Statistica Sinica **24** (2014), 1487-1504
doi:http://dx.doi.org/10.5705/ss.2013.063

JOINT MODELING OF LONGITUDINAL DATA WITH INFORMATIVE OBSERVATION TIMES AND DROPOUTS

Miao Han[1], Xinyuan Song[2], Liuquan Sun[1] and Lei Liu[3]

[1] *Chinese Academy of Sciences,* [2] *The Chinese University of Hong Kong and* [3] *Northwestern University*

Abstract: In many longitudinal studies, the response process is correlated with observation times and dropout. We propose a joint modeling for analysis of longitudinal data with informative observation times and dropout. We specify a semiparametric linear regression model for the longitudinal process, and accelerated time models for the observation and the dropout processes, while leaving the distributional form and dependent structure unspecified. Estimating equation approaches are developed for parameter estimation, and the resulting estimators are shown to be consistent and asymptotically normal. In addition, some numerical procedures are provided for model checking. The finite sample behavior of the proposed estimators is evaluated through simulation studies, and an application to a medical cost study of chronic heart failure patients from the University of Virginia Health System is provided.

Key words and phrases: Artificial censoring, estimating equations, informative dropout, informative observation times, joint modeling, longitudinal data.

1. Introduction

Longitudinal data frequently occur in medical follow-up studies and the like. Here subjects often selectively miss their visits or return at non-scheduled points in time and, as a result, observation times are irregular, and can be correlated with the longitudinal responses. In recent years, longitudinal data with informative observation times have attracted considerable attention (Lin, Scharfstein, and Rosenheck (2004)); Sun et al. (2005); Huang, Wang, and Zhang (2006); Ryu et al. (2007); Liang, Lu, and Ying (2009); Song, Mu, and Sun (2012); Zhao, Tong, and Sun (2012)). Thus, Lin, Scharfstein, and Rosenheck (2004) proposed a class of inverse intensity-of-visit process-weighted estimators in a typical marginal regression model, Sun et al. (2005) presented a conditional model, where the longitudinal outcome is assumed to depend only on the past observation history, and Liang, Lu, and Ying (2009) suggested a joint model for analysis of the longitudinal outcomes through two latent variables. All these methods require the assumption that censoring time is non-informative about the longitudinal response variable of interest.

In practice, there exist informative dropouts or dependent censoring such as death that stops the follow-up. For example, patients in a severe disease stage often die in a shorter period, and the cost for each visit tends to be higher than that of patients in a mild disease stage (Liu, Huang, and O'Quigley (2008)). A number of methods have been developed for handling informative dropout under the assumption that observation times are non-informative (Wu and Carroll (1988); Wulfsohn and Tsiatis (1997); Wang and Taylor (2001); Roy and Lin (2002); Lin and Ying (2003); Brown, Ibrahim, and DeGruttola (2005); Jin et al. (2006); Liu and Ying (2007); Ding and Wang (2008); Li, Hu, and Greene (2009)). For example, Lin and Ying (2003) proposed a joint modeling of longitudinal data with informative dropouts in which an accelerated failure time model is used to model the dropout process, while Liu and Ying (2007) proposed a combination of a linear mixed effects model for the longitudinal measurements and a semiparametric transformation model for the dropout process.

It is common when informative observation times and dropouts occur that observation and follow-up times are dependent on the response variable. A motivating example is a medical cost study of heart failure patients treated at the University of Virginia Health System. For these data, Liu, Huang, and O'Quigley (2008) and Sun et al. (2012) showed that the longitudinal medical costs could be correlated with both hospital visit times and dropout. There is limited research about this. For example, Sun, Sun, and Liu (2007) considered a joint model for the longitudinal process, the observation process, and the dropout process via a shared latent variable and assumed that the observation process is a nonhomogeneous Poisson process. Liu, Huang, and O'Quigley (2008) studied a joint random effects model for longitudinal data with informative observation times and a dependent terminal event, where the distributions of the random effects are specified. He, Tong, and Sun (2009) developed some shared frailty models to analyze panel count data with correlated observation and follow-up times, where one random effect was required to be normally distributed. Sun et al. (2012) proposed a joint modeling of longitudinal data with both informative observation times and a dependent terminal event via two latent variables.

We propose a semiparametric regression model for formulating the joint distribution of the longitudinal process, the observation process, and the dropout process. We specify a semiparametric linear regression model for the longitudinal process, and the accelerated time models for the observation and the dropout processes, while leaving the distributional form and dependent structure unspecified. The proposed model generalizes the approach of Lin and Ying (2003) to take informative observation times into account.

The rest of the article is organized as follows. Section 2 introduces notation and the model specification. Section 3 presents the proposed methods and asymptotic analysis. In Section 4, we develop a technique for checking the adequacy of the proposed model. Section 5 reports some results from simulation studies for evaluating the proposed methods. An application to the medical cost data for chronic heart failure patients from the clinical data repository at the University of Virginia Health System is provided in Section 6, and some concluding remarks are given in Section 7.

2. Model Specification

For a longitudinal study, let $Y(t)$ be the response at time t, and Z be a $p \times 1$ vector of covariates. Let D be the informative dropout time, C be the independent censoring time, and write $T = D \wedge C$, and $\delta = I(D \leq C)$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. Let $N(t)$ be the counting process denoting the number of the observation times before or at time t. The longitudinal process $Y(t)$ is observed only at the time points where $N(t)$ jumps, $t \leq T$. We assume that C is independent of $Y(\cdot)$ and $N(\cdot)$ conditional on Z, whereas D is allowed to depend on $Y(\cdot)$ and $N(\cdot)$, even conditional on Z. Let $\{Y_i(\cdot), N_i(\cdot), D_i, C_i, Z_i\}(i = 1, \ldots, n)$ be n independent replicates of $\{Y(\cdot), N(\cdot), D, C, Z\}$.

For the longitudinal process, the observation process, and the dropout time, we assume that there exist unknown constant vectors γ_0, β_0, and η_0 such that, for given Z and t, the random vectors $\{Y(te^{\gamma_0' Z}) - \beta_0' Z, N(te^{\gamma_0' Z}), De^{-\eta_0' Z}\}'$ have a common but unspecified joint distribution. Specifically,

$$\begin{pmatrix} Y_i(te^{\gamma_0' Z_i}) - \beta_0' Z_i \\ N_i(te^{\gamma_0' Z_i}) \\ D_i e^{-\eta_0' Z_i} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} Y_0(t) \\ N_0(t) \\ D_0 \end{pmatrix}, \tag{2.1}$$

where $\{Y_0(t), N_0(t), D_0\}'$ has an arbitrary distribution, and $\stackrel{d}{=}$ means equal in distribution. Under this joint model, the marginal distribution of $Y(t)$ satisfies a semiparametric linear regression model through scale-change for t:

$$E\{Y(te^{\gamma_0' Z})|Z\} = \mu_0(t) + \beta_0' Z,$$

where $\mu_0(t) = E\{Y_0(t)\}$, while the marginal distributions of $N(t)$ and D satisfy an accelerated time model for counting process (Lin, Wei, and Ying (1998)) and survival data (Tsiatis (1990)), respectively. The independent censoring time C is allowed to depend on Z in an arbitrary manner. Our proposed model is different from that of Lin and Ying (2003) in that $N(t)$ and $Y(t)$ have an arbitrary joint distribution in the sense of (2.1).

Note that $E\{Y(te^{\gamma_0' Z})|Z\} = \mu_0(t) + \beta_0' Z$ is equivalent to $E\{Y(t)|Z\} = \mu_0(te^{-\gamma_0' Z}) + \beta_0' Z$. Here γ_0 and β_0 can be interpreted as measuring two different effects the covariates might have on the longitudinal process: γ_0 to reflect the magnitude and direction of the scale-change of the mean function of $Y(t)$, β_0 to characterize the additive effect after adjusting the scale-change of the mean function of $Y(t)$. If Z consists of a single treatment indicator and say $e^{\gamma_0} = 2$, β_0 is the mean difference between the treatment response at time $2t$ and the control response at time t, which implies that β_0 is the additive effect after adjusting the scale-change of the mean function process in the treatment and the control groups. Thus, under (2.1), the treatment has both the effect of the scale change and the additive effect.

For a random sample of n subjects, the observed data consist of $\{Y_i(t)dN_i(t)$, T_i, δ_i, Z_i, $N_i(t), 0 \le t \le T_i, i = 1, \ldots, n\}$, where $T_i = D_i \wedge C_i$ and $\delta_i = I(D_i \le C_i)$. Our main interest is in estimating β_0.

3. Inference Procedures

To construct valid estimating functions for model parameters under informative dropout, we adopt the technique of artificial censoring (Ghosh and Lin (2003); Lin and Ying (2003)), which provides a way to create homogeneity for observations under comparison. Let Ω be the set of all possible values of Z. Define $T_i^*(\gamma, \eta) = T_i e^{-\eta' Z_i + d}$ and $\Delta_i(t; \gamma, \eta) = I\{T_i^*(\gamma, \eta) \ge t\}$, where $d = \inf_{z \in \Omega}(\eta - \gamma)' z$, and $d_0 = \inf_{z \in \Omega}(\eta_0 - \gamma_0)' z$. Under (2.1),

$$E\Big\{[Y_i(te^{\gamma_0' Z_i}) - \beta_0' Z_i] dN_i(te^{\gamma_0' Z_i}) | D_i e^{-\eta_0' Z_i + d_0} \ge t, Z_i\Big\} = d\mathcal{A}_0(t). \tag{3.1}$$

Define

$$M_i(t; \beta, \gamma, \eta, \mathcal{A}) = \int_0^t \Delta_i(s; \gamma, \eta) \Big\{[Y_i(se^{\gamma' Z_i}) - \beta' Z_i] dN_i(se^{\gamma' Z_i}) - d\mathcal{A}(s)\Big\}.$$

By (3.1) and the assumption that C_i is independent of $Y_i(\cdot)$ and $N_i(\cdot)$ conditional on Z_i, we have

$$\begin{aligned} & E\{M_i(t; \beta_0, \gamma_0, \eta_0, \mathcal{A}_0)\} \\ =& E\int_0^t \Delta_i(s; \gamma_0, \eta_0) \Big[E\big\{[Y_i(se^{\gamma_0' Z_i}) - \beta_0' Z_i] dN_i(se^{\gamma_0' Z_i}) | T_i e^{-\eta_0' Z_i + d_0} \ge s, Z_i\big\} - d\mathcal{A}_0(s)\Big] \\ =& E\int_0^t \Delta_i(s; \gamma_0, \eta_0) \Big[E\big\{[Y_i(se^{\gamma_0' Z_i}) - \beta_0' Z_i] dN_i(se^{\gamma_0' Z_i}) | D_i e^{-\eta_0' Z_i + d_0} \ge s, Z_i\big\} - d\mathcal{A}_0(s)\Big] \\ =& 0, \end{aligned} \tag{3.2}$$

which implies that $M_i(t; \beta_0, \gamma_0, \eta_0, \mathcal{A}_0)$ has zero mean. Thus, for given (β, γ, η), a reasonable estimator for $\mathcal{A}_0(t)$ is the solution to

$$\sum_{i=1}^n M_i(t; \beta, \gamma, \eta, \mathcal{A}) = 0.$$

Denote this estimator by $\hat{\mathcal{A}}(t;\beta,\gamma,\eta)$, and observe that

$$\hat{\mathcal{A}}(t;\beta,\gamma,\eta)=\sum_{i=1}^{n}\int_{0}^{t}\frac{\Delta_i(s;\gamma,\eta)[Y_i(se^{\gamma' Z_i})-\beta' Z_i]dN_i(se^{\gamma' Z_i})}{\sum_{j=1}^{n}\Delta_j(s;\gamma,\eta)}.$$

In view of (3.2), for given γ and η, to estimate β, by applying the generalized estimating equation approach (Liang and Zeger (1986)) and replacing $\mathcal{A}_0(t)$ by its estimator, we specify an estimating function for β_0:

$$\begin{aligned}U_1(\beta;\gamma,\eta)=\sum_{i=1}^{n}\int_{0}^{\infty}&W(t)\{Z_i-\bar{Z}(t;\gamma,\eta)\}\Big[Y_i(te^{\gamma' Z_i})-\bar{Y}^*(t;\gamma,\eta)\\&-\beta'\{Z_i-\bar{Z}(t;\gamma,\eta)\}\Big]\Delta_i(t;\gamma,\eta)dN_i(t^{\gamma' Z_i}),\end{aligned}\tag{3.3}$$

where

$$\bar{Z}(t;\gamma,\eta)=\frac{\sum_{i=1}^{n}\Delta_i(t;\gamma,\eta)Z_i}{\sum_{j=1}^{n}\Delta_j(t;\gamma,\eta)},$$
$$\bar{Y}^*(t;\gamma,\eta)=\frac{\sum_{i=1}^{n}\Delta_i(t;\gamma,\eta)Y_i^*(te^{\gamma' Z_i})}{\sum_{j=1}^{n}\Delta_j(t;\gamma,\eta)},$$

$Y_i^*(t)$ the measurement of Y_i taken at the time point nearest to t, and $W(t)$ a possibly data-dependent weight function.

In reality, γ_0 and η_0 are unknown. Note that $N_i(t)$ is subject to informative dropout by D_i. Then for given η, using the approach of Ghosh and Lin (2003), we propose an estimating function for γ_0:

$$U_2(\gamma;\eta)=\sum_{i=1}^{n}\int_{0}^{\infty}\{Z_i-\bar{Z}(t;\gamma,\eta)\}\Delta_i(t;\gamma,\eta)dN_i(te^{\gamma' Z_i}).\tag{3.4}$$

Since D_i is only subject to independent censoring by C_i, existing methods for the accelerated time model can be used to estimate η_0. Thus, if $\tilde{T}_i(\eta)=T_i e^{-\eta' Z_i}$ and $N_i^D(t;\eta)=\delta_i I\{\tilde{T}_i(\eta)\le t\}$ $(i=1,\ldots,n)$, η_0 can be consistently estimated from the estimating function

$$U_3(\eta)=\sum_{i=1}^{n}\int_{0}^{\infty}\{Z_i-\bar{Z}^D(t;\eta)\}dN_i^D(t;\eta),\tag{3.5}$$

where

$$\bar{Z}^D(t;\eta)=\frac{\sum_{i=1}^{n}I\{\tilde{T}_i(\eta)\ge t\}Z_i}{\sum_{j=1}^{n}I\{\tilde{T}_j(\eta)\ge t\}}.$$

A zero-crossing of $U_3(\eta)$, $\hat{\eta}$, is consistent and asymptotically normal (Tsiatis (1990); Lai and Ying (1991)). For given $\hat{\eta}$, let $\hat{\gamma}$ be a zero-crossing of $U_2(\gamma;\hat{\eta})$.

For given $\hat{\gamma}$ and $\hat{\eta}$, we estimate β_0 by solving $U_1(\beta;\hat{\gamma},\hat{\eta}) = 0$. The resulting estimator has the explicit expression

$$\hat{\beta} = \Big[\sum_{i=1}^{n}\int_0^\infty W(t)\{Z_i - \bar{Z}(t;\hat{\gamma},\hat{\eta})\}^{\otimes 2}\Delta_i(t;\hat{\gamma},\hat{\eta})dN_i(te^{\hat{\gamma}'Z_i})\Big]^{-1}$$
$$\times\sum_{i=1}^{n}\int_0^\infty W(t)\{Z_i - \bar{Z}(t;\hat{\gamma},\hat{\eta})\}\{Y_i(te^{\hat{\gamma}'Z_i}) - \bar{Y}^*(t;\hat{\gamma},\hat{\eta})\}\Delta_i(t;\hat{\gamma},\hat{\eta})dN_i(te^{\hat{\gamma}'Z_i}),$$

where for a vector a, $a^{\otimes 2} = aa'$. If γ_0 is 0, $\hat{\beta}$ is the estimator of Lin and Ying (2003). By using the Law of Large Numbers and the consistency of $\hat{\gamma}$ and $\hat{\eta}$, one can show that $\hat{\beta}$ is consistent.

Take $\theta = (\beta',\gamma',\eta')'$, $\hat{\theta} = (\hat{\beta}',\hat{\gamma}',\hat{\eta}')'$, and let θ_0 be the true value of θ. Then we show in the Appendix that $\hat{\theta}$ is asymptotically normal. However, it is difficult to estimate the asymptotic covariance matrix of $\hat{\theta}$ because $\hat{\gamma}$ and $\hat{\eta}$ are rank-based estimators. To proceed, we adopt a resampling technique due to Parzen, Wei, and Ying (1994). Let

$$\hat{M}_i(t;\theta) = \int_0^t \Delta_i(s;\gamma,\eta)\Big[\{Y_i(se^{\gamma'Z_i}) - \beta'Z_i\}dN_i(se^{\gamma'Z_i}) - d\hat{\mathcal{A}}(s;\beta,\gamma,\eta)\Big],$$
$$\hat{\mathcal{M}}_i(t;\gamma,\eta) = \int_0^t \Delta_i(s;\gamma,\eta)\{dN_i(se^{\gamma'Z_i}) - d\hat{\Lambda}(s;\gamma,\eta)\},$$
$$\hat{\mathcal{M}}_i^D(t;\eta) = N_i^D(t;\eta) - \int_0^t I\{\tilde{T}_i(\eta)\geq s\}d\hat{\Lambda}^D(s;\eta),$$

where

$$\hat{\Lambda}(t;\gamma,\eta) = \sum_{i=1}^{n}\int_0^t \frac{\Delta_i(s;\gamma,\eta)dN_i(se^{\gamma'Z_i})}{\sum_{j=1}^{n}\Delta_j(s;\gamma,\eta)},$$
$$\hat{\Lambda}^D(t;\eta) = \sum_{i=1}^{n}\int_0^t \frac{dN_i^D(s;\eta)}{\sum_{j=1}^{n} I\{\tilde{T}_j(\eta)\geq s\}}.$$

Let

$$\hat{\Psi}_{1i} = \int_0^\infty W(t)\{Z_i - \bar{Z}(t;\hat{\gamma},\hat{\eta})\}\Big[d\hat{M}_i(t;\hat{\theta}) - \{\bar{Y}^*(t;\hat{\gamma},\hat{\eta})$$
$$-\hat{\beta}'\bar{Z}(t;\hat{\gamma},\hat{\eta})\}d\hat{\mathcal{M}}_i(t;\hat{\gamma},\hat{\eta})\Big],$$
$$\hat{\Psi}_{2i} = \int_0^\infty \{Z_i - \bar{Z}(t;\hat{\gamma},\hat{\eta})\}d\hat{\mathcal{M}}_i(t;\hat{\gamma},\hat{\eta}),$$
$$\hat{\Psi}_{3i} = \int_0^\infty \{Z_i - \bar{Z}^D(t;\hat{\eta})\}d\hat{\mathcal{M}}_i^D(t;\hat{\eta}),$$

and let $\hat{\theta}^* = (\hat{\beta}^{*\prime},\hat{\gamma}^{*\prime},\hat{\eta}^{*\prime})'$ be the solution to

$$U(\theta) = \sum_{i=1}^{n} \hat{\Psi}_i G_i, \tag{3.6}$$

where $U(\theta) = (U_1(\beta;\gamma,\eta)', U_2(\gamma;\eta)', U_3(\eta)')'$, $\hat{\Psi}_i = (\hat{\Psi}'_{1i}, \hat{\Psi}'_{2i}, \hat{\Psi}'_{3i})'$, and $(G_1, \ldots, G_n)$ are independent standard normal variables independent of the observed data. Applying the arguments of Parzen, Wei, and Ying (1994) and Lin, Wei, and Ying (1998), the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta_0)$ can be approximated by the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ given the observed data. To approximate the distribution of $\hat{\theta}$, we produce a large number of realizations of $\hat{\theta}^*$ by repeatedly generating the random samples $(G_1, \ldots, G_n)$, while fixing the data $\{Y_i(\cdot)dN_i(\cdot), T_i, \delta_i, Z_i, N_i(\cdot)\}$ $(i = 1, \ldots, n)$ at their observed values. The covariance matrix of $\hat{\theta}$ is then approximated by the empirical covariance matrix of $\hat{\theta}^*$, and confidence intervals for θ_0 can be constructed accordingly.

4. Model Checking

We propose some simple graphical and numerical procedures for assessing the adequacy of the proposed model. Model (2.1) implies that, for $i = 1, \ldots, n$, the $D_i e^{-\eta_0' Z_i}$ have a common marginal distribution, the $E\{dN_i(te^{\gamma_0' Z_i})|D_i e^{-\eta_0' Z_i} \geq t\}$ have a common value, and the $E\big\{[Y_i(te^{\gamma_0' Z_i}) - \beta_0' Z_i]dN_i(te^{\gamma_0' Z_i})|D_i e^{-\eta_0' Z_i} \geq t\big\}$ have a common value. To verify these conditions, following Lin, Robins, and Wei (1996) and Ghosh and Lin (2003), we consider the residual processes,

$$\mathcal{F}_1(t;\eta) = n^{-1/2} \sum_{i=1}^{n} Z_i \hat{\mathcal{M}}_i^D(t;\eta),$$
$$\mathcal{F}_2(t;\gamma,\eta) = n^{-1/2} \sum_{i=1}^{n} Z_i \hat{\mathcal{M}}_i(t;\gamma,\eta),$$
$$\mathcal{F}_3(t;\theta) = n^{-1/2} \sum_{i=1}^{n} Z_i \hat{M}_i(t;\theta).$$

Let $\mathcal{F}(t;\theta) = (\mathcal{F}_1(t;\eta)', \mathcal{F}_2(t;\gamma,\eta)', \mathcal{F}_3(t;\theta)')'$. It can be shown following the argument of Lin, Robins, and Wei (1996) that the null distributions of $\mathcal{F}(t;\hat{\theta})$ is approximated by the conditional distributions of $\hat{\mathcal{F}}(t) = (\hat{\mathcal{F}}_1(t)', \hat{\mathcal{F}}_2(t)', \hat{\mathcal{F}}_3(t)')'$, where

$$\hat{\mathcal{F}}_1(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \{Z_i - \bar{Z}^D(s;\hat{\eta})\} d\hat{\mathcal{M}}_i^D(s;\hat{\eta}) G_i - \mathcal{F}_1(t;\hat{\eta}^*) + \mathcal{F}_1(t;\hat{\eta}),$$

$$\hat{\mathcal{F}}_2(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \{Z_i - \bar{Z}(s;\hat{\gamma},\hat{\eta})\} d\hat{\mathcal{M}}_i(s;\hat{\gamma},\hat{\eta}) G_i - \mathcal{F}_2(t;\hat{\gamma}^*,\hat{\eta}^*) + \mathcal{F}_2(t;\hat{\gamma},\hat{\eta}),$$

$$\hat{\mathcal{F}}_3(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \{Z_i - \bar{Z}(s;\hat{\gamma},\hat{\eta})\} d\hat{M}_i(s;\hat{\theta}) G_i - \mathcal{F}_3(t;\hat{\theta}^*) + \mathcal{F}_3(t;\hat{\theta}).$$

Thus, one can obtain a large number of realizations from $\hat{\mathcal{F}}(t)$ by repeatedly generating the standard normal random sample $(G_1, \ldots, G_n)$ while fixing the observation data. Moreover, $\mathcal{F}(t;\hat{\theta})$ can be plotted along with a few realizations $\hat{\mathcal{F}}(t)$. The validity of approximating $\mathcal{F}(t;\hat{\theta})$ by $\hat{\mathcal{F}}(t)$ depends on the correct specification of (2.1). Hence, an unusual pattern of $\mathcal{F}(t;\hat{\theta})$ compared to the realizations of $\hat{\mathcal{F}}(t)$ would suggest a lack-of-fit of (2.1). Since $\mathcal{F}(t;\hat{\theta})$ is expected to fluctuate randomly around 0, a formal lack-of-fit test can be constructed based on the supremum statistics $\sup_t \|\mathcal{F}_1(t;\hat{\eta})\|$, $\sup_t \|\mathcal{F}_2(t;\hat{\gamma},\hat{\eta})\|$, and $\sup_t \|\mathcal{F}_3(t;\hat{\theta})\|$. The p-values of these tests can be obtained by comparing their observed values with a large number of realizations from $\hat{\mathcal{F}}_1(t)$, $\hat{\mathcal{F}}_2(t)$, and $\hat{\mathcal{F}}_3(t)$, respectively, with (2.1) rejected if one of these p-values is smaller than a prespecified level.

5. Simulation Studies

We conducted simulation studies to assess the performance of the proposed estimators with the focus on estimating β_0. We considered that Z was generated from a Bernoulli distribution with success probability 0.5 or that Z was generated from a uniform distribution on $(0, 1)$. Let v be a gamma random variable with mean 1 and variance σ^2, where $\sigma^2 = 0$, 1 or 4; given v, the dependent censoring time D_0 and the baseline gap time between every two successive observations were generated as independent exponential random variables with hazard rates v and $4v$, respectively. We took $Y_0(t)$ as $Y_0(t) = a_0 + a_1 \sin t + \epsilon(t)$, where $a_0 = 0$, $a_1 = 1$, and $\epsilon(t)$ is normal with mean 0 and variance $v/(1 + v)$ for all t.

We multiplied D_0 and the baseline gap times by $e^{\eta_0 Z}$ and $e^{\gamma_0 Z}$ to produce the informative dropout time and gap times associated with Z, where $\eta_0 = 0.25$ and $\gamma_0 = \log(3)$, and set $Y(t) = \beta_0 Z + Y_0(te^{-\gamma_0 Z})$, where $\beta_0 = 1$. The dependence among $Y(t)$, $N(t)$, and D was induced by the common random effect v. When $\sigma^2 = 0$, the three related processes were independent. The independent censoring time C was generated from a uniform distribution on $(0,\ \tau)$ with $\tau = 5$ or 20. For each simulation study, we set $W(t) = 1$. The results presented were based on 1000 replications with sample sizes $n = 100$ and 200. For each data set, the resampling distribution was based on 100 realizations; this was found to be adequate.

Table 1 gives the simulation results on the estimates of β_0. In the table, Bias is the sampling mean of the estimate minus the true value, SE is the sampling standard error of the estimate, SEE is the sampling mean of standard error estimate, and CP is the 95% empirical coverage probability for β_0 based on the normal approximation. Table 1 shows that the proposed method performed well for both informative ($\sigma^2 \neq 0$) and non-informative ($\sigma^2 = 0$) cases. Thus, the proposed estimators were virtually unbiased, the estimated standard errors were close to the empirical standard errors, and the coverage probabilities of the 95%

Table 1. Simulation results for the estimation of $\beta_0 = 1$.

n	Z	σ^2	τ	Bias	SE	SEE	CP
100	Bernoulli(0.5)	0	5	-0.0045	0.1015	0.1051	0.951
			20	-0.0030	0.0903	0.0950	0.948
		1	5	-0.0085	0.1311	0.1343	0.946
			20	-0.0008	0.0981	0.1066	0.956
		4	5	0.0083	0.1925	0.2076	0.948
			20	0.0067	0.1510	0.1691	0.962
	Uniform(0,1)	0	5	-0.0013	0.1751	0.1954	0.944
			20	-0.0053	0.1617	0.1749	0.957
		1	5	0.0008	0.2299	0.2496	0.945
			20	-0.0022	0.1899	0.2023	0.958
		4	5	0.0077	0.3565	0.4001	0.953
			20	0.0032	0.2916	0.3217	0.951
200	Bernoulli(0.5)	0	5	-0.0034	0.0700	0.0707	0.949
			20	0.0000	0.0643	0.0655	0.945
		1	5	-0.0019	0.0916	0.0924	0.946
			20	-0.0030	0.0706	0.0719	0.956
		4	5	0.0043	0.1435	0.1401	0.938
			20	0.0017	0.1114	0.1147	0.947
	Uniform(0,1)	0	5	-0.0018	0.1202	0.1272	0.951
			20	0.0017	0.1137	0.1173	0.945
		1	5	-0.0063	0.1645	0.1673	0.954
			20	0.0017	0.1220	0.1330	0.966
		4	5	-0.0048	0.2458	0.2598	0.949
			20	-0.0024	0.1880	0.2107	0.961

confidence intervals were reasonable. The performance of the proposed estimator improved when the sample size increased from 100 to 200. We also considered other setups and the results were similar.

For comparison, we considered the method of Lin and Ying (2003) (denoted by LY) using the same setup as in Table 1, but with $n = 200$. In these situations, the models were misspecified for LY's method even when $\sigma^2 = 0$, because $\gamma_0 \neq 0$. Thus we considered another model, M1, that was correctly specified for LY's method: as in Lin and Ying (2003), $Y(t) = \sin(t) + \beta_0 Z + \epsilon(t)$, where Z was Bernoulli with success probability 0.5, $\epsilon(t)$ was normal with mean ϕ and variance 1 for all t, and ϕ was a standard normal random variable. Here $N(t)$ was a Poisson process with intensity rate $\psi e^{\gamma_0 Z}$ with $\gamma_0 = 0.5$, and ψ was an independent gamma random variable with mean 1 and variance 0.25; D was set to be $D_0 e^{\eta_0 Z}$ with $\eta_0 = 0.5$, where D_0 was the exponential of a normal random variable with mean ϕ and variance 1; C was a uniform variable on $(0, 30)$. Table 2 gives the results for $\beta_0 = 1$. It can be seen that LY's method yielded consistent

Table 2. Comparison results on estimation of $\beta_0 = 1$ with $n = 200$.

Z	σ^2	τ	Ours Bias	Ours SE	LY Bias	LY SE
Bernoulli(0.5)	0	5	0.0014	0.0712	-0.2705	0.0621
		20	-0.0028	0.0640	-0.2364	0.0698
	1	5	-0.0047	0.0906	-0.2523	0.0809
		20	-0.0068	0.0692	-0.0960	0.0886
	4	5	0.0046	0.1409	-0.2451	0.1631
		20	0.0024	0.1076	-0.1055	0.2181
Uniform(0,1)	0	5	-0.0036	0.1246	-0.2825	0.1032
		20	0.0042	0.1106	-0.2481	0.1035
	1	5	0.0036	0.1591	-0.2540	0.1289
		20	-0.0005	0.1237	-0.0939	0.1305
	4	5	-0.0064	0.2345	-0.2411	0.2891
		20	0.0017	0.1847	-0.1136	0.3424
M1			0.0630	0.2618	0.0008	0.2535

estimators when the observation times were noninformative, while the proposed estimator showed some bias. However, the LY estimator was highly biased when the observation times were informative.

The bias of the proposed estimator under model M1 was caused by a scale-change of $Y(t)$ from t to $te^{\gamma_0 Z}$. Thus, we conducted an additional simulation study with $\gamma_0 = 0$, using the same setup as in Table 1 but with $\gamma_0 = 0$ and $n = 200$. Model M2 was the same as M1 but with $\gamma_0 = 0$. The results for $\beta_0 = 1$ are in Table 3. As anticipated, the LY estimator was unbiased under $\sigma^2 = 0$ and M2, which was correctly specified for LY's method. Under these conditions, both methods provided reasonable and comparable estimates for β_0. However, when $\sigma^2 \neq 0$, the LY estimator appeared biased, especially when σ^2 was large, while the proposed estimator was essentially unbiased for all cases considered. We considered other setups and the results were similar.

To evaluate the loss of efficiency by artificial censoring, we considered the case of no informative observation times and dropout. Using the setup of Table 1 with $\sigma^2 = 0$, we compared the proposed method and the naive method that simply replaced $I\{T_i^*(\gamma, \eta) \geq t\}$ with $I\{T_i e^{-\gamma' Z_i} \geq t\}$ in our method. Table 4 gives the results for $\beta_0 = 1$. Both methods provided reasonable estimates, and the variances of our method were only slightly larger than those of the naive method. This amounts to little loss of efficiency by artificial censoring when there were no informative observation times and dropout.

To investigate the performance of model checking method, we conducted some simulations to assess the size and power of tests based on $\sup_t \|\mathcal{F}_1(t; \hat{\eta})\|$,

Table 3. Comparison results on estimation of $\beta_0 = 1$ with $\gamma_0 = 0$ and $n = 200$.

			Ours		LY	
Z	σ^2	τ	Bias	SE	Bias	SE
Bernoulli(0.5)	0	5	-0.0007	0.0436	0.0010	0.0407
		20	-0.0010	0.0370	0.0001	0.0367
	1	5	0.0030	0.0596	-0.0132	0.0569
		20	-0.0019	0.0483	-0.0129	0.0474
	4	5	0.0047	0.0959	-0.0599	0.1679
		20	0.0002	0.0853	-0.0593	0.1762
Uniform(0,1)	0	5	-0.0053	0.0853	-0.0033	0.0707
		20	-0.0044	0.0694	-0.0025	0.0667
	1	5	-0.0015	0.1044	-0.0156	0.1043
		20	-0.0032	0.0895	-0.0120	0.0891
	4	5	0.0060	0.1845	-0.0451	0.2839
		20	-0.0019	0.1505	-0.0594	0.3228
M2			-0.0025	0.2551	-0.0013	0.2522

Table 4. Comparison results on estimation of $\beta_0 = 1$ with $\sigma^2 = 0$.

			Ours		Naive	
n	Z	τ	Bias	SE	Bias	SE
100	Bernoulli(0.5)	5	-0.0038	0.1052	-0.0036	0.0940
		20	-0.0028	0.0906	-0.0027	0.0804
	Uniform(0,1)	5	0.0031	0.1705	0.0011	0.1471
		20	0.0005	0.1586	-0.0007	0.1292
200	Bernoulli(0.5)	5	-0.0037	0.0692	-0.0040	0.0625
		20	-0.0020	0.0638	-0.0006	0.0578
	Uniform(0,1)	5	-0.0025	0.1229	0.0007	0.1062
		20	-0.0007	0.1087	-0.0001	0.0937

$\sup_t \|\mathcal{F}_2(t;\hat{\gamma},\hat{\eta})\|$, and $\sup_t \|\mathcal{F}_3(t;\hat{\theta})\|$ with $n = 800$. We only considered that Z followed a Bernoulli distribution with success probability 0.5. We assumed that the longitudinal process, the observation process, and the dropout time were as

$$\begin{pmatrix} Y_i(te^{\gamma_0' Z_i} + 0.25kZ_i) - \beta_0' Z_i e^{0.5kZ_i} \\ N_i(te^{\gamma_0' Z_i} + 0.25kZ_i) \\ (D_i - 0.1kZ_i)e^{-\eta_0' Z_i} \end{pmatrix} \overset{d}{=} \begin{pmatrix} Y_0(t) \\ N_0(t) \\ D_0 \end{pmatrix}$$

with $k = 0$, 1, 2, 3, and 4. All other setups are the same as in Table 1 with $\tau = 20$. We considered the null hypothesis H_0 as $k = 0$. Table 5 reports the empirical sizes and powers of the proposed tests at the significance level of 0.05. In the Table 5, T1, T2, and T3 denote tests based on $\sup_t \|\mathcal{F}_1(t;\hat{\eta})\|$, $\sup_t \|\mathcal{F}_2(t;\hat{\gamma},\hat{\eta})\|$, and $\sup_t \|\mathcal{F}_3(t;\hat{\theta})\|$, respectively. The results show that the empirical sizes were

Table 5. The empirical sizes and powers of the model tests with $n = 800$.

σ^2	Test	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
0	T1	0.0520	0.3520	0.9520	1.0000	1.0000
	T2	0.0540	0.9580	1.0000	1.0000	1.0000
	T3	0.0410	0.9930	1.0000	1.0000	1.0000
1	T1	0.0440	0.1670	0.7370	0.9890	1.0000
	T2	0.0520	0.6740	0.9870	0.9990	1.0000
	T3	0.0440	0.7910	1.0000	1.0000	1.0000
4	T1	0.0420	0.0820	0.4200	0.7300	0.8750
	T2	0.0460	0.3590	0.7540	0.8870	0.9340
	T3	0.0540	0.1220	0.3470	0.7430	0.9760

close to nominal, and the tests had reasonable powers to detect deviations from the null hypothesis, with power increasing as the value of k increased.

We conducted some simulations to assess the sensitivity with respect to the dependence assumption. We assumed that the longitudinal process, the observation process, and the dropout time were given by

$$\begin{pmatrix} Y_i(te^{\gamma_0' Z_i} + 0.25k_3 Z_i) - \beta_0' Z_i e^{0.5k_3 Z_i} \\ N_i(te^{\gamma_0' Z_i} + 0.25k_2 Z_i) \\ (D_i - 0.1k_1 Z_i)e^{-\eta_0' Z_i} \end{pmatrix} \overset{d}{=} \begin{pmatrix} Y_0(t) \\ N_0(t) \\ D_0 \end{pmatrix}$$

with $(k_1, k_2, k_3) = (4, 0, 0)$, $(0, 4, 0)$ and $(0, 0, 4)$. All other setups are the same as in Table 5. We considered the null hypothesis H_0 as $(k_1, k_2, k_3) = (0, 0, 0)$. Table 6 presents the empirical sizes and powers of the proposed tests at the significance level of 0.05. The results indicate that the model checking method was sensitive to the dependence assumption. In particular, the T1 results affected the performance of T2 and T3, and T2 affected T3. We suggest the three supremum tests be conducted in the order T1, T2, and T3; if T1 is rejected, it does not make sense to run the others.

6. An Application

We illustrate the proposed methods using the medical cost data of chronic heart failure patients treated at the University of Virginia Health System. These data have been analyzed by Liu, Huang, and O'Quigley (2008) and Sun et al. (2012) In the study, there were 1,475 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. For each patient, the observed information included the clinical visit times (in month), and the medical cost for each hospital visit. In addition, three

Table 6. The empirical sizes and powers of the sensitivity to the dependence assumption.

σ^2	Test	$(k_1, k_2, k_3) = (4, 0, 0)$	$(k_1, k_2, k_3) = (0, 4, 0)$	$(k_1, k_2, k_3) = (0, 0, 4)$
0	T1	1.0000	0.0450	0.0440
	T2	0.0420	1.0000	0.0520
	T3	0.0460	1.0000	0.0720
1	T1	1.0000	0.0460	0.0430
	T2	0.2370	1.0000	0.0530
	T3	0.0690	1.0000	0.2610
4	T1	0.8860	0.0440	0.0450
	T2	0.4190	0.9730	0.0490
	T3	0.1280	0.4980	0.0770

Table 7. Joint analysis of the medical cost data for heart failure patients.

	β			γ			η		
	Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value
White	-0.3310	0.1138	0.0036	0.0966	0.0488	0.0475	0.3869	0.2236	0.0835
Age	-0.1046	0.0533	0.0495	-0.0763	0.0270	0.0048	-0.7286	0.1081	0.0000

Note: Est. is the estimate of the parameter, and SE is the standard error estimate.

baseline covariates were measured: gender, race, and age. Preliminary studies showed that patients visiting the hospital more often tended to pay more for each visit, and these patients also had a higher mortality rate. Since gender had been shown to have no effect on the medical cost and the hospital visits (Liu, Huang, and O'Quigley (2008); Sun et al. (2012)), we focused on the effects of race and age on the medical cost with observation times and death.

For patient i, let $Y_i(t)$ be the log-transformed cost. Let Z_{i1} be a binary indicator of race (white=1, nonwhite=0), and Z_{i2} denote the age group, taking values 0, 1, and 2 for 60-69, 70-79, and 80-89 years, respectively. We took the data to be described by (2.1), and used 100 realizations to estimate the asymptotic variance. We also took $W(t) = 1$. The results are summarized in Table 7. They suggest that both race and age had significant effects on the medical cost for each visit. Specifically, white patients tended to be at less risk for hospitalization and had less medical costs at each visit. Older people were more likely to visit the hospital and had lower medical cost. In addition, on average, the times to hospitalizations for white patients were 1.1 times longer than those of nonwhite patients, and white patients lived 1.47 times as long as nonwhite patients. Older patients tended to have higher hazard compared to younger patients. These findings are similar to those obtained by Liu, Huang, and O'Quigley (2008) and Sun et al. (2012).

We applied our model checking techniques to assess the adequacy of (2.1) for these data. We calculated $\mathcal{F}(t;\hat{\theta})$ and obtained $\sup_t \|\mathcal{F}_1(t;\hat{\eta})\| = 0.4092$ with a p-value of 0.28, $\sup_t \|\mathcal{F}_2(t;\hat{\gamma},\hat{\eta})\| = 1.5606$ with a p-value of 0.33, and $\sup_t \|\mathcal{F}_3(t;\hat{\theta})\| = 11.0798$ with a p-value of 0.29, based on 100 realizations of the corresponding statistics $\sup_t \|\hat{\mathcal{F}}_1(t)\|$, $\sup_t \|\hat{\mathcal{F}}_2(t)\|$ and $\sup_t \|\hat{\mathcal{F}}_3(t)\|$, respectively. These indicated no evidence of model misspecification.

7. Concluding Remarks

The proposed estimator may be inefficient if the range of $(\eta_0 - \gamma_0)'Z$ is very large, because then there would be excessive artificial censoring. We can use stratification to alleviate this problem (e.g., Lin and Ying (2003)). For example, if the large range of $(\eta_0 - \gamma_0)'Z$ is caused by some extreme values in a particular component of Z, we would stratify the sample on this covariate: one stratum containing the middle values of this covariate, and the other containing the rest. The estimating functions for regression parameters can be constructed separately in the two strata by using different values of d. Then we can combine the two estimating functions by adding them together.

Model (2.1) assumes the same set of covariates Z for the longitudinal process, the observation process, and the dropout process. The proposed estimation procedure can be extended in a straightforward manner to deal with different set of covariates for them. We have assumed that the covariates are time-independent. If not, we can specify that, for given $Z(t)$ and t, the random vector $\{Y(\int_0^t e^{\gamma_0' Z(u)}du) - \beta_0' Z(t), N(\int_0^t e^{\gamma_0' Z(u)}du), \int_0^D e^{-\eta_0' Z(u)}du\}'$ has an arbitrary common joint distribution. If $\Omega(t)$ is the set of all possible values of $Z(t)$ for each $t \geq 0$, we take $T_i^*(\gamma,\eta)$ as $\int_0^{T_i} e^{-\eta' Z_i(t)+d}dt$, where $d = \inf_{z(t)\in\Omega(t), t\geq 0}\{(\eta-\gamma)'z(t)\}$. With these modifications, applying Lin and Ying (1995) and Lin, Wei, and Ying (1998), the results in Section 3 continue to hold for time-dependent covariates. This will be reported elsewhere.

Here we assume that the rescaling time for the longitudinal process is completely determined by the rate of the observation process, and this has the benefit of always having observed values for $Y(\cdot)$ when $N(\cdot)$ jumps. This assumption can be violated. In general, we can assume that there exist unknown constant vectors γ_0, β_0, and η_0 such that, for given Z and t, the random vectors $\{Y(t) - \beta_0' Z, N(te^{\gamma_0' Z}), De^{-\eta_0' Z}\}'$ have a common, but completely unspecified, joint distribution. The proposed estimation procedure for model (2.1) cannot be extended in a straightforward manner to deal with this case. It is a challenging problem and requires further research.

In practice, some covariates have the effect of a scale change, and others can have an additive effect. Here, let Z and X are $p \times 1$ and $q \times 1$ vectors of covariates, respectively, where covariates Z have the effect of a scale change, and those in X have an additive effect. Then we can assume the marginal model

$$E\{Y(te^{\gamma_0' Z})|Z, X\} = \mu_0(t) + \beta_0' X.$$

A more general model can be seen in Gray and Brookmeyer (1998). Model (2.1) can also be modified accordingly, and the proposed estimation procedure can be extended in a straightforward manner to deal with this case.

Acknowledgement

The authors thank the Co-Editor, Jeng-Min Chiou, an associate editor, and two referees for their insightful comments and suggestions that greatly improved the article. This research was partly supported by the National Natural Science Foundation of China Grants (No. 11231010, 11171330 and 11021161), Key Laboratory of RCSDS, CAS (No.2008DP173182), GRF 404711 from the Research Grant Council of the Hong Kong Special Administration Region, and R01 HS 020263 from Agency of Healthcare Research and Quality of USA.

Appendix: Proofs of Asymptotic Results

Under (2.1), the conditional expectations $E\{dN_i(te^{\gamma_0' Z_i})|D_i e^{-\eta_0' Z_i + d_0} \geq t, Z_i\}$ $(i = 1, \ldots, n)$ have a common value

$$E\{dN_i(te^{\gamma_0' Z_i})|D_i e^{-\eta_0' Z_i + d_0} \geq t, Z_i\} = d\Lambda_0(t). \tag{A.1}$$

Let

$$\mathcal{M}_i(t) = \int_0^t \Delta_i(s; \gamma_0, \eta_0)\{dN_i(se^{\gamma_0' Z_i}) - d\Lambda_0(s)\}.$$

By (A.1) and the assumption that C_i is independent of $N_i(\cdot)$ conditional on Z_i, we have that $\mathcal{M}_i(t)$ is a zero-mean process (Ghosh and Lin (2003)). Let $\Lambda_0^D(t)$ be the common cumulative hazard function for $D_i e^{-\eta_0' Z_i} (i = 1, \ldots, n)$ so that

$$\mathcal{M}_i^D(t) = N_i^D(t; \eta_0) - \int_0^t I\{\tilde{T}_i(\eta_0) \geq t\} d\Lambda_0^D(s)$$

is a martingale process (Tsiatis (1990)). Let $M_i(t) = M_i(t; \beta_0, \gamma_0, \eta_0, \mathcal{A}_0)$, see (3.2). Simple manipulation yields

$$U_1(\beta_0; \gamma_0, \eta_0) = \sum_{i=1}^n \int_0^\infty W(t)\{Z_i - \bar{Z}(t; \gamma_0, \eta_0)\} \times \Big[dM_i(t) - \{\bar{Y}^*(t; \gamma_0, \eta_0) - \beta_0' \bar{Z}(t; \gamma_0, \eta_0)\} d\mathcal{M}_i(t)\Big],$$

$$U_2(\gamma_0;\eta_0) = \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}(t;\gamma_0,\eta_0)\} d\mathcal{M}_i(t),$$
$$U_3(\eta_0) = \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}^D(t;\eta_0)\} d\mathcal{M}_i^D(t).$$

Using Lemma 1 of Appendix A.1 of Lin and Ying (2001), we have

$$n^{-1/2}U_1(\beta_0;\gamma_0,\eta_0) = n^{-1/2}\sum_{i=1}^n \int_0^\infty w(t)\{Z_i - \bar{z}(t)\} \times \Big[dM_i(t) - \{\bar{y}^*(t) - \beta_0'\bar{z}(t)\}d\mathcal{M}_i(t)\Big] + o_p(1), \tag{A.2}$$
$$n^{-1/2}U_2(\gamma_0;\eta_0) = n^{-1/2}\sum_{i=1}^n \int_0^\infty \{Z_i - \bar{z}(t)\}d\mathcal{M}_i(t) + o_p(1), \tag{A.3}$$
$$n^{-1/2}U_3(\eta_0) = n^{-1/2}\sum_{i=1}^n \int_0^\infty \{Z_i - \bar{z}^D(t)\}d\mathcal{M}_i^D(t) + o_p(1), \tag{A.4}$$

where $w(t)$, $\bar{z}(t)$, $\bar{y}^*(t)$ and $\bar{z}^D(t)$ are the limits of $W(t)$, $\bar{Z}(t;\gamma_0,\eta_0)$, $\bar{Y}^*(t;\gamma_0,\eta_0)$, and $\bar{Z}^D(t;\eta_0)$, respectively. Note that $U(\theta) = (U_1(\beta;\gamma,\eta)', U_2(\gamma;\eta)', U_3(\eta)')'$, see (3.6). Since the right hand sides of (A.2), (A.3), and (A.4) consist of sums of independent random vectors plus asymptotically negligible terms, the Multivariate Central Limit Theorem implies that $n^{-1/2}U(\theta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix $\Sigma = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n E\{\Psi_i^{\otimes 2}\}$, where $\Psi_i = (\Psi_{1i}', \Psi_{2i}', \Psi_{3i}')'$,

$$\Psi_{1i} = \int_0^\infty w(t)\{Z_i - \bar{z}(t)\}\Big[dM_i(t) - \{\bar{y}^*(t) - \beta_0'\bar{z}(t)\}d\mathcal{M}_i(t)\Big],$$
$$\Psi_{2i} = \int_0^\infty \{Z_i - \bar{z}(t)\}d\mathcal{M}_i(t),$$
$$\Psi_{3i} = \int_0^\infty \{Z_i - \bar{z}^D(t)\}d\mathcal{M}_i^D(t).$$

Applying the technique of Ying (1993) and Lin, Wei, and Ying (1998), we can show that for θ in a small neighborhood of θ_0,

$$n^{-1/2}U(\theta) = n^{-1/2}U(\theta_0) + An^{1/2}(\theta - \theta_0) + o_p(1),$$

where A is the asymptotic slope matrix of $n^{-1}U(\theta_0)$. It then follows that $n^{-1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $A^{-1}\Sigma A^{-1}$.

References

Brown, E. R., Ibrahim, J. G. and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* **61**, 64-73.

Ding, J. and Wang, J. L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* **64**, 546-556.

Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59**, 877-885.

Gray, S. and Brookmeyer, R. (1998). Estimating a treatment effect from multidimensional longitudinal data. *Biometrics* **54**, 976-988.

He, X., Tong, X. and Sun, J. (2009). Semiparametric analysis of panel count data with correlated observation and follow-up times. *Lifetime Data Analysis* **15**, 177-196.

Huang, C. Y., Wang, M. C. and Zhang, Y. (2006). Analyzing panel count data with informative observation times. *Biometrika* **93**, 763-775.

Jin, Z., Liu, M. L., Steven, A. and Ying, Z. (2006). Analysis of longitudinal health-related quality of life data with terminal events. *Lifetime Data Analysis* **12**, 169-190.

Lai, T. L. and Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531-556.

Li, L., Hu, B. and Greene, T. (2009). A semiparametric joint model for longitudinal and survival data with application to hemodialysis study. *Biometrics* **65**, 737-745.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Liang, Y., Lu, W. and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics* **65**, 377-384.

Lin, D. Y., Robins, J. and Wei, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83**, 381-393.

Lin, D. Y., Wei, L. J. and Ying, Z. (1998). Accelerated failure time models for counting process. *Biometrika* **85**, 605-618.

Lin, D. Y. and Ying, Z. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *J. Statist. Plann. Inference* **44**, 47-63.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 1045-1056.

Lin, D. Y. and Ying, Z. (2003).. Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics* **4**, 385-398.

Lin, H., Scharfstein, D. O. and Rosenheck, D. O. (2004). Analysis of longitudinal data with irregular outcome-dependent follow-up. *J. Roy. Statist. Soc. Ser. B* **66**, 791-813.

Liu, L., Huang, X. and 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.

Liu, M. and Ying, Z. (2007). Joint analysis of longitudinal data with informative right censoring. *Biometrics* **63**, 363-371.

Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341-350.

Roy, J. and Lin, X. H. (2002). Analysis of multivariate longitudinal outcomes with nonignorable dropout and missing covariates: changes in methadone treatment practices. *J. Amer. Statist. Assoc.* **97**, 40-52.

Ryu, D., Sinha, D., Mallick, B., Lipsitz, S. R. and Lipshultz, S. E. (2007). Longitudinal studies with outcome-dependent follow-up: models and Bayesian regression. *J. Amer. Statist. Assoc.* **102**, 952-961.

Song, X., Mu, X. and Sun, L. (2012). Regression analysis of longitudinal data with time-dependent covariates and informative observation times. *Scand. J. Statist.* **39**, 248-258.

Sun, J., Park, D-H, Sun, L. and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *J. Amer. Statist. Assoc.* **100**, 882-889.

Sun, J., Sun, L. and Liu, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *J. Amer. Statist. Assoc.* **102**, 1397-1406.

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *J. Amer. Statist. Assoc.* **107**, 688-700.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-372.

Wang, Y. and Taylor, M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. Amer. Statist. Assoc.* **96**, 895-905.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175-188.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.

Zhao, X., Tong, X. and Sun, L. (2012). Joint analysis of longitudinal data with dependent observation times. *Statist. Sinica* **22**, 317-336.

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P.R. China.

E-mail: hanmiao10@mails.ucas.ac.cn

Department of Statistics, The Chinese University of Hong Kong, Hong Kong.

E-mail: xysong@sta.cuhk.edu.hk

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P.R. China.

E-mail: slq@amt.ac.cn

Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA.

E-mail: lei.liu@northwestern.edu

(Received March 2013; accepted September 2013)