Statistica Sinica **23** (2013), 571-593
doi:http://dx.doi.org/10.5705/ss.2011.285

A NEW INFERENCE APPROACH FOR JOINT MODELS OF LONGITUDINAL DATA WITH INFORMATIVE OBSERVATION AND CENSORING TIMES

Jie Zhou[1], Xingqiu Zhao[2] and Liuquan Sun[3]

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

Abstract: For the analysis of longitudinal data, Liang, Lu, and Ying (Biometrics (2009)) proposed a novel joint model to capture the relation between the longitudinal response process and the observation times through latent variables, and developed an estimation procedure under the assumptions that the distributions of the latent variables are specified and the censoring times are noninformative. This may not be true in practice, and here we propose a new estimation procedure for their model that does not require these assumptions. Estimating equation approaches are developed for parameter estimation, and the resulting estimators are shown to be consistent and asymptotically normal. In addition, some procedures are presented for model selection and model checking. Simulation studies demonstrate that the proposed method performs well and an application to a bladder cancer study is provided.

Key words and phrases: Estimating equations, informative observation and censoring times, joint modeling; latent variables, longitudinal data, model selection.

1. Introduction

Longitudinal data arise frequently in many studies, such as medical follow-up studies and observational investigations. Various methods for analyzing these data have been developed; see Laird and Ware (1982); Diggle, Liang, and Zeger (1994); Lin and Ying (2001); Fitzmaurice, Laird, and Ware (2004); Fan and Li (2004). Diggle, Liang, and Zeger (1994) summarized the commonly used methods including estimating equation and random effect model approaches. Lin and Ying (2001) and Fan and Li (2004) discussed general semiparametric analysis of longitudinal data. All of these methods need a basic assumption that the observation and censoring times are noninformative to the longitudinal response variable.

In many applications, longitudinal processes are subject to nonignorable dropout or informative censoring; this has been considered by Wu and Carroll (1988); Follmann and Wu (1995); Wulfsohn and Tsiatis (1997); Bycott and Taylor (1998); Henderson, Diggle, and Dobson (2000); Wang and Taylor (2001);

Roy and Lin (2002); Lin and Ying (2003); Tsiatis and Davidian (2004); Brown, Ibrahim, and Degruttola (2005); Liu and Ying (2007); Ding and Wang (2008); Li, Hu, and Greene (2009). In these literatures, observation times are assumed to be noninformative, but the response process may still be informed by observation times, even given the covariates. More detailed discussion on this situation can be found in Lin, Scharfstein, and Rosenheck (2004); Sun et al. (2005); Huang, Wang, and Zhang (2006); Ryu et al. (2007); Liang, Lu, and Ying (2009); Zhao, Tong, and Sun (2012). For example, Lin, Scharfstein, and Rosenheck (2004) considered a marginal regression model and proposed a class of inverse intensity-of-visit process-weighted estimators; Sun et al. (2005) proposed a joint model and developed some estimating equation-based estimators; Liang, Lu, and Ying (2009) suggested a joint model via latent variables and proposed an estimating equation based on conditional expectations of the latent variable. All these methods are designed for the situations where either the censoring or observation times are informative, but not both.

A common situation where informative observation and censoring times occur is when times are response variable-dependent. Examples include a bladder cancer study (Byar (1980)) where the occurrence of bladder tumors of a patient may be related to clinical visit times subject to dropout times or death, and a set of longitudinal data from a study of children with acute lymphoblastic leukemia that involves correlated response and observation processes subject to censoring (Lipsitz et al. (2002)). However, there is little limited research on this kind of situation. Thus, Sun, Sun, and Liu (2007) presented a joint model for the longitudinal process, the observation process and the censoring time via a shared latent variable and Liu, Huang, and O'Quigley (2008) proposed a joint random effects model for the longitudinal process, the informative observation times, and a dependent terminal event. It is well known that when the assumption of non-informative observation times or noninformative censoring time is violated, the methods relying on such assumption may yield biased results. The purpose here is to propose a new inference procedure for a class of joint models of longitudinal data with informative observation times as well as informative censoring time. We borrow the joint random effect model for the longitudinal process and the observations times proposed by Liang, Lu, and Ying (2009), and develop a approach that does not rely on the assumptions they require.

The remainder of this paper is organized as follows. Joint modeling of the longitudinal response, the observation time, and the censoring time through a latent variable is presented in Section 2. In Section 3, inference procedures about regression parameters of interest are proposed, and their asymptotic properties are established. In Section 4, we propose a focused information criterion for model selection and discuss the assessment of the models described in Section

2. Some numerical results from simulation studies for evaluating our methods are reported in Section 5. An application of the proposed methodology to the bladder cancer study is presented in Section 6, and some concluding remarks are made in Section 7.

2. Model Specifications

Consider a longitudinal study involving n independent subjects. For subject i, let $Y_i(t)$ denote the longitudinal response process of interest and $X_i(t)$ be the $p \times 1$ vector of possibly time-dependent covariates. In addition, let C_i be the censoring time and $N_i(t)$ the counting process denoting the number of observation times before or at time t. The longitudinal process $Y_i(t)$ is observed only at the time points where $N_i(t)$ jumps for $t \leq C_i$.

Following Liang, Lu, and Ying (2009), we consider a semiparametric mixed random effect model for the response process:

$$Y_i(t) = \mu_0(t) + \beta_0' X_i(t) + u_i' Z_i(t) + \varepsilon_i(t) , \tag{2.1}$$

where $\mu_0(t)$ is an unspecified smooth function of t, β_0 is a vector of unknown regression parameters, $Z_i(t)$ is a q-dimensional subvector of $(1, X_i(t)')'$, u_i is a q-dimensional subject-specific random effects, and $\varepsilon_i(t)$ is a measurement error process. For identifiability of (2.1), the random effects u_i are assumed to have zero mean.

For the observation time process, we assume that, conditional on $X_i(\cdot)$ and a latent variable v_i, $N_i(\cdot)$ is a Poisson process with intensity function

$$d\Lambda_i(t) = v_i \exp\{\gamma_0' W_i\} d\Lambda_0(t) , \tag{2.2}$$

where $\Lambda_0(t)$ is an unspecified baseline cumulative intensity function, W_i is an $r-$dimensional time-independent subvector of $X_i(t)$, and γ_0 is a vector of unknown regression parameters. For identifiability of (2.2), we assume that v_i is nonnegative and has mean 1 conditional on $X_i(\cdot)$.

For the joint modeling and analysis of the longitudinal model (2.1) and the observation time model (2.2), we assume that the association between the two random effects u_i and v_i is formulized as $E(u_i|v_i, X_i(\cdot)) = \theta_0(v_i - 1)$, where θ_0 is a q-dimensional parameter. It is also assumed that the censoring time C_i can depend on u_i, v_i, and $X_i(\cdot)$ in an arbitrary way but, conditional on v_i and $X_i(\cdot)$, $Y_i(\cdot)$, $N_i(\cdot)$ and C_i are mutually independent. In addition, we assume that $E(\varepsilon_i(t)|v_i, X_i(\cdot)) = 0$.

Remark 1. We allow for a unit component in $Z_i(t)$ to make it more general, and then many joint models via latent variables are included (e.g., Sun, Sun, and Liu (2007)). For simplicity, we only consider a frailty model with time-independent covariates in (2.2) for the observation process. It is noteworthy that time-dependent covariates can be included in this model with a more complicated estimation method (Sun, Song, and Zhou (2011)).

Remark 2. The linear relationship between u_i and v_i is assumed here for computational simplicity. In fact, the proposed method can be extended to the case that $E(u_i|v_i, X_i(\cdot)) = f(v_i;\theta_0)$, where $f(v_i;\theta_0)$ is a $q-$dimensional vector with each component a polynomial in v_i.

3. Estimation of Regression Parameters

Our main interest is to estimate β_0. Note that with the assumptions on u_i and $\varepsilon_i(t)$, (2.1) implies that

$$E(Y_i(t)|X_i(\cdot), v_i) = \mu_0(t) + \beta_0' X_i(t) + \theta_0' Z_i(t)(v_i - 1).$$

If v_i can be observed and γ_0 is known, take $X_i^*(t) = (X_i(t)', Z_i(t)(v_i - 1))'$, and

$$\bar{X}^*(t;\gamma) = \frac{\sum_{i=1}^n \Delta_i(t) v_i \exp\{\gamma' W_i\} X_i^*(t)}{\sum_{i=1}^n \Delta_i(t) v_i \exp\{\gamma' W_i\}},$$

where $\Delta_i(t) = I(C_i \geq t)$. Then, following the approach of Lin and Ying (2001), we can estimate β_0 and θ_0 using the estimating equation $U(\beta,\theta;\gamma_0) = 0$, where

$$U(\beta,\theta;\gamma) = n^{-1} \sum_{i=1}^n \int_0^\tau Q(t)[X_i^*(t) - \bar{X}^*(t;\gamma)]\{Y_i(t) - X_i(t)'\beta - (v_i - 1)\theta' Z_i(t)\} \times \Delta_i(t) dN_i(t), \tag{3.1}$$

with the weight function $Q(t)$.

In practice, v_i cannot be observed and γ_0 is unknown. Under (2.2), given the random effect v_i and covariate $X_i(\cdot)$, the observation process is a nonhomogeneous Poisson process. Let m_i denote the total number of observations for subject i before censoring C_i. It follows that, given v_i, $X_i(\cdot)$, and C_i, m_i has a Poisson distribution with mean $v_i \Lambda_0(C_i) e^{\gamma_0' W_i}$. Following Sun, Sun, and Liu (2007), let $F(t) = \Lambda_0(t)/\Lambda_0(\tau)$, $\alpha_1 = \log \Lambda_0(\tau)$ and $\alpha_0 = (\alpha_1, \gamma_0')'$, where τ is the end point of the study. Then $F(t)$ and α_0 can be estimated by

$$\hat{F}(t) = \prod_{t < s \leq \tau} \left(1 - \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n \Delta_i(s) N_i(s)}\right),$$

and the solution to the estimating equation

$$n^{-1} \sum_{i=1}^n W_i^* \Big(\frac{m_i}{\hat{F}(C_i)} - \exp\{\alpha' W_i^*\}\Big) = 0 \tag{3.2}$$

with $W_i^* = (1, W_i')'$, respectively.

Note that $E(m_i|X_i(\cdot), C_i, v_i) = v_i\Lambda_0(C_i)e^{\gamma_0' W_i}$, so it is natural to estimate v_i by

$$\hat{V}_i = \frac{m_i}{\hat{\Lambda}_0(C_i)e^{\hat{\gamma}' W_i}},$$

where $\hat{\Lambda}_0(t) = \hat{F}(t)\exp\{\hat{\alpha}_1\}$. Replacing v_i by $\hat{V}_i$ in (3.1), we obtain a plug-in estimating equation, but it usually provides a biased estimator because such a plug-in estimating equation has a nonzero mean.

Here is an adjustment of the plug-in estimating equation. Take

$$h_k(m) = \prod_{i=1}^{k}(m - i + 1), \quad \tilde{h}_{k+1}(m) = \prod_{i=2}^{k+1}(m - i + 1), \quad \text{for } k \geq 1.$$

It is easy to show that $Eh_k(m) = \lambda^k$ for a Poisson distribution random variable m with mean λ. Note that given $X_i(\cdot), C_i$, and v_i, m_i follows a Poisson distribution with mean $\lambda = v_i\Lambda_0(C_i)e^{\gamma_0' W_i}$, so, $v_i^k = E(h_k(m_i)|X_i(\cdot), C_i, v_i)\{\Lambda_0(C_i)e^{\gamma_0' W_i}\}^{-k}$, $k \geq 1$. Since

$$E(\Delta_i(t)dN_i(t)|X_i(\cdot), v_i, m_i, C_i) = \Delta_i(t)m_i\Lambda_0(C_i)^{-1}d\Lambda_0(t),$$

we have

$$E\left\{\frac{\tilde{h}_{k+1}(m_i)}{\{\Lambda_0(C_i)e^{\gamma_0' W_i}\}^k}\Delta_i(t)dN_i(t) - v_i^k\Delta_i(t)dN_i(t)\Big|X_i(\cdot), v_i\right\} = 0, \quad \text{for} \quad k \geq 1.$$

Motivated by this, we can construct unbiased estimating functions to estimate β_0 and θ_0. Define

$$\begin{aligned}
U_1(\beta, \theta; \Lambda_0, \gamma_0) &= \sum_{i=1}^{n}\int_0^{\tau} Q(t)\{X_i(t) - \bar{X}(t)\}\{Y_i(t) - \beta' X_i(t) - \theta' Z_i(t)(V_{i1} - 1)\} \\
&\quad \times\Delta_i(t)dN_i(t), \\
U_2(\beta, \theta; \Lambda_0, \gamma_0) &= \sum_{i=1}^{n}\int_0^{\tau} Q(t)\Big[\{Z_i(t)(V_{i1} - 1) - \bar{Z}(t)\}\{Y_i(t) - \beta' X_i(t)\} \\
&\quad -\theta' Z_i(t)\{Z_i(t)(V_{i2} - 2V_{i1} + 1) - \bar{Z}(t)(V_{i1} - 1)\}\Big]\Delta_i(t)dN_i(t),
\end{aligned}$$

where $V_{ik} = \tilde{h}_{k+1}(m_i)\{\Lambda_0(C_i)e^{\gamma_0' W_i}\}^{-k}$,

$$\bar{X}(t) = \frac{\sum_{i=1}^{n}\Delta_i(t)m_i\Lambda_0(C_i)^{-1}X_i(t)}{\sum_{i=1}^{n}\Delta_i(t)m_i\Lambda_0(C_i)^{-1}},$$

and

$$\bar{Z}(t) = \frac{\sum_{i=1}^{n} \Delta_i(t) m_i \Lambda_0(C_i)^{-1} Z_i(t)(V_{i1} - 1)}{\sum_{i=1}^{n} \Delta_i(t) m_i \Lambda_0(C_i)^{-1}}.$$

It is easy to show that $E\{U_i(\beta_0, \theta_0; \Lambda_0, \gamma_0)\} = 0$, $i = 1, 2$. Thus, β_0 and θ_0 can be estimated by the estimating function $U(\beta, \theta)$, where $U(\beta, \theta) = (U_1'(\beta, \theta; \hat{\Lambda}_0, \hat{\gamma}), U_2'(\beta, \theta; \hat{\Lambda}_0, \hat{\gamma}))'$. Let $\hat{\beta}$ and $\hat{\theta}$ be solution to $U(\beta, \theta) = 0$. To establish the asymptotic normality of $\hat{\beta}$ and $\hat{\theta}$, we let P_{1n}, P_{2n} and P_{3n} be the empirical distributions of $(X_i, C_i, m_i, T_{i1}, \ldots, T_{i,m_i})$, (X_i, C_i, m_i) and $(X_i, C_i, m_i, Y_i, T_{i1}, \ldots, T_{i,m_i})$, respectively. Also let $\tilde{V}_{ik}$, $\tilde{X}(t)$ and $\tilde{Z}(t)$ be defined in the same way as V_{ik}, $\bar{X}(t)$ and $\bar{Z}(t)$ with Λ_0 and γ_0 replaced by $\hat{\Lambda}_0$ and $\hat{\gamma}$. Let

$$\hat{\mathcal{A}}(t) = \int_0^t \frac{\sum_{i=1}^{n} \{Y_i(u) - \hat{\beta}' X_i(u) - \hat{\theta}' Z_i(u)(\tilde{V}_{i1} - 1)\} dN_i(u)}{\sum_{i=1}^{n} \Delta_i(t) m_i \hat{\Lambda}(C_i)^{-1}},$$

$$\hat{H}(t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} I(T_{ij} \le t), \quad \hat{R}(t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} I(T_{ij} \le t \le C_i),$$

$$\hat{\kappa}_i(t) = \sum_{j=1}^{m_i} \Big\{ \int_t^{\tau} \frac{I(T_{ij} \le u \le C_i) d\hat{H}(u)}{\hat{R}^2(u)} - \frac{I(t < T_{ij} \le \tau)}{\hat{R}(T_{ij})} \Big\},$$

$$\hat{e}_i = W_i^* \Big[\frac{m_i}{\hat{F}(C_i)} - \exp\{\hat{\alpha}' W_i^*\} \Big] - \int \frac{w^* m \hat{\kappa}_i(c) dP_{2n}(w^*, c, m)}{\hat{F}(c)},$$

$$d\hat{M}_i(t) = \big\{ Y_i(t) - \hat{\beta}' X_i(t) - \hat{\theta}' Z_i(t)(\tilde{V}_{i1} - 1) \big\} \Delta_i(t) dN_i(t) \\ -\Delta_i(t) m_i \hat{\Lambda}_0(C_i)^{-1} d\hat{\mathcal{A}}_0(t),$$

and

$$\hat{D}_1 = n^{-1} \sum_{i=1}^{n} \exp\{\hat{\alpha}' W_i^*\} W_i^{*\otimes 2},$$

where $v^{\otimes 2} = vv'$ for a vector v.

Furthermore, let $\hat{\phi}_{1i}$ denote the vector $\hat{D}_1^{-1} \hat{e}_i$ without the first entry and $\hat{\phi}_{2i}$ denote the first entry of $\hat{D}_1^{-1} \hat{e}_i$. Set $\hat{\varphi}_i(t) = \hat{\kappa}_i(t) + \hat{\phi}_{2i}$, $\hat{b}_i(c, w) = \hat{\varphi}_i(c) + \hat{\phi}_{1i}' w$, and $\hat{\xi}_i = (\hat{\xi}_{1i}', \hat{\xi}_{2i})'$, where

$$\hat{\xi}_{1i} = \int_0^{\tau} Q(t) \{X_i(t) - \tilde{X}(t)\} d\hat{M}_i(t) \\ + \int_0^{\tau} Q(t) \Big[\int \{x(t) - \tilde{X}(t)\} \frac{m}{\hat{\Lambda}_0(c)} \hat{\varphi}_i(c) I(c \ge t) dP_{2n}(x, c, m) \Big] d\hat{\mathcal{A}}_0(t) \\ + \int \sum_{l=1}^{m} Q(t_l) \{x(t_l) - \tilde{X}(t_l)\} \frac{\tilde{h}_2(m)}{\hat{\Lambda}_0(c) e^{\hat{\gamma}' w}} \hat{\theta}' z(t_l) \hat{b}_i(c, w) dP_{1n}(x, c, m, t_1, \ldots, t_m),$$

$$
\begin{aligned}
\hat{\xi}_{2i} = & \int_0^\tau Q(t)\{Z_i(t)(\tilde{V}_{i1}-1)-\tilde{Z}(t)\}\Big[\big\{Y_i(t)-\hat{\beta}'X_i(t)\big\}\Delta_i(t)dN_i(t) \\
& -\Delta_i(t)\frac{m_i}{\hat{\Lambda}_0(C_i)}d\hat{A}_0(t)\Big] \\
& -\int_0^\tau Q(t)\hat{\theta}'Z_i(t)\big\{Z_i(t)(\tilde{V}_{i2}-2\tilde{V}_{i1}+1)-\tilde{Z}(t)(\tilde{V}_{i1}-1)\big\}\Delta_i(t)dN_i(t) \\
& +\int_0^\tau Q(t)\Big[\int \frac{m\tilde{h}_2(m)}{\hat{\Lambda}_0(c)^2 e^{\hat{\gamma}'w}} z(t)\hat{b}_i(c,w)I(c\geq t)dP_{2n}(x,m,c)d\hat{\mathcal{A}}_0(t) \\
& +\int_0^\tau Q(t)\Big[\int\{z(t)(\frac{\tilde{h}_2(m)}{\hat{\Lambda}_0(c)e^{\hat{\gamma}'w}}-1)-\tilde{Z}(t)\}\frac{m}{\hat{\Lambda}_0(c)}\hat{\varphi}_i(c)I(c\geq t)dP_{2n}(x,c,m)\Big] \\
& \quad\cdot d\hat{\mathcal{A}}_0(t)-\int\Big[\sum_{u=1}^m Q(t_u)\frac{\tilde{h}_2(m)}{\hat{\Lambda}_0(c)e^{\hat{\gamma}'w}}z(t_u)\big[y(t_u)-\hat{\beta}'x(t_u)\big]\hat{b}_i(c,w)\Big] \\
& \quad\cdot dP_{3n}(x,c,m,y,t_1,\ldots,t_m) \\
& +\int\sum_{u=1}^m Q(t_u)\hat{\theta}'z(t_u)\Big[\frac{2\tilde{h}_3(m)z(t_u)}{\{\hat{\Lambda}_0(c)e^{\hat{\gamma}'w}\}^2}-\frac{\tilde{h}_2(m)(2z(t_u)+\tilde{Z}(t_u))}{\hat{\Lambda}_0(c)e^{\hat{\gamma}'w}}\Big]\hat{b}_i(c,w)dP_{1n}.
\end{aligned}
$$

Theorem 1. *Under the regularity conditions* (R1)$-$(R4) *stated in the Appendix,* $n^{1/2}(\hat{\beta}-\beta_0)$ *and* $n^{1/2}(\hat{\theta}-\theta_0)$ *have asymptotically a joint normal distribution with mean zero and a covariance matrix that can be consistently estimated by* $\hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$*, where*

$$
\hat{A}=\begin{pmatrix}\hat{A}_{11} & \hat{A}_{12}\\ \hat{A}_{12}' & \hat{A}_{22}\end{pmatrix},\qquad \hat{\Sigma}=n^{-1}\sum_{i=1}^n\hat{\xi}_i^{\otimes 2},
$$

$$
\hat{A}_{11}=n^{-1}\sum_{i=1}^n\int_0^\tau Q(t)\{X_i(t)-\tilde{X}(t)\}^{\otimes 2}\Delta_i(t)dN_i(t)\ ,
$$

$$
\hat{A}_{12}=n^{-1}\sum_{i=1}^n\int_0^\tau Q(t)\big\{X_i(t)-\tilde{X}(t)\big\}\big\{Z_i(t)(\tilde{V}_{i1}-1)-\tilde{Z}(t)\big\}\Delta_i(t)dN_i(t)\ ,
$$

$$
\begin{aligned}
\hat{A}_{22}=n^{-1}\sum_{i=1}^n\int_0^\tau Q(t)\Big\{&(\tilde{V}_{i2}-\tilde{V}_{i1}^2)Z_i(t)Z_i'(t)+\{Z_i(t)(\tilde{V}_{i1}-1)-\tilde{Z}(t)\}^{\otimes 2}\Big\}\\
&\cdot\Delta_i(t)dN_i(t)\ ,
\end{aligned}
$$

and $\hat{\xi}_i$ *is as defined above.*

4. Model Selection and Model Checking

In this section, we consider the choice of the random effect covariates and the assessment of the models described in the previous sections.

4.1. Model selection

Suppose that we have a vector of covariates $X_i(t)$ in hand, but it is hard to decide which one to be included in the random effect covariates. In practice, it may be known that some part of $X_i(t)$ does not have random effects, but we are not sure about the rest. Let $Z_i(t)$ be the part of $(1, X_i(t)')'$ which may have random effects. The purpose here is to inculde the right part of $Z_i(t)$ in the model. More specifically, model selection tools are proposed to evaluate how appropriate the setup is for the association between longitudinal outcomes and informative observation and dropout processes, given that $\beta_0' X_i(t)$ is correctly pre-specified in model (2.1). Since our main interest is in estimation of covariate effect β_0, it is natural to develop a model selection method focused in this way. Note that the focused information criterion (FIC, see Claeskens and Hjort (2008, Chap. 6)) serves this purpose very well but the focused parameter needs to be of one-dimension in the literature. We generalize FIC to adapt the current problem. Let

$$\mathrm{FIC(S)} = \sum_{j=1}^{p} nE(\hat{\beta}_{Sj} - \beta_{0j})^2,$$

where $\hat{\beta}_{Sj}$ is the jth component of $\hat{\beta}_S$, the estimator of β_0 under model S. We suggest choosing a model by minimizing FIC(S).

Noting that the model selection procedure considered here does not affect the estimation of parameters in the observation time model, we use the same notation for all models. Thus we denote by $(\hat{\beta}, \hat{\theta})$ the estimators of the full model, and by $(\hat{\beta}_S, \hat{\theta}_S)$ those for model S, where S is a subset of $\{1, 2, \cdots, q\}$, and model S represents the model with random effect covariate $Z_S(t)$, the components of $Z(t)$ with indices belonging to S. Note that when S is the empty set, the model has no random effect covariates. Let Π_S be the projection matrix such that $\Pi_S(\beta', \theta')' = (\beta', \theta_S')'$ and take $A_S = \Pi_S A \Pi_S'$, $\Sigma_S = \Pi_S \Sigma \Pi_S'$, where A and Σ are defined in the Appendix. Let Σ_2 be the q-dimensional matrix in the lower right corner of $A^{-1}\Sigma A^{-1}$, Σ_1^S be the p-dimensional matrix in the top left corner of $A_S^{-1}\Sigma_S A_S^{-1}$, and Q_S be the top p rows of $A_S^{-1}\Pi_S A$. Let model P_n be the nth model, under which (2.1) holds for $i = 1, \cdots, n$, with $E(u_i | v_i, X_i(\cdot)) = (\theta_0 + \delta/\sqrt{n})(v_i - 1)$. Some properties in a local misspecification setting are summarized here.

Theorem 2. *Under* (R1)−(R4)*, we have, under* P_n*,*

$$D_n \equiv \sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(\delta, \Sigma_2),$$
$$\sqrt{n}(\hat{\beta}_S - \beta_0) \Rightarrow N(Q_S\delta, \Sigma_1^S).$$

The results of this theorem suggest an asymptotic evaluation of FIC(S) in the local misspecification setting. Thus, under P_n, FIC(S) is approximately

$$\mathrm{MSE}(S,\delta)=\sum_{j=1}^{p} Q_{Sj}\delta\delta' Q_{Sj}'+tr(\Sigma_1^S),$$

where Q_{Sj} represents the jth row of Q_S and $tr(\Sigma_1^S)$ is the trace of Σ_1^S. Note that $D_nD_n'-\Sigma_2$ is an asymptotically unbiased estimator of $\delta\delta'$. Let $\hat{D}_n=\sqrt{n}\hat{\theta}$, and $\hat{\Sigma}_2$, $\hat{Q}_S$, and $\hat{\Sigma}_1^S$, be consistent estimators of Σ_2, Q_S, and Σ_1^S, respectively. Then FIC(S) can be estimated by

$$\widehat{\mathrm{FIC}}(\mathrm{S})=\sum_{j=1}^{p}\max\{\hat{Q}_{Sj}(\hat{D}_n\hat{D}_n'-\hat{\Sigma}_2)\hat{Q}_{Sj}',0\}+tr(\hat{\Sigma}_1^S).$$

Remark 3. $\hat{\Sigma}_1^S$ can be obtained from $\hat{\Sigma}$ as in the definition of Σ_1^S. Note that the evaluation of $\hat{\Sigma}$ depends on $\hat{\beta}$ and $\hat{\theta}$. Then $\hat{\Sigma}_1^S$ can also be obtained by substituting $\hat{\beta}_S$ and $\hat{\theta}_S$ into the expression of $\hat{\Sigma}$.

Other model selection methods such as the Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) could also be considered, where AIC $=2k/n+\log(RSS/n)$ and BIC $=k\log(n)/n+\log(RSS/n)$, k is the dimension of θ_S, and $RSS=\sum_{i=1}^n \hat{M}_i(C_i)^2$ is the sum of squared residuals. However, these selection methods are not designed for obtaining a good estimate for a focused parameter, and hence they are not appropriate for the purpose of getting a good estimator for β_0. In the simulation section, we will compare the proposed model selection method with the AIC and BIC criteria.

4.2. Model checking

In this subsection, we propose a test statistic for model assessment. To check model (2.2), we can use some discussion and simple approaches of Huang and Wang (2004) for recurrent event data with informative censoring. Here we focus on checking the goodness of fit of model (2.1). Following Lin et al. (2000), we consider the cumulative sums of residuals:

$$\mathcal{F}(t,x)=n^{-1/2}\sum_{i=1}^{n}\int_0^t I(X_i(u)\le x)d\hat{M}_i(u),$$

where the event $\{X_i(u)\le x\}$ means that each of the p components of $X_i(u)$ is no larger than the respective component of x.

Take the null hypothesis H_0 to be the correct specification of model (2.1) under the assumption that the random component $u_i'Z_i(t)$ and model (2.2) are correctly specified. We show in the Appendix that, under H_0, the null distribution of $\mathcal{F}(t,x)$ can be approximated by a zero-mean Gaussian process

$$\tilde{\mathcal{F}}(t,x) = n^{-1/2}\sum_{i=1}^{n}\left[\int_0^t \big\{I(X_i(u)\le x) - \bar{I}(u,x)\big\} d\hat{M}_i(u) + \hat{\psi}_i(t,x) + \hat{\Gamma}(t,x)'\hat{\xi}_i\right], \tag{4.1}$$

where

$$\bar{I}(t,x) = \frac{\sum_{i=1}^n \Delta_i(t) m_i \hat{\Lambda}(C_i)^{-1} I(X_i(t)\le x)}{\sum_{i=1}^n \Delta_i(t) m_i \hat{\Lambda}(C_i)^{-1}},$$

$$\hat{\psi}_i(t,x) = \int_0^t \Big[\int \{I(\underset{\sim}{x}(u)\le x) - \bar{I}(u,x)\}\frac{m}{\hat{\Lambda}_0(c)}\hat{\varphi}_i(c) I(c\ge u) dP_{2n}(\underset{\sim}{x},c,m)\Big] d\hat{\mathcal{A}}_0(u)$$
$$+ \int \sum_{l=1}^{m}\{I(\underset{\sim}{x}(t_l)\le x) - \bar{I}(t_l,x)\}\frac{\tilde{h}_2(m) I(t_l\le t)}{\hat{\Lambda}_0(c) e^{\hat{\gamma}'w}}\hat{\theta}' z(t_l)\hat{b}_i(c,w)$$
$$\cdot dP_{1n}(\underset{\sim}{x},c,m,t_1,\ldots,t_m),$$

$$\hat{\Gamma}(t,x) = \frac{1}{n}\sum_{i=1}^n \int_0^t \big\{I(X_i(u)\le x) - \bar{I}(u,x)\big\}\begin{pmatrix} X_i(u) \\ Z_i(u)(\tilde{V}_{i1}-1)\end{pmatrix} dN_i(u).$$

Note that it is difficult to estimate the asymptotic covariance function of $\mathcal{F}(t,x)$ analytically. We appeal to the resampling approach. Let $(G_1,\ldots,G_n)$ be independent standard normal variables independent of the data. Then it can be shown that the null distribution of $\mathcal{F}(t,x)$ can be approximated by the conditional distribution of

$$\hat{\mathcal{F}}(t,x) = n^{-1/2}\sum_{i=1}^{n}\Big[\int_0^t \big\{I(X_i(u)\le x) - \bar{I}(u,x)\big\} d\hat{M}_i(u) + \hat{\psi}_i(t,x) + \hat{\Gamma}(t,x)'\hat{\xi}_i\Big] G_i.$$

Thus, one can obtain realizations from $\hat{\mathcal{F}}(t,x)$ by repeatedly generating the standard normal random sample $(G_1,\ldots,G_n)$ while fixing the observed data. Since $\mathcal{F}(t,x)$ is expected to fluctuate randomly around 0 under H_0, a formal lack-of-fit test may be constructed based on the supremum statistic $\sup_{t,x}|\mathcal{F}(t,x)|$, with which the p-value can be obtained by comparing the observed value of $\sup_{t,x}|\mathcal{F}(t,x)|$ to a large number of realizations from $\sup_{t,x}|\hat{\mathcal{F}}(t,x)|$.

5. Simulation Studies

Simulation studies were conducted to examine the finite sample properties of the proposed estimators. In the study, the covariates X_{1i} and X_{2i} were generated from a Bernoulli distribution with success probability 0.5 and a uniform distribution $U(0,1)$, respectively. The latent variable v_i followed a gamma distribution with mean 1 and variance 0.5. The censoring time C_i was generated from the minimum of C_i^* and $\tau = 4$, where C_i^* follows $U(1,5)$ or $U(1, 1+v^{-1})$, representing independent or dependent censoring. Given v_i, the observation times were generated from a Poisson process with intensity $cv_i \exp\{0.2X_{1i} - 0.5X_{2i}\}$, with $c = 1.2$ and 2.3 corresponding to the independent and dependent censoring. The average number of observations per subject was about 3 for both cases. The longitudinal response was generated as

$$Y_i(t) = 1 + 0.5t + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i Z_i + \epsilon_i(t),$$

where $Z_i = X_{1i}$, $u_i = \theta(v_i - 1) + N(0,1)$, $\epsilon_i(t)$ is normal with mean ψ_i and standard deviation 0.5 for all t, and ψ_i is a standard normal random variable. We set $\beta_2 = 1$, $\beta_1 = 1, -1$ and $\theta = 1, -1, 0$. All simulations were repeated 1,000 times.

The simulation results for estimation of β_1 and β_2 are reported in Table 1 for two cases of independent and dependent censoring and two sample sizes. Each part in the table includes the biases (BIAS) given by the sample means of proposed estimates minus the true values, the sample standard errors (SSE) of the estimates $\hat{\beta}$, the means of the estimated standard errors (ESE) of $\hat{\beta}$, and the empirical 95% coverage probabilities (CP) for β. It can be seen that the proposed estimation procedures performed well for the situations considered here. Specifically, the biases of the proposed estimators are close to zero, the proposed variance estimation procedure provides good estimates, and the 95% empirical coverage probabilities based on a normal approximation seem reasonable. It is interesting the estimation results seem to be better when the censoring times are related to the response and observation processes. The reason may be that the observation numbers are more stable for the dependent censoring case, since a subject with larger intensity tends to be censored earlier. Other choices for the latent variable, such as a log-normal distribution and a combination of gamma and log-normal distribution, were also considered. The simulation results were similar and are not presented here.

An additional simulation study was conducted for comparison with the methods of Liang, Lu, and Ying (2009) (denoted by LLY) and Lin and Ying (2001) (denoted by LY). We considered the same setups as above. Only the simulation results for $\beta_1 = 1$ are presented in Table 2. Note that LY considered the classic model for the situation where both observation and censoring times are conditionally independent given covariates; LLY considered a more general model that allows for informative observation times when the censoring time is conditionally independent given covariates. The simulation results reveal reasonable performance of the three methods. That is, the proposed method works well for all cases considered, but may lose efficiency when the models of LLY or LY hold; the LLY and LY methods may lead to biased estimates when the corresponding

Table 1. Simulation results for β_1 and β_2.

			β_1				β_2			
β_1	θ	n	BIAS	SSE	ESE	CP	BIAS	SSE	ESE	CP
Independent Censoring										
1	1	200	-0.0560	0.3539	0.3295	0.946	0.0120	0.5406	0.5244	0.937
		300	-0.0275	0.2505	0.2261	0.934	0.0011	0.4116	0.4036	0.931
	-1	200	0.0601	0.3334	0.3112	0.941	0.0120	0.5205	0.5106	0.937
		300	0.0409	0.2449	0.2291	0.940	-0.0047	0.4277	0.4053	0.939
	0	200	-0.0009	0.2673	0.2570	0.934	0.0028	0.4738	0.4356	0.925
		300	-0.0053	0.2100	0.2068	0.937	-0.0201	0.3668	0.3576	0.945
-1	1	200	-0.0685	0.3082	0.2877	0.936	0.0195	0.5296	0.5048	0.931
		300	-0.0329	0.2397	0.2243	0.938	0.0091	0.4259	0.4026	0.934
	-1	200	0.0554	0.3049	0.2872	0.936	0.0226	0.5353	0.5015	0.936
		300	0.0370	0.2341	0.2241	0.946	0.0012	0.4275	0.4008	0.929
	0	200	0.0048	0.2635	0.2553	0.942	-0.0110	0.4562	0.4337	0.933
		300	0.0108	0.2107	0.2068	0.944	0.0127	0.3692	0.3592	0.941
Dependent Censoring										
1	1	200	-0.0280	0.2531	0.2419	0.938	0.0064	0.4634	0.4555	0.943
		300	-0.0272	0.2045	0.1951	0.940	0.0098	0.3798	0.3663	0.938
	-1	200	0.0481	0.2585	0.2444	0.947	-0.0133	0.4794	0.4509	0.929
		300	0.0151	0.1988	0.1934	0.937	-0.0158	0.3766	0.3631	0.934
	0	200	-0.0040	0.2293	0.2231	0.935	0.0076	0.4088	0.3940	0.942
		300	0.0068	0.1816	0.1792	0.945	-0.0101	0.3404	0.3240	0.933
-1	1	200	-0.0491	0.2480	0.2478	0.952	0.0415	0.5461	0.4895	0.951
		300	-0.0176	0.2054	0.1941	0.948	-0.0086	0.3779	0.3649	0.939
	-1	200	0.0337	0.3192	0.2823	0.940	-0.0090	0.4901	0.4816	0.947
		300	0.0321	0.2050	0.1951	0.930	-0.0054	0.3571	0.3669	0.955
	0	200	0.0137	0.2324	0.2237	0.949	-0.0163	0.4214	0.3981	0.938
		300	0.0046	0.1811	0.1785	0.949	0.0085	0.3277	0.3223	0.945

independent conditions are violated. Specifically, as shown in Table 2, the LY estimator seems to be biased when the observation times or the censoring time is informative and LLY estimator seems to be biased when the censoring time is informative. The simulation results for other setups were similar and are not presented here.

We also conducted some simulation studies to evaluate the performance of the proposed model selection method. For comparison, AIC and BIC methods were also considered. The data were generated using the same setups as before, except that here we only took $\beta_1 = 1$, with $\theta = 1, 2, 4$, and 8. The random effect covariate Z_i was initialized to be $(1, X_{1i}, X_{2i})'$. To assess the performance of three model selection methods, we calculated two numbers: the average numbers of zero-estimated coefficients whose true values were zero (labeled as 'Correct'), and the average numbers of zero-estimated coefficients whose true values were non-zero (labeled as 'Incorrect'). Noting that the FIC method was designed for a

Table 2. Simulation results for comparison.

			Independent Censoring				Dependent Censoring			
			β_1		β_2		β_1		β_2	
n	θ	Method	BIAS	SSE	BIAS	SSE	BIAS	SSE	BIAS	SSE
200	1	ZZS	-0.0642	0.3419	0.0162	0.5372	-0.0437	0.2544	0.0024	0.4768
		LLY	-0.0003	0.2662	0.0102	0.4942	-0.1546	0.2550	0.0255	0.4650
		LY	0.4841	0.2934	0.0056	0.5326	0.3308	0.2524	0.0096	0.4631
	-1	ZZS	0.0491	0.3049	-0.0200	0.4994	0.0317	0.2517	-0.0136	0.5611
		LLY	-0.0111	0.2682	-0.0176	0.4779	-0.0195	0.2320	-0.0328	0.4441
		LY	-0.4906	0.2934	-0.0190	0.5293	-0.3478	0.2506	-0.0163	0.4777
	0	ZZS	-0.0132	0.2548	0.0053	0.4369	0.0080	0.2358	-0.0137	0.4000
		LLY	-0.0147	0.2387	0.0019	0.4311	-0.0737	0.2339	-0.0115	0.4015
		LY	-0.0016	0.2500	-0.0019	0.4347	0.0024	0.2273	-0.0219	0.4022
300	1	ZZS	-0.0373	0.2337	0.0101	0.4270	-0.0270	0.1970	-0.0047	0.3895
		LLY	0.0019	0.2150	0.0047	0.4076	-0.1538	0.2026	0.0179	0.3799
		LY	0.5024	0.2419	0.0011	0.4421	0.3429	0.2091	0.0055	0.3813
	-1	ZZS	0.0325	0.2373	0.0045	0.4143	0.0277	0.2072	0.0043	0.3799
		LLY	-0.0041	0.2197	0.0015	0.3904	-0.0033	0.1980	-0.0138	0.3641
		LY	-0.4879	0.2378	-0.0010	0.4201	-0.3333	0.2169	0.0015	0.3953
	0	ZZS	-0.0024	0.2135	0.0133	0.3715	-0.0036	0.1789	0.0232	0.3463
		LLY	-0.0036	0.2077	0.0124	0.3711	-0.0810	0.1838	0.0249	0.3519
		LY	-0.0010	0.1957	0.0091	0.3699	-0.0079	0.1802	0.0197	0.3453

Note: ZZS stands for our proposed estimator; LLY stands for the estimator in Liang, Lu, and Ying (2009); LY stands for the estimator in Lin and Ying (2001).

better estimate of β_0, we also calculated the mean squared errors of the resulting estimator $\hat{\beta}$ under the selected model for each method. The simulation results based on 500 repetitions are reported in Tables 3 and 4. It can be seen from Table 3 that the proposed FIC method tends to select more variables into the model, while those variables that should be included in the model are rarely missed. In addition, although AIC and BIC perform better than the FIC method in terms of the 'Correct' number, their performances are worse than the FIC method with respect to the 'Incorrect' number. Also their estimated 'Incorrect' numbers are away from the true value, zero, and this could lead to serious problems. Furthermore, the results in Table 4 show that the FIC method yielded smaller mean squared errors of $\hat{\beta}$ than AIC and BIC, that is, the FIC method led to a better estimate of β_0. The two evaluations of $\widehat{\text{FIC}}$ gave similar results for all cases under consideration.

6. An Application

We applied the proposed methods to the bladder cancer data that have been analyzed by Sun et al. (2005), Sun, Sun, and Liu (2007) and Liang, Lu, and Ying (2009), among others. This study was conducted by the Veterans

Table 3. Simulation results for model selection.

		Average number of zero coefficients							
		Correct				Incorrect			
n	θ	$\widehat{FIC}_1$	$\widehat{FIC}_2$	AIC	BIC	$\widehat{FIC}_1$	$\widehat{FIC}_2$	AIC	BIC
		Independent Censoring							
200	1	0.990	1.026	1.358	1.584	0.238	0.232	0.504	0.542
	2	0.864	0.912	1.382	1.580	0.080	0.064	0.464	0.480
	4	0.832	0.906	1.472	1.668	0.036	0.040	0.406	0.414
	8	0.732	0.866	1.538	1.678	0.050	0.046	0.402	0.406
300	1	0.822	0.868	1.256	1.546	0.090	0.080	0.404	0.434
	2	0.694	0.720	1.316	1.572	0.014	0.016	0.356	0.364
	4	0.630	0.690	1.400	1.622	0.008	0.008	0.348	0.352
	8	0.610	0.648	1.458	1.648	0.002	0.002	0.282	0.286
		Dependent Censoring							
200	1	0.816	0.960	1.522	1.706	0.196	0.246	0.664	0.694
	2	0.638	0.822	1.532	1.772	0.070	0.084	0.560	0.574
	4	0.560	0.746	1.618	1.760	0.060	0.068	0.572	0.578
	8	0.500	0.696	1.666	1.786	0.020	0.042	0.534	0.542
300	1	0.548	0.640	1.466	1.690	0.074	0.090	0.496	0.538
	2	0.372	0.484	1.636	1.800	0.022	0.020	0.528	0.548
	4	0.360	0.438	1.590	1.744	0.006	0.016	0.512	0.512
	8	0.272	0.388	1.668	1.804	0.006	0.012	0.442	0.450

Note: $\widehat{FIC}_1$ uses $\hat{\beta}$ and $\hat{\theta}$ in the estimation of Σ_1^S; $\widehat{FIC}_2$ uses $\hat{\beta}_S$ and $\hat{\theta}_S$ in the estimation of Σ_1^S (see Remark 3). The true numbers for 'Correct' and 'Incorrect' are 2 and 0, respectively.

Table 4. Mean squared errors for $\hat{\beta}$ resulted from different model selection methods.

		Independent Censoring				Dependent Censoring			
n	θ	$\widehat{FIC}_1$	$\widehat{FIC}_2$	AIC	BIC	$\widehat{FIC}_1$	$\widehat{FIC}_2$	AIC	BIC
200	1	0.387	0.383	0.460	0.452	0.266	0.272	0.328	0.333
	2	0.592	0.556	0.932	0.938	0.422	0.422	0.599	0.590
	4	1.589	1.578	2.645	2.701	1.099	1.136	1.751	1.777
	8	5.599	5.211	8.672	8.889	3.633	3.699	6.629	6.651
300	1	0.258	0.252	0.324	0.333	0.192	0.185	0.228	0.230
	2	0.373	0.351	0.619	0.612	0.276	0.284	0.458	0.470
	4	1.132	1.088	1.928	1.931	0.692	0.664	1.397	1.409
	8	3.300	3.147	6.399	6.370	2.216	2.041	4.803	4.857

Note: $\widehat{FIC}_1$ uses $\hat{\beta}$ and $\hat{\theta}$ in the estimation of Σ_1^S; $\widehat{FIC}_2$ uses $\hat{\beta}_S$ and $\hat{\theta}_S$ in the estimation of Σ_1^S (see Remark 3).

Administration Cooperative Urological Research Group. At the beginning of the study, the patients were randomly assigned to placebo and thiotepa treatment groups. For each patient, the observed information includes the clinical visit or observation times (in month), and the number of bladder tumors that occurred

Table 5. Model selection for the bladder tumor data.

Indicator	$\widehat{FIC}_1$	$\widehat{FIC}_2$	AIC	BIC	Indicator	$\widehat{FIC}_1$	$\widehat{FIC}_2$	AIC	BIC
(0 0 0)	0.5241	0.4324	1.5906	1.5906	(1 0 0)	0.5363	0.4920	1.5796	1.6083
(0 0 1)	0.5922	0.5950	1.5861	1.6148	(1 0 1)	0.7131	0.7237	1.6142	1.6717
(0 1 0)	0.4910*	0.4232*	1.5265*	1.5552*	(1 1 0)	0.6633	0.6563	1.6441	1.7016
(0 1 1)	9.7263	8.7640	1.5430	1.6005	(1 1 1)	0.7895	0.7895	1.6267	1.7129

Note: 'Indicator' is for inclusion of $(1, X_{1i}, X_{2i})$ in Z_i, for example, (0 0 1) means only X_{2i} is included in Z_i. $\widehat{FIC}_1$ uses $\hat{\beta}$ and $\hat{\theta}$ in the estimation of Σ_1^S; $\widehat{FIC}_2$ uses $\hat{\beta}_S$ and $\hat{\theta}_S$ in the estimation of Σ_1^S (see Remark 3). "*" corresponds to the minimum value.

between clinical visits. The data include 85 bladder cancer patients, 47 in the placebo group and 38 in the thiotepa treatment group. Two baseline covariates were measured: the number of initial tumors before entering the study and the size of the largest initial tumor. Here we focus on the effects of thiotepa treatment and the number of initial tumors on the tumor recurrence process in the presence of both informative observation times and a dependent terminal event.

For the analysis, we take $Y_i(t)$ as the natural logarithm of the number of observed tumors at time t plus 1 to avoid 0, $i = 1, \ldots, 85$. Let $X_{i1} = 1$ if the patient was in the thiotepa group and 0 if the patient was in the placebo group, and X_{i2} to be the logarithm of the number of the initial tumors plus 1. Let τ be the longest observation time (being 53 months). To choose the random effect covariate Z_i, we applied the model selection methods proposed in Section 4 with an initial choice of $Z_i = (1, X_{i1}, X_{i2})'$. The values of $\widehat{\mathrm{FIC}}(S)$, AIC, and BIC for different submodels S are presented in Table 5, and all of the methods suggested $Z_i = X_{i1}$. The application of the proposed method in Section 3 with $Q(t) = 1$ yielded $\hat{\beta}_1 = -0.1451$ and $\hat{\beta}_2 = 0.1958$ with the estimated standard errors of 0.0482 and 0.0515, respectively. These results imply that both the thiotepa treatment and initial number of tumors have significant effects on the tumor occurrence process. In particular, the thiotepa treatment significantly reduced the bladder tumor occurrence rate, and the patients with the higher number of initial tumors tend to have a higher tumor occurrence rate. In addition, the clinical visit process seems to be related to the thiotepa treatment, but not to the initial number of tumors. Moreover $\hat{\theta} = -0.1373$, with estimated standard error 0.0703, shows that the tumor recurrence process and the observation process were significantly negatively associated. These results are consistent with those obtained by Sun et al. (2005) and Liang, Lu, and Ying (2009).

The comparison of of our approach with LY's and LLY's methods is summarized in Table 6. The LY estimate for the treatment effect is significantly overestimated when compared to the other two approaches. The LLY estimate and ours agree with each other; this can be explained by the fact that the censoring time may be noninformative in this study.

Table 6. Estimation results with the bladder tumor data.

	β_1		β_2		θ	
Method	Estimate	ESE	Estimate	ESE	Estimate	ESE
ZZS	-0.145	0.048	0.196	0.052	-0.137	0.0703
LLY	-0.127	0.051	0.190	0.051	-0.091	0.037
LY	-0.182	0.046	0.189	0.050	–	–

We also applied the model checking techniques presented in Section 5 to assess the adequacy of model (2.1) for the bladder cancer data. We calculated the statistic $\mathcal{F}(x,t)$ and found $\sup_{x,t}|\mathcal{F}(x,t)| = 1.4488$ with p-value of 0.959, based on 1,000 realizations, indicating that model (2.1) fits the data well.

7. Concluding Remarks

We have proposed a joint modeling approach for analyzing longitudinal data via latent variables when both observation times and censoring times are informative. The joint models are more flexible in the sense that the distributions of the latent variables are left unspecified. An estimating equation approach was proposed for parameter estimation, which yields consistent and asymptotically normal estimators. We also provided a focused information criterion for model selection and an assessment of model checking. Our estimation procedure can be easily implemented. Simulation results suggest that the proposed estimation approach performs well, and an illustrative example was provided.

In the joint models, we have assumed that $E(u_i|v_i, X_i(\cdot)) = \theta_0(v_i - 1)$, a linear form, see Section 5. In fact, as long as $E(u_i|v_i, X_i(\cdot))$ is a polynomial in v_i, unbiased estimating equations can be constructed. The estimation procedure can be extended to this case easily. Noting that polynomials can be used to approximate continuous functions, this extension is useful, but a high order of the polynomial may lead to something unstable. The simple linear form may be a good choice for small or moderate sample sizes.

For the model selection, the traditional focused information criterion is based on the maximum likelihood estimation. Here we extended it to the estimating equation-based approach. Further studies are needed.

Acknowledgement

The authors would like to thank the Editor, an associate editor, and two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. The second author's research was partly supported by the Research Grant Council of Hong Kong (PolyU 5032/09P). The third author's research was partly supported by the National Natural Science Foundation of China Grants (No. 11171330, 10971015 and 11021161) and Key Laboratory of RCSDS, CAS (No. 2008DP173182).

Appendix : Proofs of Asymptotic Results

We use the notation of the text, and all limits are taken as $n \to \infty$. Let $\bar{x}(t)$ and $\bar{z}^*(t)$ be the limit of $\tilde{X}(t)$ and $\tilde{Z}(t)$, respectively. Write $Z_i^*(t) = Z_i(t)(v_i - 1)$.

To study the asymptotic distributions of $\hat{\beta}$ and $\hat{\theta}$, we need the following regularity conditions.

(R1) $P(C \geq \tau, v > 0) > 0$, $P(C > \tau_\delta) = 1$, where $\tau_\delta = \inf\{t : \Lambda_0(t) > \delta\}$ for some $\delta > 0$, and $E\{N(\tau)^2\} < \infty$.

(R2) $G(t) = E\{vI(C \geq t)\exp(\gamma_0' W)\}$ is a continuous function for $t \in [0, \tau]$.

(R3) The weight function $Q(t)$ has bounded variation and converges to a deterministic function $q(t)$ in probability uniformly in $t \in [0, \tau]$;

(R4) A is nonsingular, where $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}$,

$$A_{11} = E\Big\{ \int_0^\tau q(t)\{X_i(t) - \bar{x}(t)\}^{\otimes 2} \Delta_i(t) dN_i(t) \Big\},$$
$$A_{22} = E\Big\{ \int_0^\tau q(t)\{Z_i^* - \bar{z}^*(t)\}^{\otimes 2} \Delta_i(t) dN_i(t) \Big\},$$
$$A_{12} = E\Big\{ \int_0^\tau q(t)\{X_i(t) - \bar{x}(t)\}\{Z_i^* - \bar{z}^*(t)\} \Delta_i(t) dN_i(t) \Big\}.$$

Define $R(t) = G(t)\Lambda_0(t)$, $H(t) = \int_0^t G(u) d\Lambda_0(u)$, $D_1 = E\{\exp\{\alpha_0' W_i^*\} W_i^{*\otimes 2}\}$,

$$\kappa_i(t) = \sum_{j=1}^{m_i} \Big\{ \int_t^\tau \frac{I(T_{ij} \leq u \leq C_i) dH(u)}{R^2(u)} - \frac{I(t < T_{ij} \leq \tau)}{R(T_{ij})} \Big\},$$
$$e_i = W_i^* \Big[\frac{m_i}{F(C_i)} - \exp\{\alpha_0' W_i^*\} \Big] - \int \frac{w^* m \kappa_i(c) dP_1(w^*, c, m)}{F(c)}.$$

where $P_1(w^*, c, m)$ is the joint probability measure of (W_i^*, C_i, m_i). Let ϕ_{1i} denote the vector $D_1^{-1} e_i$ without the first entry and ϕ_{2i} denote the first entry of $D_1^{-1} e_i$. Set $\varphi_i(t) = \kappa_i(t) + \phi_{2i}$, and $b_i(c, w) = \varphi_i(c) + \phi_{1i}' w$.

Proof of Theorem 1. Under (R1) and (R2), it follows from Wang and Taylor (2001) that

$$n^{1/2}\{\hat{\Lambda}_0(t) - \Lambda_0(t)\} = n^{-1/2} \Lambda_0(t) \sum_{i=1}^n \varphi_i(t) + o_p(1), \tag{A.1}$$

$$n^{1/2}\{\hat{\gamma} - \gamma_0\} = n^{-1/2} \sum_{i=1}^n \phi_{1i} + o_p(1). \tag{A.2}$$

If

$$dM_i(t) = \big\{Y_i(t) - \beta_0' X_i(t) - \theta_0' Z_i(t)(V_{i1} - 1)\big\}\Delta_i(t) dN_i(t) - \Delta_i(t) m_i \Lambda_0(C_i)^{-1} d\mathcal{A}_0(t),$$

then $M_i(t)$ is a zero-mean process. Hence, using the functional version of the Law of Large Numbers and Lemma A.1 of Lin and Ying (2001), we get

$$\begin{aligned} & n^{-1/2} U_1(\beta_0, \theta_0; \hat{\Lambda}_0, \hat{\gamma}) \\ & = n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\{X_i(t) - \bar{x}(t)\} dM_i(t) \\ & \quad - n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\{X_i(t) - \bar{x}(t)\}\big\{\tilde{V}_{i1} - V_{i1}\big\}\theta_0' Z_i(t)\Delta_i(t) dN_i(t) \\ & \quad - n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\{X_i(t) - \bar{x}(t)\} m_i \big\{\hat{\Lambda}_0(C_i)^{-1} - \Lambda_0(C_i)^{-1}\big\}\Delta_i(t) d\mathcal{A}_0(t) + o_p(1). \end{aligned} \tag{A.3}$$

Using (A.1), (A.2), and a Taylor series expansion, we have

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\{X_i(t) - \bar{x}(t)\}\{\tilde{V}_{i1} - V_{i1}\}\theta_0' Z_i(t)\Delta_i(t) dN_i(t) \\ & = -n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\{X_i(t) - \bar{x}(t)\} \frac{\tilde{h}_2(m_i)}{\Lambda_0(C_i) e^{\gamma_0' W_i}} \\ & \quad \times \Big[(\hat{\gamma} - \gamma_0)' W_i + \Lambda_0(C_i)^{-1}\{\hat{\Lambda}_0(C_i) - \Lambda_0(C_i)\}\Big]\theta_0' Z_i(t)\Delta_i(t) dN_i(t) + o_p(1) \\ & = -n^{-1/2} \sum_{i=1}^{n} \int \sum_{l=1}^{m} q(t_l)\{x(t_l) - \bar{x}(t_l)\} \frac{\tilde{h}_2(m)\theta_0' z(t_l)}{\Lambda_0(c) e^{\gamma_0' w}} b_i(c, w) \\ & \quad dP_1(x, c, m, t_1, \ldots, t_m) + o_p(1), \end{aligned} \tag{A.4}$$

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\{X_i(t) - \bar{x}(t)\} m_i \{\hat{\Lambda}_0(C_i)^{-1} - \Lambda_0(C_i)^{-1}\}\Delta_i(t) d\mathcal{A}_0(t) \\ & = -n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\{X_i(t) - \bar{x}(t)\} \frac{m_i}{\Lambda_0(C_i)^2}\{\hat{\Lambda}_0(C_i) - \Lambda_0(C_i)\}\Delta_i(t) d\mathcal{A}_0(t) + o_p(1) \\ & = -n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} q(t)\Big[\int \{x(t) - \bar{x}(t)\} \frac{m}{\Lambda_0(c)} \varphi_i(c) I(c \geq t) dP_2(x, c, m)\Big] d\mathcal{A}_0(t) \\ & \quad + o_p(1), \end{aligned} \tag{A.5}$$

where $P_1(x, c, m, t_1, \ldots, t_m)$ and $P_2(x, c, m)$ is the joint probability measure of $(X_i, C_i, m_i, T_{i1}, \ldots, T_{i,m_i})$ and (X_i, C_i, m_i), respectively. Combining (A.3)–(A.5), we obtain

$$n^{-1/2}U_1(\beta_0,\theta_0) = n^{-1/2}\sum_{i=1}^{n}\xi_{1i} + o_p(1), \tag{A.6}$$

where

$$\begin{aligned}\xi_{1i} =& \int_0^\tau q(t)\{X_i(t)-\bar{x}(t)\}dM_i(t)\\ &+\int_0^\tau q(t)\Big[\int\{x(t)-\bar{x}(t)\}\frac{m}{\Lambda_0(c)}\varphi_i(c)I(c\geq t)dP_2(x,c,m)\Big]d\mathcal{A}_0(t)\\ &+\int\sum_{l=1}^{m}q(t_l)\{x(t_l)-\bar{x}(t_l)\}\frac{\tilde{h}_2(m)}{\Lambda_0(c)e^{\gamma_0' w}}\theta_0' z(t_l)b_i(c,w)dP_1(x,c,m,t_1,\ldots,t_m).\end{aligned}$$

Following similar arguments as in the proof of (A.6), we obtain

$$\begin{aligned}&n^{-1/2}U_2(\beta_0,\theta_0;\hat{\Lambda}_0,\hat{\gamma})\\ &= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau q(t)\Big[\{Z_i(t)(V_{i1}-1)-\bar{z}(t)\}\{Y_i(t)-\beta_0'X_i(t)\}\\ &\quad -\theta_0'Z_i(t)\{Z_i(t)(V_{i2}-2V_{i1}+1)-\bar{z}(t)(V_{i1}-1)\}\Big]\Delta_i(t)dN_i(t)\\ &\quad +n^{-1/2}\sum_{i=1}^{n}\int_0^\tau q(t)\big\{\tilde{V}_{i1}-V_{i1}\big\}Z_i(t)\big[Y_i(t)-\beta_0'X_i(t)\big]\Delta_i(t)dN_i(t)\\ &\quad -n^{-1/2}\sum_{i=1}^{n}\int_0^\tau q(t)\theta_0'Z_i(t)Z_i(t)\big\{\tilde{V}_{i2}-V_{i2}-2(\tilde{V}_{i1}-V_{i1})\big\}'Z_i(t)\Delta_i(t)dN_i(t)\\ &\quad +n^{-1/2}\sum_{i=1}^{n}\int_0^\tau q(t)\bar{z}(t)\theta_0'Z_i(t)\big\{\tilde{V}_{i1}-V_{i1}\big\}\Delta_i(t)dN_i(t)\\ &\quad -n^{-1/2}\sum_{i=1}^{n}\int_0^\tau q(t)\{\tilde{Z}(t)-\bar{z}(t)\}\big[Y_i(t)-\beta_0'X_i(t)-\theta'Z_i(t)(V_{i1}-1)\big]\Delta_i(t)dN_i(t)\\ &\quad +o_p(1)\\ &= n^{-1/2}\sum_{i=1}^{n}\xi_{2i}+o_p(1), \end{aligned}\tag{A.7}$$

where

$$\begin{aligned}\xi_{2i} =& \int_0^\tau q(t)\{Z_i(t)(V_{i1}-1)-\bar{z}(t)\}\Big[\{Y_i(t)-\beta_0'X_i(t)\}dN_i(t)-\frac{m_i}{\Lambda_0(C_i)}\Delta_i(t)d\mathcal{A}_0(t)\Big]\\ &-\int_0^\tau q(t)\big\{\theta_0'Z_i(t)\{Z_i(t)(V_{i2}-2V_{i1}+1)-\bar{z}(t)(V_{i1}-1)\}\Delta_i(t)dN_i(t)\\ &+\int_0^\tau q(t)\int\frac{m\tilde{h}_2(m)}{\Lambda_0(c)^2e^{\gamma_0' w}}z(t)b_i(c,w)I(c\geq t)dP_2(x,m,c)d\mathcal{A}_0(t)\end{aligned}$$

$$+\int_0^\tau q(t)\Big[\int \{z(t)(\frac{\tilde{h}_2(m)}{\Lambda_0(c)e^{\gamma_0' w}}-1)-\bar{z}(t)\}\frac{m}{\Lambda_0(c)}\varphi_i(c)I(c\geq t)dP_2(x,c,m)\Big] d\mathcal{A}_0(t)$$
$$-\int\Big[\sum_{u=1}^m q(t_u)\frac{\tilde{h}_2(m)}{\Lambda_0(c)e^{\gamma_0' w}}z(t_u)\big[y(t_u)-\beta_0' x(t_u)\big]b_i(c,w)\Big] dP_3(x,c,m,y,t_1,\ldots,t_m)$$
$$+\int\sum_{u=1}^m q(t_u)\theta_0' z(t_u)\Big[\frac{2z(t_u)\tilde{h}_3(m)}{\{\Lambda_0(c)e^{\gamma_0' w}\}^2}-\frac{(2z(t_u)+\bar{z}(t_u))\tilde{h}_2(m)}{\Lambda_0(c)e^{\gamma_0' w}}\Big]b_i(c,w) dP_1(x,c,m,t_1,\ldots,t_m).$$

Thus, by (A.6), (A.7) and the Multivariate Central Limit Theorem , $n^{-1/2}U(\beta_0, \theta_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\Sigma = E\xi_i^{\otimes 2}$, where $\xi_i = (\xi_{1i}', \xi_{2i}')'$. Note that $-n^{-1}\partial U(\beta_0,\theta_0)/\partial(\beta',\theta')$ converges in probability to A as defined in (R4). A Taylor expansion of $U(\hat{\beta},\hat{\theta})$ at $U(\beta_0,\theta_0)$ gives

$$n^{1/2}\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\theta}-\theta_0\end{pmatrix} = A^{-1}n^{-1/2}U(\beta_0,\theta_0)+o_p(1). \tag{A.8}$$

Thus, $n^{1/2}(\hat{\beta}-\beta_0)$ and $n^{1/2}(\hat{\theta}-\theta_0)$ have asymptotically a joint normal distribution with mean zero and covariance matrix $A^{-1}\Sigma A^{-1}$.

Proof of Theorem 2. If

$$dM_i(t,\delta) = \Big\{Y_i(t)-\beta_0' X_i(t)-\Big(\theta_0+\frac{\delta}{\sqrt{n}}\Big)' Z_i(t)(V_{i1}-1)\Big\}dN_i(t) -\Delta_i(t)m_i\Lambda_0(C_i)^{-1}d\mathcal{A}_0(t),$$

then $M_i(t,\delta)$ is a zero-mean process under P_n. Hence, using the functional version of the Law of Large Numbers and Lemma A.1 of Lin and Ying (2001), we get

$$\begin{aligned}
&n^{-1/2}U_1(\beta_0,\theta_0;\hat{\Lambda}_0,\hat{\gamma})\\
&=n^{-1/2}\sum_{i=1}^n\int_0^\tau q(t)\{X_i(t)-\bar{x}(t)\}dM_i(t,\delta)+A_{12}\delta\\
&\quad -n^{-1/2}\sum_{i=1}^n\int_0^\tau q(t)\{X_i(t)-\bar{x}(t)\}\big\{\tilde{V}_{i1}-V_{i1}\big\}\theta_0' Z_i(t)\Delta_i(t)dN_i(t)\\
&\quad -n^{-1/2}\sum_{i=1}^n\int_0^\tau q(t)\{X_i(t)-\bar{x}(t)\}m_i\big\{\hat{\Lambda}_0(C_i)^{-1}-\Lambda_0(C_i)^{-1}\big\}\Delta_i(t)d\mathcal{A}_0(t)+o_p(1).
\end{aligned}$$

The last two terms have the same approximations as in the proof of Theorem 1. Hence

$$n^{-1/2}U_1(\beta_0, \theta_0; \hat{\Lambda}_0, \hat{\gamma}) = n^{-1/2}\sum_{i=1}^{n}\xi_{1i}(\delta) + A_{12}\delta + o_p(1),$$

where $\xi_{1i}(\delta)$ is defined in the same way as ξ_{1i} except that $dM_i(t)$ is replaced by $dM_i(t, \delta)$. Similarly, we have

$$n^{-1/2}U_2(\beta_0, \theta_0; \hat{\Lambda}_0, \hat{\gamma}) = n^{-1/2}\sum_{i=1}^{n}\xi_{2i}(\delta) + A_{22}\delta,$$

where $\xi_{2i}(\delta)$ is defined in the same way as ξ_{2i} except that the second term is replaced by

$$\int_0^\tau q(t)\Big(\theta_0 + \frac{\delta}{\sqrt{n}}\Big)' Z_i(t)\{Z_i(t)(V_{i2} - 2V_{i1} + 1) - \bar{z}(t)(V_{i1} - 1)\}\Delta_i(t)dN_i(t).$$

The proof can be completed by using a Taylor expansion and Theorem 2.8.10 (the Functional Central Limit Theorem) of van der Vaart and Wellner (1996).

Proof of (4.1). Write

$$\begin{aligned}
\mathcal{F}(t,x) &= n^{-1/2}\sum_{i=1}^{n}\int_0^t I(X_i(u) \le x)d\hat{M}_i(u) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^t \big\{I(X_i(u) \le x) - \bar{I}(u,x)\big\}d\hat{M}_i(u) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^t \big\{I(X_i(u) \le x) - \bar{I}(u,x)\big\}\big\{Y_i(t) - \hat{\beta}'X_i(t) \\
&\quad -\hat{\theta}'Z_i(t)(\tilde{V}_{i1} - 1)\big\}\Delta_i(t)dN_i(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^t \big\{I(X_i(u) \le x) - \bar{I}(u,x)\big\}\big\{Y_i(t) - \beta_0'X_i(t) \\
&\quad -\theta_0'Z_i(t)(\tilde{V}_{i1} - 1)\big\}\Delta_i(t)dN_i(t) - \hat{\Gamma}(t,x)'\sqrt{n}((\hat{\beta} - \beta_0)', (\hat{\theta} - \theta_0)')'.
\end{aligned}$$

Then the approximation of the null distribution of $\mathcal{F}(t,x)$ follows from arguments similar to those used in the proof of the asymptotic approximation of $n^{-1/2}U_1(\beta_0, \theta_0; \hat{\Lambda}_0, \hat{\gamma})$.

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.

Bycott, P. and Taylor, J. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statist. Medicine* **17**, 2061-2077.

Byar, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology* (Edited by M. Pavane-Macaluso, P. H. Smith, F. Edsmyr), 363-370. Plenum, New York.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging.* Cambridge University Press, Cambridge.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *The Analysis of Longitudinal Data.* Oxford University Press, Oxford, England.

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

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.* **99**, 710-723.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2004). *Applied Longitudinal Analysis.* John Wiley and Sons, New York.

Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151-168.

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **4**, 465-480.

Huang, C. Y. and Wang, M. C. (2004). Joint modeling and estimation of recurrent event processes and failure time. *J. Amer. Statist. Assoc.* **99**, 1153-1165.

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

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

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, 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. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 103-126.

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

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711-730.

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.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R. and Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome- dependent follow-up. *Biometrics* **58**, 621-630.

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.

Roy, J. and Lin, X. (2002). Analysis of multivariate longitudinal outcomes with nonignorable dropouts 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.

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. and Zhou, J. (2011). Regression analysis of longitudinal data with time-dependent covariates in the presence of informative observation and censoring times. *J. Statist. Plann. Inference* **141**, 2902-2919.

Tsiatis, A. and Davidian, M. (2004). Joint modelling of longitudinal and time-to-event data: An overview. *Statist. Sinica* **14**, 809-834.

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

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.

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

School of Mathematical Sciences, Capital Normal University, Beijing 100048, China.

E-mail: zhoujie@amss.ac.cn

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.

E-mail: xingqiu.zhao@polyu.edu.hk

Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100190, China.

E-mail: slq@amt.ac.cn

(Received November 2011; accepted May 2012)