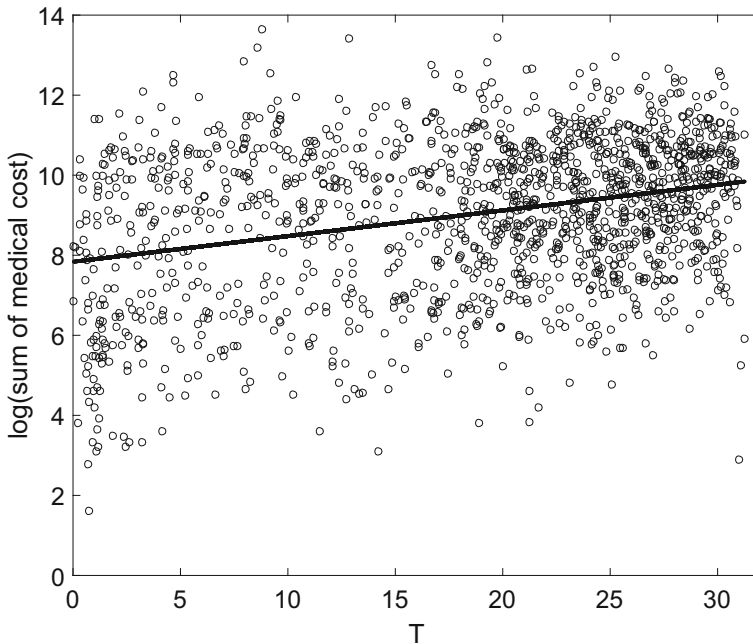

Fig. 4 Scatter plot of the log sum of medical cost versus the observed survival time

medical cost and the hospital visits [10,17], here we focused on the effects of race and age on the actual monetary expense of the hospital with informative observation times and a dependent terminal event.

As in Liu et al. [10] and Sun et al. [17], we defined $Y_i(t)$ as the log-transformed cost. For covariates, let X_{i1} be a binary indicator of race (white $= 1$, nonwhite $= 0$),

Table 4 Analysis results for the medical cost data of heart failure patients

	γ		β		Correlation		
	White	Age	White	Age	η_1	η_2	θ
Our method							
Est	0.7489	− 0.8565	− 0.1005	0.1168	− 4.2311	2.3297	0.3634
SEE	0.7515	0.2829	0.0370	0.0215	2.1014	1.1065	0.0480
p-value	0.3190	0.0025	0.0066	0.0000	0.0441	0.0353	0.0000
LLY's method							
Est	− 0.4287	− 0.2234	− 0.0834	0.0798	0.0112	0.0011	–
SEE	0.1205	0.0855	0.0349	0.0203	0.0091	0.0075	–
p-value	0.0003	0.0090	0.0132	0.0000	0.2174	0.8813	–

Est is the estimate of the parameter, *SEE* is the standard error estimate

and X_{i2} denote the age group, taking values 0, 1, and 2 for 60–69, 70–79, and 80–89 years, respectively. Let τ be the longest follow-up time. The asymptotic variance was estimated by the bootstrap method with 100 bootstrap samples. We chose $Z_i = (X_{i1}, X_{i2})^T$ in model (1), because the race and age are significantly related to the hospital visits. The analysis results are summarized in Table 4. For the hospital process, both age and race are significantly related to the hospital visiting, which is in line with the result obtained by Liu et al. [10] (denoted by LHO). In particular, older patients were more likely to visit the hospital and had lower medical cost. White patients visited hospital at less risk and tended to have less medical costs at each visit. For the cost process, we found that age had a significant effect on the medical cost for each visit, but race did not seem to be directly related to the medical cost. Although the results of the LHO's method implied that age was only marginally significant at 2% level, their estimator for the race effect was significantly different from ours, and the direction of the race effect is reversed. One possible reason is the misspecification of the assumption that the two random effects are independently normally in the LHO's method. The estimate $\hat{\theta} = 0.3634$ (p value < 0.0001) indicates that there was a significantly positive association between the hospital visits and the death. That is, patients who tended to visit hospital more frequently had a higher mortality rate. In addition, in view of (10), this estimate suggests that a patient who is known to die at time t is expected to have more than 1.3 times as many hospital visits as a patient with identical covariates who has not died by the time t. Moreover, based on the estimate $\hat{\eta}$, older patients who visited the hospital more often tended to pay more for each visit, but white patients visiting the hospital more often tended to pay less for each visit. These results are basically consistent with those obtained by the LHO's method.

For comparison, we also analyzed the data with the LLY's method, regarding the terminal event as an independent censoring time. The comparison results are provided in the second half of Table 4. For the cost process, the effects of race and age estimated based on the LLY's method are substantially smaller, and the direction of the race effect is even reversed. For the hospital process, LLY's method also produced the smaller effects of race and age. Moreover, the association parameters estimated using LLY's

method are significantly different from ours. This is because the LLY's method ignores the dependent terminal event, and thus yields biased estimates.

6 Discussion

In this article, we proposed a joint modeling for analyzing longitudinal data with informative observation times and a dependent terminal event via two latent variables. The joint model is more comprehensive and flexible in that it does not assume that the observation process is a nonhomogeneous Poisson process and allows that some covariates have random effects. An estimating equation approach was developed for parameter estimation, which yielded consistent and asymptotically normal estimators. The simulation results showed that the proposed estimation approach performs well, and the method was robust to misspecification of the frailty distribution for the situations considered.

Here we have assumed that the covariate histories $\{X_i(t) : 0 \leq t \leq T_i\}$ are observed, and hence $\bar{X}(t; \beta)$ is well defined. In practice, however, the covariate histories are typically measured discretely, and are often available at the observation times. Thus, some smoothing procedure is needed to interpolate and approximate $\bar{X}(t; \beta)$. For this, as discussed in Lin and Ying [8], by using the singleton nearest-neighbor method, we may approximate $\bar{X}(t; \beta)$ by

$$\bar{X}^*(t; \beta) = \frac{\sum_{j=1}^n \Delta_j(t) X_j^*(t) \exp\{\beta^T X_j^*(t)\}}{\sum_{j=1}^n \Delta_j(t) \exp\{\beta^T X_j^*(t)\}},$$

where $X_j^*(t)$ is the measurement of $X_j(\cdot)$ at the time point nearest to t. Other choices of $\bar{X}^*(t; \beta)$ would be the nearest two-neighbor or two-left-right-neighbor average [8]. Also, we may approximate $\bar{X}(t; \beta)$ by other smoothing methods such as a linear smoother or a nonparametric smoother [18]. A similar approximate method can be used for $\bar{\hat{B}}(t)$. It would be worthwhile to further address this issue both theoretically and numerically.

In the joint models, we have assumed that the relationship between the latent variables has a linear form. In fact, as long as $E(u_i|v_i, X_i(t))$ is a polynomial in v_i, the estimation procedure can be directly extended to this case. This extension is useful because any continuous function can be approximated by polynomials. However, a high order of polynomial may lead to unstable approximation. Thus, a simple linear form may be a good choice for small or moderate sample sizes. It would be desirable to develop a method for the case that $E\{u_i|v_i, X_i(t)\}$ is an unspecified function of v_i, that is, the dependence structure between the random effects u_i and the frailty v_i is left unspecified [17]. Nevertheless, this extension is highly nontrivial and it requires further investigation.

Note that models (2) and (3) allow a positive association between the observation process and the terminal event. Although these models fit the example discussed in Sect. 5 well, the negative association may exist. Based on the discussion of Kalbfleisch et al. [3], the proposed model can be generalized to allow for a negative association

between the observation process and the terminal event. For this case, we specify v_i^{-1} as the frailty in model (2) and retain v_i as the frailty in models (1) and (3), where the frailty v_i is assumed to follow a gamma distribution with mean 1 and variance $\theta < 1$. It can be checked that

$$\psi_i^*(t) = E\{v_i^{-1}|X_i(t), D_i \geq t\} = \frac{\psi_i(t)}{1 - \theta}.$$

By replacing $\psi_i(t)$ and $1 + \theta$ with $\psi_i^*(t)^{-1}$ and $1 - \theta$, respectively, in U_2, U_3, and U_5, the same estimating equations can be constructed as in the previous sections. A more general approach is to generalize model (2) to

$$d\Lambda_R(t|v_i) = v_i^\sigma \exp\{\beta_0^T X_i(t)\}d\Lambda_0^R(t),$$

where σ is an unknown parameter. Estimation of σ in this model is a challenging problem and requires substantial efforts in the future. Finally, the proposed estimation procedure was developed on the basis of the generalized estimating equation approach. The efficiency of the resulting estimators is worthy of further investigation.

Appendix : Proof of Asymptotic Results

In order to study the asymptotic properties of the proposed estimators, we need the following regularity conditions:

(C1) $\{Y_i(\cdot), N_i^R(\cdot), N_i^D(\cdot), T_i, X_i(\cdot), i = 1, \ldots, n\}$ are independent and identically distributed.

(C2) $E\{N_i^R(\tau)\}$ is bounded, and $\Pr(T_i \geq \tau) > 0$.

(C3) $X_i(t)$ is almost surely of bounded variation on $[0, \tau]$.

(C4) There exist a compact set $\mathcal{B}$ of ξ_0 such that for $\xi \in \mathcal{B}$, $\Gamma(\xi)$ is nonsingular, where $\Gamma(\xi)$ is the limit of $-\partial \tilde{U}(\xi)/\partial \xi^T$ with $\tilde{U}(\xi)$ defined in (A.3).

For notational simplicity, henceforth we omit superscript t in the covariate $X_i(t)$. Let $\tilde{\Lambda}_0^D(t)$ and $\tilde{\Lambda}_0^R(t)$ denote the solutions to the following estimating equations

$$\frac{1}{n}\sum_{i=1}^n \left[\psi_i(t; \alpha_0, \theta_0, \Lambda_0^D)dN_i^D(t) - \Delta_i(t)\exp\{\alpha_0^T X_i\}d\Lambda_0^D(t)\right] = 0,$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \left[\psi_i(t; \alpha_0, \theta_0, \Lambda_0^D) dN_i^R(t) - \Delta_i(t) \exp\{\beta_0^T X_i\} d\Lambda_0^R(t) \right] = 0.$$

Proof of Theorem 1

Define

$$\tilde{\Psi}_1(t) = \frac{\theta_0}{n} \sum_{i=1}^{n} \int_0^t \exp\{\alpha_0^T X_i\} \frac{dN_i^D(u)}{S_0(u; \alpha_0)},$$

where $S_0(t; \alpha) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i(t) \exp\{\alpha^T X_i\}$. By the definition of $\tilde{\Lambda}_0^D(t)$, it can be checked that

$$\tilde{\Lambda}_0^D(t) - \Lambda_0^D(t) = \sum_{i=1}^{n} \int_0^t [\tilde{\Lambda}_0^D(u) - \Lambda_0^D(u)] d\tilde{\Psi}_1(u) + \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{dM_i^D(u)}{S_0(u; \alpha_0)},$$

which is a linear Volterra integral equation, and the solution is

$$\tilde{\Lambda}_0^D(t) - \Lambda_0^D(t) = \frac{1}{\tilde{P}(t)} \int_0^t \tilde{P}(u-) \frac{\sum_{i=1}^{n} dM_i^D(u)}{n S_0(u; \alpha_0)},$$

where $\tilde{P}(t) = \prod_{s \leq t} \{1 - d\tilde{\Psi}_1(s)\}$ is the product-integral of $\tilde{\Psi}_1(s)$ over $[0, t]$. Using the uniform convergence of $\tilde{\Lambda}_0^D(t)$, the uniform strong law of large numbers [12] and Lemma A.1 of Lin and Ying [8], we get that uniformly in $t \in [0, \tau]$,

$$\tilde{\Lambda}_0^D(t) - \Lambda_0^D(t) = \frac{1}{n} \sum_{i=1}^{n} \phi_{1i}(t) + o_p\left(n^{-1/2}\right), \tag{A.1}$$

where

$$\phi_{1i}(t) = \frac{1}{P(t)} \int_0^t P(u-) \frac{dM_i^D(u)}{s_0(u; \alpha_0)},$$

and $P(t)$ and $s_0(t; \alpha)$ are the limits of $\tilde{P}(t)$ and $S_0(t; \alpha)$, respectively. Let

$$\tilde{\Psi}_2(t) = \frac{\theta_0}{n} \sum_{i=1}^{n} \int_0^t \exp\{\beta_0^T X_i\} \frac{dN_i^R(u)}{S_0(u; \beta_0)},$$

and $\Psi_2(t)$ be the limit of $\tilde{\Psi}_2(t)$. Note that

$$\tilde{\Lambda}_0^R(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{\psi_i(u; \tilde{\Lambda}_0^D) \mathrm{d}N_i^R(u)}{S_0(u; \beta_0)},$$

where $\psi_i(t; \Lambda_0^D) = \psi_i(t; \alpha_0, \theta_0, \Lambda_0^D)$. It then follows from (A.1) that

$$\tilde{\Lambda}_0^R(t) - \Lambda_0^R(t) = \int_0^t [\tilde{\Lambda}_0^D(u) - \Lambda_0^D(u)] \mathrm{d}\tilde{\Psi}_2(u) + \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{\mathrm{d}M_i^R(u)}{S_0(u; \beta_0)}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \phi_{2i}(t) + o_p(n^{-1/2}), \tag{A.2}$$

where

$$\phi_{2i}(t) = \int_0^t \phi_{1i}(u) d\Psi_2(u) + \int_0^t \frac{\mathrm{d}M_i^R(u)}{s_0(u)}.$$

Let $\omega_{2i}(t; \xi)$ be defined as $\omega_{2i}(t)$ with $\psi_i(t)$ replaced by $\psi_i(t; \alpha, \theta, \tilde{\Lambda}_0^D)$, and

$$\tilde{U}(\xi) = \left(\tilde{U}_1(\xi)^T, \tilde{U}_2(\xi)^T, \tilde{U}_3(\xi), \tilde{U}_4(\xi)^T \right)^T, \tag{A.3}$$

where

$$\tilde{U}_1(\xi) = \sum_{i=1}^{n} \int_0^\tau \left\{ X_i - \bar{X}(t; \alpha) \right\} \psi_i(t; \alpha, \theta, \tilde{\Lambda}_0^D) \mathrm{d}N_i^D(t),$$

$$\tilde{U}_2(\xi) = \sum_{i=1}^{n} \int_0^\tau \left\{ X_i - \bar{X}(t; \beta) \right\} \psi_i(t; \alpha, \theta, \tilde{\Lambda}_0^D) \mathrm{d}N_i^R(t),$$

$$\tilde{U}_3(\xi) = \sum_{i=1}^{n} \int_0^\tau \left\{ N_i^R(t) - (\theta + 1)Q(t)\omega_{2i}(t; \xi) \right\} \mathrm{d}N_i^D(t)$$

and

$$\tilde{U}_4(\xi) = \sum_{i=1}^{n} \int_0^\tau \left(\begin{array}{c} X_i - \bar{X}(t; \beta) \\ B_i(t; \tilde{\Lambda}_0^D) - \tilde{B}(t; \xi) \end{array} \right) \left[Y_i(t)\psi_i(t; \alpha, \theta, \tilde{\Lambda}_0^D) \mathrm{d}N_i^R(t) \right.$$

$$\left. - \left\{ \gamma^T X_i + \eta^T B_i(t; \tilde{\Lambda}_0^D) \right\} \exp\{\beta^T X_i\} \mathrm{d}\tilde{\Lambda}_0^R(t) \right],$$

with $B_i(t; \tilde{\Lambda}_0^D) = B_i(t; \alpha, \theta, \tilde{\Lambda}_0^D)$, and

$$\tilde{B}(t; \xi) = \frac{\sum_{j=1}^{n} \Delta_j(t) B_j(t; \tilde{\Lambda}_0^D) \exp\{\beta^T X_j\}}{\sum_{j=1}^{n} \Delta_j(t) \exp\{\beta^T X_j\}}.$$

Note that

$$\tilde{U}_1(\xi_0) = \sum_{i=1}^{n} \int_0^\tau \left\{ X_i - \bar{X}(t; \alpha_0) \right\} \psi_i(t; \tilde{\Lambda}_0^D) \mathrm{d} N_i^R(t)$$

$$= \theta_0 \sum_{i=1}^{n} \int_0^\tau \Delta_i(t) \exp\{\alpha_0^T X_i\} \left\{ X_i - \bar{X}(t; \alpha_0) \right\} \left\{ \tilde{\Lambda}_0^D(t) - \Lambda_0^D(t) \right\} \mathrm{d} N_i^D(t)$$

$$+ \sum_{i=1}^{n} \int_0^\tau \left\{ X_i - \bar{X}(t; \alpha_0) \right\} \mathrm{d} M_i^D(t),$$

where $\psi_i(t; \tilde{\Lambda}_0^D) = \psi_i(t; \alpha_0, \theta_0, \tilde{\Lambda}_0^D)$. Let

$$\tilde{H}_1(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \theta_0 \Delta_i(u) \exp\{\alpha_0^T X_i\} \left\{ X_i - \bar{X}(u; \alpha_0) \right\} \mathrm{d} N_i^D(u),$$

and $\bar{x}(t; \alpha_0)$ and $H_1(t)$ are the limits of $\bar{X}(t; \alpha_0)$ and $\tilde{H}_1(t)$, respectively. Similarly to (A.2), we have

$$\tilde{U}_1(\xi_0) = \sum_{i=1}^{n} \vartheta_{1i} + o_p\left(n^{1/2}\right), \tag{A.4}$$

where

$$\vartheta_{1i} = \int_0^\tau \left\{ X_i - \bar{x}(t; \alpha_0) \right\} \mathrm{d} M_i^D(t) + \int_0^\tau \phi_{1i}(t) \mathrm{d} H_1(t).$$

Likewise,

$$\tilde{U}_2(\xi_0) = \sum_{i=1}^{n} \vartheta_{2i} + o_p(n^{1/2}), \tag{A.5}$$

where

$$\vartheta_{2i} = \int_0^\tau \left\{ X_i - \bar{x}(t; \beta_0) \right\} \mathrm{d} M_i^R(t) + \int_0^\tau \phi_{1i}(t) \mathrm{d} H_2(t),$$

and $\bar{x}(t; \beta_0)$ and $H_2(t)$ are the limits of $\bar{X}(t; \beta_0)$ and $\tilde{H}_2(t)$ with

$$\tilde{H}_2(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \theta_0 \Delta_i(u) \exp\{\alpha_0^T X_i\} \left\{ X_i - \bar{X}(u; \beta_0) \right\} \mathrm{d} N_i^R(u).$$

Let

$$\omega_{2i}^*(t) = \psi_i^{-1}(t; \Lambda_0^D) \int_0^t \exp\{\beta_0^T X_i\} d\Lambda_0^R(u),$$

$$\tilde{\omega}_{2i}(t) = \psi_i^{-1}(t; \tilde{\Lambda}_0^D) \int_0^t \exp\{\beta_0^T X_i\} d\tilde{\Lambda}_0^R(u),$$

$$\tilde{Q}(t) = \frac{\sum_{i=1}^n \tilde{\omega}_{2i}^{-1}(t) \Delta_i^*(t) N_i^R(t)}{\sum_{i=1}^n \Delta_i^*(t)},$$

and $q(t)$ be the limit of $\tilde{Q}(t)$. Note that

$$
\begin{aligned}
\tilde{U}_3(\xi_0) = \sum_{i=1}^n \int_0^\tau \left\{ N_i^R(t) - (\theta_0 + 1)\omega_{2i}^*(t) q(t) \right\} dN_i^D(t) \\
- (\theta_0 + 1) \sum_{i=1}^n \int_0^\tau \tilde{Q}(t) \left\{ \tilde{\omega}_{2i}(t) - \omega_{2i}^*(t) \right\} dN_i^D(t) \\
- (\theta_0 + 1) \sum_{i=1}^n \int_0^\tau \left\{ \tilde{Q}(t) - q(t) \right\} \omega_{2i}^*(t) dN_i^D(t).
\end{aligned}
\tag{A.6}
$$

As in the proof of (A.4), the second term on the right-hand side of (A.6) equals

$$
\begin{aligned}
&- (\theta_0 + 1) \sum_{i=1}^n \int_0^\tau \left[\tilde{Q}(t) \{\psi_i^{-1}(t; \tilde{\Lambda}_0^D) - \psi_i^{-1}(t; \Lambda_0^D)\} \int_0^t \exp\{\beta_0^T X_i\} d\Lambda_0^R(u) \right] dN_i^D(t) \\
&- (\theta_0 + 1) \sum_{i=1}^n \int_0^\tau \tilde{Q}(t) \psi_i^{-1}(t; \tilde{\Lambda}_0^D) \exp\{\beta_0^T X_i\} \left[\tilde{\Lambda}_0^R(t) - \Lambda_0^R(t) \right] dN_i^D(t) \\
&= (\theta_0 + 1) \sum_{i=1}^n \int_0^\tau \left[\phi_{1i}(t) dH_3(t) - \phi_{2i}(t) dH_4(t) \right] + o_p(n^{1/2}),
\end{aligned}
\tag{A.7}
$$

where $H_3(t)$ and $H_4(t)$ are the limits of $\tilde{H}_3(t)$ and $\tilde{H}_4(t)$ with

$$\tilde{H}_3(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \theta_0 \left[\tilde{Q}(u) \psi_i^{-1}(u; \tilde{\Lambda}_0^D) \omega_{2i}^*(u) \right] dN_i^D(u)$$

and

$$\tilde{H}_4(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \tilde{Q}(u) \psi_i^{-1}(u; \tilde{\Lambda}_0^D) dN_i^D(u).$$

Let $d\Phi(t) = E\{\omega_{2i}^*(t)dN_i^D(t)\}$,

$$\tilde{H}_5(t) = \frac{\theta_0 \sum_{i=1}^n \psi_i^{-1}(t; \tilde{\Lambda}_0^D)\tilde{\omega}_{2i}^{-1}(t)\Delta_i^*(t)N_i^R(t)}{\sum_{i=1}^n \Delta_i^*(t)},$$

$$\tilde{H}_6(t) = \frac{\sum_{i=1}^n \psi_i^{-1}(t; \tilde{\Lambda}_0^D)\tilde{\omega}_{2i}^{-1}(t)\omega_{2i}^{*-1}(t)\Delta_i^*(t)N_i^R(t)}{\sum_{i=1}^n \Delta_i^*(t)},$$

and $H_5(t)$ and $H_6(t)$ be the limits of $\tilde{H}_5(t)$ and $\tilde{H}_6(t)$, respectively. In a similar manner, the third term on the right-hand side of (A.6) is

$$(\theta_0 + 1) \sum_{i=1}^n \int_0^\tau \left[\phi_{1i}(t)H_5(t) + \phi_{2i}(t)H_6(t)\right]d\Phi(t)$$

$$- (\theta_0 + 1) \sum_{i=1}^n \int_0^\tau \left[\frac{\omega_{2i}^{*-1}(t)\Delta_i^*(t)N_i^R(t)}{E\{\Delta_i^*(t)\}} - \frac{q(t)}{E\{\Delta_i^*(t)\}}\Delta_i^*(t)\right]d\Phi(t) + o_p(n^{1/2}).$$

$$(A.8)$$

Thus, it follows from (A.6)–(A.8) that

$$\tilde{U}_3(\xi_0) = \sum_{i=1}^n \vartheta_{3i} + o_p(n^{1/2}), \tag{A.9}$$

where

$$\vartheta_{3i} = \int_0^\tau \left\{N_i^R(t) - (\theta_0 + 1)\omega_{2i}^*(t)q(t)\right\}dN_i^D(t)$$

$$- (\theta_0 + 1) \int_0^\tau \left[\phi_{1i}(t)dH_3(t) + \phi_{2i}(t)dH_4(t)\right]$$

$$+ (\theta_0 + 1) \int_0^\tau \left[\phi_{1i}(t)H_5(t) + \phi_{2i}(t)H_6(t)\right]d\Phi(t)$$

$$- (\theta_0 + 1) \int_0^\tau \left[\frac{\omega_{2i}^{*-1}(t)\Delta_i^*(t)N_i^R(t)}{E\{\Delta_i^*(t)\}} - \frac{q(t)\Delta_i^*(t)}{E\{\Delta_i^*(t)\}}\right]d\Phi(t).$$

Let $B_i(t) = B_i(t; \alpha_0, \theta_0, \Lambda_0^D)$, and

$$\bar{B}(t) = \frac{\sum_{j=1}^n \Delta_j(t)B_j(t)\exp\{\beta_0^T X_j\}}{\sum_{j=1}^n \Delta_j(t)\exp\{\beta_0^T X_j\}}.$$

Note that

$$\tilde{U}_4(\xi_0) = R_1 + R_2 + R_3, \tag{A.10}$$

where

$$R_1 = \sum_{i=1}^{n} \int_0^{\tau} \begin{pmatrix} X_i - \bar{X}(t;\beta_0) \\ B_i(t) - \bar{B}(t) \end{pmatrix} dM_i(t),$$

$$R_2 = \sum_{i=1}^{n} \int_0^{\tau} \begin{pmatrix} X_i - \bar{X}(t;\beta_0) \\ B_i(t;\tilde{\Lambda}_0^D) - \bar{B}(t) \end{pmatrix} \{d\tilde{M}_i(t) - dM_i(t)\}$$

$$R_3 = \sum_{i=1}^{n} \int_0^{\tau} \left[\begin{pmatrix} X_i - \bar{X}(t;\beta_0) \\ B_i(t;\tilde{\Lambda}_0^D) - \bar{B}(t) \end{pmatrix} - \begin{pmatrix} X_i - \bar{X}(t;\beta_0) \\ B_i(t) - \bar{B}(t) \end{pmatrix} \right] dM_i(t),$$

and

$$d\tilde{M}_i(t) = Y_i(t)\psi_i(t;\tilde{\Lambda}_0^D)dN_i^R(t) - \Delta_i(t)\{\gamma_0^T X_i + \eta_0^T B_i(t;\tilde{\Lambda}_0^D)\} \exp\{\beta_0^T X_i\}d\tilde{\Lambda}_0^R(t)$$
$$- \Delta_i(t)\exp\{\beta_0^T X_i\}d\mathcal{A}_0(t).$$

For R_1 of (A.10), using the Lemma A.1 of Lin and Ying [8], we have

$$R_1 = \sum_{i=1}^{n} \int_0^{\tau} \begin{pmatrix} X_i - \bar{x}(t;\beta_0) \\ B_i(t) - \bar{b}(t) \end{pmatrix} dM_i(t) + o_p(n^{1/2}),$$

where $\bar{b}(t)$ is the limit of $\bar{B}(t)$. For R_2, we get

$$R_2 = \sum_{i=1}^{n} \int_0^{\tau} \begin{pmatrix} X_i - \bar{X}(t;\beta_0) \\ B_i(t;\tilde{\Lambda}_0^D) - \bar{B}(t) \end{pmatrix} \Big\{ Y_i(t)\big[\psi_i(t;\tilde{\Lambda}_0^D) - \psi_i(t;\Lambda_0^D)\big]dN_i^R(t)$$
$$- \Delta_i(t)\gamma_0^T X_i \exp\{\beta_0^T X_i\}d\{\tilde{\Lambda}_0^R(t) - \Lambda_0^R(t)\}$$
$$- \Delta_i(t)\eta_0^T X_i \exp\{\beta_0^T X_i\}\big[B_i(t;\tilde{\Lambda}_0^D)d\tilde{\Lambda}_0^R(t) - B_i(t)d\Lambda_0^R(t)\big] \Big\}.$$

It then follows from (A.1) and (A.2) that

$$R_2 = \sum_{i=1}^{n} \int_0^{\tau} \Big\{ \phi_{1i}(t)dG_1(t) - G_2(t)d\phi_{2i}(t) \Big\} + o_p(n^{1/2}),$$

where $G_1(t)$ and $G_2(t)$ are the limits of $\tilde{G}_1(t)$ and $\tilde{G}_2(t)$, respectively, with

$$\tilde{G}_1(t) = \frac{1}{n}\sum_{i=1}^{n} \int_0^{t} \begin{pmatrix} X_i - \bar{X}(u;\beta_0) \\ B_i(u;\tilde{\Lambda}_0^D) - \bar{\bar{B}}(u) \end{pmatrix} \theta_0 Y_i(u)\exp\{\alpha_0^T X_i\}dN_i^R(u)$$
$$+ \frac{1}{n}\sum_{i=1}^{n} \int_0^{t} \begin{pmatrix} X_i - \bar{X}(u;\beta_0) \\ B_i(u;\tilde{\Lambda}_0^D) - \bar{\bar{B}}(u) \end{pmatrix} \frac{(\theta_0^2 + \theta_0)\Delta_i(u)\eta_0^T Z_i \exp\{\alpha^T X_i\}}{\psi_i(u;\tilde{\Lambda}_0^D)\psi_i(u)} d\tilde{\Lambda}_0^R(u),$$
$$\tilde{G}_2(t) = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} X_i - \bar{X}(t;\beta_0) \\ B_i(t;\tilde{\Lambda}_0^D) - \bar{\bar{B}}(t) \end{pmatrix} \Delta_i(t)\gamma_0^T X_i \exp\{\beta_0^T X_i\}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} X_i - \bar{X}(t; \beta_0) \\ B_i(t; \tilde{\Lambda}_0^D) - \bar{\bar{B}}(t) \end{pmatrix} \Delta_i(t) \eta_0^T B_i(t) \exp\{\beta_0^T X_i\}.$$

By the Lemma A.1 of Lin and Ying [8], it can be shown that $R_3 = o_p(n^{1/2})$. Thus, we have

$$\tilde{U}_4(\xi_0) = \sum_{i=1}^{n} \vartheta_{4i} + o_p(n^{1/2}), \tag{A.11}$$

where

$$\vartheta_{4i} = \int_0^\tau \begin{pmatrix} X_i - \bar{x}(t; \beta_0) \\ B_i(t) - \bar{b}(t) \end{pmatrix} dM_i(t) + \int_0^\tau \left\{ \phi_{1i}(t) dG_1(t) - G_3(t) d\phi_{2i}(t) \right\}.$$

Let $\vartheta_i = (\vartheta_{1i}^T, \vartheta_{2i}^T, \vartheta_{3i}, \vartheta_{4i}^T)^T$, and $\Gamma = \Gamma(\xi_0)$ defined in condition (C4). Then it follows from (A.4), (A.5), (A.9), (A.11), and the Taylor expansion that

$$n^{1/2}(\hat{\xi} - \xi_0) = n^{-1/2} \Gamma^{-1} \sum_{i=1}^{n} \vartheta_i + o_p(1).$$

By the multivariate central limit theorem, $n^{1/2}(\hat{\xi} - \xi_0)$ is asymptotically normal with mean zero and covariance matrix $\Gamma^{-1} \Sigma (\Gamma^T)^{-1}$, where $\Sigma = E\{\vartheta_i \vartheta_i^T\}$.

References

1. Cai N, Lu W, Zhang H (2012) Time-varying latent effect model for longitudinal data with informative observation times. Biometrics 68:1093–1102
2. Kalbfleisch J, Prentice R (2002) The statistical analysis of failure time data, 2nd edn. Wiley, New York
3. Kalbfleisch J, Schaubel D, Ye Y, Gong Q (2013) An estimating function approach to the analysis of recurrent and terminal events. Biometrics 69:366–374
4. Liang K, Zeger S (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22
5. Liang Y, Lu W, Ying Z (2009) Joint modeling and analysis of longitudinal data with informative observation times. Biometrics 65:377–384
6. Lin DY, Wei LJ, Yang I, Ying Z (2000) Semiparametric regression for the mean and rate function of recurrent events. J R Stat Soc Ser B 69:711–730
7. Lin DY, Ying Z (1994) Semiparametric analysis of the additive risk model. Biometrika 81:61–71
8. Lin DY, Ying Z (2001) Semiparametric and nonparametric regression analysis of longitudinal data. J Am Stat Assoc 96:103–126
9. Lin H, Scharfstein D, Rosenheck D (2004) Analysis of longitudinal data with irregular outcome-dependent follow-up. J R Stat Soc Ser B 66:791–813
10. Liu L, Huang X, O'Quigley J (2008) Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. Biometrics 64:950–958
11. Liu L, Wolfe R, Kalfleisch J (2007) A shared random effects model for censored medical costs and mortality. Stat Med 26:139–155
12. Pollard D (1990) Empirical processes: theory and applications. Institute of Mathematical Statistics, Hayward

13. Ryu D, Sinha D, Mallick B, Lipsitz S, Lipshultz S (2007) Longitudinal studies with outcome-dependent follow-up: models and Bayesian regression. J Am Stat Assoc 102:952–961
14. Song X, Mu X, Sun L (2012) Regression analysis of longitudinal data with time-dependent covariates and informative observation times. Scand J Stat 39:248–258
15. Sun J, Park D-H, Sun L, Zhao X (2005) Semiparametric regression analysis of longitudinal data with informative observation times. J Am Stat Assoc 100:882–889
16. Sun J, Sun L, Liu D (2007) Regression analysis of longitudinal data in the presence of informative observation and censoring times. J Am Stat Assoc 102:1397–1406
17. Sun L, Song X, Zhou J, Liu L (2012) Joint analysis of longitudinal data with informative observation times and a dependent terminal event. J Am Stat Assoc 107:688–700
18. Wang J-L, Wang W (2001) Discussion on "Semiparametric and nonparametric regression analysis of longitudinal data" by D. Y. Lin and Z. Ying. J Am Stat Assoc 96:119–123
19. Ye Y, Kalbfleisch J, Schaubel D (2007) Semiparametric analysis of correlated recurrent and terminal events. Biometrics 63:78–87
20. Zhao X, Tong X, Sun L (2012) Joint analysis of longitudinal data with dependent observation times. Stat Sin 22:317–336
21. Zeng D, Chen Q, Ibrahim JG (2009) Gamma frailty transformation models for multivariate survival times. Biometrika 96:277–291
22. Zhou J, Zhao X, Sun L (2013) A new inference approach for joint models of longitudinal data with informative observation and censoring times. Stat Sin 23:571–593