

Additive-Multiplicative Rates Model for Recurrent Event Data with Intermittently Observed Time-Dependent Covariates

TIANMENG LYU¹, XIANGHUA LUO^{2,3}, AND YIFEI SUN^{4,*}

¹*Novartis Pharmaceuticals Corporation, East Hanover, NJ, U.S.A.*

²*Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, U.S.A.*

³*Biostatistics Core, Masonic Cancer Center, University of Minnesota, Minneapolis, MN, U.S.A.*

⁴*Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, U.S.A.*

Abstract

Regression methods, including the proportional rates model and additive rates model, have been proposed to evaluate the effect of covariates on the risk of recurrent events. These two models have different assumptions on the form of the covariate effects. A more flexible model, the additive-multiplicative rates model, is considered to allow the covariates to have both additive and multiplicative effects on the marginal rate of recurrent event process. However, its use is limited to the cases where the time-dependent covariates are monitored continuously throughout the follow-up time. In practice, time-dependent covariates are often only measured intermittently, which renders the current estimation method for the additive-multiplicative rates model inapplicable. In this paper, we propose a semiparametric estimator for the regression coefficients of the additive-multiplicative rates model to allow intermittently observed time-dependent covariates. We present the simulation results for the comparison between the proposed method and the simple methods, including last covariate carried forward and linear interpolation, and apply the proposed method to an epidemiologic study aiming to evaluate the effect of time-varying streptococcal infections on the risk of pharyngitis among school children. The R package implementing the proposed method is available at www.github.com/TianmengL/rectime.

Keywords *additive-multiplicative rates model; kernel smoothing; recurrent events; semiparametric method; time-dependent covariates*

1 Introduction

In various clinical and biomedical studies, the event of interest may happen multiple times, which is referred to as a recurrent event. Examples include recurrent bleedings in patients with hematologic malignancies (Stanworth et al., 2015) and recurrent cardiovascular events in subjects with diabetes (Van Der Heijden et al., 2013). During the follow-up of recurrent events, it is common to have repeated measurements of time-dependent covariates, and it is often of interest to investigate the effect of such covariates on the occurrence of recurrent events. Our motivation is from an observational study about pharyngitis among school children. Pharyngitis is often caused by viruses, but some bacteria, including streptococci, can cause pharyngitis as well. The goal of this study was to explore the effect of streptococci on the risk of pharyngitis. In this study, weekly visits were scheduled to monitor the recurrent occurrence of pharyngitis, and the status of streptococci infection was determined for those diagnosed with pharyngitis. In the meantime, monthly

*Corresponding author. Email: ys3072@cumc.columbia.edu.

visits were scheduled for each participant to monitor the streptococci infection status regularly.

For the analysis of recurrent events, Prentice et al. (1981) and Andersen and Gill (1982) proposed the multiplicative model on the intensity function, which is interpreted as the instantaneous risk of event conditioning on the event history. To achieve a better interpretation and to allow flexible dependence structure among the recurrent events, various authors (Pepe and Cai, 1993; Lawless et al., 1997; Lin et al., 2000) discussed the regression models on mean or rate function of recurrent event process and assumed that the covariate effects were in a multiplicative fashion. Alternatively, the covariate effects may add to the baseline function. Liu and Wu (2011) considered the additive intensity model, and Schaubel et al. (2006) proposed the additive rates model where the covariate effects were in absolute forms. The multiplicative and additive models have different assumptions on the relationship between the covariate effects and the event process; thus it is desirable to estimate both types of effects under a general model setting. For univariate survival data where the event of interest only happens once, Lin and Ying (1995) proposed the additive-multiplicative model for the hazard function and Scheike and Zhang (2002) considered the Cox-Aalen model to allow the covariate effects to be time-varying. Recently, Cai et al. (2017a) studied modeling additive and multiplicative effects simultaneously in the mean residual life function. For recurrent event data, Han et al. (2016) proposed the additive-multiplicative models focusing on the markers contingent on recurrent events with an informative terminal event. Liu et al. (2010) considered the additive-multiplicative rates models for recurrent events, which allows the covariates to have both multiplicative and additive effects on the rate function of the recurrent event process.

Although the additive-multiplicative rates model allows the covariates to be time-dependent, it requires the observation of the complete history of the covariates, which is nearly impossible in practice. Typically, the time-dependent covariates are measured intermittently during the follow-up. One way to deal with the infrequently updated covariates is to predict the missing covariate values by smoothing the observed values, as was discussed in Raboud et al. (1993); Tsiatis et al. (1995); Boscardin et al. (1998); Bycott and Taylor (1998); Dafni and Tsiatis (1998) and summarized in Andersen and Liestøl (2003). Another commonly used approach is to jointly model the longitudinal covariate process and the event process. The joint models of repeated measured longitudinal data and time-to-event data have been studied extensively. Many authors considered the joint models of the two processes through latent random effects, including Wulfsohn and Tsiatis (1997); Xu and Zeger (2001); Vonesh et al. (2006), among others. In the setting of recurrent event data, Henderson et al. (2000) proposed to model the relationship between the time-dependent covariates and recurrent events by a latent Gaussian process, and Li (2016) considered the joint model of the recurrent event process and the binary covariate process. More complex models which considered the covariate process, recurrent event process and the terminal event simultaneously also have been studied, including Kim et al. (2012) and Cai et al. (2017b) among others. To our knowledge, little research has been done to explore the additive-multiplicative rates model with intermittently observed time-dependent covariates.

Recently, Li et al. (2016) and Cao et al. (2015) have proposed kernel smoothed estimators for the proportional rates model and hazard model, respectively, with intermittently observed time-dependent covariates. Specifically, Cao et al. (2015) considered the scenario where the time-dependent covariates were not measured at event times but during regular visits only and Li et al. (2016) focused on the case where the covariates were measured at both recurrent times and regular visits. Lyu et al. (2021) and Sun et al. (2021a) considered kernel weighted estimation procedures for additive rates model and additive hazard regression model, respectively. In addition, Cao and Fine (2021) proposed a weighted last covariate carried forward approach

for proportional hazard model with time-dependent covariates not observed at failure times and Sun et al. (2021b) proposed an estimation procedure based on inverse-rate-weighting and kernel-smoothing to estimate the proportional rates model with intermittently observed time-dependent covariates measured at informative clinical visits. In this paper, we propose to extend the kernel smoothing method to the parameter estimation of the additive-multiplicative rates model. The proposed estimator is expected to be accurate; it relies on fewer assumptions of the underlying covariate and recurrent event processes.

The rest of the paper is organized as follows. The additive-multiplicative rates model and the proposed estimator are introduced in Section 2. Simulation studies to evaluate the performance of the proposed estimator are presented in Section 3. Section 4 includes the analysis of the pharyngitis data as introduced in the motivating example. Finally, a concluding remark is included in Section 5.

2 Model and the Proposed Estimator

2.1 Additive-Multiplicative Rates Model

Suppose that n subjects are recruited in a study. Let $i = 1, \dots, n$ index the subjects. Let $N_i^*(t)$ denote the number of events that subject i has experienced at or prior to time t when there is no censoring. Let $\mathbf{W}_i(t) = (\mathbf{Z}_i(t)^\top, \mathbf{X}_i(t)^\top)^\top$ be a $p \times 1$ vector of possibly time-dependent covariates and let $\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0^\top, \boldsymbol{\beta}_0^\top)^\top$ be the corresponding true regression parameters. Following Liu et al. (2010), we assume the rate function of the counting process $N_i^*(t)$ has the following form,

$$\lambda(t|\mathbf{W}_i(t)) = g\{\boldsymbol{\gamma}_0^\top \mathbf{Z}_i(t)\} + h\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(t)\}\lambda_0(t), \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline rate function, and the link functions g and h are assumed to be known. Specifically, if we let $g(x) = x$ and $h(x) = \exp(x)$, then model (1) becomes

$$\lambda(t|\mathbf{W}_i(t)) = \boldsymbol{\gamma}_0^\top \mathbf{Z}_i(t) + \exp\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(t)\}\lambda_0(t). \tag{2}$$

Therefore, the model can be regarded as a generalization of the semiparametric additive rates model and proportional rates model for recurrent event process.

Let C_i denote the censoring time of subject i and we assume that C_i is independent of the counting process $N_i^*(t)$ given $\mathbf{W}_i(t)$ in the sense that $E\{dN_i^*(t)|\mathbf{W}_i(t), C_i \geq t\} = E\{dN_i^*(t)|\mathbf{W}_i(t)\}$. Define $Y_i(t) = I(C_i \geq t)$ and denote by $N_i(t) = N_i^*(\min(t, C_i))$ the observed number of events. Let $[0, \tau]$ be a pre-specified time interval of interest, and the recurrent event process could potentially be observed beyond τ . For model estimation, we define the following process,

$$M_i(t, \boldsymbol{\theta}) = N_i(t) - \int_0^t Y_i(u) [g\{\boldsymbol{\gamma}^\top \mathbf{Z}_i(u)\}du + h\{\boldsymbol{\beta}^\top \mathbf{X}_i(u)\}d\mu_0(u)],$$

where $\mu_0(t) = \int_0^t \lambda_0(u)du$ is the baseline mean function. Following Lin and Ying (1995), Liu et al. (2010) proposed the following estimating function for model (1):

$$U(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{\mathbf{D}_i(\boldsymbol{\theta}, u) - \bar{\mathbf{D}}(\boldsymbol{\theta}, u)\}dM_i(u, \boldsymbol{\theta}), \tag{3}$$

where $\mathbf{D}_i(\boldsymbol{\theta}, t)$ is a p -dimensional smooth process involving $\mathbf{W}_i(t)$ and $\boldsymbol{\theta}$, and

$$\bar{\mathbf{D}}(\boldsymbol{\theta}, t) = \frac{\sum_{i=1}^n Y_i(t)\mathbf{D}_i(\boldsymbol{\theta}, t)h\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}}{\sum_{i=1}^n Y_i(t)h\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}}.$$

According to Lin and Ying (1995) and Liu et al. (2010), a possible choice for $\mathbf{D}_i(\boldsymbol{\theta}, t)$ is

$$\mathbf{D}_i(\boldsymbol{\theta}, t) = \begin{pmatrix} \frac{g'\{\boldsymbol{\gamma}^\top \mathbf{Z}_i(t)\} \mathbf{Z}_i(t)}{h\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}} \\ \frac{h'\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\} \mathbf{X}_i(t)}{h\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}} \end{pmatrix}.$$

For a given $\boldsymbol{\theta}$, by solving $\sum_{i=1}^n \int_0^t dM_i(u, \boldsymbol{\theta}) = 0$, the baseline mean function $\mu_0(t)$ in (3) can be estimated by

$$\widehat{\mu}_0(t, \boldsymbol{\theta}) = \int_0^t \frac{\sum_{i=1}^n [dN_i(u) - Y_i(u)g\{\boldsymbol{\gamma}^\top \mathbf{Z}_i(u)\}du]}{\sum_{i=1}^n Y_i(u)h\{\boldsymbol{\beta}^\top \mathbf{X}_i(u)\}}, \quad (4)$$

thus the estimating function in (3) becomes

$$U(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{\mathbf{D}_i(\boldsymbol{\theta}, u) - \bar{\mathbf{D}}(\boldsymbol{\theta}, u)\} [dN_i(u) - Y_i(u)g\{\boldsymbol{\gamma}^\top \mathbf{Z}_i(u)\}du]. \quad (5)$$

Since we are mostly interested in simultaneously modeling the relative and absolute difference in rate function due to the covariates, we focus on the model in (2) in the remainder of this paper. Under model (2), we have

$$\mathbf{D}_i(\boldsymbol{\theta}, t) = \begin{pmatrix} \mathbf{Z}_i(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(t)\} \\ \mathbf{X}_i(t) \end{pmatrix},$$

and thus

$$\bar{\mathbf{D}}(\boldsymbol{\theta}, t) = \begin{pmatrix} \frac{\sum_{i=1}^n Y_i(t) \mathbf{Z}_i(t)}{\sum_{i=1}^n Y_i(t) \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}} \\ \frac{\sum_{i=1}^n Y_i(t) \mathbf{X}_i(t) \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}}{\sum_{i=1}^n Y_i(t) \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}} \end{pmatrix}.$$

The estimating function in (5) can be written as

$$\begin{aligned} U(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{\mathbf{D}_i(\boldsymbol{\theta}, u) - \bar{\mathbf{D}}(\boldsymbol{\theta}, u)\} dN_i(u) - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{\mathbf{D}_i(\boldsymbol{\theta}, u) - \bar{\mathbf{D}}(\boldsymbol{\theta}, u)\} Y_i(u) \boldsymbol{\gamma}^\top \mathbf{Z}_i(u) du \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{\mathbf{D}_i(\boldsymbol{\theta}, u) - \bar{\mathbf{D}}(\boldsymbol{\theta}, u)\} dN_i(u) \\ &\quad - \int_0^\tau \frac{1}{n} \sum_{i=1}^n Y_i(u) \left\{ \frac{\sum_{j=1}^n \mathbf{D}_j(\boldsymbol{\theta}, u) Y_j(u) \mathbf{Z}_j^\top(u)}{\sum_{j=1}^n Y_j(u)} - \bar{\mathbf{D}}(\boldsymbol{\theta}, u) \frac{\sum_{j=1}^n Y_j(u) \mathbf{Z}_j^\top(u)}{\sum_{j=1}^n Y_j(u)} \right\} du \boldsymbol{\gamma}. \end{aligned} \quad (6)$$

Then the estimator $\widehat{\boldsymbol{\theta}}$ can be obtained by solving $U(\boldsymbol{\theta}) = \mathbf{0}$.

2.2 Estimation When Covariates Are Intermittently Observed

When time-dependent covariates are present, the evaluation of the estimating function requires the time-dependent covariates to be observed continuously throughout the entire follow-up time. However, as is illustrated in the motivating example, time-dependent covariates were measured at intermittent visits, including the event times and monthly regular visits. Specifically, denote by $O_i(t)$ the number of measurements of the covariates at regular visits at or prior to time t , and we have $dO_i(t) = 1$ when the i th subject has a regular visit at t . Moreover, we assume that the observation process $O_i(\cdot)$ is independent of $\{\mathbf{W}_i(\cdot), N_i(\cdot), C_i\}$, $i = 1, \dots, n$. The observed data are $\{N_i(t), O_i(t), \mathbf{W}_i(t) dN_i(t), \mathbf{W}_i(t) dO_i(t), C_i; 0 \leq t \leq C_i, i = 1, \dots, n\}$, which are assumed to be independent and identically distributed. Hence the values of covariates between measurement

times are unknown, which renders that the estimating function in (6) cannot be evaluated based on the observed data only. A simple approach to deal with the intermittently observed time-dependent covariates is to impute the missing values by carrying forward the last observed value. However, this approach imposes a strong assumption that the covariate processes are step functions and thus is expected to introduce bias in the model estimation in both survival and recurrent event analysis (Prentice, 1982; Faucett et al., 1998; Cao et al., 2015; Li et al., 2016; Lyu et al., 2021). Another possible approach is to impute the missing values between two observation times by linear interpolation. The linear interpolation method assumes that the covariate value is a linear function of time between every two adjacent observation times. However, this assumption may not hold in practice, especially for binary covariates. As an alternative approach, estimation methods based on kernel smoothing to deal with intermittently observed time-dependent covariates have gained growing interest recently (Cao et al., 2015; Li et al., 2016). In what follows, we propose a novel semiparametric estimator by applying the kernel smoothing method to deal with intermittently observed time-dependent covariates in the additive-multiplicative rates model. In contrast with methods that impute the missing values of each individual, the proposed method estimates the mean covariate processes via kernel smoothing and is expected to yield lower bias.

First, we show that the estimating function in (6) can be written as a function of empirical processes. Given a vector \mathbf{a} , we define the operator \otimes such that $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^\top$. For $k = 0, 1$, we define

$$\begin{aligned} S_z^{(k)}(t) &= n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{Z}_i(t)^{\otimes k}, \\ S_x^{(k)}(t, \boldsymbol{\beta}) &= n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{X}_i(t)^{\otimes k} \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}, \\ S_{z2x}(t, \boldsymbol{\beta}) &= n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{Z}_i(t)^{\otimes 2} \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}, \\ S_{zx}(t) &= n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{X}_i(t) \mathbf{Z}_i^\top(t). \end{aligned}$$

Then the estimating function in (6) can be re-expressed as $U(\boldsymbol{\theta}) = (U_1(\boldsymbol{\theta})^\top, U_2(\boldsymbol{\theta})^\top)^\top$, where

$$\begin{aligned} U_1(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(u) \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(u)\} dN_i(u) - \int_0^\tau \frac{S_z^{(1)}(u)}{S_x^{(0)}(u, \boldsymbol{\beta})} \left\{ \frac{1}{n} \sum_{i=1}^n dN_i(u) \right\} \\ &\quad - \int_0^\tau \frac{1}{n} \sum_{i=1}^n Y_i(u) \left\{ \frac{S_{z2x}(u, \boldsymbol{\beta})}{S_z^{(0)}(u)} - \frac{S_z^{(1)}(u)}{S_x^{(0)}(u, \boldsymbol{\beta})} \left(\frac{S_z^{(1)}(u)}{S_z^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}, \\ U_2(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \mathbf{X}_i(u) dN_i(u) - \int_0^\tau \frac{S_x^{(1)}(u, \boldsymbol{\beta})}{S_x^{(0)}(u, \boldsymbol{\beta})} \left\{ \frac{1}{n} \sum_{i=1}^n dN_i(u) \right\} \\ &\quad - \int_0^\tau \frac{1}{n} \sum_{i=1}^n Y_i(u) \left\{ \frac{S_{zx}(u)}{S_z^{(0)}(u)} - \frac{S_x^{(1)}(u, \boldsymbol{\beta})}{S_x^{(0)}(u, \boldsymbol{\beta})} \left(\frac{S_z^{(1)}(u)}{S_z^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}. \end{aligned} \tag{7}$$

In equation (7), the quantities $n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(t)\} dN_i(u)$ and $n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{X}_i(u) dN_i(u)$ can be easily calculated because the covariates are observed when recurrent events occur, as shown in the motivating example. However, calculating the ratios

$S_z^{(1)}(t)/S_x^{(0)}(t, \boldsymbol{\beta})$, $S_{z2x}(t, \boldsymbol{\beta})/S_z^{(0)}(t)$, $S_z^{(1)}(t)/S_z^{(0)}(t)$, $S_{zx}(t)/S_z^{(0)}(t)$, and $S_x^{(1)}(t, \boldsymbol{\beta})/S_x^{(0)}(t, \boldsymbol{\beta})$ requires the covariate process $\mathbf{W}(\cdot)$ to be observed throughout the follow-up period. Nevertheless, in real applications, it is usually not feasible to continuously monitor the covariates and the covariates are only measured intermittently.

Next, we present how to approximate these ratios based on the observed data by applying the kernel smoothing method. It can be easily seen that the processes $S_z^{(k)}(t)$, $S_x^{(k)}(t, \boldsymbol{\beta})$, $S_{z2x}(t, \boldsymbol{\beta})$, and $S_{zx}(t)$, converge in probability to the limiting functions $s_z^{(k)}(t) = E\{Y_i(t)\mathbf{Z}_i(t)^{\otimes k}\}$, $s_x^{(k)}(t, \boldsymbol{\beta}) = E[Y_i(t)\mathbf{X}_i(t)^{\otimes k} \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}]$, $s_{z2x}(t, \boldsymbol{\beta}) = E[Y_i(t)\mathbf{Z}_i(t)^{\otimes 2} \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}]$, and $s_{zx}(t) = E\{Y_i(t)\mathbf{X}_i(t)\mathbf{Z}_i^\top(t)\}$, respectively. With intermittently observed covariate data, we approximate the unknown ratios using kernel smoothing estimators that converge to the same target functions. Consider the following smoothed processes,

$$\begin{aligned}\widehat{S}_{z,h}^{(k)}(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^\infty K_h(t-u) Y_i(u) \mathbf{Z}_i(u)^{\otimes k} dO_i(u), \\ \widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n \int_0^\infty K_h(t-u) Y_i(u) \mathbf{X}_i(u)^{\otimes k} \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(u)\} dO_i(u), \\ \widehat{S}_{z2x,h}(t, \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n \int_0^\infty K_h(t-u) Y_i(u) \mathbf{Z}_i(u)^{\otimes 2} \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(u)\} dO_i(u), \\ \widehat{S}_{zx,h}(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^\infty K_h(t-u) Y_i(u) \mathbf{X}_i(u) \mathbf{Z}_i(u)^\top dO_i(u),\end{aligned}\tag{8}$$

where $K_h(t) = K(t/h)/h$, K is a second order kernel function with a bounded support on $[-1, 1]$, and h is the bandwidth parameter. For bandwidth selection, as is shown in the Appendix, the bandwidth $h = cn^{-v}$, where $1/4 < v < 1/2$, so we chose $h = cn^{-1/3}$ where the constant c was chosen by following the cross-validation approach as presented in Lyu et al. (2021).

We define $m(t) = E\{O_i(t)\}$, then it can be shown that the kernel smoothed processes, $\widehat{S}_{z,h}^{(k)}(t)$, $\widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta})$, $\widehat{S}_{z2x,h}(t, \boldsymbol{\beta})$, and $\widehat{S}_{zx,h}(t)$, converge in probability to $s_z^{(k)}(t)m(t)$, $s_x^{(k)}(t, \boldsymbol{\beta})m(t)$, $s_{z2x}(t, \boldsymbol{\beta})m(t)$, $s_{zx}(t)m(t)$, respectively. Since $m(t)$ cancels out in the ratios, the ratios of the kernel smoothed processes converge to the ratios of the corresponding limiting processes. Thus, the ratios $S_z^{(1)}(t)/S_x^{(0)}(t, \boldsymbol{\beta})$, $S_{z2x}(t, \boldsymbol{\beta})/S_z^{(0)}(t)$, $S_z^{(1)}(t)/S_z^{(0)}(t)$, $S_{zx}(t)/S_z^{(0)}(t)$ and $S_x^{(1)}(t, \boldsymbol{\beta})/S_x^{(0)}(t, \boldsymbol{\beta})$ in (7) can be replaced by the kernel smoothed counterparts. Moreover, to account for estimation bias near the boundary $t = 0$ due to positive observation times, we set $\widehat{S}_{z,h}^{(k)}(t) = \widehat{S}_{z,h}^{(k)}(h)$, $\widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta}) = \widehat{S}_{x,h}^{(k)}(h, \boldsymbol{\beta})$, $\widehat{S}_{z2x,h}(t, \boldsymbol{\beta}) = \widehat{S}_{z2x,h}(h, \boldsymbol{\beta})$, $\widehat{S}_{zx,h}(t) = \widehat{S}_{zx,h}(h)$ for $t \in [0, h)$. Then the proposed estimating function is $\widehat{U}(\boldsymbol{\theta}) = \begin{pmatrix} \widehat{U}_1(\boldsymbol{\theta}) \\ \widehat{U}_2(\boldsymbol{\theta}) \end{pmatrix}$, where

$$\begin{aligned}\widehat{U}_1(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(u) \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(u)\} dN_i(u) - \int_0^\tau \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta})} \left\{ \frac{1}{n} \sum_{i=1}^n dN_i(u) \right\} \\ &\quad - \int_0^\tau \frac{1}{n} \sum_{i=1}^n Y_i(u) \left\{ \frac{\widehat{S}_{z2x,h}(u, \boldsymbol{\beta})}{\widehat{S}_{z,h}^{(0)}(u)} - \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta})} \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}, \\ \widehat{U}_2(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \mathbf{X}_i(u) dN_i(u) - \int_0^\tau \frac{\widehat{S}_{x,h}^{(1)}(u, \boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta})} \left\{ \frac{1}{n} \sum_{i=1}^n dN_i(u) \right\} \\ &\quad - \int_0^\tau \frac{1}{n} \sum_{i=1}^n Y_i(u) \left\{ \frac{\widehat{S}_{zx,h}(u)}{\widehat{S}_{z,h}^{(0)}(u)} - \frac{\widehat{S}_{x,h}^{(1)}(u, \boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta})} \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}.\end{aligned}$$

The proposed estimator $\widehat{\boldsymbol{\theta}}_h = (\widehat{\boldsymbol{\gamma}}_h^\top, \widehat{\boldsymbol{\beta}}_h^\top)^\top$ can be obtained by solving $\widehat{U}(\boldsymbol{\theta}) = \mathbf{0}$. The large sample properties of $\widehat{\boldsymbol{\theta}}_h$ are summarized below. A detailed proof is included in the Appendix.

Theorem 1. *Under regularity conditions 1-10 in the Appendix, $\widehat{\boldsymbol{\theta}}_h$ converges in probability to $\boldsymbol{\theta}_0$. Moreover, as $n \rightarrow \infty$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}_0)$ converges in distribution to a zero mean normal random variable with variance $A(\boldsymbol{\theta}_0)^{-1}V(\boldsymbol{\theta}_0)\{A(\boldsymbol{\theta}_0)^{-1}\}^\top$, where $A(\boldsymbol{\theta}_0)$ and $V(\boldsymbol{\theta}_0)$ are defined in the Appendix.*

The asymptotic variance of $\widehat{\boldsymbol{\theta}}_h$ involves unknown nuisance functions that need to be estimated using kernel smoothing methods. Hence bootstrap is recommended for variance estimation because of its better finite-sample performance.

For the estimation of the baseline mean function, we show that the estimator in (4) can be written as

$$\begin{aligned} \widehat{\mu}_0(t, \boldsymbol{\theta}) &= \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\boldsymbol{\gamma}^\top \mathbf{Z}_i(u)du\}}{\sum_{i=1}^n Y_i(u) \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(u)\}} \\ &= \sum_{i=1}^n \int_0^t \frac{S_z^{(0)}(u)}{S_x^{(0)}(u, \boldsymbol{\beta})} \frac{1}{\sum_{i=1}^n Y_i(u)} dN_i(u) - \int_0^t \boldsymbol{\gamma}^\top \frac{S_z^{(1)}(u)}{S_x^{(0)}(u, \boldsymbol{\beta})} du. \end{aligned}$$

Thus, the baseline mean function can be estimated by

$$\widehat{\mu}_{0,h}(t, \widehat{\boldsymbol{\theta}}_h) = \sum_{i=1}^n \int_0^t \frac{\widehat{S}_{z,h}^{(0)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \widehat{\boldsymbol{\beta}}_h)} \frac{1}{\sum_{i=1}^n Y_i(u)} dN_i(u) - \int_0^t \widehat{\boldsymbol{\gamma}}_h^\top \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \widehat{\boldsymbol{\beta}}_h)} du.$$

Following Liu et al. (2010), to ensure that the estimated baseline mean function is monotone, an alternative estimator is $\widetilde{\mu}_{0,h}(t, \widehat{\boldsymbol{\theta}}_h) = \max_{0 \leq u \leq t} \widehat{\mu}_{0,h}(u, \widehat{\boldsymbol{\theta}}_h)$.

3 Simulation

Simulation studies were conducted to evaluate the performance of the proposed method. 1000 data replicates with sample size 100 and 200 were generated for each simulation scenario. The resampling size in the bootstrap method for variance estimation was set to be 50. Since the rate model does not fully specify the probability feature, we simulated the recurrent events based on the following intensity model

$$\lambda\{t|\mathbf{W}_i(t), u_i\} = u_i [\gamma_0 Z_i(t) + \exp\{\beta_0 X_i(t)\}\lambda_0(t)], \tag{9}$$

where u_i is the frailty variable with mean 1 and variance σ^2 . The frailty variable induces the within-subject correlations. We explored two distributions of the frailty variable: gamma distribution and log-normal distribution, and two values of the variance $\sigma^2 = 0.2, 0.4$, for different levels of correlations. The baseline intensity function $\lambda_0(t) = 0.3I(t \leq 10) + 0.5I(10 < t \leq 20)$. Note that the intensity model in (9) implies the rate model $\lambda\{t|\mathbf{W}_i(t)\} = \gamma_0 Z_i(t) + \exp\{\beta_0 X_i(t)\}\lambda_0(t)$. To evaluate the proposed method on different types of covariates, we let $Z_i(t)$ be a continuous covariate and $X_i(t)$ be a binary covariate. $Z_i(t)$ was simulated by a linear function of time t as $Z_i(t) = b_{0i} + b_{1i}t$. The random intercept b_{0i} was generated from a normal distribution with mean 0.5 and variance 0.05. The random slope was simulated from a normal distribution with mean -0.05 and variance $5 \cdot 10^{-4}$. With a negative mean of the random slope, the covariate $Z(\cdot)$ has a decreasing time trend at the population level. For the binary covariate $X_i(t)$, we first generated

the baseline $X_i(0)$ from a Bernoulli distribution with probability 0.2. Then the binary covariate process was assumed to alternate between states 0 and 1. We assumed that the duration of state 0 of subject i followed an exponential distribution with rate function $1/(\xi_i g(t))$ and the duration of state 1 followed an exponential distribution with rate $1/\xi_i$, where ξ_i was a subject-specific random effect which followed a gamma distribution with mean 1 and variance 0.25. The value of $g(t)$ was 4 for $t \in [0, 10]$ and changed to 6 afterwards, which indicates a decreasing time trend at the population level. For the values of regression coefficients, we considered three scenarios: (1) the true model included both an additive part and a multiplicative part: $\beta_0 = 0.5, \gamma_0 = 0.2$; (2) the additive-multiplicative model degenerated to the additive rates model: $\beta_0 = 0, \gamma_0 = 0.2$; (3) the additive-multiplicative degenerated to the proportional rates model: $\beta_0 = 0.5, \gamma_0 = 0$.

In all scenarios, we assumed that the covariates of a subject were always observed at the event times of the same subject. For the regular visits, we assumed that the covariates were measured at the baseline visit (time 0) and at each pre-scheduled regular visit per unit time interval, which means that there were 20 regular visits in time period $(0, 20]$ of each subject. The time of the regular visit in each unit time interval was simulated from a uniform distribution from 0 to 1. The censoring time was randomly generated from a uniform distribution from 0 to 20. Both the event times and regular visits happened after the censoring time were dropped.

We applied the proposed method to the simulated data and compared the results with those from the LCCF and linear interpolation approaches. We used the Epanechnikov kernel function in the proposed kernel smoothing method. The results for simulated datasets with gamma or lognormal frailty are presented in Tables 1 and 2, respectively. We provide the relative bias (Bias) and Monte Carlo standard deviation (SD) of the point estimations for each estimation method that we compared. For the proposed method, we also report the average of the estimated standard errors by bootstrap method (ASE) and the coverage percentage (CP). When the true model is the additive-multiplicative model ($\beta_0 = 0.5, \gamma_0 = 0.2$), the LCCF method gives biased estimations for both β and γ . The linear interpolation method has small bias for the estimation of γ , which is likely due to the linear feature of covariate $Z(\cdot)$, but gives biased estimations for β . The proposed method provides virtually unbiased estimations for both regression coefficients. The ASEs are close to the empirical SDs and the coverage percentages are around 95%. As the variance of the frailty increases from 0.2 to 0.4, the Monte Carlo SDs of the estimations increases. When the true model degenerates to the additive model or the multiplicative model with one significant covariate ($\beta = 0, \gamma = 0.2$; $\beta = 0.5, \gamma = 0$), the proposed method provides virtually unbiased estimations for both coefficients, which indicates the robustness of the additive-multiplicative model.

4 Real Data Analysis

In this section, we analyzed the Indian pharyngitis data (Jose et al., 2018) using the proposed method. Pharyngitis is the infection of the back of the throat and it can be caused by viruses or bacteria. When the cause is group A streptococcus (GAS), pharyngitis is also known as strep throat. The symptoms of GAS pharyngitis include sore throat, fever, nausea and it may cause some rare but serious diseases including rheumatic heart disease if left untreated. GAS pharyngitis is common in children from age 5 to age 15 and can be transmitted through saliva or nasal secretions. Meanwhile, other bacteria, for example, group G streptococcus (GGS), may cause pharyngitis with similar clinical symptoms as well. In the motivating example, 307 school children aged 7 to 11 years old in a rural area in Velore, India were recruited to investigate

Table 1: Simulation results: the frailty followed a gamma distribution; n is the sample size; σ^2 is the variance of the frailty distribution; \bar{m} is the average number of recurrent events; Bias is the relative bias computed by dividing the difference of the mean of the 1000 estimated parameters and the true value by the true value (if the true value is 0, Bias is the mean of the 1000 estimated parameters); SD is the standard deviation of the 1000 estimated values; ASE is the mean of the 1000 estimated standard errors by bootstrap method; CP is the proportion of 95% confidence intervals covering the true value.

n	σ^2	\bar{m}		LCCF		Linear		Proposed			
				Bias	SD	Bias	SD	Bias	SD	ASE	CP
$\beta = 0.5, \gamma = 0.2$											
100	0.2	4.23	β	-0.136	0.126	0.145	0.158	0.012	0.144	0.148	0.933
			γ	-0.575	0.109	0.009	0.104	0.034	0.105	0.103	0.937
	0.4	4.24	β	-0.131	0.139	0.150	0.176	0.011	0.157	0.155	0.948
			γ	-0.621	0.128	-0.033	0.123	-0.013	0.123	0.123	0.938
200	0.2	4.25	β	-0.133	0.092	0.153	0.114	0.009	0.103	0.102	0.929
			γ	-0.570	0.082	0.006	0.078	0.021	0.078	0.073	0.931
	0.4	4.24	β	-0.139	0.099	0.143	0.124	0.000	0.111	0.108	0.946
			γ	-0.602	0.096	-0.023	0.092	-0.009	0.092	0.088	0.935
$\beta = 0, \gamma = 0.2$											
100	0.2	3.82	β	-0.012	0.148	-0.005	0.195	-0.003	0.169	0.172	0.953
			γ	-0.540	0.105	-0.008	0.100	0.033	0.101	0.100	0.934
	0.4	3.83	β	-0.016	0.163	-0.011	0.216	-0.006	0.181	0.183	0.946
			γ	-0.526	0.128	0.008	0.123	0.051	0.123	0.119	0.938
200	0.2	3.84	β	-0.010	0.101	-0.002	0.131	-0.002	0.113	0.120	0.952
			γ	-0.546	0.075	-0.021	0.073	0.016	0.073	0.070	0.943
	0.4	3.82	β	-0.017	0.106	-0.009	0.140	-0.006	0.119	0.125	0.968
			γ	-0.550	0.092	-0.024	0.089	0.012	0.089	0.086	0.937
$\beta = 0.5, \gamma = 0$											
100	0.2	3.92	β	-0.148	0.122	0.143	0.152	0.007	0.136	0.143	0.955
			γ	-0.112	0.108	0.001	0.103	0.006	0.103	0.099	0.931
	0.4	3.90	β	-0.163	0.136	0.121	0.172	-0.021	0.153	0.150	0.938
			γ	-0.124	0.131	-0.011	0.124	-0.006	0.125	0.120	0.930
200	0.2	3.91	β	-0.138	0.086	0.157	0.107	0.013	0.097	0.098	0.952
			γ	-0.113	0.074	-0.002	0.070	0.002	0.071	0.071	0.940
	0.4	3.91	β	-0.136	0.093	0.161	0.117	0.015	0.103	0.104	0.949
			γ	-0.113	0.096	-0.002	0.090	0.002	0.090	0.086	0.929

the relationship between streptococcal infections and the risk of pharyngitis. Each child was examined weekly for the symptoms of pharyngitis. For those who were diagnosed with pharyngitis, throat cultures were obtained to test if GAS and GGS were positive. In the meantime, to monitor the streptococci status regularly, monthly regular visits were scheduled for each child. Since the regular visits were pre-scheduled, it is reasonable to assume that they are independent of the recurrent event, censoring and covariate processes, as is required by the proposed method.

Table 2: Simulation results: the frailty followed a lognormal distribution; n is the sample size; σ^2 is the variance of the frailty distribution; \bar{m} is the average number of recurrent events; Bias is the relative bias computed by dividing the difference of the mean of the 1000 estimated parameters and the true value by the true value (if the true value is 0, Bias is the mean of the 1000 estimated parameters); SD is the standard deviation of the 1000 estimated values; ASE is the mean of the 1000 estimated standard errors by bootstrap method; CP is the proportion of 95% confidence intervals covering the true value.

n	σ^2	\bar{m}		LCCF		Linear		Proposed			
				Bias	SD	Bias	SD	Bias	SD	ASE	CP
$\beta = 0.5, \gamma = 0.2$											
100	0.2	4.24	β	-0.141	0.130	0.141	0.161	0.004	0.145	0.147	0.950
			γ	-0.605	0.110	-0.021	0.106	0.005	0.106	0.102	0.930
	0.4	4.22	β	-0.141	0.135	0.140	0.167	0.004	0.150	0.157	0.954
			γ	-0.575	0.134	0.010	0.129	0.032	0.130	0.123	0.938
200	0.2	4.25	β	-0.141	0.088	0.141	0.110	0.006	0.100	0.102	0.943
			γ	-0.560	0.078	0.015	0.076	0.030	0.076	0.074	0.940
	0.4	4.24	β	-0.130	0.093	0.152	0.114	0.007	0.102	0.107	0.959
			γ	-0.584	0.093	-0.008	0.090	0.010	0.090	0.088	0.943
$\beta = 0, \gamma = 0.2$											
100	0.2	3.83	β	-0.029	0.160	-0.027	0.210	-0.019	0.177	0.173	0.934
			γ	-0.554	0.105	-0.020	0.101	0.026	0.102	0.099	0.940
	0.4	3.82	β	-0.016	0.164	-0.012	0.215	-0.008	0.181	0.180	0.947
			γ	-0.568	0.130	-0.034	0.126	0.010	0.127	0.117	0.932
200	0.2	3.82	β	-0.019	0.107	-0.013	0.140	-0.010	0.120	0.120	0.944
			γ	-0.524	0.073	-0.001	0.071	0.037	0.071	0.070	0.939
	0.4	3.82	β	-0.014	0.110	-0.007	0.144	-0.003	0.122	0.125	0.941
			γ	-0.559	0.092	-0.033	0.089	0.001	0.089	0.085	0.940
$\beta = 0.5, \gamma = 0$											
100	0.2	3.90	β	-0.149	0.118	0.142	0.147	0.007	0.134	0.140	0.954
			γ	-0.112	0.108	0.001	0.102	0.006	0.103	0.099	0.931
	0.4	3.91	β	-0.155	0.129	0.133	0.161	-0.009	0.146	0.150	0.951
			γ	-0.111	0.127	0.002	0.120	0.007	0.121	0.119	0.947
200	0.2	3.91	β	-0.146	0.086	0.148	0.108	0.005	0.098	0.098	0.957
			γ	-0.110	0.075	0.001	0.071	0.004	0.071	0.070	0.944
	0.4	3.91	β	-0.147	0.093	0.146	0.116	0.000	0.103	0.104	0.948
			γ	-0.113	0.092	-0.002	0.087	0.003	0.087	0.086	0.945

We considered the following four candidate models: (1) both covariates have multiplicative effects (Model MM); (2) both covariates have additive effects (Model AA); (3) GAS has an additive effect and GGS has a multiplicative effect (Model AM); (4) GAS has a multiplicative effect and GGS has an additive effect (Model MA). The same approach as described in Section 2 was applied to select the bandwidth used in the proposed kernel smoothing method. Table 3 shows the regression results of the four models. All four models suggest that the presence of

Table 3: Analysis of Indian pharyngitis data. AM is the additive-multiplicative rates model which includes GAS in the additive part and GGS in the multiplicative part; MA is the additive-multiplicative rates model which includes GAS in the multiplicative part and GGS in the additive part; MM is the proportional rates model; AA is the additive rates model. Est is the estimated regression coefficient; SE is the standard error estimated by bootstrap with resampling size 100; CI is the 95% confidence interval.

Model	GAS			GGS		
	Est	SE	95% CI	Est	SE	95% CI
MM	0.418	0.117	(0.189, 0.647)	0.146	0.118	(-0.085, 0.377)
AA	0.067	0.020	(0.028, 0.106)	0.020	0.017	(-0.013, 0.053)
AM	0.067	0.019	(0.030, 0.104)	0.148	0.121	(-0.089, 0.385)
MA	0.437	0.120	(0.202, 0.672)	0.021	0.023	(-0.024, 0.066)

GAS increases the risk of pharyngitis, while the effect of GGS is not significant in any model. In light of the same direction of the effect and the similar statistical significance from the four candidate models, the choice would be based on study objectives and the interpretation of the effect. As discussed in Schaubel et al. (2006), in certain settings, the absolute covariate effect is of more interest than the relative covariate effect. For instance, the former can directly provide information for predicting change in event rate attributable to a covariate, while the latter would need information on the baseline rate.

5 Discussion

In this paper, we proposed a semiparametric estimator for the regression coefficients of the additive-multiplicative rates model to deal with the intermittently observed time-dependent covariates. The additive-multiplicative rates model generalizes the proportional rates model and additive rates model, and hence allows some covariates to have multiplicative effects on the risk of recurrent events and others to have additive effects. The proposed method applies the nonparametric kernel smoothing approach to estimate the mean processes of the time-dependent covariates, thus it does not rely on any assumption of the covariate distribution or any specification of the covariate process and is expected to be more robust. The proposed method requires that the rate function $m(t)$ of the observation time process is positive and bounded on $[0, \tau]$. As the observations of time-dependent covariates become more dense, we expect better performance of the proposed method.

In this paper, we assume that the observation process is independent of the covariates, the recurrent event process, and the censoring time. When such an independence assumption is violated, the proposed method may yield biased estimation. When the rate function of the observation process is determined by observed covariates, one can assume a proportional rate model on the observation process and construct an inverse-rate-weighting in the kernel smoothing estimator for the outcome model, following the arguments of Sun et al. (2021b). Extending their method to additive-multiplicative rates model is a future research direction.

In practice, a problem is to determine the covariates included in the additive part $\mathbf{Z}(t)$ and the multiplicative part $\mathbf{X}(t)$. As discussed in Liu et al. (2010), if a covariate is expected to greatly influence the risk difference or the researchers are most interested in the absolute risks,

then it should be included in $\mathbf{Z}(t)$. Otherwise, if a covariate is expected to strongly influence the risk ratios or the researchers are interested in the relative risks, then it should be included in $\mathbf{X}(t)$. If the underlying biological process is not clear and the number of covariates is small, an alternative way is to consider all the possible candidate models and then develop rigorous model selection approach to determine the best model. The mean-square-type distance measure between the observed and expected recurrences implemented in Liu et al. (2010) could be a model selection criteria for the additive and/or multiplicative model structure. However, it cannot be directly applied to data with intermittently observed time-dependent covariates. Future research on model selection procedures for intermittently observed time-dependent covariates is warranted.

An R package has been developed to implement the proposed estimator for the additive-multiplicative rates model with intermittently observed time-dependent covariates, as well as the estimators for the proportional rates model and the additive rates model with such covariates. The R package `rectime` is available at www.github.com/TianmengL/rectime.

Supplementary Material

The supplementary material includes the R code that implements the proposed methods. It also includes an example file to illustrate how to simulate data and estimate model parameters using the provided code files.

Appendix

Regularity Conditions

1. The observed data $\{N_i(t), O_i(t), \mathbf{W}_i(t)dN_i(t), \mathbf{W}_i(t)dO_i(t), C_i; 0 \leq t \leq C_i, i = 1, \dots, n\}$ are assumed to be independent and identically distributed.
2. $N_i(\tau)$ is bounded. Define $\lambda^c(t)dt = E\{dN_i(t)\}$, and $\lambda^c(\cdot)$ is of bounded variation.
3. The true parameter $\boldsymbol{\theta}_0$ is in a compact set $\Theta \subset \mathcal{R}^p$ and the baseline rate function $\lambda_0(t)$ is absolutely continuous.
4. For each element in the covariate vector $\mathbf{W}_i(t)$, the covariate process $W_{ij}(t)$ has uniformly bounded total variation, namely, $\int_0^\tau |dW_{ij}(t)| + |W_{ij}(0)| \leq M$ for some $M > 0$ for all i and j . Without loss of generality, we assume $W_{ij}(t) \geq 0$.
5. The censoring time C_i is independent of $N_i^*(\cdot)$ conditional on $\mathbf{W}_i(\cdot)$ in the sense that $E\{dN_i^*(t)|\mathbf{W}_i(t), C_i \geq t\} = E\{dN_i^*(t)|\mathbf{W}_i(t)\}$, and $G(\tau) = P(C_i \geq \tau) > 0$.
6. The functions $s_{z2x}(t, \boldsymbol{\beta}) = E[Y_i(t)\mathbf{Z}_i(t)^{\otimes 2} \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}]$, $s_{zx}(t) = E[Y_i(t)\mathbf{X}_i(t)\mathbf{Z}_i(t)^\top]$, $s_z^{(k)}(t) = E[Y_i(t)\mathbf{Z}_i(t)^k]$, and $s_x^{(k)}(t, \boldsymbol{\beta}) = E[Y_i(t)\mathbf{X}_i(t)^k \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}]$, $k = 0, 1$, have bounded second derivatives for $t \in [0, \tau]$.
7. The observation time process $O_i(\cdot)$ is independent of $\{\mathbf{W}_i(\cdot), N_i(\cdot), C_i\}$ and is bounded. Moreover, its rate function $m(t)$ is positive and has bounded second derivative for $t \in [0, \tau]$.
8. The matrix $A(\boldsymbol{\theta}_0) = -E\left\{\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\}$ is nonsingular.
9. The kernel function $K(\cdot)$ is a symmetric density function with bounded support on $[-1, 1]$ which satisfies: $\int_{-1}^1 K(t)dt = 1$, $\int_{-1}^1 tK(t)dt = 0$, and $\int_{-1}^1 t^2K(t)dt$ is a positive constant.
10. The bandwidth $h = cn^{-v}$, where $1/4 < v < 1/2$ and $c > 0$ is some constant.

Proof of Consistency

To show the consistency of $\widehat{\boldsymbol{\theta}}_h$, it is sufficient to prove that the processes that constitute the estimating function $\widehat{U}(\boldsymbol{\theta})$, including $n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(u) \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(u)\} dN_i(u)$, $n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{X}_i(u) dN_i(u)$, $n^{-1} \sum_{i=1}^n dN_i(u)$, $\widehat{S}_{z,h}^{(k)}(t)$, $\widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta})$, $k = 0, 1$, $\widehat{S}_{z2x,h}(t, \boldsymbol{\beta})$ and $\widehat{S}_{zx,h}(t)$, converge to their limits uniformly. We know $\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0^\top, \boldsymbol{\beta}_0^\top)^\top$ where $\boldsymbol{\gamma}_0$ is a $m \times 1$ vector, $\boldsymbol{\beta}_0$ is a $q \times 1$ vector and $m + q = p$. Since $\boldsymbol{\theta}_0$ is in a compact set Θ in \mathcal{R}^p by assumption 3, $\boldsymbol{\beta}_0$ is contained in a compact set \mathcal{B} in \mathcal{R}^q . The function classes $\mathcal{F}_{z,k} = \{\int_0^t Y(u) \mathbf{Z}(u)^{\otimes k} dO(u) : t \in [0, \tau]\}$, $\mathcal{F}_{zx} = \{\int_0^t Y(u) \mathbf{X}(u) \mathbf{Z}(u)^\top dO(u) : t \in [0, \tau]\}$ are monotone and $\mathcal{F}_{x,k} = \{\int_0^t Y(u) \mathbf{X}(u)^{\otimes k} \exp\{\boldsymbol{\beta}^\top \mathbf{X}(u)\} dO(u) : \boldsymbol{\beta} \in \mathcal{B}, t \in [0, \tau]\}$, $\mathcal{F}_{z2x} = \{\int_0^t Y(u) \mathbf{Z}(u)^{\otimes 2} \exp\{-\boldsymbol{\beta}^\top \mathbf{X}(u)\} dO(u) : \boldsymbol{\beta} \in \mathcal{B}, t \in [0, \tau]\}$ have bracketing number of polynomial order.

According to Theorem 2.14.9 in Van Der Vaart and Wellner (1996) and following similar steps in the Appendix of Li et al. (2016), we can show that $\sup_{t \in [h, \tau]} |\widehat{S}_{z,h}^{(k)}(t) - E\{\widehat{S}_{z,h}^{(k)}(t)\}|$, $\sup_{\boldsymbol{\beta} \in \mathcal{B}, t \in [h, \tau]} |\widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta}) - E\{\widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta})\}|$, $\sup_{\boldsymbol{\beta} \in \mathcal{B}, t \in [h, \tau]} |\widehat{S}_{z2x,h}^{(k)}(t, \boldsymbol{\beta}) - E\{\widehat{S}_{z2x,h}^{(k)}(t, \boldsymbol{\beta})\}|$, $\sup_{t \in [h, \tau]} |\widehat{S}_{zx,h}^{(k)}(t) - E\{\widehat{S}_{zx,h}^{(k)}(t)\}|$ converge in probability to 0 when $nh^2 \rightarrow \infty$.

It can be seen that

$$\begin{aligned} \sup_{t \in [h, \tau]} |E\{\widehat{S}_{z,h}^{(k)}(t)\} - s_z^{(k)}(t)m(t)| &= O(h^2), \\ \sup_{\boldsymbol{\beta} \in \mathcal{B}, t \in [h, \tau]} |E\{\widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta})\} - s_x^{(k)}(t, \boldsymbol{\beta})m(t)| &= O(h^2), \\ \sup_{\boldsymbol{\beta} \in \mathcal{B}, t \in [h, \tau]} |E\{\widehat{S}_{z2x,h}^{(k)}(t, \boldsymbol{\beta})\} - s_{z2x}^{(k)}(t, \boldsymbol{\beta})m(t)| &= O(h^2), \\ \sup_{t \in [h, \tau]} |E\{\widehat{S}_{zx,h}^{(k)}(t)\} - s_{zx}^{(k)}(t)m(t)| &= O(h^2), \end{aligned}$$

and

$$\begin{aligned} \sup_{t \in [0, h]} |s_z^{(k)}(t)m(t) - s_z^{(k)}(h)m(h)| &= O(h), \\ \sup_{\boldsymbol{\beta} \in \mathcal{B}, t \in [0, h]} |s_x^{(k)}(t, \boldsymbol{\beta})m(t) - s_x^{(k)}(h, \boldsymbol{\beta})m(h)| &= O(h), \\ \sup_{\boldsymbol{\beta} \in \mathcal{B}, t \in [0, h]} |s_{z2x}^{(k)}(t, \boldsymbol{\beta})m(t) - s_{z2x}^{(k)}(h, \boldsymbol{\beta})m(h)| &= O(h), \\ \sup_{t \in [0, h]} |s_{zx}^{(k)}(t)m(t) - s_{zx}^{(k)}(h)m(h)| &= O(h). \end{aligned}$$

Then the uniform consistency of $\widehat{S}_{z,h}^{(k)}(t)$, $\widehat{S}_{x,h}^{(k)}(t, \boldsymbol{\beta})$, $k = 0, 1$, $\widehat{S}_{z2x,h}(t, \boldsymbol{\beta})$ and $\widehat{S}_{zx,h}(t)$ have been proved. By the law of large numbers, we can show that $n^{-1} \sum_{i=1}^n N_i(t)$, $n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(u) \times \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(u)\} dN_i(u)$, and $n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{X}_i(u) dN_i(u)$ converge in probability to $E\{N_i(t)\}$, $\int_0^\tau E\{\mathbf{Z}_i(t) \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(t)\} dN_i(t)\}$, and $\int_0^\tau E\{\mathbf{X}_i(t) dN_i(t)\}$, respectively.

Therefore, the proposed estimator $\widehat{\boldsymbol{\theta}}_h$ converges to the true parameter $\boldsymbol{\theta}_0$ in probability.

Proof of Asymptotic Normality

We prove the asymptotic normality of $\sqrt{n}\widehat{U}(\boldsymbol{\theta}_0) = (\sqrt{n}\widehat{U}_1(\boldsymbol{\theta}_0)^\top, \sqrt{n}\widehat{U}_2(\boldsymbol{\theta}_0)^\top)^\top$. We show that $\sqrt{n}\widehat{U}_1(\boldsymbol{\theta}_0)$ has the form

$$\begin{aligned} \sqrt{n}\widehat{U}_1(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(u) \exp\{-\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} dN_i(u) - \frac{1}{\sqrt{n}} \int_0^\tau \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta}_0)} \left\{ \sum_{i=1}^n dN_i(u) \right\} \\ &\quad - \frac{1}{\sqrt{n}} \int_0^\tau \sum_{i=1}^n Y_i(u) \left\{ \frac{\widehat{S}_{z2x,h}(u, \boldsymbol{\beta}_0)}{\widehat{S}_{z,h}^{(0)}(u)} \right\} du \boldsymbol{\gamma}_0 \\ &\quad + \frac{1}{\sqrt{n}} \int_0^\tau \sum_{i=1}^n Y_i(u) \left\{ \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta}_0)} \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}_0 \\ &\stackrel{\text{def}}{=} I_1 + I_2 + I_3 + I_4. \end{aligned}$$

Then we have

$$\begin{aligned} I_2 &= -\frac{1}{\sqrt{n}} \int_0^\tau \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta}_0)} \left\{ \sum_{i=1}^n dN_i(u) \right\} \\ &= -\sqrt{n} \int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} d \left\{ \frac{1}{n} \sum_{i=1}^n N_i(u) - \mathbb{E}\{N_i(u)\} \right\} - \sqrt{n} \int_0^\tau \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta}_0)} d\mathbb{E}\{N_i(u)\} \\ &\quad + o_p(1) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau Y_i(u) \left(\mathbf{Z}_i(u) - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \exp\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} \right) \frac{\lambda^c(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0) m(u)} dO_i(u) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} dN_i(u) + o_p(1), \\ I_3 &= -\frac{1}{\sqrt{n}} \int_0^\tau \sum_{i=1}^n Y_i(u) \left\{ \frac{\widehat{S}_{z2x,h}(u, \boldsymbol{\beta}_0)}{\widehat{S}_{z,h}^{(0)}(u)} \right\} du \boldsymbol{\gamma}_0 \\ &= -\sqrt{n} \int_0^\tau \frac{s_{z2x,h}(u, \boldsymbol{\beta}_0)}{s_z^{(0)}(u)} \left\{ \frac{1}{n} \sum_{i=1}^n Y_i(u) - s_z^{(0)}(u) \right\} du \boldsymbol{\gamma}_0 - \sqrt{n} \int_0^\tau \left\{ \frac{\widehat{S}_{z2x,h}(u, \boldsymbol{\beta}_0)}{\widehat{S}_{z,h}^{(0)}(u)} \right\} s_z^{(0)}(u) du \boldsymbol{\gamma}_0 \\ &\quad + o_p(1) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau Y_i(u) \left(\mathbf{Z}_i(u)^{\otimes 2} \exp\{-\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} - \frac{s_{z2x,h}(u, \boldsymbol{\beta}_0)}{s_z^{(0)}(u)} \right) \frac{1}{m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau Y_i(u) \frac{s_{z2x,h}(u, \boldsymbol{\beta}_0)}{s_z^{(0)}(u)} du \boldsymbol{\gamma}_0 + o_p(1), \end{aligned}$$

and

$$\begin{aligned}
 I_4 &= \frac{1}{\sqrt{n}} \int_0^\tau \sum_{i=1}^n Y_i(u) \left\{ \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta}_0)} \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}_0 \\
 &= \sqrt{n} \int_0^\tau \left\{ \frac{1}{n} \sum_{i=1}^n Y_i(u) - s_z^{(0)}(u) \right\} \left\{ \frac{s_{z,h}^{(1)}(u)}{s_{x,h}^{(0)}(u, \boldsymbol{\beta}_0)} \left(\frac{s_{z,h}^{(1)}(u)}{s_{z,h}^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}_0 \\
 &\quad + \sqrt{n} \int_0^\tau s_z^{(0)}(u) \left\{ \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u, \boldsymbol{\beta}_0)} \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}_0 + o_p(1) \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} Y_i(u) \left\{ \mathbf{Z}_i(u)^\top - \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} \frac{1}{m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau Y_i(u) \left(\mathbf{Z}_i(u) - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \exp\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} \right) \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \frac{s_z^{(0)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0) m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau Y_i(u) \left\{ \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}_0 + o_p(1).
 \end{aligned}$$

Thus we have

$$\sqrt{n} \widehat{U}_1(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i^{(1)}(\boldsymbol{\theta}_0) + o_p(1),$$

where $\phi_i^{(1)}(\boldsymbol{\theta}_0) = \int_0^\tau \mathbf{Z}_i(u) \exp\{-\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} dN_i(u) + \phi_{2i}^{(1)}(\boldsymbol{\theta}_0) + \phi_{3i}^{(1)}(\boldsymbol{\theta}_0) + \phi_{4i}^{(1)}(\boldsymbol{\theta}_0) + o_p(1)$ and

$$\begin{aligned}
 \phi_{2i}^{(1)}(\boldsymbol{\theta}_0) &= - \int_0^\tau Y_i(u) \left(\mathbf{Z}_i(u) - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \exp\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} \right) \frac{\lambda^c(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0) m(u)} dO_i(u) \\
 &\quad - \int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} dN_i(u), \\
 \phi_{3i}^{(1)}(\boldsymbol{\theta}_0) &= - \int_0^\tau Y_i(u) \left(\mathbf{Z}_i(u)^{\otimes 2} \exp\{-\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} - \frac{s_{z2x,h}(u, \boldsymbol{\beta}_0)}{s_z^{(0)}(u)} \right) \frac{1}{m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\
 &\quad - \int_0^\tau Y_i(u) \frac{s_{z2x}(u, \boldsymbol{\beta}_0)}{s_z^{(0)}(u)} du \boldsymbol{\gamma}_0, \\
 \phi_{4i}^{(1)}(\boldsymbol{\theta}_0) &= \int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} Y_i(u) \left\{ \mathbf{Z}_i(u)^\top - \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} \frac{1}{m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\
 &\quad + \int_0^\tau Y_i(u) \left(\mathbf{Z}_i(u) - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \exp\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} \right) \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \frac{s_z^{(0)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0) m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\
 &\quad + \int_0^\tau Y_i(u) \left\{ \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}_0.
 \end{aligned}$$

Similarly, it can be shown that $\sqrt{n} \widehat{U}_2(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i^{(2)}(\boldsymbol{\theta}_0) + o_p(1)$, where $\phi_i^{(2)}(\boldsymbol{\theta}_0) =$

$\int_0^\tau \mathbf{X}_i(u) dN_i(u) + \phi_{2i}^{(2)}(\boldsymbol{\theta}_0) + \phi_{3i}^{(2)}(\boldsymbol{\theta}_0) + \phi_{4i}^{(2)}(\boldsymbol{\theta}_0) + o_p(1)$, and

$$\begin{aligned} \phi_{2i}^{(2)}(\boldsymbol{\theta}_0) &= - \int_0^\tau Y_i(u) \left(\mathbf{X}_i(u) - \frac{s_x^{(1)}(u, \boldsymbol{\beta}_0)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \right) \exp\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\} \frac{\lambda^c(u)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)m(u)} dO_i(u) \\ &\quad - \int_0^\tau \frac{s_x^{(1)}(u, \boldsymbol{\beta}_0)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} dN_i(u), \\ \phi_{3i}^{(2)}(\boldsymbol{\theta}_0) &= - \int_0^\tau Y_i(u) \left(\mathbf{X}_i(u) \mathbf{Z}_i(u)^\top - \frac{s_{zx}(u)}{s_z^{(0)}(u)} \right) \frac{1}{m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\ &\quad - \int_0^\tau Y_i(u) \frac{s_{zx}(u)}{s_z^{(0)}(u)} du \boldsymbol{\gamma}_0, \\ \phi_{4i}^{(2)}(\boldsymbol{\theta}_0) &= \int_0^\tau \frac{s_x^{(1)}(u, \boldsymbol{\beta}_0)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} Y_i(u) \left\{ \mathbf{Z}_i(u)^\top - \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} \frac{1}{m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\ &\quad + \int_0^\tau Y_i(u) \left(\mathbf{X}_i(u) - \frac{s_x^{(1)}(u, \boldsymbol{\beta}_0)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \right) \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \frac{s_z^{(0)}(u) \exp\{\boldsymbol{\beta}_0^\top \mathbf{X}_i(u)\}}{s_x^{(0)}(u, \boldsymbol{\beta}_0)m(u)} dO_i(u) \boldsymbol{\gamma}_0 \\ &\quad + \int_0^\tau Y_i(u) \left\{ \frac{s_x^{(1)}(u, \boldsymbol{\beta}_0)}{s_x^{(0)}(u, \boldsymbol{\beta}_0)} \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} du \boldsymbol{\gamma}_0. \end{aligned}$$

Define $\phi_i(\boldsymbol{\theta}_0) = (\phi_i^{(1)}(\boldsymbol{\theta}_0)^\top, \phi_i^{(2)}(\boldsymbol{\theta}_0)^\top)^\top$, then we have $\sqrt{n}\widehat{U}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i(\boldsymbol{\theta}_0) + o_p(1)$.

Define $A(\boldsymbol{\theta}) = E\left\{-\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} = \begin{pmatrix} E\left\{-\frac{\partial U_1(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}\right\} & E\left\{-\frac{\partial U_1(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\right\} \\ E\left\{-\frac{\partial U_2(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}\right\} & E\left\{-\frac{\partial U_2(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\right\} \end{pmatrix} = \begin{pmatrix} A_{11}(\boldsymbol{\theta}) & A_{12}(\boldsymbol{\theta}) \\ A_{21}(\boldsymbol{\theta}) & A_{22}(\boldsymbol{\theta}) \end{pmatrix}$. We define $s_{z2x}^{(2)}(t, \boldsymbol{\theta}) = E[Y_i(t)\mathbf{Z}_i(t)\mathbf{X}_i(t)^\top \{\boldsymbol{\gamma}^\top \mathbf{Z}_i(t)\} \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}]$ and $s_x^{(2)}(t, \boldsymbol{\beta}) = E[Y_i(t)\mathbf{X}_i(t)^{\otimes 2} \times \exp\{\boldsymbol{\beta}^\top \mathbf{X}_i(t)\}]$, then we have

$$\begin{aligned} A_{11}(\boldsymbol{\theta}) &= \int_0^\tau s_z^{(0)}(u) \left\{ \frac{s_{z2x}(u, \boldsymbol{\beta})}{s_z^{(0)}(u)} - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta})} \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} du, \\ A_{12}(\boldsymbol{\theta}) &= E \left[\int_0^\tau \mathbf{Z}_i(u) \mathbf{X}_i(u)^\top \exp\{-\boldsymbol{\beta}^\top \mathbf{X}_i(u)\} dN_i(u) \right] - \int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta})} \left(\frac{s_x^{(1)}(u, \boldsymbol{\beta})}{s_x^{(0)}(u, \boldsymbol{\beta})} \right)^\top \lambda^c(u) du \\ &\quad - \int_0^\tau s_z^{(0)}(u) \left\{ \frac{s_{z2x}^{(2)}(u, \boldsymbol{\theta})}{s_z^{(0)}(u)} - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u, \boldsymbol{\beta})} \left(\frac{s_x^{(1)}(u, \boldsymbol{\beta})}{s_x^{(0)}(u, \boldsymbol{\beta})} \right)^\top \boldsymbol{\gamma}^\top \frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right\} du, \\ A_{21}(\boldsymbol{\theta}) &= \int_0^\tau s_z^{(0)}(u) \left\{ \frac{s_{zx}(u)}{s_z^{(0)}(u)} - \frac{s_x^{(1)}(u, \boldsymbol{\beta})}{s_x^{(0)}(u, \boldsymbol{\beta})} \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\top \right\} du, \\ A_{22}(\boldsymbol{\theta}) &= \int_0^\tau \left\{ \frac{s_x^{(2)}(u, \boldsymbol{\beta})}{s_x^{(0)}(u, \boldsymbol{\beta})} - \left(\frac{s_x^{(1)}(u, \boldsymbol{\beta})}{s_x^{(0)}(u, \boldsymbol{\beta})} \right)^{\otimes 2} \right\} \lambda^c(u) du \\ &\quad - \int_0^\tau s_z^{(0)}(u) \left\{ \frac{s_x^{(2)}(u, \boldsymbol{\beta})}{s_x^{(0)}(u, \boldsymbol{\beta})} - \left(\frac{s_x^{(1)}(u, \boldsymbol{\beta})}{s_x^{(0)}(u, \boldsymbol{\beta})} \right)^{\otimes 2} \right\} \boldsymbol{\gamma}^\top \frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} du. \end{aligned}$$

Also, we have $\widehat{A}(\boldsymbol{\theta}) = -\frac{\partial \widehat{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -\frac{\partial \widehat{U}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} & -\frac{\partial \widehat{U}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ -\frac{\partial \widehat{U}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} & -\frac{\partial \widehat{U}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \widehat{A}_{11}(\boldsymbol{\theta}) & \widehat{A}_{12}(\boldsymbol{\theta}) \\ \widehat{A}_{21}(\boldsymbol{\theta}) & \widehat{A}_{22}(\boldsymbol{\theta}) \end{pmatrix}$. By Taylor expansion,

we have $\widehat{U}(\widehat{\boldsymbol{\theta}}_h) - \widehat{U}(\boldsymbol{\theta}_0) = \widehat{A}(\boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}^*$ is on the line segment between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_h$. Since we can show that $\widehat{A}(\boldsymbol{\theta}^*)$ converges to $A(\boldsymbol{\theta}_0)$, we have $\sqrt{n}(\widehat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}_0)$ converges to a normal distribution with mean zero and variance $A(\boldsymbol{\theta}_0)^{-1}V(\boldsymbol{\theta}_0)\{A(\boldsymbol{\theta}_0)^{-1}\}^\top$, where $V(\boldsymbol{\theta}_0) = E\{\phi_1(\boldsymbol{\theta}_0)\phi_1(\boldsymbol{\theta}_0)^\top\}$.

Acknowledgement

The authors thank Dr. Dean Follmann for kindly sharing the India pharyngitis study data. They also thank the University of Minnesota Supercomputing Institute for providing computing resources.

Funding

Lyu and Luo were partially supported by the U.S. National Institutes of Health (R03MH112895). Luo was also supported by the U.S. National Institutes of Health (P30CA077598).

References

- Andersen PK, Gill RD (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4): 1100–1120.
- Andersen PK, Liestøl K (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics*, 4(4): 633–649.
- Boscardin WJ, Taylor JM, Law N (1998). Longitudinal models for aids marker data. *Statistical Methods in Medical Research*, 7(1): 13–27.
- Bycott P, Taylor J (1998). A comparison of smoothing techniques for cd4 data measured with error in a time-dependent cox proportional hazards model. *Statistics in Medicine*, 17(18): 2061–2077.
- Cai J, He H, Song X, Sun L (2017a). An additive-multiplicative mean residual life model for right-censored data. *Biometrical Journal*, 59(3): 579–592.
- Cai Q, Wang MC, Chan KCG (2017b). Joint modeling of longitudinal, recurrent events and failure time data for survivor's population. *Biometrics*, 73(4): 1150–1160.
- Cao H, Churpek MM, Zeng D, Fine JP (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. *Journal of the American Statistical Association*, 110(511): 1187–1196.
- Cao H, Fine JP (2021). On the proportional hazards model with last observation carried forward covariates. *Annals of the Institute of Statistical Mathematics*, 73(1): 115–134.
- Dafni UG, Tsiatis AA (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 54(4): 1445–1462.
- Faucett CL, Schenker N, Elashoff RM (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association*, 93(442): 427–437.
- Han M, Song X, Sun L, Liu L (2016). An additive-multiplicative mean model for marker data contingent on recurrent event with an informative terminal event. *Statistica Sinica*, 26(3): 1197–1218.
- Henderson R, Diggle P, Dobson A (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4): 465–480.
- Jose JJM, Brahmadathan KN, Abraham VJ, Huang CY, Morens D, Hoe NP, et al. (2018). Streptococcal group a, c and g pharyngitis in school children: A prospective cohort study in southern India. *Epidemiology & Infection*, 146(7): 848–853.
- Kim S, Zeng D, Chambless L, Li Y (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in Biosciences*, 4(2): 262–281.

- Lawless JF, Nadeau C, Cook RJ (1997). Analysis of mean and rate functions for recurrent events. In: Lin DY and Fleming TR (eds.), *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, 37–49. Springer.
- Li S (2016). Joint modeling of recurrent event processes and intermittently observed time-varying binary covariate processes. *Lifetime Data Analysis*, 22(1): 145–160.
- Li S, Sun Y, Huang CY, Follmann DA, Krause R (2016). Recurrent event data analysis with intermittently observed time-varying covariates. *Statistics in Medicine*, 35(18): 3049–3065.
- Lin DY, Wei LJ, Yang I, Ying Z (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4): 711–730.
- Lin DY, Ying Z (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics*, 23(5): 1712–1734.
- Liu Y, Wu Y, Cai J, Zhou H (2010). Additive-multiplicative rates model for recurrent events. *Lifetime Data Analysis*, 16(3): 353–373.
- Liu YY, Wu YS (2011). Semiparametric additive intensity model with frailty for recurrent events. *Acta Mathematica Sinica, English Series*, 27(9): 1831.
- Lyu T, Luo X, Huang CY, Sun Y (2021). Additive rates model for recurrent event data with intermittently observed time-dependent covariates. *Statistical Methods in Medical Research*, 30(10): 2239–2255.
- Pepe MS, Cai J (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, 88(423): 811–820.
- Prentice RL (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2): 331–342.
- Prentice RL, Williams BJ, Peterson AV (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2): 373–379.
- Raboud J, Reid N, Coates RA, Farewell VT (1993). Estimating risks of progressing to aids when covariates are measured with error. *Journal of the Royal Statistical Society. Series A*, 156(3): 393–406.
- Schaubel DE, Zeng D, Cai J (2006). A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis*, 12(4): 389–406.
- Scheike TH, Zhang MJ (2002). An additive-multiplicative Cox-Aalen regression model. *Scandinavian Journal of Statistics*, 29(1): 75–88.
- Stanworth SJ, Hudson CL, Estcourt LJ, Johnson RJ, Wood EM (2015). Risk of bleeding and use of platelet transfusions in patients with hematologic malignancies: Recurrent event analysis. *Haematologica*, 100(6): 740–747.
- Sun X, Song X, Sun L (2021a). Additive hazard regression of event history studies with intermittently measured covariates. *Canadian Journal of Statistics*, doi: <https://doi.org/10.1002/cjs.11630>.
- Sun Y, McCulloch CE, Marr KA, Huang CY (2021b). Recurrent events analysis with data collected at informative clinical visits in electronic health records. *Journal of the American Statistical Association*, 116(534): 594–604.
- Tsiatis AA, Degruittola V, Wulfsohn MS (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and cd4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90(429): 27–37.
- Van Der Heijden AA, van't Riet E, Bot SD, Cannegieter SC, Stehouwer CD, Baan CA, et

- al. (2013). Risk of a recurrent cardiovascular event in individuals with type 2 diabetes or intermediate hyperglycemia. *Diabetes Care*, 36(11): 3498–3502.
- Van Der Vaart AW, Wellner JA (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.
- Vonesh EF, Greene T, Schluchter MD (2006). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*, 25(1): 143–163.
- Wulfsohn MS, Tsiatis AA (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1): 330–339.
- Xu J, Zeger SL (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C*, 50(3): 375–387.