



Published in final edited form as:

Stat Med. 2016 August 15; 35(18): 3049–3065. doi:10.1002/sim.6901.

Recurrent Event Data Analysis With Intermittently Observed Time-Varying Covariates

Shanshan Li^a, Yifei Sun^b, Chiung-Yu Huang^{b,c,*}, Dean A. Follmann^d, and Richard Krause^d

^aDepartment of Biostatistics, Indiana University Fairbanks School of Public Health, Indianapolis, IN 46202, U.S.A

^bDepartment of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, U. S. A

^cDivision of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21205, U. S. A

^dNational Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20817 U. S. A

Abstract

Although recurrent event data analysis is a rapidly evolving area of research, rigorous studies on estimation of the effects of intermittently observed time-varying covariates on the risk of recurrent events have been lacking. Existing methods for analyzing recurrent event data usually require that the covariate processes are observed throughout the entire follow-up period. However, covariates are often observed periodically rather than continuously. We propose a novel semiparametric estimator for the regression parameters in the popular proportional rate model. The proposed estimator is based on an estimated score function where we kernel smooth the mean covariate process. We show that the proposed semiparametric estimator is asymptotically unbiased, normally distributed and derive the asymptotic variance. Simulation studies are conducted to compare the performance of the proposed estimator and the simple methods carrying forward the last covariates. The different methods are applied to an observational study designed to assess the effect of Group A streptococcus (GAS) on pharyngitis among school children in India.

Keywords

Estimating equations; Kernel smoothing; Partial likelihood; Recurrent events; Survival analysis

1. Introduction

In many epidemiology and biomedical settings, data on risk factors and events that occur repeatedly over time are collected. Modeling and estimation of covariate effects on the occurrence of recurrent events has been a much discussed topic in the past few decades; see [1] and [2] for comprehensive reviews. Statistical methods for recurrent event data analysis usually require that the covariate processes are observed throughout the entire follow-up

*Correspondence to: Chiung-Yu Huang, Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21205, U. S. A. cyhuang@jhu.edu.

period, continuing to the end of the study or until loss to follow-up [3, 4, 5, 6, 7]. In many applications, however, the values of time-varying covariates are only observed periodically. As a result, the missing covariates in the estimation functions need to be replaced with some estimated values. For continuous measures, one can obtain the predicted value at any time by assuming a mixed effect model or by smoothing the neighboring observed covariate values. See [8] for a nice summary. However, these approaches do not apply for binary measures, as there is not a ‘smoothed’ estimate for binary data. A natural alternative that works for both continuous and binary measures is the last covariate carried forward (LCCF) approach. Under LCCF, the last known value of the covariate is used forward in time until a new value is measured. Thus the true covariate process is approximated by a step function with jumps at the measurement times. Analogous to the bias induced by the LCCF method for time to event analysis [9, 10, 11], the LCCF method is expected to lead to biased estimation in the recurrent event data analysis.

The particular research that motivated this work is an observational study designed to evaluate the effect of Group A streptococcus (GAS) on the risk of developing pharyngitis (sore throat). Pharyngitis is most frequently due to viruses, but several bacteria, including Group A streptococci (GAS), persist as a common cause of pharyngitis even in the era of antibiotics. GAS pharyngitis is more prevalent in children than adults, and mainly occurs in winter and early spring. A World Health Organization report estimates that there are over 616 million new cases per year of GAS pharyngitis, of which over 550 million occur in less developed countries. To study the effect of GAS on pharyngitis, a total of 305 school children were recruited in a rural area near Vellore, India. During the follow-up period, cases of pharyngitis were identified weekly. Throat swabs were obtained on those with pharyngitis to identify the presence of GAS. Additionally, monthly throat cultures were obtained on the school children to determine the GAS carriage rate. In our analysis, occurrences of pharyngitis are the recurrent events of interest and GAS colonization status is a time-varying covariate.

We apply the popular semiparametric proportional rate model [12, 13, 3] to evaluate the effect of GAS colonization on pharyngitis. This model allows for an arbitrary baseline rate function of pharyngitis, and is an analogue to Cox regression for recurrent events. The parameters in the proportional rate model can be estimated by maximizing the pseudo-partial likelihood function. To properly construct the pseudo-partial likelihood, the GAS colonization status must be known exactly in everyone who is still under observation in the study whenever someone has an event. However, in the Indian pharyngitis study, the GAS status were observed monthly while the events of pharyngitis were assessed weekly. As a result, the covariate values at each event time were observed for children who had pharyngitis, but were possibly missing for other children in the corresponding risk set. Naively imputing the missing value with the last observed covariate may lead to biased estimation. Instead, we propose to solve an estimated score equation that is constructed by kernel smoothing the observed covariate values collected around each event time. The proposed estimator can be applied to handle both continuous and binary covariate processes, and it is shown to be asymptotically unbiased and normally distributed. We derive the asymptotic variance by properly incorporating the uncertainty of the estimated covariate functions in the estimation procedure. Simulations are conducted to compare the bias and

efficiency of the proposed method with that of the LCCF approach. We apply the methods to the Indian study of GAS and pharyngitis.

2. Models and Methods

2.1. Semiparametric Proportional Rate Model

Let subscript i be the index for a subject, $i = 1, 2, \dots, n$. For subject i , let $N_i^*(t)$ be the number of recurrent events occurring at or before time $t \in [0, \tau]$, where the recurrent events could potentially be observed beyond a prespecified time point τ . Thus the counting process $N_i^*(t)$ has a jump of size one when an event (such as the sore throat in the Indian pharyngitis study) occurs. Many authors, including [14] and [15], have considered modeling the intensity function of the underlying recurrent event process $\{N_i^*(\cdot), t \in [0, \tau]\}$, where the intensity function is the instantaneous risk of event occurrence conditioning on the preceding event history. Recent research has focused on modeling the marginal rate function of the recurrent event process. The rate function $\lambda_i(t)$ of $N_i^*(\cdot)$, defined by

$\lambda_i(t)dt = E\{dN_i^*(t)\} = E\{N_i^*(t^- + dt) - N_i^*(t^-)\}$, is the risk of experiencing recurrent events in the small time interval $[t, t + dt)$ without conditioning on the preceding event history.

Thus, statistical modeling of the rate function allows for an arbitrary dependence structure among recurrent events. In many public health and biomedical studies, modeling the rate function is preferred for analysis, especially in identifying treatment effects and risk factors, because the regression parameters have a direct marginal interpretation.

Let $Z_i(t)$ be a $p \times 1$ vector of covariates of interest. The proportional rate model [12, 13, 3] assumes that, conditioning on $Z_i(t)$, the rate function for subject i at time t is given by

$$\lambda\{t|Z_i(t)\} = \lambda_0(t) \exp\{\beta^\top Z_i(t)\}, \quad (1)$$

where β is a $p \times 1$ vector of the regression parameters and $\lambda_0(t)$ is an arbitrary baseline rate function. The regression parameter β_j is interpreted as the logarithm of the ratio of the rate function at time t for every unit increase in the j th explanatory variable. For ease of discussion, we shall assume that the explanatory variable, $Z_i(t)$, is a univariate time-dependent covariate process evolving in the time interval $[0, \tau]$, that is, $p = 1$. Extensions of the proposed estimator to multivariate covariate processes are straightforward.

In most applications, the underlying counting process $N_i^*(\cdot)$ is subject to censoring due to loss to follow-up or end of the study. Let C_i denote the time to loss to follow-up or end of the study for subject i , $i = 1, \dots, n$. Hence C_i is the censoring time and is observed for all study subjects. We assume that C_i is independent of N_i^* given Z_i in the sense that

$E\{dN_i^*(t)|Z_i(t), C_i > t\} = E\{dN_i^*(t)|Z_i(t)\}$. Define the counting process $N_i(t) = N_i^*(t \wedge C_i)$, where $a \wedge b = \min(a, b)$. Let $Y_i(t) = I\{C_i > t\}$ be the indicator for a subject being under observation at time t . As suggested by [3], the regression parameter β in (1) can be estimated by maximizing the log pseudo-partial likelihood:

$$l(\beta) = \sum_{i=1}^n \int_0^\tau \left(\beta Z_i(t) - \log \left[\sum_{k=1}^n Y_k(t) \exp\{\beta Z_k(t)\} \right] \right) dN_i(t),$$

which is equivalent to solving the (normalized) pseudo-partial score function

$$U(\beta) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ Z_i(t) - \frac{S^{(1)}(t, \beta)}{S^{(0)}(t, \beta)} \right\} dN_i(t) \quad (2)$$

for zero, with $S^{(k)}(t, \beta) = n^{-1} \sum_{i=1}^n Y_i(t) Z_i(t)^k \exp\{\beta Z_i(t)\}$, $k = 0, 1, 2$. The construction of the pseudo-partial score function requires that $Z_i(t)$ is completely observed on $[0, C_i]$.

In practice, however, covariates are typically measured periodically. An ad-hoc method to deal with missing covariate values at time t is to fill in the missing values with the last observed covariate values before time t . In studies of recurrent events, it is common that covariate values are collected when an event occurs in addition to regular follow-up visits. Thus there are two possible ways to carry forward the last observation. One is to carry forward all covariates regardless of whether the measurement is from an event or from the regular follow-up visits. For example, in the Indian pharyngitis study, one may carry forward the status of GAS colonization collected at the time of the pharyngitis event as well as the GAS carriage status collected at the monthly visits. We call this all covariates carried forward (ACCF) method. A second possibility is to carry forward only regular follow-up visits, that is, GAS carriage data in the Indian pharyngitis study. We term this the carriage covariates carried forward (CCCF) method. Research has shown that the last covariate carried forward approach leads to biased inferential results of the covariate effects in the survival setting [9, 10]. To our knowledge no previous work in the literature investigated the bias induced by the ACCF method or the CCCF method in recurrent event data analysis.

2.2. The Proposed Estimator for Binary Covariates

In this section, we propose new statistical methods for recurrent event data analysis when the data on the time-varying covariates are collected at regular discrete time points in all subjects as well as at the exact times of an event, for the subject having the event. As noted by many authors, including [16], the pseudo-partial score function U in (2) is a functional of four empirical processes $n^{-1} \sum_{i=1}^n Z_i(t) dN_i(t)$, $n^{-1} \sum_{i=1}^n dN_i(t)$, $S^{(0)}(t, \beta)$, and $S^{(1)}(t, \beta)$. In the Indian pharyngitis study, throat swabs were obtained on those with pharyngitis in order to identify the presence of GAS, that is, covariates were collected at event visits. Therefore, under our setting, $Z_i(t)$ is observed when $N_i(\cdot)$ jumps at t and the first two empirical processes $n^{-1} \sum_{i=1}^n Z_i(t) dN_i(t)$ and $n^{-1} \sum_{i=1}^n dN_i(t)$ are always observed. On the other hand, the last two stochastic processes usually involve missing covariate values. We propose to replace $S^{(0)}(t, \beta)$ and $S^{(1)}(t, \beta)$ with estimators that converge to the same limits. We will show that the new estimating equation converges to the same limit as U for

fixed β , which ensures that the solution of the estimated score function is a consistent estimator of β .

To see this, we first consider the simple case where the time-dependent covariate $Z_{\lambda}(t)$ is a dichotomous random variable, while the extension to continuous covariate processes will be presented in Section 2.3. As an example, $Z_{\lambda}(t) = 1$ indicates a positive throat culture for GAS at time t and $Z_{\lambda}(t) = 0$ otherwise. Let $r(t) = E[Z_{\lambda}(t)]$ denote the population average of the covariate at time t . We will assume for now that C_i and $Z_{\lambda}(t)$ are independent – this assumption will be relaxed later in Theorem 1. Let $G(t)$ denote the survival function of C_i . It follows from the law of large numbers that

$$S^{(0)}(t, \beta) = n^{-1} \sum_{i=1}^n Y_i(t) \exp\{\beta Z_i(t)\} \rightarrow [e^{\beta \cdot 1} \text{pr}\{Z_i(t)=1\} + e^{\beta \cdot 0} \text{pr}\{Z_i(t)=0\}] G(t) = \{e^{\beta r(t)} + 1 - r(t)\} G(t)$$

and, similarly, $S^{(1)}(t, \beta) = n^{-1} \sum_{i=1}^n Y_i(t) Z_i(t) \exp\{\beta Z_i(t)\} \rightarrow e^{\beta r(t)} G(t)$ in probability for fixed β and $t \in [0, \tau]$ as $n \rightarrow \infty$. As a result, we have

$$\frac{S^{(1)}(t, \beta)}{S^{(0)}(t, \beta)} \rightarrow \frac{e^{\beta r(t)}}{e^{\beta r(t)} + 1 - r(t)}$$

in probability as $n \rightarrow \infty$. Thus the limiting function of U can be estimated consistently if a consistent estimator of $r(t)$ can be obtained using available data.

Intuitively, one may estimate $r(t)$, that is, the prevalence rate of GAS colonization at time t in the data example, by dividing the number of positive throat cultures collected around time t by the total number of swabs collected around time t . We consider employing a kernel estimator for $r(t)$ which computes a locally weighted average of the covariate values. Let $O_{\lambda}(t)$ denote the cumulative number of measurements collected at regular visits before and at time t for the λ th subject. Note that a regular visit can also be an event visit as the patient may be sick at regular visits, that is, $O_{\lambda}(\cdot)$ and $N_{\lambda}(\cdot)$ are allowed to jump at the same time point. In many applications it is reasonable to assume that $O_{\lambda}(\cdot)$ is independent of the time-dependent covariate $Z_{\lambda}(\cdot)$ and the censoring time C_i . For the Indian pharyngitis study, $O_{\lambda}(t)$ is a function with unit steps at the monthly carriage visits for the λ th child. Let $K_h(t) = h^{-1} K(t/h)$ be a kernel function with bandwidth h that satisfies $\int_{-\infty}^{\infty} K(t) dt = 1$ and $\int_{-\infty}^{\infty} t K(t) dt = 0$. A kernel estimator for $r(t)$ is given by

$$\hat{r}_h(t) = \frac{n^{-1} \sum_{i=1}^n \int_0^{\tau} K_h(t-u) Y_i(u) Z_i(u) dO_i(u)}{n^{-1} \sum_{i=1}^n \int_0^{\tau} K_h(t-u) Y_i(u) dO_i(u)}, \quad t \in [h, \tau-h]. \quad (3)$$

To avoid bias estimation in the boundary region, we set $r_h^{\wedge}(t) = r_h^{\wedge}(h)$ for $t \in [0, h)$ and $r_h^{\wedge}(t) = r_h^{\wedge}(\tau - h)$ for $t \in (\tau - h, \tau]$. If the uniform kernel is employed, that is, $K(t) = 2^{-1} I(|t| \leq 1)$, the denominator of $r_h^{\wedge}(t)$ is the total number of swabs collected in the time window $[t - h, t + h]$, while the numerator is the number of positive throat cultures in the same time window. In this case, $r_h^{\wedge}(t)$ is simply the proportion of positive throat cultures in the interval $[t - h, t + h]$ and thus is obviously a reasonable approximation of the prevalence rate of GAS colonization

at time point t . The uniform kernel weights all observations in the window equally, even though observations closer to t should be more informative about $r(t)$ than distant ones. In practice, one may consider non-uniform kernel functions, such as the Gaussian kernel, that weights observations of covariates according to their distance in time.

Define the covariate collection function $m(t)$ by $m(t)dt = E\{dO_1(t)\}$, thus $m(t)$ is the instantaneous “risk” of a regular visit occurring at time t . We can show that, provided $m(t) > 0$ for $t \in [0, \tau]$ and under the regularity conditions given in the appendix,

$n^{-1} \sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) Z_i(u) dO_i(u)$ converges in probability to $r(t)m(t)G(t)$ uniformly in t as $h \rightarrow 0$ and $nh^2 \rightarrow \infty$. Similarly, we have $n^{-1} \sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) dO_i(u)$ uniformly converges in probability to $m(t)G(t)$ as $h \rightarrow 0$ and $nh^2 \rightarrow \infty$. Thus the uniform consistency of $r_h^\wedge(t)$ follows directly from Slutsky’s Theorem as $h \rightarrow 0$ and $nh^2 \rightarrow \infty$.

The proposed estimator $r_h^\wedge(t)$ can be viewed an extension of the Nadaraya-Watson estimator [17, 18], except that the denominator and numerator of $r_h^\wedge(t)$ are not based on independent observations, because each study subject may have more than one observed covariate value. The bandwidth parameter h controls bias as well as the degree of smoothness in the estimated prevalence rate function $r_h^\wedge(t)$: a small bandwidth leads to a smaller bias but a greater variance, while a large bandwidth leads to a greater bias but a smaller variance [19].

To estimate β , we construct the estimated score function

$$\hat{U}_h(\beta) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ Z_i(t) - \frac{e^{\beta \hat{r}_h(t)}}{e^{\beta \hat{r}_h(t)} + 1 - \hat{r}_h(t)} \right\} dN_i(t). \quad (4)$$

Let $\mathcal{U}(\beta)$ be the limit of the pseudo-partial likelihood $\mathcal{U}(\beta)$ in (2) with the complete covariate data. It follows directly from the consistency of $r_h^\wedge(t)$ that $\hat{U}_h(\beta)$ converges in probability to $\mathcal{U}(\beta)$. Let β_h be the solution of $\hat{U}_h(\beta) = 0$. Because the true parameter β is the unique solution of $\mathcal{U}(\beta) = 0$, one can show that β_h is a consistent estimator for β under some regularity conditions. The large sample properties of β_h are studied rigorously in Theorem 1 of Section 2.3.

Note that although the proportional rate model (1) postulates the risk of experiencing recurrent events at time t given the covariate history up to t , the proposed estimation procedure borrows information from covariate values beyond time t to derive a consistent estimate of $r(t)$ at time t . The validity of the proposed approach relies on the fact that the proportional rate model is formulated on the basis of the rate function, that is, the risk of experiencing recurrent events unconditioning on the event history. Under this marginal model, the estimated score function and the pseudo-partial score function converge to the same limit, provided that $r(t)$ can be estimated consistently.

Interestingly, in the special case where the expected value of the covariate process over all individuals is known to be constant, that is, $r(t) = r$, one can show that

$$U(\beta) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ Z_i(t) - \frac{e^{\beta r}}{e^{\beta r} + 1 - r} \right\} dN_i(t) + o_p(n^{-1/2}).$$

Thus, solving $U(\beta) = 0$ yields

$$e^\beta = \frac{\sum_{i=1}^n \int_0^\tau Z_i(t) dN_i(t)}{\sum_{i=1}^n \int_0^\tau \{1 - Z_i(t)\} dN_i(t)} \times \frac{1-r}{r} + o_p(n^{-1/2}), \tag{5}$$

where $n_1 = \sum_{i=1}^n \int_0^\tau Z_i(t) dN_i(t)$ and $n_0 = \sum_{i=1}^n \int_0^\tau \{1 - Z_i(t)\} dN_i(t)$ are the numbers of positive and negative throat cultures at sick visits in the Indian pharyngitis study. By replacing r in (5) with the proportion of positive throat cultures at all monthly visits, it is interesting to see that $\exp(\beta)$ can be consistently estimated by a simple cross-product ratio

$$\frac{n_1}{n_0} \times \frac{z_0}{z_1}, \tag{6}$$

where $z_1 = \sum_{i=1}^n \int_0^\tau Z_i(t) dO_i(t)$ and $z_0 = \sum_{i=1}^n \int_0^\tau \{1 - Z_i(t)\} dO_i(t)$ are the numbers of positive and negative throat cultures at the regular follow-up visits. In other words, the log cross-product ratio, $\log\{(n_1 z_0)/(n_0 z_1)\}$, is a consistent estimator for β .

2.3. Extensions and Bandwidth Selection

So far we have focused on the case where $Z_i(t)$ is a univariate binary covariate process. It is straightforward to extend the idea of the estimated score function to accommodate continuous covariate processes. Let $s^{(k)}(t, \beta) = E[Y_i(t) Z_i(t)^k \exp\{\beta Z_i(t)\}]$ be the limiting function of $S^{(k)}(t, \beta)$, $k = 0, 1, 2$. Intuitively, for fixed β , we can consistently estimate $s^{(k)}(t, \beta)$ with a Nadaraya-Watson-type estimator

$$\frac{n^{-1} \sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) Z_i(u)^k \exp\{\beta Z_i(u)\} dO_i(u)}{n^{-1} \sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) dO_i(u)}, \quad t \in [h, \tau-h].$$

Define $\hat{S}_h^{(k)}(t, \beta) = n^{-1} \sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) Z_i(u)^k \exp\{\beta Z_i(u)\} dO_i(u)$. To avoid the bias in the boundary region, we set $\hat{S}_h^{(k)}(t, \beta) = \hat{S}_h^{(k)}(h, \beta)$ for $t \in [0, h]$, and set

$\hat{S}_h^{(k)}(t, \beta) = \hat{S}_h^{(k)}(\tau-h, \beta)$ for $t \in (\tau-h, \tau]$. In the Appendix, we show that $\hat{S}_h^{(k)}(t, \beta)$ converges uniformly in probability to their limits $s^{(k)}(t, \beta) m(t)$, where $m(t) dt = E[dO_i(t)]$.

Thus, by Slutsky's Theorem, the ratio of $\hat{S}_h^{(1)}(t, \beta)$ and $\hat{S}_h^{(0)}(t, \beta)$ converges to $\mathcal{E}(t, \beta) = s^{(1)}(t, \beta)/s^{(0)}(t, \beta)$, that is,

$$\hat{\mathcal{E}}_h(t, \beta) = \frac{n^{-1} \sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) Z_i(u) \exp\{\beta Z_i(u)\} dO_i(u)}{n^{-1} \sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) \exp\{\beta Z_i(u)\} dO_i(u)} \rightarrow \mathcal{E}(t, \beta) \quad (7)$$

uniformly in probability as $h \rightarrow 0$ and $nh^2 \rightarrow \infty$. It is easy to verify that $\hat{\mathcal{E}}_h(t, \beta)$ reduces to $\mathcal{E}^\beta r_{\hat{\lambda}}(t) / \{\mathcal{E}^\beta r_{\hat{\lambda}}(t) + 1 - r_{\hat{\lambda}}(t)\}$ when $Z_{\hat{\lambda}}(t)$ is a univariate binary covariate process.

We propose to replace $\mathcal{S}^{(1)}(t, \beta) / \mathcal{S}^{(0)}(t, \beta)$ with $\hat{\mathcal{E}}_h(t, \beta)$ in (2) and estimate β by solving $\hat{U}_h(\beta) = 0$, where

$$\hat{U}_h(\beta) = n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \hat{\mathcal{E}}_h(t, \beta)\} dN_i(t). \quad (8)$$

Because $\hat{U}_h(\beta)$ is composed of four empirical processes that converge in probability to their limits uniformly in $t \in [0, \tau]$, we have $\sup_{\beta \in B} |\hat{U}_h(\beta) - \mathcal{U}(\beta)| \rightarrow 0$ in probability as $h \rightarrow 0$ and $nh^2 \rightarrow \infty$. Furthermore, we show in the appendix that \hat{U}_h can be expressed as the sum of asymptotically *i.i.d.* random variables $\hat{U}_h(\beta) = n^{-1} \sum_{i=1}^n \psi_i(\beta) + o_p(n^{-1/2})$, where $\psi_i(\beta)$ is defined in the Appendix. Thus $n^{1/2} \hat{U}_h(\beta)$ is asymptotically normal with variance $\Omega(\beta) = \text{var}\{\psi_1(\beta)\}$. Let $\hat{\beta}_h$ be the solution of $\hat{U}_h(\beta) = 0$ and define $\Gamma(\beta) = -\mathcal{U}(\beta) / \beta$. Theorem 1 summarizes the large sample distribution properties of $\hat{\beta}_h$, with proofs given in the appendix. Note that, though the explicit form of the variance estimate is given in the theorem, a bootstrap variance estimate can be used for convenience.

Theorem 1—Under conditions (A1)–(A9), $\hat{\beta}_h$ is a consistent estimator of the true parameter β_0 and $\sqrt{n}(\hat{\beta}_h - \beta_0)$ converges to a mean zero normal distribution with variance $\Sigma(\beta_0) = \Gamma(\beta_0)^{-1} \Omega(\beta_0) \Gamma(\beta_0)^{-1}$, provided $h = O(n^{-\nu})$, with $1/4 < \nu < 1/2$.

As suggested by Theorem 1, the asymptotic distribution of the proposed estimator does not depend on the choice of bandwidth as long as the bandwidth condition is satisfied, that is, $h = O(n^{-\nu})$ with $1/4 < \nu < 1/2$. In our work, following [20], we use a K-fold cross-validation method for bandwidth selection, and use minus logarithm of the partial likelihood function as the prediction error criterion. Specifically, let D_k , $k = 1, \dots, K$, denote a partition of the dataset. For a fixed h and the k th subgroup of the data, define

$$PE_k(h) = - \sum_{i \in D_k} \int_0^\tau \left(\hat{\beta}_{(-k)} Z_i(t) - \log \left[\frac{\sum_{j \in D_k} \int_0^\tau K_h(t-u) Y_j(u) \exp\{\hat{\beta}_{(-k)} Z_j(u)\} dO_j(u)}{\sum_{j \in D_k} \int_0^\tau K_h(t-u) Y_j(u) dO_j(u)} \right] \right) dN_i(t),$$

where $\hat{\beta}_{(-k)}$ is estimated using data from individuals not in D_k with bandwidth h . The total prediction error function can be obtained as $PE(h) = \sum_{k=1}^K PE_k(h)$, and we choose the optimal bandwidth h by minimizing $PE(h)$.

3. Simulations

We conduct a series of numerical simulation studies to evaluate the finite-sample performance of different estimators with moderate sample size. For each simulation, we generate 1000 simulated datasets, each with 300 subjects. In the first set of simulation studies, we consider the scenarios where $Z_{\lambda}(t)$ is a binary covariate process taking 0 or 1 for values, where $Z_{\lambda}(0) = 1$ with Bernoulli probability $p = 0.2$. We generate the binary covariate process from a multistate process, where the value of the covariate process alternates between 0 and 1. In other words, the multistate process consists of two periods of states which correspond to GAS negative and GAS positive periods of time in the data example. For subject i , the duration of state 0 is generated using a random variable with hazard function $\xi_i g(t)$, and the duration of state 1 is generated using a random variable with hazard ξ_i , where ξ_i follows a gamma distribution with mean 1 and variance 0.25. The recurrent events of a subject are generated from a proportional intensity model, where, conditional on the subject-specific random effect γ_i , the intensity of the recurrent event process for subject i is $\lambda(t) = \lambda_0(t) \exp\{\beta Z_{\lambda}(t) + \gamma_i\}$ with γ_i being generated from a normal distribution with mean 0 and variance 0.25. We set $\beta = 0.5$ and $\lambda_0^*(t) = 0.1I(t \leq 10) + 0.5I(10 < t \leq 20)$. Integrating out the random effect, we obtain the proportional rate model $\lambda(t) = \lambda_0(t) \exp\{\beta Z_{\lambda}(t)\}$ with $\lambda_0(t) = \lambda_0^*(t) \exp(0.25/2)$.

We compare the performances of four estimators of β : (a) CPR, the cross-product ratio estimator, (b) ESF, the estimated score function approach, (c) ACCF, the rate ratio estimator with all carriage and event covariates carried forward, (d) CCCF, the rate ratio estimator with carriage covariates carried forward. We consider two scenarios: $g(t) = 4$ for $t \in [0, 20]$ and $g(t) = 4I(t \leq 10) + 6I(10 < t \leq 20)$. The prevalence rate of GAS for the former scenario is 20% at any time point, while for the latter scenario the prevalence rate is 20% for $t \leq 10$ and drops to 14% for $t > 10$. Each subject has 20 scheduled visits on $[0, 20]$, with one visit per unit time interval. The time of visit in each interval is uniformly distributed. We evaluate the behavior of the estimators under various degrees of missingness, where the probability of missing a pre-scheduled visit is set to be 0%, 20%, 40%, and 60% for all visits. The last observed regular visit is treated as the censoring time so the recurrent events are only observed up to the last observed regular visit. To estimate the pseudo-partial score function, we employ the Gaussian kernel with the quartiles of the kernel density function at $\pm 0.25h$, where for each simulated dataset the bandwidth h is chosen by applying the 10-fold cross-validation method described in Section 2.3.

Table 1 summarizes the empirical bias and the empirical standard deviation of 1000 estimated regression parameters, the estimated asymptotic standard error, the bootstrap standard error based on 500 bootstrapped samples, the coverage rate of the 95% bootstrap confidence interval, and the relative efficiency that compares the mean square error of an estimator to that of the rate ratio estimator under the perfect scenario where the covariate process is monitored continuously throughout the entire study period. When $g(t) = 4$, there is no time trend in the covariates. The constant prevalence rate assumption holds, and thus the CPR estimator is consistent. The bias and relative efficiency of ESF is comparable to that of CPR. The ACCF method yields substantial bias, especially when the missing probability is high, while the CCCF method remains consistent. Some intuition for this behavior can be

developed. Suppose prevalence is constant. At any time t with $\beta > 0$, an estimate of the prevalence rate using all covariate data will tend to be biased too high as events just prior to t will tend to have $Z = 1$. An estimate of the prevalence just using the carriage data collected at the monthly visits will not be biased and this differential behavior seems to have consequences for estimation of β as well. For all four estimators, the bootstrap standard errors are very close to the empirical standard deviations and the coverage rates of the 95% bootstrap confidence intervals are close to the nominal level (0.95), supporting the appropriateness of the bootstrap approach. Moreover, for the proposed ESF estimator, the asymptotic standard errors are close to the empirical standard deviations.

In the scenarios where $g(t) = 4I(t < 10) + 6I(10 < t < 20)$, there is a decreasing time trend in the covariate process. The simple CPR estimator is substantially biased because the steady state assumption is violated. Moreover, both the ACCF and CCCF methods are also biased and their bias increases with the missing probability. The bias of the proposed ESF estimator is small, and its relative efficiency is much higher than its competitors. Both the bootstrap standard errors and the asymptotic standard errors of the ESF estimator track the empirical standard deviations well.

The second set of simulation studies examines the performance of ESF, ACCF, CCCF methods in scenarios where $Z_i(t)$ is continuous. The recurrent events, the censoring times, and the missing probabilities are simulated under the same model as in the first set of simulations. For subject i , we set $Z_i(t) = b_{0i} + b_{1i}t$, where the random intercept b_{0i} and the random slope b_{1i} are generated from a bivariate normal distribution. We assume that the random intercept has mean 1 and variance 0.1, and the correlation between b_{0i} and b_{1i} is set to be 0.2. The left panel of Table 2 shows the simulation results where b_{1i} has zero mean and a variance 0.002. When b_{1i} has mean zero, there is no time trend in the covariates. The proposed ESF approach performs well in that the bias is small and its relative efficiency is very high. The CCCF method is apparently biased, while the ACCF method possesses a smaller bias. This is because, on the individual level, the covariate value is either increasing or decreasing with time. Carrying forward the measurement closer in time will tend to have smaller bias, therefore, the ACCF method performs better than the CCCF method. The right panel of Table 2 shows the simulation results when the random slope b_{1i} has mean -0.05 and variance 0.002, thus with a time trend. As expected, the proposed ESF method has the highest relative efficiency with small bias. Both ACCF and CCCF are biased and the bias increases with probability of missingness. Both estimators tend to be biased towards a smaller value due to the decreasing nature of the covariate process.

In order to assess the sensitivity of the proposed estimation procedure to the bandwidth selection, we have compared the simulation results with different choices of bandwidth (results not shown). It is found that the estimated regression coefficients are very similar, differing only in the third decimal place. In addition, as pointed out by an anonymous reviewer, the condition $m(t) > 0$ for all t may not be satisfied in some studies. To compare the proposed estimator with its competitors in the case where $m(t) > 0$ fails to hold, we have conducted additional simulation where we set the timing of regular visits to be fixed (results not shown). It is found that the proposed estimator also shows smaller bias and similar/better relative efficiency (in terms of mean square errors) when compared with its competitors. In

other words, our estimator still outperforms its competitors even when this technical assumption $m(t) > 0$ fails to hold.

All analyses are performed in R, version 3.1.0. The computation environment is a multi-core Linux cluster with more than 680 cores running in the average of 2.5 GHz speed and 4.4 TB of memory. In the simulations reported above, it takes a few seconds to obtain parameter estimation for the proposed estimator with a given bandwidth h , and it takes less than 5 mins to do the bandwidth selection using cross-validation. The other methods also obtain the estimates in a few seconds.

4. Data Analysis

Pharyngitis is one of the most common reasons patients seek the care of a physician, and GAS is one of the common causes of pharyngitis. A meta-analysis of the prevalence of GAS in the pharynx revealed that children with pharyngitis had a GAS prevalence rate of 37% (CI: 32% – 43%), and children with no clinical evidence of infection had a 12% prevalence rate of GAS (CI: 9% – 14%) [21]. Because GAS pharyngitis is a communicable disease, family members and school classmates of the patient are frequently infected. Furthermore, patients with viral pharyngitis, which is more common than streptococcal pharyngitis, may have incidental GAS pharyngeal carriage, with a rate similar to that of the symptom-free children. Streptococcal pharyngitis in most cases cannot be distinguished on clinical grounds from viral pharyngitis. For this reason the throat culture for the detection of Group A streptococci remains the diagnostic gold standard.

We analyze data from the Indian pharyngitis study to evaluate the effect of GAS colonization on the risk of developing pharyngitis. Between March 2002 and March 2004, 305 school children from Vellore India, aged between 7 and 11, were examined weekly for pharyngitis. For those with pharyngitis, a throat culture was obtained to identify the presence of GAS. Additionally, monthly throat cultures were obtained on all the study children to determine the prevalence of GAS. Note that, although the regular visits were scheduled on a monthly basis, the actual observation times were irregularly spaced across subjects to balance the workload. Similarly, although the event visits were scheduled weekly, the actual observation times were irregularly spaced. Therefore, the “continuous observation times” assumption is approximately satisfied in our application. All patients with GAS infections were treated with antibiotics, which usually shortens the infectious period to 24 hours. In case the antibiotic therapy was not effective, a two-week rule was applied to determine an episode of pharyngitis, that is, a pharyngitis event occurred within 14 days after a previous episode was considered as the same episode. We fit the proportional rate model [3] with GAS colonization status as a time-varying covariate and applied the methods discussed in Section 2 to estimate the covariate effect. The time origin for the recurrent event analysis is set to be the first day of the study, that is, March 11, 2002. In general, the assumption that $O_{\lambda}(\cdot)$ is independent of $\{N_i^*(\cdot), Z_{\lambda}(\cdot), C_i\}$ is reasonably met in the Indian pharyngitis study, because the monthly GAS carriage visits were prescheduled and thus not informative about the underlying covariate processes and event processes. On the other hand, the event visits occurred whenever a child had a sore throat, thus the observed GAS colonization status are more likely to be positive if GAS infection is associated with a higher risk of pharyngitis.

Table 3 presents a tally of outcomes based on the regular carriage visits and the pharyngitis event visits. Over the 2 years of the study, 641 pharyngitis events occurred or roughly one per child per year. With 2827 monthly visits, the carriage data corresponds to a total followup of about 236 children years. About 17% of the throat cultures collected at pharyngitis visits were GAS positive, while 11% of the throat cultures obtained at the carriage visits were GAS positive. Figure 1 presents the estimated prevalence rate of GAS colonization using data from the monthly carriage visits. It is seen that the GAS prevalence rate was changing over time in the observation period, with a range from 0.05 to 0.2.

We report the rate ratio estimates using four different methods. The simple CPR method estimates a rate ratio of 1.61 with a 95% confidence interval (CI) of (1.31–1.99). However, this estimator is likely to be biased because GAS colonization varies with season. The ACCF method that carries forward all available covariate data yields a rate ratio of 1.28 (95% CI: 1.01–1.58), while the CCCF method that carries forward only carriage covariate yields a rate ratio of 1.37 (95% CI: 1.01–1.76). Finally, applying 10-fold cross-validation for the proposed ESF approach with a Gaussian kernel, we estimate a rate ratio of 1.46 (95% CI: 1.18–1.86). Confidence intervals for all four methods are obtained by the percentile method of the nonparametric bootstrap for clustered data with 1000 bootstrap samples, where the sampling unit is the child. Similarly as observed in the simulation studies, the CCCF and the proposed approach give comparable results, while the other two estimators appear to yield results deviated from the proposed estimator. Since the simulations in the previous section show that the proposed estimated score function approach is flexible and robust, we advocate the proposed approach for inference here. We conclude that there is significant effect of GAS on the risk of developing pharyngitis among school children. This suggests that population methods of control such as vaccine development may be worth considering for this population. The risk of pharyngitis increases by 46% (CI: 18%–86%) for a child who is colonized with GAS.

5. Remarks

This article aims to estimate the effect of covariates on recurrent events when time-varying covariates are measured at pre-scheduled follow-up visits (regular visits) and in subjects when they have an event (event visits). The former type of visits carries no information about the underlying recurrent event process, whereas the event-based sampling of covariate values typically provides a biased representation of the underlying covariate process. If higher values of a covariate are associated with an event, then event based sampling will tend to get larger covariates than from the underlying process. This is a subtle but important point and often leads to substantial bias using the “obvious” method of last-covariate value carried forward, as we have shown.

We propose to solve an estimated score function which estimates the mean covariate process in the pseudo-partial score function by kernel smoothing the observed covariate data at regular visits. While the idea of applying the smoothing technique to individual components in an estimating equation is not brand new in longitudinal data analysis and survival analysis, see for example [22], we believe that applying the smoothing technique to tackle

the commonly encountered problem of not observing covariate values for every individual in the risk sets is innovative in recurrent event analysis.

Though illustrated with univariate covariate process, the proposed method can be extended to multivariate covariate processes by employing multivariate kernel smoothing techniques in a straightforward manner. Note that the bandwidths for different covariates in multivariate kernel smoothing are allowed to be different. Overall, the estimated score function approach is the most flexible and robust approach as it can accommodate any kind of covariate process. The cross-product ratio estimator and the last covariate carried forward approach require more restrictive assumptions for unbiased and efficient inference.

An alternative approach to deal with intermittently observed time-dependent covariates is to jointly model the recurrent event process and the covariate process. This requires postulation of a specific stochastic model for the covariate process. For binary covariate process, one can assume a transition probability model [10]. For continuous covariate process, one can assume a mixed effects model [11, 23]. The validity of the joint modeling approach relies on the assumptions of the covariate process. It may lead to substantial bias if the model for the covariate process does not hold. Another alternative approach, as mentioned by one reviewer, is to fill in the missing covariates using simple smoothing techniques such as taking the mean of the neighboring values [24]. This is a simple and straightforward approach, however, it has a few drawbacks. First of all, it is not applicable to the binary covariate measures. Secondly, it is more computationally intensive than our proposed approach, as it requires estimation of the missing covariate values for each individual. Thirdly, because the covariate estimates are based on individual data, this approach may lead to substantial bias when the individual observations are sparse in time. Lastly, it lacks theoretical justification.

For future research, as suggested by one reviewer, it would be interesting to investigate the problem of infrequently updated covariate values in other recurrent event data analysis settings. For example, the self-controlled case series (SCCS) method is an alternative method to analyze recurrent event data when the event occurrence rate is low [25]. The two-state infection model [2] and the multi-state model [26] can be used in studies where individuals experience more than one types of events. Existing estimation procedures for these models require the covariate process be observed continuously throughout the observation period. It would be interesting to develop approaches to deal with missing covariate values under these different settings. Further investigation is warranted.

Acknowledgments

The authors are grateful to the editor, associate editor, and referees for their comments which helped improve this paper. This research was supported by the National Cancer Institute grant R01CA193888.

References

1. Andersen, PK.; Borgan, O.; Gill, RD.; Keiding, N. *Statistical Models Based on Counting Processes*. Springer-Verlag; New York: 1993.
2. Cook, RJ.; Lawless, JF. *The Statistical Analysis of Recurrent Events*. New York: Springer; 2007.

3. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Ser B.* 2000; 62:711–730.
4. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics.* 2004; 60(3):747–756. [PubMed: 15339298]
5. Schaubel DE, Cai J. Analysis of clustered recurrent event data with application to hospitalization rates among renal failure patients. *Biostatistics.* 2005; 6(3):404–419. [PubMed: 15831581]
6. Farrington CP, Whitaker HJ. Semiparametric analysis of case series data. *Journal of the Royal Statistical Society, Series C.* 2006; 55:553–594.
7. Huang CY, Qin J, Wang MC. Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics.* 2010; 66:39–49. [PubMed: 19459839]
8. Andersen PK, Liestol K. Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics.* 2003; 4(4):633–649. [PubMed: 14557116]
9. Prentice RL. Covariate measurement errors and parameter estimates in a failure time regression time. *Biometrika.* 1982; 69:331–342.
10. Faucett CL, Schenker N, Elashoff RM. Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association.* 1998; 93:427–437.
11. Tsiatis AA, Davidian M. Joint modelling of longitudinal and time-to-event data: An overview. *Statistica Sinica.* 2004; 14:809–834.
12. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics.* 1995; 37(2):158–168.
13. Lawless, JF.; Nadeau, C.; Cook, RJ. *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis.* New York: Springer; 1997. Analysis of mean and rate functions for recurrent events; p. 37-49.
14. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika.* 1981; 68:373–379.
15. Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *Annals of Statistics.* 1982; 10:1100–1120.
16. Huang CY, Wang MC. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association.* 2004; 99:1153–1165. [PubMed: 24068850]
17. Nadaraya EA. On estimating regression. *Theory of Probability and its Applications.* 1964; 9:141–142.
18. Waston GS. Smooth regression analysis. *The Indian Journal of Statistics, Series A.* 1964; 26:359–372.
19. Silverman, BW. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall; 1986.
20. Tian L, Zucker D, Wei L. On the Cox model with time-varying regression coefficients. *Journal of the American statistical Association.* 2005; 100(469):172–183.
21. Shaikh N, Leonard E, Martin JM. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: A meta-analysis. *Pediatrics.* 2010; 126:557–564.
22. Sun Y. Estimation of semiparametric regression model with longitudinal data. *Lifetime Data Analysis.* 2010; 16(2):271–298. [PubMed: 19890712]
23. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society, Series C.* 2009; 58(1):65–81.
24. Raboud J, Reid N, Coates R, Farewell V. Estimating risks of progressing to aids when covariates are measured with error. *Journal of the Royal Statistical Society, Series A.* 1993:393–406.
25. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Statistical methods in medical research.* 2009; 18:7–26. [PubMed: 18562396]
26. Tom BD, Farewell VT. Intermittent observation of time-dependent explanatory variables: a multistate modelling approach. *Statistics in Medicine.* 2011; 30(30):3520–3531. [PubMed: 22139873]
27. Van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes.* Springer; 1996.

28. Nan B, Wellner JA. A general semiparametric Z-estimation approach for case-cohort studies. *Statistica Sinica*. 2013; 23(3):1155–1180. [PubMed: 24489449]
29. Kosorok, MR. Introduction to Empirical Processes and Semiparametric Inference. New York: Springer; 2007.
30. Mammen E, Nielsen JP. A general approach to the predictability issue in survival analysis with applications. *Biometrika*. 2007; 94(4):873–892.

Appendix

Proof of Theorem 1

Assumptions

- (A1) $\{N_i(\cdot), O_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}$ are independent and identically distributed.
- (A2) $N_i(\tau)$ is bounded. $\lambda^c(\cdot)$, the rate function of $N_i(\cdot)$, is of bounded variation.
- (A3) The true parameter β_0 lies in a compact set \mathcal{B} in \mathcal{R} , and the baseline rate function $\lambda_0(t)$ is absolutely continuous.
- (A4) The covariate process $Z_i(t)$ has uniformly bounded total variation, namely, $\int_0^\tau |dZ_i(t)| + |Z_i(0)| \leq c$ for some $c > 0$ for all i . Without loss of generality, we assume $Z_i(t) \geq 0$.
- (A5) The censoring time C_i is independent of $N_i^*(\cdot)$ conditional on $Z_i(\cdot)$ with $G(\tau) = \text{pr}(C_i \leq \tau) > 0$.
- (A6) The functions $s^{(k)}(t, \beta) = E[Y_i(t)Z_i(t)^k \exp\{\beta Z_i(t)\}]$, $k = 0, 1$, have bounded second derivatives for $t \in [0, \tau]$.
- (A7) The observation time process $O_i(\cdot)$ is independent of $\{N_i^*(\cdot), Z_i(\cdot), C_i\}$ and is bounded. Moreover, the covariate collection function $m(t)$ at time t , defined by $m(t)dt = E[dO_i(t)]$, is positive and has bounded second derivative for $t \in [0, \tau]$.
- (A8) The kernel function $K(\cdot)$ is a symmetric density function with a bounded support.
- (A9) $h = \alpha n^{-\nu}$, where $1/4 < \nu < 1/2$.

We first show the uniform consistency of $\hat{\beta}_h$ when $h \rightarrow 0$, and $nh^2 \rightarrow \infty$. Because the functional defined by \hat{U}_h is continuous with respect to the supreme norm topology, it is sufficient to show that the four processes

$n^{-1} \sum_{i=1}^n \int_0^\tau Z_i(t) dN_i(t)$, $n^{-1} \sum_{i=1}^n N_i(t)$, $\hat{S}_h^{(1)}(t, \beta)$ and $\hat{S}_h^{(0)}(t, \beta)$ converge in probability to their limits uniformly for $\beta \in \mathcal{B}$ and $t \in [0, \tau]$. For $k = 0, 1$, the function classes

$\mathcal{F}_k = \{\int_0^t Y(u)Z(u)^k \exp\{\beta Z(u)\} dO(u) : \beta \in \mathcal{B}, t \in [0, \tau]\}$ have bracketing number $N_{[]}(\varepsilon, \mathcal{F}_k, L_2(P))$ of polynomial order $1/\varepsilon^4$. Define

$\hat{R}^{(k)}(t, \beta) = n^{-1} \sum_{i=1}^n \int_0^t Y_i(u)Z_i(u)^k \exp\{\beta Z_i(u)\} dO_i(u)$ and

$r^{(k)}(t, \beta) = E[\int_0^t Y(u)Z(u)^k \exp\{\beta Z(u)\} dO(u)]$, by Theorem 2.14.9 in [27], we have

$$P\left(\sup_{\beta \in \mathcal{B}, t \in [0, \tau]} \sqrt{n} |\hat{R}^{(k)}(t, \beta) - r^{(k)}(t, \beta)| > x\right) \leq c_k x^{v_k} e^{-2x^2}, \tag{9}$$

where c_k and v_k are constants that depend on k . Since for $t \in [h, \tau - h]$,

$\hat{S}_h^{(k)}(t, \beta) = \int_0^\tau K_h(t-u) \hat{R}^{(k)}(du, \beta)$, through integration by part, we have

$$\begin{aligned} \sup_{\beta \in \mathcal{B}, t \in [h, \tau-h]} |\hat{S}_h^{(k)}(t, \beta) - E\{\hat{S}_h^{(k)}(t, \beta)\}| &= \sup_{\beta \in \mathcal{B}, t \in [h, \tau-h]} \left| \int_{t-h}^{t+h} \{\hat{R}^{(k)}(u, \beta) - r^{(k)}(u, \beta)\} dK_h(t-u) \right| \\ &\leq h^{-1} \sup_{\beta \in \mathcal{B}, t \in [0, \tau]} |\hat{R}^{(k)}(t, \beta) - r^{(k)}(t, \beta)| \cdot V(K), \end{aligned}$$

where $V(K)$ is the variation of the kernel function K .

By equation (9), for any $\varepsilon > 0$, we have

$$\begin{aligned} P(h^{-1} \sup_{\beta \in \mathcal{B}, t \in [0, \tau]} |\hat{R}^{(k)}(t, \beta) - r^{(k)}(t, \beta)| > \varepsilon) &= P\left(\sup_{\beta \in \mathcal{B}, t \in [0, \tau]} \sqrt{n} |\hat{R}^{(k)}(t, \beta) - r^{(k)}(t, \beta)| > \sqrt{nh}\varepsilon\right) \\ &< c_k (\sqrt{nh}\varepsilon)^{v_k} e^{-2nh^2\varepsilon^2} \rightarrow 0, \text{ as } nh^2 \rightarrow \infty. \end{aligned}$$

Therefore, $\sup_{\beta \in \mathcal{B}, t \in [h, \tau-h]} |\hat{S}_h^{(k)}(t, \beta) - E\{\hat{S}_h^{(k)}(t, \beta)\}|$ converge to 0 in probability as $nh^2 \rightarrow \infty$. Also, since

$$\sup_{\beta \in \mathcal{B}, t \in [h, \tau-h]} \left| E\{\hat{S}_h^{(k)}(t, \beta)\} - s^{(k)}(t, \beta)m(t) \right| = O(h^2), \quad \sup_{\beta \in \mathcal{B}, t \in [0, h]} |s^{(k)}(t, \beta)m(t) - s^{(k)}(h, \beta)m(h)| = O(h)$$

, and $\sup_{\beta \in \mathcal{B}, t \in [0, h]} |s^{(k)}(t, \beta)m(t) - s^{(k)}(h, \beta)m(h)| = O(h)$, and $\sup_{\beta \in \mathcal{B}, t \in [\tau-h, \tau]} |s^{(k)}(t, \beta)m(t) - s^{(k)}(\tau-h, \beta)m(\tau-h)| = O(h)$, the uniform consistency of $\hat{S}_h^{(k)}(t, \beta)$ holds. Moreover, the

monotone bounded stochastic process $n^{-1} \sum_{i=1}^n N_i(t)$ converges in probability to its limits $\Lambda^c(t) = E\{N_i(t)\}$ as $n \rightarrow \infty$. By the law of large numbers for *i.i.d.* random variables,

$n^{-1} \sum_{i=1}^n \int_0^\tau Z_i(t) dN_i(t)$ converges in probability to $\int_0^\tau E\{Z_i(t) dN_i(t)\}$. Hence by Lemma 2.1 of [28], we have $\hat{\beta}_h$ converges in probability to β_0 .

Next, we prove the asymptotic normality of $\sqrt{n} \hat{U}_h(\beta_0)$, which can be written as

$$\begin{aligned} \sqrt{n} \hat{U}_h(\beta_0) &= n^{-1/2} \sum_{i=1}^n \left\{ \int_0^\tau Z_i(t) dN_i(t) - \int_0^\tau \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} dN_i(t) \right\} \\ &\quad + n^{-1/2} \sum_{i=1}^n \left\{ \int_0^\tau \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} dN_i(t) - \int_0^\tau \frac{\hat{S}_h^{(1)}(t, \beta_0)}{\hat{S}_h^{(0)}(t, \beta_0)} dN_i(t) \right\}. \end{aligned} \tag{10}$$

Because $N(t)$ is a bounded monotone stochastic process, $n^{-1/2} \sum_{i=1}^n \{N_i(t) - \Lambda^c(t)\}$

converges weakly to a zero mean tight Gaussian process. Moreover, $\left\{ \frac{s^{(1)}(t, \beta_0) - \hat{S}_h^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0) - \hat{S}_h^{(0)}(t, \beta_0)} \right\}$ has total bounded variation and converges in probability to 0. By Lemma 4.2 of [29], the second term on the right hand side of Equation (10) is

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n \left\{ \int_0^\tau \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} dN_i(t) - \int_0^\tau \frac{\hat{S}_h^{(1)}(t, \beta_0)}{\hat{S}_h^{(0)}(t, \beta_0)} dN_i(t) \right\} &= -n^{1/2} \int_0^\tau \left\{ \frac{\hat{S}_h^{(1)}(t, \beta_0)}{\hat{S}_h^{(0)}(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right\} \lambda^c(t) dt + o_p(1) \\ &= -n^{1/2} \int_0^\tau \left\{ \frac{\hat{S}_h^{(1)}(t, \beta_0)}{\hat{S}_h^{(0)}(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)m(t)}{\hat{S}_h^{(0)}(t, \beta_0)} \right\} \lambda^c(t) dt - n^{-1/2} \int_0^\tau \left\{ \frac{s^{(1)}(t, \beta_0)m(t)}{\hat{S}_h^{(0)}(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right\} \lambda^c(t) dt + o_p(1) \\ &\stackrel{def}{=} I + II + o_p(1). \end{aligned}$$

For ease of presentation, we introduce a few new notations. Let g be a function of bounded variation on $[0, \tau]$. Define $S_i^{(k)}(t, \beta) = \int_0^\tau K_h(t-u) Y_i(u) Z_i(u)^k \exp\{Z_i(u)\beta\} dO_i(u)$ for $t \in [h, \tau-h]$, and $S_i^{(k)}(t, \beta) = S_i^{(k)}(h, \beta)$ for $t \in [0, h]$, $S_i^{(k)}(t, \beta) = S_i^{(k)}(\tau-h, \beta)$ for $t \in [\tau-h, \tau]$.

Thus $\hat{S}_h^{(k)}(t, \beta) = n^{-1} \sum_{i=1}^n S_i^{(k)}(t, \beta)$. We also establish the following two properties in order to prove the asymptotic distribution of I and II . (Detailed proof for the two properties can be found in the end of the Appendix.)

- i. It can be shown that the function classes $\{ \int_0^u g(t) S_i^{(k)}(t, \beta) dt \}$ and $\{ \int_0^u g(t) Y_i(t) Z_i(t)^k \exp\{\beta_0 Z_i(t)\} dO_i(t) \}$ are bounded and monotone in u and thus is Donsker. By the functional central limit theorem, for $u \in [0, \tau]$, we have

$$\sqrt{n} \int_0^u g(t) \hat{S}_h^{(k)}(t, \beta_0) dt - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^u g(t) Y_i(t) Z_i(t)^k \exp\{\beta_0 Z_i(t)\} dO_i(t) = O(\sqrt{nh^2}) + O_p(\sqrt{h}).$$

Note that the right hand side of the above equation will be $o_p(1)$ if $nh^4 \rightarrow 0$ and $h \rightarrow 0$.

- ii. Under condition (A9), along the same line of [30], can be shown that $\sqrt{n} \int_0^\tau g(t) \{ \hat{S}^{(0)}(t, \beta_0)^{-1} - s^{(0)}(t, \beta_0)^{-1} m(t)^{-1} \} \{ \hat{S}^{(k)}(t, \beta_0) - s^{(k)}(t, \beta_0) m(t) \} dt = o_p(1)$

For I , we have

$$\begin{aligned} I &= -n^{1/2} \int_0^\tau \left\{ \frac{\hat{S}_h^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)m(t)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right\} \lambda^c(t) dt + o_p(1) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \int_0^\tau \frac{\lambda^c(t)}{s^{(0)}(t, \beta_0)m(t)} Y_i(t) Z_i(t) \exp\{\beta_0 Z_i(t)\} dO_i(t) - \int_0^\tau \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \lambda^c(t) dt \right\} + o_p(1), \end{aligned}$$

where the first equation holds by using (ii) and the second equation holds from (i). Similarly, for II , we have

$$II = \sqrt{n} \int_0^\tau s^{(1)}(t, \beta_0) \frac{\hat{S}_h^{(0)}(t, \beta_0) - s^{(0)}(t, \beta_0)m(t)}{s^{(0)}(t, \beta_0)^2 m(t)} \lambda^c(t) dt + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \int_0^\tau \frac{s^{(1)}(t, \beta_0) \lambda^c(t)}{s^{(0)}(t, \beta_0)^2 m(t)} Y_i(t) \exp\{\beta_0 Z_i(t)\} dO_i(t) - \int_0^\tau \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \lambda^c(t) dt \right\} + o_p(1).$$

Hence we have $\sqrt{n} \hat{U}_h(\beta_0) = n^{-1/2} \sum_{i=1}^n \psi(\beta_0) + o_p(1)$, where

$$\psi(\beta_0) = \int_0^\tau \{Z_i(t) - \mathcal{E}(t, \beta_0)\} dN_i(t) - \int_0^\tau Y_i(t) \{Z_i(t) - \mathcal{E}(t, \beta_0)\} \exp\{\beta_0 Z_i(t)\} \{s^{(0)}(t, \beta_0)m(t)\}^{-1} \lambda^c(t) dO_i(t).$$

Define $\Gamma_{\hat{h}}(\beta) = - \hat{U}_h(\beta) / \beta$ and $\Gamma(\beta) = - U(\beta) / \beta$, that is,

$$\hat{\Gamma}_h(\beta) = n^{-1} \sum_{i=1}^n \int_0^\tau \left[\left\{ \frac{\hat{S}_h^{(1)}(t, \beta)}{\hat{S}_h^{(0)}(t, \beta)} \right\}^2 - \frac{\hat{S}_h^{(2)}(t, \beta)}{\hat{S}_h^{(0)}(t, \beta)} \right] dN_i(t)$$

and

$$\Gamma(\beta) = \int_0^\tau \left[\left\{ \frac{s^{(1)}(t, \beta)}{s^{(0)}(t, \beta)} \right\}^2 - \frac{s^{(2)}(t, \beta)}{s^{(0)}(t, \beta)} \right] \lambda^c(t) dt.$$

Arguing as before, we can show that $\Gamma_{\hat{h}}(\beta)$ converges to $\Gamma(\beta)$ in probability for $\beta \in \mathcal{B}$. By applying Cauchy–Schwarz inequality, it can be shown that both $\Gamma_{\hat{h}}(\beta)$ and $\Gamma(\beta)$ are positive definite. Applying a Taylor series expansion, we have $\hat{U}_h(\beta_{\hat{h}}) - \hat{U}_h(\beta_0) = \Gamma_{\hat{h}}(\beta^*) (\beta_{\hat{h}} - \beta_0)$, where β^* lies on the line segment between $\beta_{\hat{h}}$ and β_0 . It follows the consistency of $\beta_{\hat{h}}$ for β_0 as well as the continuity of $\Gamma(\beta)$ at β_0 that $\Gamma_{\hat{h}}(\beta^*)$ converges to $\Gamma(\beta_0)$ in probability. Hence by Slutsky’s Theorem, $n^{1/2}(\beta_{\hat{h}} - \beta_0)$ converges to a mean zero normal distribution with variance $\Gamma(\beta_0)^{-1} \Omega(\beta_0) \Gamma(\beta_0)^{-1}$, where $\Omega(\beta_0) = E\{\psi_1(\beta_0)^2\}$

Proof of (i)—Let $R_i^{(k)}(t, \beta)$ denote the i th term of $R^{(k)}(t, \beta)$, that is,

$$R_i^{(k)}(t, \beta) = \int_0^t Y_i(u) Z_i(u)^k \exp\{Z_i(u)\beta\} dO_i(u)$$

Let $a_i(u)$ denote the i th term of the left hand side of the equation in property (i), that is, $a_i(u) = \int_0^u g(t) S_i^{(k)}(t, \beta_0) dt - \int_0^u g(t) dR_i^{(k)}(t, \beta_0)$.

We aim to show $n^{-1/2} \sum_{i=1}^n a_i(u) = O(\sqrt{nh^2}) + O_p(\sqrt{n})$.

We first show that $\sup_{u \in [0, h]} |Ea_i(u)| = O(h^2)$ and $\sup_{u \in [0, h]} E|a_i(u)| = O(h)$. Let $f_1^{(k)}(t)$ denote the derivative of $s^{(k)}(t, \beta_0)m(t)$ with respect to t . For $0 < u < h$, we have

$$\begin{aligned}
 & \sup_{u \in [0, h]} |Ea_i(u)| = \sup_{u \in [0, h]} \left| E \int_0^u g(t) S_i^{(k)}(t, \beta_0) dt - E \int_0^u g(t) dR_i^{(k)}(t, \beta_0) \right| \\
 & = \sup_{u \in [0, h]} \left| E \int_0^u \int_0^{2h} K_h(h-s) g(t) dR_i^{(k)}(s, \beta_0) dt - E \int_0^u g(t) dR_i^{(k)}(t, \beta_0) \right| \\
 & = \sup_{u \in [0, h]} \left| \int_0^u g(t) \int_0^{2h} K_h(h-s) dr^{(k)}(s, \beta_0) dt - \int_0^u g(t) dr^{(k)}(t, \beta_0) \right| \\
 & \leq \sup_{u \in [0, h]} \left| \int_0^u g(t) s^{(k)}(h, \beta_0) m(h) dt - \int_0^u g(t) s^{(k)}(t, \beta_0) m(t) dt \right| + O(h^2) \\
 & \leq \sup_{u \in [0, h]} \left| \int_0^u g(t) f_1^{(k)}(h)(h-t) dt \right| + O(h^2) \\
 & = O(h^2).
 \end{aligned} \tag{11}$$

And

$$\sup_{u \in [0, h]} E|a_i(u)| \leq h \cdot \left\{ \sup_{t \in [0, h]} E|g(t) S_i^{(k)}(t, \beta_0)| + \sup_{t \in [0, h]} |g(t) s_i^{(k)}(t, \beta_0) m(t)| \right\} = O(h). \tag{12}$$

Similarly, we can show $\sup_{u \in [\tau-h, \tau]} |Ea(u)| = O(h^2)$, and $\sup_{u \in [\tau-h, \tau]} E|a(u)| = O(h)$.

For $h < u < \tau - h$, it can be shown that

$$\begin{aligned}
 & \sup_{u \in [h, \tau-h]} |Ea_i(u)| = \sup_{u \in [h, \tau-h]} \left| E \int_0^u g(t) S_i^{(k)}(t, \beta_0) dt - E \int_0^u g(t) dR_i^{(k)}(t, \beta_0) \right| \\
 & = \sup_{u \in [h, \tau-h]} \left| E \int_0^h g(t) S_i^{(k)}(t, \beta_0) dt - E \int_0^h g(t) dR_i^{(k)}(t, \beta_0) \right| + \\
 & \quad \sup_{u \in [h, \tau-h]} \left| E \int_h^u \int_{t-h}^{t+h} \{g(t) - g(s)\} K_h(t-s) dR_i^{(k)}(s, \beta_0) dt \right| + \\
 & \quad \sup_{u \in [h, \tau-h]} \left| E \int_0^{2h} \int_h^{s+h} K_h(t-s) dt g(s) dR_i^{(k)}(s, \beta_0) - E \int_h^{2h} g(s) dR_i^{(k)}(s, \beta_0) \right| + \\
 & \quad \sup_{u \in [h, \tau-h]} \left| E \int_{u-h}^u \int_{s+h}^u K_h(t-s) dt g(s) dR_i^{(k)}(s, \beta_0) + E \int_u^{u+h} \int_{s-h}^u K_h(t-s) dt g(s) dR_i^{(k)}(s, \beta_0) \right| \\
 & = Ia + Ib + Ic + Id
 \end{aligned}$$

Note that $Ia = O(h^2)$ from (11). Also, we have

$$Ib \leq \sup_{u \in [h, \tau-h]} \left| \int_h^u \int_{t-h}^{t+h} \{g(t) - g(s)\} K_h(t-s) s^{(k)}(t, \beta_0) m(t) ds dt \right| + O(h^2) = O(h^2).$$

Let $f_2^{(k)}(t)$ denote the derivative of $g(t) s^{(k)}(t, \beta_0) m(t)$, then

$$\begin{aligned}
 Ic &= \sup_{u \in [h, \tau-h]} \left| E \int_0^{2h} \int_h^{s+h} K_h(t-s) dt g(s) dR_i^{(k)}(s, \beta_0) - E \int_h^{2h} g(s) dR_i^{(k)}(s, \beta_0) \right| \\
 &= \sup_{u \in [h, \tau-h]} \left| \int_0^{2h} \int_{1-s/h}^1 K(x) dx g(s) s^{(k)}(s, \beta_0) m(s) ds - \int_h^{2h} g(u) s^{(k)}(s, \beta_0) m(s) ds \right| \\
 &= \sup_{u \in [h, \tau-h]} \left| \int_0^h \int_{1-s/h}^1 K(x) dx g(s) s^{(k)}(s, \beta_0) m(s) ds + \int_h^{2h} \int_{1-s/h}^{-1} K(x) dx g(s) s^{(k)}(s, \beta_0) m(s) ds \right| \\
 &= \sup_{u \in [h, \tau-h]} \left| \int_0^h \int_{1-s/h}^1 K(x) dx \left\{ g(s) s^{(k)}(s, \beta_0) m(s) - g(0) s^{(k)}(0, \beta_0) m(0) \right\} ds - \int_h^{2h} \int_{-1}^{1-s/h} K(x) dx \left\{ g(s) s^{(k)}(s, \beta_0) m(s) - g(0) s^{(k)}(0, \beta_0) m(0) \right\} ds \right| \\
 &\leq \sup_{u \in [h, \tau-h]} \left| \int_0^h \int_{1-s/h}^1 K(x) dx f_2^{(k)}(0) \cdot s ds - \int_h^{2h} \int_{-1}^{1-s/h} K(x) dx f_2^{(k)}(0) \cdot s ds \right| + O(h^2) \\
 &= O(h^2).
 \end{aligned}$$

$$\begin{aligned}
 E(Id) &= \sup_{u \in [h, \tau-h]} \left| E \int_{u-h}^u \int_{s+h}^u K_h(t-s) dt dg(s) dR_i^{(k)}(s) + E \int_u^{u+h} \int_{s-h}^u K_h(t-s) dt dg(s) dR_i^{(k)}(s) \right| \\
 &= \sup_{u \in [h, \tau-h]} \left| - \int_{u-h}^u \int_{(u-s)/h}^1 K(x) dx g(s) u^{(k)}(s, \beta_0) m(s) ds + \int_u^{u+h} \int_{-1}^{(u-s)/h} K(x) dx g(s) u^{(k)}(s, \beta_0) m(s) ds \right| \\
 &= \sup_{u \in [h, \tau-h]} \left| - \int_{u-h}^u \int_{(u-s)/h}^1 K(x) dx \left\{ g(s) u^{(k)}(s, \beta_0) m(s) - g(u) u^{(k)}(u, \beta_0) m(u) \right\} ds + \int_u^{u+h} \int_{-1}^{(u-s)/h} K(x) dx \left\{ g(s) u^{(k)}(s, \beta_0) m(s) - g(u) u^{(k)}(u, \beta_0) m(u) \right\} ds \right| \\
 &= O(h^2).
 \end{aligned}$$

Therefore, we have $\sup_{u \in [h, \tau-h]} |Ea(u)| = O(h^2)$. Following similar steps as above, we can prove $\sup_{u \in [h, \tau-h]} |Ea(u)| = O(h)$.

Combining the results for $u \in [h, \tau-h]$, $u \in [0, h]$ and $u \in [\tau-h, \tau]$, we have $\sup_{u \in [0, \tau]} |Ea(u)| = O(h^2)$ and $\sup_{u \in [0, \tau]} |Ea(u)| = O(h)$. Since $\int_0^u g(t) dR_i^{(k)}(t, \beta_0) dt$ and $\int_0^u g(t) dR_i^{(k)}(t, \beta_0)$ are bounded monotone functions in u , $n^{-1/2} \sum_{i=1}^n [a_i(u) - E\{a_i(u)\}]$ ($0 \leq u \leq \tau$) converge weakly to a zero-mean Gaussian process with variance function less than $M_1 h$, where M_1 is a constant. Thus we have $n^{-1/2} \sum_{i=1}^n a_i(u) = O(\sqrt{nh^2}) + O_p(\sqrt{h})$.

Proof of (ii)—We now give the proof of (ii) using similar arguments as in [30]. To ensure a non-zero denominator $\hat{S}_h^{(0)}$, we define $\tilde{S}_h^{(0)}(t, \beta) = \hat{S}_h^{(0)}(t, \beta)$ when $\hat{S}_h^{(0)}(t, \beta) > a/\log(n)$ and $\tilde{S}_h^{(0)}(t, \beta) = a/\log(n)$ if $\hat{S}_h^{(0)}(t, \beta) \leq a/\log(n)$, where a is a constant. Then we have $\sqrt{n} \sup_{t \in [0, \tau]} |\tilde{S}_h^{(0)}(t, \beta_0) - \hat{S}_h^{(0)}(t, \beta_0)| = o_p(1)$ by calculating its L_2 -norm using Equation (9).

Define $\hat{S}_{-i}^{(0)}(t, \beta_0) = n^{-1} \sum_{k \neq i} \int_0^\tau K_h(t-u) Y_k(u) \exp\{Z_k(u)\beta_0\} dO_k(u)$ and $\hat{S}_{-ij}^{(0)}(t, \beta_0) = n^{-1} \sum_{k \neq i, j} \int_0^\tau K_h(t-u) Y_k(u) \exp\{Z_k(u)\beta_0\} dO_k(u)$. We further define $\tilde{S}_{-i}^{(0)}(t, \beta_0) = I(\hat{S}_{-i}^{(0)}(t, \beta_0) > a/\log n) \cdot \hat{S}_{-i}^{(0)}(t, \beta_0) + I(\hat{S}_{-i}^{(0)}(t, \beta_0) \leq a/\log n) \cdot a/\log n$, and

$\tilde{S}_{-ij}^{(0)}(t, \beta_0) = I(\hat{S}_{-ij}^{(0)}(t, \beta_0) > a/\log n) \cdot \hat{S}_{-ij}^{(0)}(t, \beta_0) + I(\hat{S}_{-ij}^{(0)}(t, \beta_0) \leq a/\log n) \cdot a/\log n$. We also

$$\begin{aligned} L_i(t) &= \int_0^t S_i^{(k)}(u, \beta_0) du \\ &\quad - \int_0^t s^{(k)}(u, \beta_0) m(u) du, \quad h_i(t) \\ &= g(t) \tilde{S}_{-i}^{(0)}(t, \beta_0)^{-1} \\ &\quad - g(t) s^{(0)}(t, \beta_0)^{-1} m(t)^{-1}, \quad h_{ij}(t) \\ &= g(t) \tilde{S}_{-ij}^{(0)}(t, \beta_0)^{-1} \end{aligned}$$

define $-g(t) s^{(0)}(t, \beta_0)^{-1} m(t)^{-1}$.

We first show

$\sqrt{n} \int_0^\tau g(t) \{ \tilde{S}_{-i}^{(0)}(t, \beta_0)^{-1} - s^{(0)}(t, \beta_0)^{-1} m(t)^{-1} \} \{ \hat{S}^{(k)}(t, \beta_0) - s^{(k)}(t, \beta_0) m(t) \} dt$, which is $n^{-1/2} \sum_{i=1}^n \int_0^\tau h_i(t) dL_i(t)$, converges in probability to 0 by proving L_2 -convergence. Note that

$$\begin{aligned} &\frac{1}{n} E \left[\sum_{i=1}^n \int_0^\tau h_i(u) dL_i(u) \right]^2 \\ &= \frac{1}{n} E \left[\sum_{i=1}^n \sum_{j=1}^n \int_0^\tau \int_0^\tau \{ h_{ij}(u) + h_i(u) - h_{ij}(u) \} \{ h_{ij}(v) + h_j(v) - h_{ij}(v) \} dL_i(u) dL_j(v) \right] \\ &= \frac{1}{n} E \left[\sum_{i=1}^n \sum_{j=1}^n \int_0^\tau \int_0^\tau h_{ij}(u) h_{ij}(v) dL_i(u) dL_j(v) + 2 \sum_{i=1}^n \sum_{j=1}^n \int_0^\tau \int_0^\tau h_{ij}(u) \{ h_j(v) - h_{ij}(v) \} dL_i(u) dL_j(v) \right. \\ &\quad \left. + \sum_{i=1}^n \sum_{j=1}^n \int_0^\tau \int_0^\tau \{ h_i(u) - h_{ij}(u) \} \{ h_j(v) - h_{ij}(v) \} dL_i(u) dL_j(v) \right] \stackrel{def}{=} Ie + If + Ig. \end{aligned}$$

We then prove Ie, If, Ic are all $o(1)$ term. To prove this we first note that $\sup_{t \in [0, \tau]} |EL_f(t)| = O(h^2)$ and $\sup_{t \in [0, \tau]} |ES_i^{(k)}(t, \beta_0) - s^{(k)}(t, \beta_0) m(t)| = O(h^2)$. Also, when $h = O(n^{-\nu})$ and $\nu < 1/2$, we have

$$E\{ \sup_{u \in [0, \tau]} |h_i(u)|^2 \} = o(1), \quad E\{ \sup_{u \in [0, \tau]} |h_{ji}(u)|^2 \} = o(1), \quad E\{ n \sup_{u \in [0, \tau]} |h_{ji}(u) - h_i(u)|^2 \} = o(1).$$

Therefore, for Ie , we have

$$\begin{aligned} Ie &= (n-1) E \{ \int_0^\tau h_{ij}(u) dL_i(u) \int_0^\tau h_{ij}(v) dL_j(v) \} + \frac{1}{n} \sum_{i=1}^n E \{ \int_0^\tau h_i(u) dL_i(u) \}^2 \\ &\leq (n-1) E \{ \sup_{u \in [0, \tau]} |h_{ij}(u)|^2 \} \left\{ \int_0^\tau |ES_i^{(k)}(u, \beta_0) - s^{(k)}(u, \beta_0) m(u)| du \right\}^2 + \frac{1}{n} \sum_{i=1}^n E \{ \int_0^\tau h_i(u) dL_i(u) \}^2 \\ &= o(1). \end{aligned}$$

For If , we have

$$\begin{aligned}
 |If| &= |2(n-1)E \{ \int_0^\tau h_{ij}(u)dL_i(u) \int_0^\tau \{h_j(v)-h_{ij}(v)\}dL_j(v) \}| \\
 &= 2(n-1) |E \int_0^\tau h_{ij}(u)dL_i(u) [EL_j(\tau)\{h_j(\tau)-h_{ij}(\tau)\} - \int_0^\tau EL_j(u)d\{h_j(u)-h_{ij}(u)\}]| \\
 &\leq o(1)+2(n-1)E \left[\left| \int_0^\tau h_{ij}(u)dL_i(u) \int_0^\tau |EL_j(u)| \left| d \left\{ \frac{n^{-1}g(u)S_i^{(0)}(u)}{\tilde{S}_{-ij}^{(0)}(u)\tilde{S}_{-j}^{(0)}(u)} \right\} \right| \right| \right] \\
 &\leq o(1)+2(n-1)n^{-1}E \left[\left| \int_0^\tau h_{ij}(u)dL_i(u) \right| \cdot O(\{\log(n)\}^3) \cdot \int_0^\tau |EL_j(u)| \left\{ S_i^{(0)}(u)|dg(u)|+g(u)|\dot{S}_i^{(0)}(u)|du+g(u)S_i^{(0)}(u)|\hat{S}_{-j}^{(0)}(u)|du \right\} \right] \\
 &= o(1)+2(n-1)n^{-1}E \left| \int_0^\tau h_{ij}(u)dL_i(u) \right| \cdot O(\{\log(n)\}^3) \cdot O(1)=o(1),
 \end{aligned}$$

where $\dot{S}(\cdot)$ denotes the derivative of $S(\cdot)$. For Ig , by Cauchy-Schwartz Inequality,

$$\begin{aligned}
 |Ig| &\leq 1/n \sum_{i=1}^n \sum_{j=1}^n |E \int_0^\tau \{h_i(u)-h_{ij}(u)\}dL_i(u) \int_0^\tau \{h_j(v)-h_{ij}(v)\}dL_j(v)| \\
 &\leq 1/n \sum_{i=1}^n \sum_{j=1}^n [E\{\int_0^\tau \{h_i(u)-h_{ij}(u)\}dL_i(u)\}^2]^{1/2} [E\{\int_0^\tau \{h_j(u)-h_{ij}(u)\}dL_j(u)\}^2]^{1/2} \\
 &\leq 1/n \sum_{i \neq j} E \sup_{u \in [0, \tau]} |h_1(u)-h_{12}(u)|^2 E\{\int_0^t dL_1(u)\}^2 = o(1).
 \end{aligned}$$

Therefore, we have $1/\sqrt{n} \sum_{i=1}^n \int_0^\tau h_i(u)dL_i(u) = o_p(1)$. Moreover, since

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau g(u) \{ \tilde{S}_h^{(0)}(u, \beta_0)^{-1} - \tilde{S}_{-i}^{(0)}(u, \beta_0)^{-1} \} dL_i(u) = o_p(1),$$

we have proved the equation

$$\sqrt{n} \int_0^\tau g(t) \{ \tilde{S}^{(0)}(t, \beta_0)^{-1} - s^{(0)}(t, \beta_0)^{-1} m(t)^{-1} \} \{ \hat{S}^{(k)}(t, \beta_0) - s^{(k)}(t, \beta_0) m(t) \} dt = o_p(1).$$

Since $\sqrt{n} \sup_{t \in [0, \tau]} | \tilde{S}_h^{(0)}(t, \beta_0) - \hat{S}_h^{(0)}(t, \beta_0) | = o_p(1)$, we have

$$\sqrt{n} \int_0^\tau g(t) \{ \hat{S}^{(0)}(t, \beta_0)^{-1} - s^{(0)}(t, \beta_0)^{-1} m(t)^{-1} \} \{ \hat{S}^{(k)}(t, \beta_0) - s^{(k)}(t, \beta_0) m(t) \} dt = o_p(1).$$

Hence (ii) is proved.

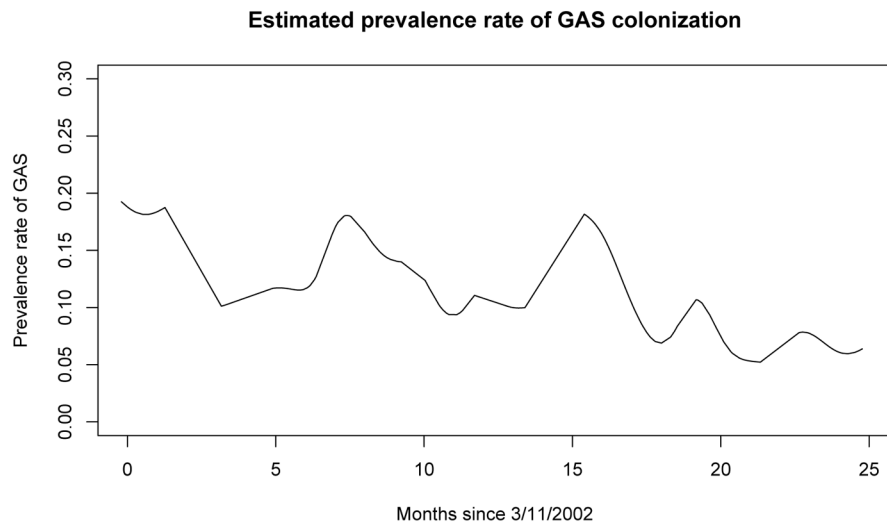


Figure 1. Estimated GAS prevalence rate in the Indian pharyngitis study.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Simulation results for β with a binary covariate process. Bias and ES are the empirical bias ($\times 1000$) and empirical standard deviation ($\times 1000$) of 1000 regression parameter estimates; ASE is the estimated asymptotic standard error ($\times 1000$) for the proposed estimator; BSE is the bootstrap standard error ($\times 1000$) based on 500 bootstrapped samples; CR is the coverage rate of the 95% bootstrap confidence interval; RE is relative efficiency given by the ratio of the empirical MSE of an estimator to that of the rate ratio estimator under the perfect scenario where the covariate process is monitored continuously until loss to follow-up.

% Missing	Covariate process without trend					Covariate process with trend						
	Bias	ES	ASE	BSE	CR	RE	Bias	ES	ASE	BSE	CR	RE
Cross-product ratio estimator												
0%	4	61		62	0.96	0.69	-125	70		68	0.51	0.15
20%	3	64		64	0.95	0.62	-127	72		70	0.50	0.15
40%	5	69		68	0.94	0.56	-128	76		74	0.56	0.14
60%	0	74		76	0.95	0.45	-133	82		82	0.58	0.12
Estimated score function approach												
0%	5	57	58	58	0.94	0.77	0	65	63	64	0.93	0.75
20%	4	61	60	62	0.95	0.66	-2	69	66	68	0.95	0.68
40%	5	68	64	69	0.95	0.53	-5	76	70	75	0.93	0.54
60%	1	79	70	81	0.94	0.39	-11	87	77	89	0.94	0.38
All covariates carried forward approach												
0%	-65	53	53	53	0.77	0.36	-91	59	58	58	0.61	0.27
20%	-90	55	54	54	0.63	0.23	-123	60	59	59	0.44	0.17
40%	-119	58	56	56	0.41	0.14	-165	63	61	61	0.21	0.10
60%	-166	61	61	61	0.22	0.08	-220	65	66	66	0.07	0.06
Carriage covariates carried forward approach												
0%	5	59	60	60	0.95	0.73	-20	67	65	65	0.92	0.63
20%	4	65	66	66	0.96	0.63	-32	73	71	71	0.91	0.49
40%	5	73	75	75	0.94	0.44	-55	82	80	80	0.87	0.34
60%	2	90	91	91	0.97	0.31	-98	103	98	98	0.80	0.16

Simulation results for $\hat{\beta}$ with a continuous covariate process. Bias and ES are the empirical bias ($\times 1000$) and empirical standard deviation ($\times 1000$) of 1000 regression parameter estimates; ASE is the estimated asymptotic standard error ($\times 1000$) for the proposed estimator; BSE is the bootstrap standard error ($\times 1000$) based on 500 bootstrapped samples; CR is the coverage rate of the 95% bootstrap confidence interval; RE is relative efficiency given by the ratio of the empirical MSE of an estimator to that of the rate ratio estimator under the perfect scenario where the covariate process is monitored continuously until loss to follow-up.

Table 2

% Missing	Covariate process without trend					Covariate process with trend						
	Bias	ES	ASE	BSE	CR	RE	Bias	ES	ASE	BSE	CR	RE
	Estimated score function approach											
0%	4	91	89	90	0.95	0.98	0	93	92	90	0.92	1.00
20%	4	92	92	90	0.95	0.96	1	95	97	92	0.93	0.96
40%	5	93	99	92	0.95	0.94	0	96	104	93	0.92	0.93
60%	6	98	108	97	0.95	0.85	4	98	114	98	0.92	0.90
	All carriage covariates carried forward approach											
0%	11	92	91	0.95	0.95	-36	95	94	0.91	0.82		
20%	10	92	91	0.95	0.96	-51	96	94	0.87	0.72		
40%	6	92	91	0.95	0.95	-78	97	95	0.81	0.56		
60%	3	93	91	0.95	0.96	-130	100	98	0.68	0.32		
	Carriage covariates carried forward approach											
0%	21	94	93	0.97	0.88	-37	97	95	0.89	0.79		
20%	27	96	94	0.95	0.83	-55	98	97	0.86	0.67		
40%	37	99	95	0.94	0.74	-89	101	100	0.80	0.48		
60%	60	106	102	0.93	0.56	-165	111	108	0.61	0.22		

Table 3

A tabulation of the number of carriage and pharyngitis visits broken down by Group A streptococcus throat colonization from 305 school-children from Vellore India followed for 2 years.

Group A Strep	Carriage Data Monthly Visit	Event Data Pharyngitis Events
Negative	2504	531
Positive	323	110
Total	2827	641

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript