Statistica Sinica Preprint No: SS-2022-0324	
Title	Semiparametric Inference for Longitudinal Data with Informative Observation Times and Terminal Event
Manuscript ID	SS-2022-0324
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0324
Complete List of Authors	Shirong Deng, Kin-yat Liu, Wen Su and Xingqiu Zhao
Corresponding Authors	Xingqiu Zhao
E-mails	xingqiu.zhao@polyu.edu.hk
Notice: Accepted version subject to English editing.	

Semiparametric Inference for Longitudinal Data with Informative Observation Times and Terminal Event

Shirong Deng, Kin-yat Liu, Wen Su and Xingqiu Zhao

Wuhan University, The Chinese University of Hong Kong,

City University of Hong Kong and The Hong Kong Polytechnic University

Abstract: In many longitudinal studies, irregularly repeated measures are often correlated with observation times. Also, there may exist a dependent terminal event such as death that stops the follow-up and is subject to right censoring. To deal with such complex data, we propose a class of flexible semiparametric marginal conditional mean models for longitudinal response processes. The new models include the interaction between the observation history and some covariates, and an unknown functional form of the length from the observation time to the terminal event time, while leaving the within-subject dependence structure of the response process and patterns of the observation process to be arbitrary. For estimation of both scalar and functional parameters in the proposed models, we develop a two-stage spline-based least squares estimation approach and establish the asymptotic properties of the proposed estimators. The performance of the proposed estimation procedure is examined by simulation studies, and a longitudinal data example is provided for illustration.

Key words and phrases: Conditional modeling; Empirical process; Informative observation times; Longitudinal data; Terminal event time.

1. Introduction

Longitudinal data occur frequently in a wide variety of settings, including epidemiological studies, clinical trials, economic applications and others. The response variables and covariates are observed repeatedly at irregular time points for different subjects under study, and the observations are independent among different subjects and may be correlated within each subject. For the analysis of such longitudinal data, various parametric and semiparametric methods have been studied by Laird and Ware (1982), Liang and Zeger (1986), Zeger and Diggle (1994) among others, and excellent reviews have been provided by Lin and Ying (2001) and Diggle et al. (2002).

A basic assumption behind the above methods is that the observation times are independent of response variable, completely or given covariates. However, such an assumption can be violated in many applications, that is, the observation times are informative to the longitudinal responses. An example can be found in the longitudinal CD4 lymphocyte counts of didanosine/zalcitabine study conducted by the Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA) (Abrams, et al., 1994; Goldman, et al. 1996). One phenomenon from some preliminary analysis is that some patients who were too ill for testing gave less visiting times, and thus, they tended to have less CD4 lymphocyte baseline counts on average, that is, the response of CD4 counts may be associated with the

observation times. Thus it is desirable to take into account these informative observation times when we perform the analysis of longitudinal data. To investigate this problem, two methods have been developed. One is the conditional modeling approach (e.g., Sun et al., 2005; Zhao et al., 2014), which directly characterized the dependence between the response and the observation times. Another one is the frailty-based approach (e.g., Sun et al., 2007; Liang et al., 2009; Zhao et al., 2012; Deng and Zhao, 2019), which used the frailties to represent the correlations between the response and the observation times.

In many longitudinal studies, especially the studies for populations with fatal diseases, there may exist a dependent terminal event such as death that stops the follow-up. Two types of approaches are widely used for longitudinal data analysis with dependent terminal events and non-informative observation times. One is the joint modeling approach (Wang and Taylor, 2001; Roy and Lin 2002; Lin and Ying 2003, among others), which used the shared random effects to indirectly model the correlations between the longitudinal response and terminal event. Another is the conditional modeling approach proposed by Kong et al. (2018), which treated the terminal event time as a covariate in a conditional model for the longitudinal response.

Furthermore, both observation times and terminal events may affect longitudinal processes. For example, in the aforementioned ddI/ddC study, the patients

in poorer health with less CD4 testing had slightly lower baseline CD4 counts on average. 40% patients died during the follow-up period, whose baseline CD4 counts were lower, on average, than those who survived (Abrams, et al., 1994; Goldman, et al. 1996). Joint modeling methods have been developed by Liu et al. (2008), Sun et al. (2012), Han et al. (2014), and others to analyze longitudinal data of this nature. These methods incorporate latent variables in models for the longitudinal response, observation times, and terminal events, thereby capturing the relationships among these variables. However, it is worth noting that these existing methods may not explicitly capture the relationship between the longitudinal response and observation times, as well as the relationship between the longitudinal response and the terminal event.

To explore a direct evaluation of the impact of observation times and a terminal event on the longitudinal response process, we propose a class of flexible semiparametric marginal conditional mean model for the longitudinal response that treat both the terminal event time and observation history as covariates, while leaving the within-subject dependence structure of the response process and patterns of the observation process to be arbitrary. Specifically, an unknown functional form for the length from the observation times to the terminal event time is assumed in the proposed model since the influence of the terminal event time on the longitudinal response would have different forms when the observation times have

different distances to the terminal event time. For example, in the ddI/ddC study, the average CD4 counts rose during the first 2 months in the ddI group but fell in the ddC group, and the counts tended to fall in the ddI group but appeared to rise in the ddC group after 2 months until the terminal event time (Abrams, et al., 1994; Goldman, et al. 1996). In addition, the interaction between the observation history and some covariates is considered in the model since the relation between the observation and response processes may vary with some covariates. For example, in the ddI/ddC study, the average CD4 counts changes from baseline are different between the ddI group and ddC group with different observation testing times. This indicates that the patients' CD4 counts and observation times are related with the treatment (Abrams, et al., 1994; Goldman, et al. 1996). This conditional modeling provides a more intuitive and meaningful interpretation while displaying the functional dependence of the response variable on the terminal event time for the observation times that have different distances to the terminal event time. For the estimation of regression parameters and nonparametric function in the proposed models, we develop a two-stage spline-based least squares estimation approach, where the nuisance conditional distribution function for the terminal event time is estimated in the first stage, and the least square loss function based on spline approximation given the nuisance parameters is minimized in the second stage.

The remainder of this paper is organized as follows. We begin in Section 2 by introducing notation and describing the semiparametric marginal conditional mean model for the longitudinal response with informative observation times and dependent terminal event. In Section 3, a two-stage spline-based least squares estimation approach is developed to estimate regression parameters and nonparametric function in the proposed models. The asymptotic properties including consistency and rate of convergence for regression parameters and nonparametric function estimators, and the asymptotic normality for regression parameter estimator are established in Section 4. The simulation results are presented in Section 5 to assess the finite-sample performance of the proposed inference procedure. Also, comparisons between the proposed method and the two-stage approach proposed by Kong et al. (2018) are conducted to illustrate the robustness of the proposed method. A real example of longitudinal data is provided to illustrate an application of the proposed method in Section 6. Some concluding remarks are made in Section 7. All technical proofs are given in the Supplemental Materials.

2. Statistical Model

Consider a longitudinal study that consists of a random sample of n subjects. For subject i, let $Y_i(t)$ denote the response variable, $\boldsymbol{X}_i$ denote a p-dimensional vector of covariates. In additional, let U_i be the terminal event time and C_i be the

censoring time $i = 1, \ldots, n$. It is assumed that the terminal event time U_i is subject to right-censored. If $U_i \leq C_i$, then U_i is observed; otherwise U_i is right-censored by C_i. The observed event time is denoted by $\tilde{U}_i = \min(U_i, C_i)$ and the censoring indicator is denoted by $\Delta_i = I(U_i \leq C_i)$, where $I(\cdot)$ is the indicator function. Suppose that $Y_i(t)$ is observed at distinct time points $T_{K_i,1} < T_{K_i,2} < \ldots < T_{K_i,K_i}$, where K_i is the total number of observations on subject i. In the following, we regard these observation times arising from an underlying counting process $N^*(t)$ characterized by $N_i(t) = \sum_{j=1}^{K_i} I\left(T_{K_i,j} \leq t\right) = N_i^*(\min(t, \tilde{U}_i))$, with $K_i = N_i^*(\tilde{U}_i)$ for subject i, $i = 1, \ldots, n$. Then, the process $Y_i(t)$ is observed only at the time points where $N_i(t)$ jumps.

Define $\mathcal{F}_{it} = \{N_i(s), 0 \leq s < t\}$. For the analysis, given $U_i = u$, $\boldsymbol{X}_i$, $\mathcal{F}_{it}$ and the covariate $\boldsymbol{W}_i$, which is allowed to be a component of the vector $\boldsymbol{X}_i$, we assume that $Y_i(t)$ follows the marginal model

$$E\{Y_i(t)|U_i = u, \boldsymbol{X}_i, \boldsymbol{W}_i, \mathcal{F}_{it}\} = \mu_0(u - t) + \boldsymbol{\beta}_0' \boldsymbol{X}_i + \boldsymbol{\alpha}_0' H(\mathcal{F}_{it}, \boldsymbol{W}_i), \quad \tau_0 \leq t \leq u, \tag{2.1}$$

where $\mu_0(\cdot)$ is an unspecified smooth function, $\boldsymbol{\beta}_0$ is a p-dimensional vector of unknown regression parameters, $\boldsymbol{\alpha}_0$ is a q-dimensional vector of regression coefficients, $H(\cdot)$ is a vector of known functions of the counting process $N_i(t)$ up to $t-$ and the covariate $\boldsymbol{W}_i$, representing the interaction between the observation

history and some covariates, and τ_0 is a positive constant which can be considered as the lower bound of observation times in practice. In particular, in longitudinal follow-up clinical studies with different treatments, $\boldsymbol{W}_i$'s can be defined as the treatment indicators, and thus $\boldsymbol{\alpha}$ represents the effect of interaction between the frequency of observation times and treatment group on the longitudinal response variable.

Our proposed model (2.1) follows the framework of the marginal mean model for longitudinal data, as described in Lin and Ying (2001). This model characterizes the marginal mean of the response process while leaving the within-subject dependence structure and distribution form of the response process unspecified. In contrast to conventional approaches, our modeling approach is distinct. We aim to consider both informative observation times and terminal events simultaneously. To achieve this, we extend Zhao et al. (2014)'s semiparametric marginal mean model for the longitudinal response by incorporating both the terminal event time and the potential effect of the observation process as covariates into the marginal mean model for the response process. Importantly, our model (2.1) encompasses the proposed mixed effects model by Kong et al. (2018) as a special case, demonstrating the flexibility and versatility of our approach. Furthermore, no additional model assumption is needed for the observation process, and the fitted conditional model can be useful for prediction in longitudinal data studies.

In model (2.1), the function $H(\cdot)$ specifies the dependence of the process $Y_i(t)$ on the observation process $N_i(t)$. The choice of H can vary depending on the specific situation and context. Following the discussion in Sun et al. (2005), a natural and simple choice for H could be $H(\mathcal{F}_{it}, \boldsymbol{W}_i) = N_i(t-)\boldsymbol{W}i$, where $Y_i(t)$ and $\mathcal{F}_{it}$ are related through the total number of observations made before time t, and this relationship can vary with the covariate $\boldsymbol{W}_i$. An alternative choice is to consider that $Y_i(t)$ depends on $\mathcal{F}_{it}$ only through a recent number of observations, for example, within a time window of length l. In this case, we can define $H(\mathcal{F}_{it}, \boldsymbol{W}_i) = (N_i(t-) - N_i(t-l))\boldsymbol{W}_i$. It is also possible to define H as a vector that combines both of the aforementioned choices. This can be useful when both the total number of observations and recent observations contain valuable information about $Y_i(t)$. In practice, we typically do not treat $H(\cdot)$ as a nonparametric parameter to avoid the complexity of the estimation method. Instead, we consider specific functional forms for H that capture the relevant information from the observation process $N_i(t)$. These choices strike a balance between flexibility and simplicity, allowing us to effectively model the relationship between $Y_i(t)$ and $\mathcal{F}_{it}$.

In addition, for inference on models (2.1), we need some basic assumptions:

(A1) conditional on $(\boldsymbol{X}, \boldsymbol{W})$, the censoring time C is independent of $N^*(\cdot)$ and $Y(\cdot)$ (C is noninformative);

(A2) U and C are conditionally independent given $(\boldsymbol{X}, \boldsymbol{W})$;

(A3) $E\{Y_i(t)|\boldsymbol{X}_i, \boldsymbol{W}_i, N_i(s), 0 \leq s \leq t, U_i, C_i\} = E\{Y_i(t)|\boldsymbol{X}_i, \boldsymbol{W}_i, \mathcal{F}_{it}, U_i, C_i\}$, which means that conditional on the covariate, the terminal event time and the censoring time, the mean of response variable at time point t is only related to the observation history before t.

The observation for each individual consists of $\mathbf{O} = (K, \tilde{T}_K, \tilde{Y}_K, \tilde{N}_K, \boldsymbol{X}, \boldsymbol{W}, \tilde{U}, \Delta)$, with $\tilde{T}_K = (T_{K,1}, \cdots, T_{K,K})$, $\tilde{Y}_K = (Y(T_{K,1}), \cdots, Y(T_{K,K}))$, and $\tilde{N}_K = (N(T_{K,1}), \cdots, N(T_{K,K}))$. Throughout this paper, we will assume that we observe n i.i.d. copies, $\mathbf{O}_1, \cdots, \mathbf{O}_n$ of $\mathbf{O}$. The main purpose here is to estimate the regression coefficients $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0')'$ and the smooth baseline mean function $\mu_0(\cdot)$ with the nuisance parameter F_0.

3. Estimation Procedure

To estimate $\boldsymbol{\theta}_0$ and μ_0 in model (2.1), a natural idea is to use the least squares loss function

$$\ell_n(\boldsymbol{\theta}, \mu) = \frac{1}{n}\sum_{i=1}^{n} I(U_i \geq T_{K_i,j}) \left\{Y_i(T_{K_i,j}) - \boldsymbol{\theta}' \boldsymbol{Z}_i(\boldsymbol{X}_i, T_{K_i,j}) - \mu(U_i - T_{K_i,j})\right\}^2,$$

where $\boldsymbol{Z}(\boldsymbol{X}, t) = (\boldsymbol{X}', H(\mathcal{F}_t, \boldsymbol{W})')'$. In practice, U_i cannot be observed for some subject i due to censoring. To solve the problem, we propose to utilize the condi-

tional expectation of $\ell_n(\boldsymbol{\theta}, \mu)$ given the observed data as a new loss function. To this end, we let $F_0(u|\boldsymbol{x})$ be the conditional cumulative distribution function of U given covariate $\boldsymbol{X} = \boldsymbol{x}$, and compute

$$
\begin{aligned}
& E\left\{\sum_{j=1}^{K} \xi(T_{K,j}) \left[Y(T_{K,j}) - \boldsymbol{\theta}'\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu(U - T_{K,j})\right]^2 \Big| \mathbf{O}\right\} \\
=& \sum_{j=1}^{K} \xi(T_{K,j}) \Bigg[\Delta \left\{Y(T_{K,j}) - \boldsymbol{\theta}'\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu(\tilde{U} - T_{K,j})\right\}^2 \\
& + (1-\Delta) \frac{\int_{\tilde{U}}^{\tau} \left\{Y(T_{K,j}) - \boldsymbol{\theta}'\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu(u - T_{K,j})\right\}^2 dF_0(u|\boldsymbol{X})}{1 - F_0(\tilde{U}|\boldsymbol{X})}\Bigg] \\
=:& L(\boldsymbol{\theta}, \mu; F_0),
\end{aligned}
$$

where $\xi(u) = I(\tilde{U} \geq u)$ and τ is a prespecified constant such that $P(\tilde{U} \geq \tau) > 0$ (see Kong et al. 2018, P16). Therefore, we define a new least square loss function as

$$
\begin{aligned}
& \mathbb{P}_n L(\boldsymbol{\theta}, \mu; F) \\
=& \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \xi_i(T_{K_i,j}) \Bigg[\Delta_i \left\{Y_i(T_{K_i,j}) - \boldsymbol{\theta}'\boldsymbol{Z}_i(\boldsymbol{X}_i, T_{K_i,j}) - \mu(\tilde{U}_i - T_{K_i,j})\right\}^2 \\
& + (1-\Delta_i) \frac{\int_{\tilde{U}_i}^{\tau} \left\{Y_i(T_{K_i,j}) - \boldsymbol{\theta}'\boldsymbol{Z}_i(\boldsymbol{X}_i, T_{K_i,j}) - \mu(u - T_{K_i,j})\right\}^2 dF(u|\boldsymbol{X}_i)}{1 - F(\tilde{U}_i|\boldsymbol{X}_i)}\Bigg],
\end{aligned} \tag{3.2}
$$

where $\mathbb{P}_n$ denotes the empirical measure. Since the loss function (3.2) involves the unknown nuisance parameter F, we propose a two-stage estimation procedure. In

stage one, we obtain an estimator $\hat{F}_n$ of F_0 based on the data $\{\tilde{U}_i, \Delta_i, \boldsymbol{X}_i\}$. In stage two, $\hat{\boldsymbol{\theta}}_n$ and $\hat{\mu}_n$ are obtained by minimizing the loss function $\mathbb{P}_n L(\boldsymbol{\theta}, \mu; \hat{F}_n)$.

In stage one, to estimate $F_0(u|\boldsymbol{x})$, we can assume a survival model such as the Cox proportional hazards model (Cox, 1972), that is, the hazard function of U has the following form

$$\lambda(u|\boldsymbol{X}) = \lambda(u) \exp\{\boldsymbol{\gamma}'\boldsymbol{X}\}, \tag{3.3}$$

where $\boldsymbol{\gamma}$ is a p-dimensional unknown regression parameter, and $\lambda(u)$ is the unknown underlying hazard function. Denote the true values of $\boldsymbol{\gamma}$ and λ in model (3.3) as $\boldsymbol{\gamma}_0$ and λ_0, respectively. The regression coefficient $\boldsymbol{\gamma}_0$ can be estimated by the partial likelihood and the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ by the Breslow estimator (Breslow, 1972), denoted by $\hat{\boldsymbol{\gamma}}_n$ and $\hat{\Lambda}_n$, respectively. As a consequence, $F_0(u|\boldsymbol{x})$ is estimated by

$$\hat{F}_n(u|\boldsymbol{X}) = 1 - \exp\{-\hat{\Lambda}_n(u) \exp(\hat{\boldsymbol{\gamma}}_n'\boldsymbol{X})\}. \tag{3.4}$$

In stage two, we propose to use B-splines to approximate $\mu_0(\cdot)$. For a finite closed interval $[0, \tau]$, let $\mathcal{I} = \{t_i\}_1^{m_n+2l}$, with $0 = t_1 = \ldots = t_l < t_{l+1} < \ldots < t_{m_n+l} < t_{m_n+l+1} = \ldots = t_{m_n+2l} = \tau$, be a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals and $m_n = O(n^\nu)$, for $0 < \nu < 1/2$. Let $\Psi_n = \Psi_{l,\mathcal{I}}$ (with order l and knots $\mathcal{I}$) be the space of polynomial splines of order l defined

in Schumaker (2007, P108, Definition 4.1). According to Schumaker (2007, P117, Corollary 4.10), there exists a local basis $\{B_{il}, 1 \le i \le q_n\}$ with $q_n = m_n + l$ such that for any $\phi \in \Psi_{l,\mathcal{I}}$, we can write $\phi(t) = \sum_{i=1}^{q_n} \eta_i B_{il}(t)$.

Under suitable smoothness assumptions, $\mu_0(\cdot)$ can be well approximated by a function $\mu_n(t)$ in $\Psi_{l,\mathcal{I}}$. Denote $\mu_n(t) = \boldsymbol{\eta}'\boldsymbol{B}_l(t)$, where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{q_n})'$ and $\boldsymbol{B}_l(t) = (B_{1l}(t), \ldots, B_{q_n l}(t))'$. Then the loss function $\mathbb{P}_n L(\boldsymbol{\theta}, \mu; \hat{F}_n)$ is approximate to $\mathbb{P}_n L^*(\boldsymbol{\theta}, \boldsymbol{\eta}; \hat{F}_n)$ with μ substituted by μ_n. Denote the minimizer of $\mathbb{P}_n L^*(\boldsymbol{\theta}, \boldsymbol{\eta}; \hat{F}_n)$ as $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\eta}}_n$, which has a closed form as follows:

$$
\begin{aligned}
\begin{pmatrix} \hat{\boldsymbol{\theta}}_n \\ \hat{\boldsymbol{\eta}}_n \end{pmatrix} = \Bigg[\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \xi_i(T_{K_i,j}) \Bigg\{ \Delta_i \begin{pmatrix} \boldsymbol{Z}_i(\boldsymbol{X}_i, T_{K_i,j}) \\ \boldsymbol{B}_l(\tilde{U}_i - T_{K_i,j}) \end{pmatrix}^{\otimes 2} \\
+ \frac{(1-\Delta_i)}{1 - \hat{F}_n(\tilde{U}_i | \boldsymbol{X}_i)} \int_{\tilde{U}_i}^{\tau} \begin{pmatrix} \boldsymbol{Z}_i(\boldsymbol{X}_i, T_{K_i,j}) \\ \boldsymbol{B}_l(u - T_{K_i,j}) \end{pmatrix}^{\otimes 2} d\hat{F}_n(u | \boldsymbol{X}_i) \Bigg\} \Bigg]^{-1} \\
\times \Bigg[\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \xi_i(T_{K_i,j}) Y_i(T_{K_i,j}) \Bigg\{ \Delta_i \begin{pmatrix} \boldsymbol{Z}_i(\boldsymbol{X}_i, T_{K_i,j}) \\ \boldsymbol{B}_l(\tilde{U}_i - T_{K_i,j}) \end{pmatrix} \\
+ \frac{(1-\Delta_i)}{1 - \hat{F}_n(\tilde{U}_i | \boldsymbol{X}_i)} \int_{\tilde{U}_i}^{\tau} \begin{pmatrix} \boldsymbol{Z}_i(\boldsymbol{X}_i, T_{K_i,j}) \\ \boldsymbol{B}_l(u - T_{K_i,j}) \end{pmatrix} d\hat{F}_n(u | \boldsymbol{X}_i) \Bigg\} \Bigg].
\end{aligned}
$$

Then the resulting estimator for $\mu_0(t)$ is $\hat{\mu}_n(t) \equiv \sum_{k=1}^{q_n} \hat{\eta}_{nk} B_{k,l}(t)$.

4. Asymptotic Results

To establish the asymptotic properties of the proposed estimators, we need the following regularity conditions.

(C1) The maximum spacing of the knots satisfies $\triangle = \max_{l+1<i<m_n+l+1} \mid t_i - t_{i-1} \mid = O(n^{-v})$. Moreover, there exists a constant $M > 0$ such that $\triangle/\delta \leq M$ uniformly in n, where $\delta = \min_{l+1<i<m_n+l+1} \mid t_i - t_{i-1} \mid$.

(C2) The parameter spaces of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$, Θ is bounded and convex on $\mathbb{R}^{p+q}$, and the true parameter $(\boldsymbol{\theta}_0, \mu_0) \in \Theta^\circ \times \mathcal{F}_r$, where Θ° is the interior of Θ, and $\mathcal{F}_r$ is the collection of bounded functions f on $[0, \tau]$ with bounded derivatives $f^{(j)}, j = 1, \cdots, k$, and the kth derivative $f^{(k)}$ satisfies the following Lipschitz continuity condition:

$$|f^{(k)}(s) - f^{(k)}(t)| \leq M|s - t|^{\zeta}, s, t \in [0, \tau],$$

where k is a positive integer, $\zeta \in (0, 1]$ such that $r = k + \zeta \geq 2$, M is a positive constant and $f^{(k)}$ is the kth derivative of function f.

(C3) $H(\cdot)$ has bounded total variations, $P(\|\boldsymbol{X}\| \leq M_1) = 1$ for a positive constant M_1, and the number of observation times K is bounded almost surely.

(C4) The study stops at a finite time $\tau > 0$, s.t. $\inf_{\boldsymbol{x}} P(U \geqslant \tau | \boldsymbol{X} = \boldsymbol{x}) = \omega_1 > 0$

for constant ω_1.

(C5) If with probability 1, $\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X},t) - h_2(U-t) = 0$ for some deterministic function h_2 and $\boldsymbol{h}_1 \in \mathbb{R}^{p+q}$, then $\boldsymbol{h}_1 = 0$ and $h_2(\cdot) = 0$.

(C6) $\boldsymbol{J} = E\left\{\sum_{j=1}^{K}\xi(T_{K,j})\left[\boldsymbol{Z}(\boldsymbol{X},T_{K,j}) - E\{\boldsymbol{Z}(\boldsymbol{X},T_{K,j})|K,T_{K,j},U,C\}\right]^{\otimes 2}\right\}$ is positive definite.

(C7) $F_0(u|\boldsymbol{X})$ is absolutely continuous with respect to Lebesgue measure, and the density function $f_0(u|\boldsymbol{X}) > f_0, u \in [0,\tau]$ for some constant $f_0 > 0$.

These are all mild conditions that could be satisfied in usual situations. Condition (C1) is similar to those required by Stone (1986) and Zhou et al. (1998); Condition (C2) is a common assumption in nonparametric smoothing estimation problems. Usually, $r = 2$ (i.e., $k = 1$ and $\zeta = 1$) should be satisfied in many situations and the requirement that $r \geq 2$ is to guarantee the desirable control of the spline approximation error rates of the first derivatives of μ_0. The boundedness conditions (C3) is easily justified in most applications. Condition (C5) is needed to establish the identifiability of the model. Condition (C6) can be interpreted that the sample covariance is asymptotically nonsingular. Condition (C7) implies that the terminal event time U given the covariates $\boldsymbol{X}$ has a strictly positive density.

For any function $\mu \in \mathcal{F}_r$, define

$$\|\mu\|_2^2 = E\left[\sum_{j=1}^{K} \xi(T_{K,j})\mu^2(U - T_{K,j})\right].$$

Then for $(\boldsymbol{\theta}_1, \mu_1), (\boldsymbol{\theta}_2, \mu_2) \in \Theta \times \mathcal{F}_r$, we define the distance as

$$d((\boldsymbol{\theta}_1, \mu_1), (\boldsymbol{\theta}_2, \mu_2)) = \{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^2 + \|\mu_1 - \mu_2\|_2^2\}^{1/2},$$

where $\|\cdot\|$ is the Euclidean norm.

Let $\|F\|_\infty = \sup_{u,\boldsymbol{x}} |F(u|\boldsymbol{x})|$. The asymptotic properties for the estimator $(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n)$ are summarized as follows.

Theorem 1 (Consistency)**.** *Suppose that* $\|\hat{F}_n - F_0\|_\infty = o_p(1)$. *Under the conditions (C1)-(C5),* $d((\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n), (\boldsymbol{\theta}_0, \mu_0)) \xrightarrow{P} 0$ *as* $n \to +\infty$.

Theorem 2 (Rate of convergence)**.** *Suppose that* $\|\hat{F}_n - F_0\|_\infty = O_p(n^{-\frac{r}{1+2r}})$. *Under the conditions (C1)-(C7),*

$$d((\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n), (\boldsymbol{\theta}_0, \mu_0)) = O_p(n^{-\min\{\nu r, \frac{1-\nu}{2}\}}).$$

Remark 1. *When* $\nu = \frac{1}{1+2r}$, $n^{-\min\{\nu r, \frac{1-\nu}{2}\}} = n^{-\frac{r}{1+2r}}$, *we conclude from Stone*

(1980, 1982) that the rate of convergence for the estimator $(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n)$ *is the optimal rate in nonparametric regression.*

For ease of exposition, we use $\{\tilde{\boldsymbol{O}}_i, i = 1, \cdots, n\}$ to represent the i.i.d. sample for estimation F_0 in stage 1.

Theorem 3 (Asymptotic normality)**.** *Suppose that* $\|\hat{F}_n - F_0\|_\infty = O_p(n^{-\frac{r}{1+2r}})$*. And assume that there exists a uniformly bounded process* $\mathcal{O}$ *and a Lipschitz function* $\tilde{g}$ *such that*

$$\sqrt{n}\int_0^\tau \psi(u;\boldsymbol{O})d\{\hat{F}_n(u|\boldsymbol{X}) - F_0(u|\boldsymbol{X})\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \tilde{\psi}(u;\boldsymbol{O})d\mathcal{O}(u;\boldsymbol{O};\tilde{\boldsymbol{O}}_i)$$

is distributed asymptotically as a normal distribution with mean zero for the integrable function ψ*, where* $\tilde{\psi} = \tilde{g} \circ \psi$ *with* $\tilde{g} \circ \psi$ *denoting the composite of functions* $\tilde{g}$ *and* ψ*. Under the conditions (C1)-(C7) and* $\frac{1}{4r} \leq \nu < \frac{1}{2}$*,* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *converges in distribution to a mean zero normal random variable with variance matrix* $\boldsymbol{J}^{-1}\boldsymbol{Q}\boldsymbol{J}^{-1}$ *as* n *tends to infinity, where* $\boldsymbol{J}$ *is defined in condition (C7), and* $Q = E[\{\psi^*(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) + m^{**}(\boldsymbol{\theta}_0, \mu_0, F_0; \tilde{\boldsymbol{O}})\}^{\otimes 2}]$ *with* ψ^* *and* m^{**} *being given in the Supplementary Materials for proving this theorem.*

From the theorem, $\hat{\boldsymbol{\theta}}_n$ achieves the standard convergence rate although the overall convergence rate of the proposed estimator is slower than $n^{-1/2}$. To make inference, we estimate the asymptotic variance of $\hat{\boldsymbol{\theta}}_n$ by the bootstrap method since

an unknown conditional expectation term $E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j})|K, T_{K,j}, U, C\}$ is involved.

C8. The information matrix of the partial likelihood for the Cox regression model at the true parameter values is positive definite.

Corollary 1. *Suppose Conditions (C1)-(C8) hold. If there exists some positive constant* ω_2 *such that* $\inf_{\boldsymbol{x}} P(C \geqslant \tau|\boldsymbol{X} = \boldsymbol{x}) = \omega_2 > 0$*, then Theorem 3 holds for* $\hat{\theta}_n$ *when* F_0 *is estimated by* $\hat{F}_n(u|\boldsymbol{X})$ *in (3.4).*

5. Simulation Study

In this section, simulation studies were conducted to assess the finite-sample properties of the proposed estimators. To illustrate the robustness of the proposed method, we compared the estimation performance of the proposed method and the two-stage semiparametric likelihood-based approach proposed by Kong et al. (2018). As both models are conditional models, this comparison provides valuable insights into the strengths of our proposed method. We generated the response process from the following two models (the proposed model (I) and Kong's model (II)):

(I) $Y_i(t) = \mu_0(U_i - t) + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha H(\mathcal{F}_{it}, W_i) + \epsilon_i(t)$, where $\mu_0(U - t) = 1/\exp(U - t)$, X_{1i} and X_{2i} were generated from Bernoulli distribution with success probability 0.5 and the uniform distribution over interval $[0, 1]$, re-

spectively, $H(\mathcal{F}_{it}, W_i) = N_i(t-)W_i$ with $W_i = X_{1i}$, and $\epsilon_i(t)$ are independent standard normal variables.

(II) $Y_i(t) = \mu_0(U_i - t, \boldsymbol{\xi}) + \beta_1 X_{1i} + \beta_2 X_{2i} + Z_i b_i + G_i(t) + \epsilon_i(t)$, where $\mu_0(U-t, \boldsymbol{\xi}) = \xi_1 e^{-(U-t-\xi_2)^2 \xi_3} + \beta_0$ with $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3) = (1, -1, 0.4)$ and $\beta_0 = 1$, X_{1i} and X_{2i} were generated as in (I), Z_i follows a standard normal distribution, b_i follows a normal distribution $N(0, \exp(\varphi))$ with $\varphi = -0.1$, $G_i(t)$ is a mean zero Gaussian process with $var(G_i(t)) = \nu(t) = \exp(\nu_0 + \nu_1 t)$ and $corr(G_i(t_1), G_i(t_2)) = \rho^{|t_1 - t_2|}$ for $(\rho, \nu_0, \nu_1) = (1/(1 + \exp(1)), 1, -1)$, and ϵ_i are independent measurement errors which follow a normal distirbution $N(0, \sigma^2)$ with $\sigma^2 = \exp(-0.5)$.

In the above setups, the hazard function of U_i is

$$\lambda(u|X_{1i}, X_{2i}) = \lambda_0(u) \exp\left(\kappa(0.5X_{1i} + X_{2i})\right),$$

where $\lambda_0(u) = u$ and κ was adjusted to achieve a desired censoring rate. The censoring time C_i was generated from the uniform distribution over interval $(\tau/2, \tau)$ with $\tau = 6$. Specifically, the censoring rates were 20% and 40% when $\kappa = -2.1$ and -3, respectively. Denote $\tilde{U}_i = \min(C_i, U_i)$.

For the generation of the observation process $N_i(t)$ given the covariate $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$, we considered two cases:

(a) The number of observation times K_i equals one plus a Poisson random variable with mean $\tilde{U}_i \exp(-X_{1i}+X_{2i})$ and the observation times $(T_{K_i1}, \ldots, T_{K_iK_i})$ were taken to be the order statistics of a random sample of size K_i from the uniform distribution over $(0, \tilde{U}_i)$.

(b) K_i follows the uniform distribution over $\{1, 2, 3\}$ when $X_{1i} = 0$ and the uniform distribution over $\{4, 5, 6\}$ otherwise, and the observation times $(T_{K_i1}, \ldots, T_{K_iK_i})$ were generated in the same way as in set-up (a).

The true parameter values used in our simulation studies were $\beta_0 = (\beta_{10}, \beta_{20})' = (-1, 1)'$, $\alpha_0 = 1$. Our results were obtained from 1000 independent runs. To estimate $\mu_0(t)$, cubic B-splines were used in computing the spline estimators, where we took the number of interior knots as $n^{1/v}$ with v being chosen by BIC criterion. The equally spaced knots are given by $U^*_{\min} + k(U^*_{\max} - U^*_{\min})/(m_n + 1), k = 0, 1, \cdots, m_n + 1$, with $U^*_{\min}$ and $U^*_{\max}$ being the respective minimum and maximum values of distinct times $\{U^*_{ij}\}$'s $(U^*_{ij} = \tilde{U}_i - T_{K_i,j}, j = 1, \cdots, K_i, i = 1, \cdots, n)$.

Table 1 presents the simulation results on estimation of $(\beta_{10}, \beta_{20}, \alpha_0)$ using the proposed method with the sample sizes $n = 100$ or 200, censoring rates 20% or 40% under models (I) and (II) with the Poisson (a) or non-Poisson (b) observation times. The simulation results include the estimated bias (BIAS) given by the average of the estimates minus the true value, the bootstrap standard errors of the estimates (BSE), the sample standard deviation of the estimates (SSE), and

the bootstrap 95% coverage probabilities (CP) obtained from 1000 independent runs. Here, we used 200 replications in bootstrap to estimate the standard errors. In Table 2, BIAS and SSE results for the estimation of (β_{10}, β_{20}) are given for comparison between the proposed method and Kong's method (Kong et al., 2018). Table 3 gives the integrated square error (ISE) results for the estimation accuracy of $\hat{\mu}_n$ in (I) and $\hat{\mu}_n^* = \mu_0(\cdot, \hat{\boldsymbol{\xi}})$ in (II), where $ISE(\hat{\mu}_n) = \int_0^{\tau/2}\{\mu_0(t) - \hat{\mu}_n(t)\}^2 dt$ (P53, Härdle et al., 2004). Figure 1 shows the estimation results of $\mu_0(t)$ under setups I with $\mu_0(t) = 1/\exp(t)$ and II with $\mu_0(t) = e^{-0.4(t+1)^2} + 1$ for the sample size $n = 100$ or 200 and cases "A" (CR=20%, Observation process (a)), "B" (CR=40%, Observation process (a)), "C" (CR=20%, Observation process (b)), and "D" (CR=40%, Observation process (b)). In Figure 1, the red solid line represents the true curve for $\mu_0(t)$, the black dashed line represents the estimated curve by our proposed method, and the blue dotted line represents the pointwise average of the estimated normal kernel function by Kong's method.

Based on our simulation results, we have the following findings: (i) Under both the proposed model setup (I) and Kong's model setup (II) for the response variable, the proposed two stage spline-based least square estimators perform well for both the non-Poisson and Poisson observation processes with different censoring rates and sample sizes. Specifically, the estimates are approximately unbiased; the sample standard errors of the estimates and the bootstrap standard errors of the

Table 1: Simulation results on estimation of $(\beta_1, \beta_2, \alpha)$ using the proposed method.

			Observation(a)				Observation(b)			
			CR = 20%		CR = 40%		CR = 20%		CR = 40%	
			$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
I	β_1	BIAS	-0.0016	-0.0009	0.0158	0.0157	0.0059	0.0011	0.0119	0.0013
		SSE	0.1570	0.1068	0.1752	0.1213	0.1710	0.1232	0.1950	0.1394
		BSE	0.1550	0.1059	0.1780	0.1206	0.1754	0.1224	0.2005	0.1379
		CP	0.9480	0.9430	0.9480	0.9430	0.9420	0.9520	0.9450	0.9340
	β_2	BIAS	0.0146	0.0217	0.0415	0.0369	0.0201	0.0166	0.0340	0.0278
		SSE	0.2125	0.1477	0.2233	0.1513	0.2110	0.1470	0.2517	0.1743
		BSE	0.2102	0.1457	0.2295	0.1553	0.2141	0.1486	0.2544	0.1739
		CP	0.9430	0.9350	0.9470	0.9420	0.9470	0.9470	0.9460	0.9430
	α	BIAS	0.0079	0.0088	0.0092	0.0056	0.0014	0.0023	0.0029	0.0058
		SSE	0.0675	0.0461	0.0787	0.0509	0.0515	0.0355	0.0599	0.0410
		BSE	0.0711	0.0456	0.0840	0.0525	0.0519	0.0361	0.0606	0.0414
		CP	0.9510	0.9440	0.9410	0.9500	0.9450	0.9510	0.9320	0.9460
II	β_1	BIAS	-0.0043	-0.0014	0.0023	-0.0044	-0.0075	-0.0016	-0.0088	-0.0030
		SSE	0.1463	0.1086	0.1804	0.1195	0.1795	0.1251	0.2009	0.1402
		BSE	0.1546	0.1058	0.1777	0.1206	0.1769	0.1229	0.2019	0.1386
		CP	0.9540	0.9510	0.9320	0.9530	0.9410	0.9400	0.9480	0.9450
	β_2	BIAS	0.0167	0.0125	0.0126	0.0151	0.0025	-0.0012	0.0032	0.0033
		SSE	0.2118	0.1430	0.2224	0.154	0.2050	0.1505	0.2485	0.1744
		BSE	0.2105	0.1455	0.2260	0.1558	0.2149	0.1486	0.2570	0.1742
		CP	0.9410	0.9420	0.9370	0.9480	0.9570	0.9520	0.9490	0.9530
	α	BIAS	0.0038	0.0051	0.0047	0.0086	0.0029	0.0022	0.0049	0.0045
		SSE	0.0655	0.0455	0.0805	0.0527	0.0530	0.0347	0.0604	0.0398
		BSE	0.0700	0.0449	0.0836	0.0525	0.0525	0.0359	0.0611	0.0413
		CP	0.9470	0.9420	0.9560	0.9360	0.9290	0.9550	0.9340	0.9520

Table 2: Comparison for the estimation of (β_1, β_2) between the proposed method and Kong's method (KNKSH).

		Method		Observation(a) CR = 20% $n = 100$	Observation(a) CR = 20% $n = 200$	Observation(a) CR = 40% $n = 100$	Observation(a) CR = 40% $n = 200$
I	β_1	Proposed	BIAS	-0.0016	-0.0009	0.0158	0.0157
			SSE	0.1570	0.1068	0.1752	0.1213
		KNKSH	BIAS	0.0349	0.0346	0.0684	0.0658
			SSE	0.1139	0.0816	0.1138	0.0788
	β_2	Proposed	BIAS	0.0146	0.0217	0.0415	0.0369
			SSE	0.2125	0.1477	0.2233	0.1513
		KNKSH	BIAS	0.0460	0.0560	0.1342	0.1270
			SSE	0.2005	0.1421	0.2057	0.1381
II	β_1	Proposed	BIAS	-0.0043	-0.0014	0.0023	-0.0044
			SSE	0.1463	0.1086	0.1804	0.1195
		KNKSH	BIAS	0.0218	0.0214	0.0353	0.0346
			SSE	0.1126	0.0814	0.1135	0.0767
	β_2	Proposed	BIAS	0.0167	0.0125	0.0126	0.0151
			SSE	0.2118	0.1430	0.2224	0.1540
		KNKSH	BIAS	0.0288	0.0379	0.0740	0.0740
			SSE	0.1998	0.1456	0.2015	0.1403

		Method		Observation(b) CR = 20% $n = 100$	Observation(b) CR = 20% $n = 200$	Observation(b) CR = 40% $n = 100$	Observation(b) CR = 40% $n = 200$
I	β_1	Proposed	BIAS	0.0059	0.0011	0.0119	0.0013
			SSE	0.1710	0.1232	0.1950	0.1394
		KNKSH	BIAS	0.0277	0.0260	0.0636	0.0611
			SSE	0.1224	0.0883	0.1320	0.0927
	β_2	Proposed	BIAS	0.0201	0.0166	0.0340	0.0278
			SSE	0.2110	0.1470	0.2517	0.1743
		KNKSH	BIAS	0.0582	0.0640	0.1300	0.1330
			SSE	0.2040	0.1461	0.2127	0.1506
II	β_1	Proposed	BIAS	-0.0075	-0.0016	-0.0088	-0.0030
			SSE	0.1795	0.1251	0.2009	0.1402
		KNKSH	BIAS	0.0178	0.0154	0.0359	0.0335
			SSE	0.1224	0.0885	0.1304	0.0909
	β_2	Proposed	BIAS	0.0025	-0.0012	0.0032	0.0033
			SSE	0.2050	0.1505	0.2485	0.1744
		KNKSH	BIAS	0.0310	0.0346	0.0657	0.0698
			SSE	0.2013	0.1417	0.2121	0.1467

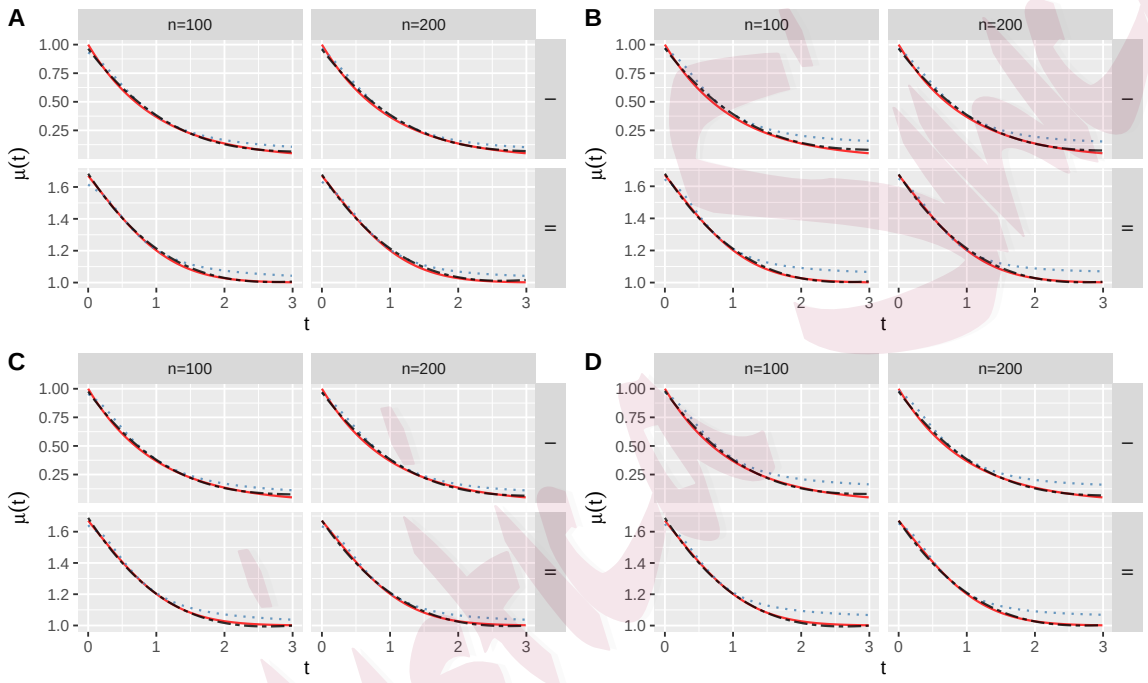

Figure 1: The estimation results of $\mu_0(t)$ under setups I with $\mu_0(t) = 1/\exp(t)$ and II with $\mu_0(t) = e^{-0.4(t+1)^2} + 1$. "A" (CR=20%, Observation process (a)), "B" (CR=40%, Observation process (a)), "C" (CR=20%, Observation process (b)), and "D" (CR=40%, Observation process (b)). Red solid (true), black dashed (proposed), blue dotted (Kong's).

Table 3: ISE comparison between the proposed method and Kong's method (KNKSH).

		Observation(a)			
		CR = 20%		CR = 40%	
Setup	Method	$n = 100$	$n = 200$	$n = 100$	$n = 200$
I	Proposed	0.0811	0.0391	0.0825	0.0369
	KNKSH	0.0878	0.0448	0.1036	0.0487
II	Proposed	0.0819	0.0384	0.0818	0.0394
	KNKSH	0.0946	0.0512	0.0994	0.0477
		Observation(b)			
I	Proposed	0.0934	0.0430	0.1060	0.0480
	KNKSH	0.1005	0.0540	0.1231	0.0632
II	Proposed	0.0964	0.0441	0.1101	0.0497
	KNKSH	0.1008	0.0518	0.1149	0.0572

proposed estimators are close to each other; the bootstrap 95% coverage rates are close to the nominal level, that is, the proposed procedure provides reasonable estimates and the normal approximation seems to be appropriate. (ii) Our estimates have biases closer to zero and smaller ISE values than Kong's method (KNKSH), and SSEs by two methods are comparable. (iii) The estimated curves of $\mu_0(t)$ are very close to their real curves with the moderate sample size, indicating that the B-splines estimator for $\mu_0(t)$ works well under all situations. While the estimated normal kernel function $\hat{\mu}_n^*$ based on Kong's 2-stage method has some deviations from the real curve, especially when the censoring rate increases.

In conclusion, simulations demonstrate that the proposed estimation procedure is robust in terms of model structures for longitudinal response and observation

processes.

6. Application

This section presents an analysis of AIDS (the acquired immunodeficiency syndrome) clinical trial by applying our proposed method. The study was initiated in December 1990 by the Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA) (Abrams, et al., 1994; Goldman, et al. 1996). This AIDS clinical trial was a multi-center, randomized, open-label, community-based clinical trial comparing the clinical efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC). In the trial, 467 HIV-infected patients who met entry conditions (either an AIDS diagnosis or two CD4 lymphocyte counts of $\leq$ 300 cells/mm^3, and leading to the intolerance of zidovudine (AZT) or the progression of disease during the therapy) were enrolled and randomly assigned to receive either ddI (500 mg per day) or ddC (2.25 mg per day), stratified by clinical unit and by AZT intolerance versus failure. 230 patients received ddI and 237 received ddC. By the end of the study, 100 patients had died in the ddI group and 88 in the ddC group, resulting in 59.7% censoring rate. Absolute CD4 lymphocyte counts were measured at baseline and at the 2-,6-, and 12-month visits (and a few at 18 months), but less frequently if the patient refused or was too ill for testing. The median length of follow-ups from the time of randomization was

16 months (ranging from 12 to 21). Data for each patient consist of survival time (months from admission to death or censoring), patients status at the follow-up time (dead = 1, alive = 0), drug (ddI =1, ddC = 0), gender (female =1, male = 0), previous opportunistic infection at the study entry PrevOI (AIDS diagnosis = 1, no AIDS diagnosis = 0), AZT stratum (AZT failure =1, AZT intolerance =0), and CD4 counts at the beginning of the study and the following visiting times. This dataset can be found in the "JM" R package.

To analyze the data, for patient i, define x_{1i} as drug, x_{2i} as gender, x_{3i} as AIDS diagnosis indicator PrevOI and x_{4i} as the AZT stratum indicator; denote $\tilde{U}_i$ as the observed event time with the indicator $\Delta_i = 1$ when death happens, 0 when censoring happens. Define the response $Y_i(t)$ to be the natural logarithm of the CD4 counts of patient i up to time t plus 1. Let $N_i(\cdot)$ represent the accumulated observation numbers of patient i over the study period. Assume that $Y_i(t)$ can be described by model (2.1) with $H(\mathcal{F}_{it}, \boldsymbol{W}_i) = N_i(t-)(X_{1i}, X_{3i})'$, meaning that the relation between CD4 counts and observation times may vary with different treatments and previous AIDS diagnosis status. Also we assume that the death time U_i follows the Cox model as in (3.3) with $\boldsymbol{X}_i = (X_{1i}, X_{3i}, X_{4i})'$. For estimation of μ_0, we use the cubic B-spline approximation.

Applying the estimation procedure proposed in the previous sections, we obtained the estimation results for the regression coefficients in Table 4. Gender and

Table 4: The estimation results for the ddI/ddC study by using our proposed method.

	Longitudinal model		Survival model	
Covariate	Coeff	Estimate (95%CI)	Coeff	Estimate (95%CI)
Drug	β_1	-0.0117(-0.1295,0.1066)	γ_1	0.2086(-0.0781,0.4953)
Gender	β_2	0.0076 (-0.2563,0.2900)		
PrevOI	β_3	-0.4803(-0.6779,-0.3111)	γ_2	1.2932(0.8483,1.7382)
AZT	β_4	-0.0503(-0.1752,0.0821)	γ_3	0.1278(-0.1889,0.4447)
Drug$\times N(t-)$	α_1	0.0724(0.0058,0.1355)		
PrevOI$\times N(t-)$	α_2	0.0024(-0.0529,0.0519)		

AZT stratum seem to have no significant impact on CD4 counts. The estimated PrevOI coefficient and its 95% confidence interval indicate that patients diagnosed with AIDS at baseline had significant lower CD4 counts compared with those without AIDS. The interaction between drug and the observation times would positively influence on the CD4 counts, that is, with more observation times, patients in the ddI group would have significantly more CD4 counts than those in the ddC group. And the interaction between PrevOI and the observation times has no significant influence on the CD4 counts. The estimated curve for μ_0 with 95% pointwise confidence interval is given in Figure 2. From this figure, it can be seen that the baseline curve for the CD4 counts has a increasing pattern when the testing time is far away from the terminal event time, while the pattern becomes to decrease when the testing time is close to the terminal event time.

In the Cox model, the estimates for coefficients indicate that zalcitabine was

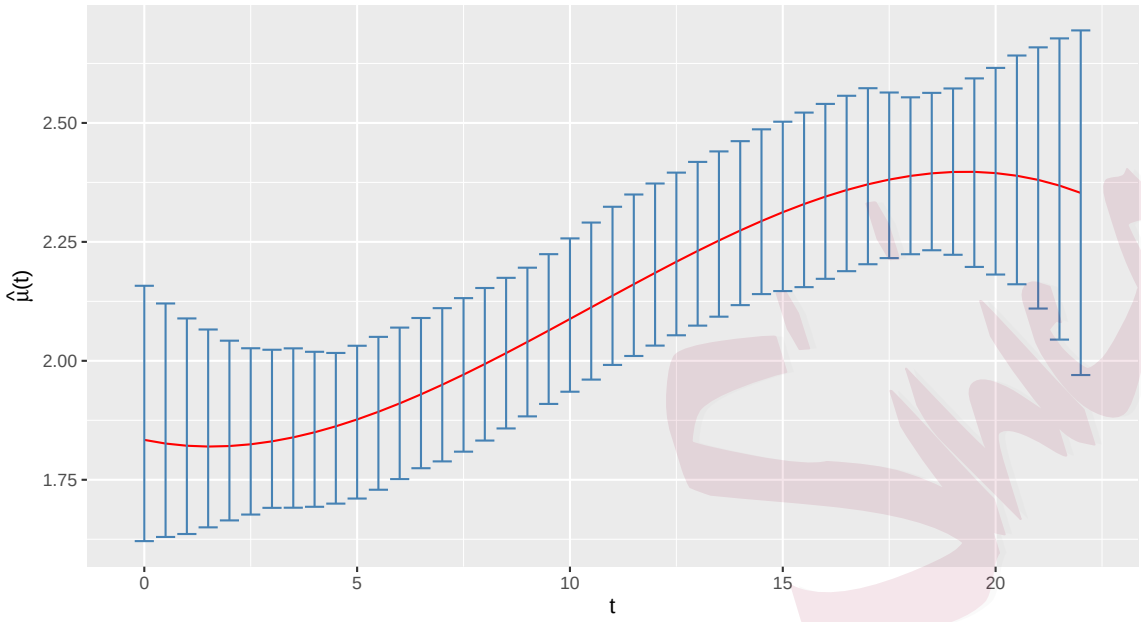

Figure 2: The estimated curve (red solid line) of μ_0 with 95% pointwise confidence interval (blue) for the ddI/ddC study.

as efficacious as didanosine in delaying death of patients, while zalcitabine may have a survival advantage than didanosine. Patients diagnosed with AIDS at baseline had significant lower survival probability compared with those without AIDS, and AZT stratum seem to have no significant impact on patients' survival time. The goodness-of-fit for the Cox model was checked for this ddI/ddC data and the corresponding goodness-of-fit empirical p-values are 0.5486, 0.8056 and 0.3656, based on $20,000$ simulated martingale residual score process (Lin et al., 1993). These results indicate that the proportional hazards model for drug, previous AIDS diagnosis indicator and AZT stratum indicator fits the data reasonably well.

7. Concluding remarks

Taking into account that both informative observation times and informative terminal event time may exist at the same time for longitudinal data, a class of new flexible semiparametric marginal conditional mean model for the longitudinal response process has been proposed. First of all, considering that the influence of the terminal event time on the longitudinal response may have different patterns, we treat the length from the observation times to the occurrence of the terminal event time as a covariate in the conditional model with an unknown functional form. Kong et al. (2018) considered a specified functional form. Second, the new model allows for the interaction between the observation history and some covariates. This is different from the joint modeling approach that uses latent variables to characterize the correlation between the response process and the observation times. Third, we leave the within-subject dependence structure of the response process and patterns of the observation process to be arbitrary, while Kong et al. (2018) specified the distributional form of a longitudinal response process with a pre-specified visit scheme.

For inference about the unknown function and regression parameters in the proposed models, a two-stage spline-based least squares estimation approach has been developed, where the nuisance conditional distribution function for the terminal event time is estimated in the first stage, and the approximate least square

loss function is used to estimate the parameters for the longitudinal model in the second stage. As demonstrated in our simulation studies, the proposed approaches are more flexible and robust with respect to the model structures for the response and observation processes.

The specific model for the terminal event time is left unspecified in our approach. However, when the estimator for the conditional distribution of the terminal event time given covariates satisfies certain asymptotic properties, we can establish the corresponding asymptotic properties of our proposed two-stage estimator. For the terminal event time, various alternative survival models can be utilized, such as the Cox model, additive hazards model, or accelerated failure time model.

Note that in our proposed model, the covariate $\boldsymbol{X}$ can indeed be time-varying. The proposed estimation procedure allows for time-varying covariates and the asymptotic properties of the estimator remain valid under this scenario.

Further research is to extend the proposed methods to other useful models such as marginal conditional varying-coefficient or nonparametric regression models for longitudinal response processes.

Supplementary Materials

The supplementary materials include the proofs of lemmas and theorems.

Acknowledgements

The authors would like to thank the Editor, the Associate Editor and the two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. This research is partly supported by the National Natural Science Foundation of China (No. 12271459, 12171374), the Research Grant Council of Hong Kong (15306521), and The Hong Kong Polytechnic University.

References

Abrams, D. I., Goldman, A. I., Launer, C., et al. (1994). A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *The New England Journal of Medicine*, **330**, 657–662.

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

Breslow, N. E. (1972). Discussion of "Regression models and life-tables" by D. R. Cox. *Journal of the Royal Statistical Society, Series B*, **34**, 216–217.

Cox, D. R. (1972). Regression models and Life Tables. *Journal of the Royal Statistical Society. Series B*, **34**, 187 - 220.

Deng, S. and Zhao, X. (2019). Covariate-adjusted regression for distorted longitudinal data with informative observation times. *Journal of the American Statistical Association*, **114**, 1241–1250.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data.* 2nd Edition. Oxford, U. K. Oxford University Press.

Goldman, A. I., Carlin, B. P., Crane, L. R., Launer, C., Korvick, J. A., Deyton, L. and Abrams, D. I. (1996). Response of CD4 lymphocytes and clinical consequences of treatment using ddI or ddC in patients with advanced HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **11**, 161–169.

Han, M., Song, X., Sun, L. and Liu, L. (2014). Joint modeling of longitudinal data with informative observation times and dropouts. *Statistica Sinica*, **24**(4), 1487–1504.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics*, **27**, 1536–1563.

Kalbfleisch, J. D., and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time data*, 2nd ed. Hoboken: Wiley.

Kong, S., Nan, B., Kalbfleisch, J. D., Saran, R., and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *Journal of the American Statistical Association*, **113**, 357–368.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer series in statistics. Springer.

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

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

Lin, D. Y., and Ying Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, **96,** 103–113.

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

Liu, L., Huang, X. and O′Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observation times and a dependent terminal event with application to medical cost data. *Biometrics*, **64**, 950–958.

Roy, J., and Lin, X. (2002), Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: changes in methadone treatment practices. *Journal of the American Statistical Association*, **97**, 40–52.

Schumaker, L. (2007), *Spline Functions: Basic Theory (Third Edition)*, Cambridge University Press.

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics*, **22**, 580–615.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, **8**, 1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10**, 1040–1053.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, **14**, 590–606.

Sun, J., Park, D., Sun, L., and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association*, **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. *Journal of the American Statistical Association*, **102** (480), 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. *Journal of the American Statistical Association*, **107** (498), 688–700.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van der Vaart, A. W. (2002). Semiparametric Statistics in *Lectures on Probability Theory and Statistics, Ecole d'Ete de Probabilites de Saint-Flour XXIX99*, ed. P. Bernard, Berlin: Springer-Verlag, pp.330–457.

van der Vaart, A. W. and Weller, J. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer-Verlag.

Wang, Y., and Taylor, M. G. (2001), Jointly modeling longitudinal and event time

data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, **96**, 895–905.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689–699.

Zhao, X., Deng, S., Liu, L., and Liu, L. (2014). Sieve estimation in semiparametric modeling of longitudinal data with informative observation times. *Biostatistics*, **15**(1), 140–153.

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

Zhou, S., Shen, X., and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Computational Statistics and Data Analysis*, **26**, 1760–1782.