Published in final edited form as:

Lifetime Data Anal. 2010 April ; 16(2): 271–298. doi:10.1007/s10985-009-9136-2.

Estimation of semiparametric regression model with longitudinal data

Yanqing Sun

Department of Mathematics and Statistics, The University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223, USA

Abstract

In a longitudinal study, an individual is followed up over a period of time. Repeated measurements on the response and some time-dependent covariates are taken at a series of sampling times. The sampling times are often irregular and depend on covariates. In this paper, we propose a sampling adjusted procedure for the estimation of the proportional mean model without having to specify a sampling model. Unlike existing procedures, the proposed method is robust to model misspecification of the sampling times. Large sample properties are investigated for the estimators of both regression coefficients and the baseline function. We show that the proposed estimation procedure is more efficient than the existing procedures. Large sample confidence intervals for the baseline function are also constructed by perturbing the estimation equations. A simulation study is conducted to examine the finite sample properties of the proposed estimators and to compare with some of the existing procedures. The method is illustrated with a data set from a recurrent bladder cancer study.

Keywords

Asymptotic efficiency; Data-driven bandwidth selection; Censored follow-up times; Kernel smoothing; Model misspecification; Panel count data; Proportional mean model; Sampling adjusted estimation; Weighted least squares estimator

1 Introduction

We consider semiparametric modeling of covariate effects on a longitudinal response process based on repeated measurements observed at a series of sampling times. Suppose that there is a random sample of n subjects. For the i th subject, let $Y_i(t)$ be the response process and let $Z_i(t)$ be the possibly time-dependent covariates of dimensions $p \times 1$ over the time interval $[0, \tau]$. For example, $Y_i(t)$ can be the response of the i th individual observed at time t , or the cumulative number of events that have occurred up to time t . We consider the following marginal proportional means model for $Y_i(t)$, $0 \leq t \leq \tau$,

$$\mu_i(t) = E\{Y_i(t) | Z_i(t)\} = \mu_0(t) \exp\{\beta^T Z_i(t)\}, \quad i=1, \dots, n \quad (1)$$

where $\mu_0(t)$ is a completely unspecified function and β is a p -dimensional vector of unknown parameters. The notation β^T represents transpose of a vector or matrix β .

In longitudinal studies, subjects are followed over a period of time and the responses are taken at different time points. Suppose that the observations of $Y_i(t)$ are taken at the sampling time points $0 \leq t_{i1} < t_{i2} < \dots < t_{in_i} \leq \tau$, where n_i is the total number of observations on the i th subject. The sampling times are often irregular and depend on covariates. In

addition, some subjects may drop out of the study early. Let $N_i(t) = \sum_{j=0}^{n_i} I(t_{ij} \leq t)$ be the number of observations taken on the i th subject by time t , where $I(\cdot)$ is the indicator function. Let C_i be the end of follow-up time or censoring time for the i th subject. Then the responses for the i th subject can only be observed at the time points before C_i . Thus $N_i(t)$ can be written as $N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is a counting process for sampling times.

Nonparametric and semiparametric modeling of longitudinal data have been studied extensively in recent years. These include semiparametric methods by Moyeed and Diggle (1994); Zeger and Diggle (1994); Liang et al. (2003), and nonparametric methods by Hoover et al. (1998); Wu et al. (1998); Wu and Zhang (2002), Wu and Liang (2004), Martinussen and Scheike (1999, 2000, 2001); Lin and Ying (2001); Scheike (2002); Sun and Wu (2005) and Fan and Li (2004). Most of the aforementioned works focus on additive time-varying coefficients regression models for the mean response.

When $Y_i(t)$, $0 \leq t \leq \tau$, is a continuously observed counting process subject to right censoring, model (1) has been studied by Andersen and Gill (1982); Lawless and Nadeau (1995), and Lin et al. (2000) among others. Data collected on the individual response processes $Y_i(t)$, $0 \leq t \leq \tau$, at a finite set of sampling times are also called panel data. Many authors have studied nonparametric estimation of panel count data in the absence of covariates; cf. Sun and Kalbfleisch (1995); Wellner and Zhang (2000); Lu et al. (2007); Hu et al. (2009). The semiparametric model (1) for the mean function allows the study of a treatment effect based on panel data, where the responses are not limited to the count data. Zhang (2002) proposed a semiparametric pseudolikelihood method for model (1) under the assumption that $Y_i(t)$ is a nonhomogeneous Poisson process. For panel count data, model (1) has been studied by Sun and Wei (2000); Cheng and Wei (2000), and Hu et al. (2003). These works generally assume that the sampling times are independent of covariates or follow a proportional mean rate model to account for possible dependence on the covariates. However, mis-specification of the sampling model may mislead the inferences for the response process. Assuming time independent covariate processes, Hu et al. (2003, Sect. 2.2) proposed estimation procedure that does not require modelling of the observation process but assuming that the observation processes are discrete and that a subject has observations at some time points with positive probability. This approach is useful for situations where the panel count processes are observed at a fixed set of time points and would require discretizing the time scale in practice when the condition is not met.

We propose a sampling adjusted procedure for estimation of model (1) without having to specify a sampling model for observation times whether it is proportional, additive or transformation mean rate model, cf. Lin et al. (2000, 2001) and Scheike (2002). The proposed method applies to longitudinal response processes not limited to count data. Consider the following general mean rate model for sampling times:

$$E \left\{ dN_i^*(t) | Z_i(t) \right\} = \alpha(t, Z_i(t)) dt = \alpha_i(t) dt, \quad (2)$$

where, other than some imposed smooth conditions, the function $\alpha(t, z)$ is completely unspecified. We focus on the statistical analysis of model (1) for modelling the mean response without having to worry about the implications of model misspecification of the sampling times. Thus our estimator for β is robust to model misspecification of the sampling

times. We also show that the proposed estimator for β is more efficient than the estimators proposed by Cheng and Wei (2000) and by Hu et al. (2003). In addition, a sampling adjusted estimator for the baseline function $\mu_0(t)$ is proposed and its pointwise confidence intervals are constructed using the perturbing method of Jin et al. (2001). In the case when some of the covariates may be time-dependent, an added advantage of the proposed method is that $Z_i(t)$ need not to be observed at all time t . Only the values at sampling times are needed, which is practically convenient since the time-dependent covariates are usually observed at the sampling times.

The rest of the paper is organized as follows. In Sect. 2.1, we propose a sampling adjusted procedure for estimation of the proportional means model. Large sample properties are investigated. Pointwise confidence intervals for the baseline function is constructed in Sect. 2.2. The asymptotic efficiency of the proposed estimator is compared to the existing procedures in Sect. 2.3. A simulation study is presented in Sect. 3. An application of the proposed method to a recurrent bladder cancer study is given in Sect. 4. All proofs are given in the Appendix.

2 Estimation of proportional mean regression model

In this section, we propose an estimation procedure for model (1) based on the observations $\{(Y_i(t_{ij}), Z_i(t_{ij})); j = 1, \dots, n_i, i = 1, \dots, n\}$. These are the values of $\{(Y_i(t), Z_i(t)), 0 \leq t \leq \tau\}$ observed at sampling times or the jump time points of $N_i(t) = N_i^*(t \wedge C_i), i = 1, \dots, n$. Assume that the processes $N_i^*(\cdot)$ and $Y_i(\cdot)$ are independent given $Z_i(\cdot)$. The asymptotic results of the proposed estimators are proved. The asymptotic efficiency is discussed.

2.1 Estimation procedures and asymptotic properties

Under model (1) and Condition A given in the Appendix, $Y_i(t) dN_i(t) - \mu_0(t) \exp\{\beta^T Z_i(t)\} dN_i(t)$ has mean zero for $0 \leq t \leq \tau$. The estimation of $\mu_0(t)$ for fixed β can be derived based on this property. Because of sparsity of the data at time t , we gather the data around a neighborhood of t through kernel smoothing. This motivates the following estimate for $\mu_0(t)$ when β is known:

$$\tilde{\mu}_0(t; \beta) = \frac{\sum_{i=1}^n \int_0^\tau K_b(t-s) Y_i(s) dN_i(s)}{\sum_{i=1}^n \int_0^\tau K_b(t-s) \exp\{\beta^T Z_i(s)\} dN_i(s)}, \quad (3)$$

where $K_b(x) = K(x/b)/b$, $K(\cdot)$ is a kernel function and b is the bandwidth. Let

$$\begin{aligned} \tilde{S}_y(t) &= n^{-1} \sum_{i=1}^n \int_0^\tau K_b(t-s) Y_i(s) dN_i(s) \\ \tilde{S}^{(k)}(t; \beta) &= n^{-1} \sum_{i=1}^n \int_0^\tau K_b(t-s) Z_i^{\otimes k}(s) \exp\{\beta^T Z_i(s)\} dN_i(s) \end{aligned}$$

for $k = 0, 1, 2$, where $Z^{\otimes 2} = ZZ^T$, $Z^{\otimes 1} = Z$ and $Z^{\otimes 0} = 1$. Define $\tilde{Z}(t; \beta) = \tilde{S}^{(1)}(t; \beta)/\tilde{S}^{(0)}(t; \beta)$. We can write $\mu_0(t; \beta) = \tilde{S}_y(t)/\tilde{S}^{(0)}(t; \beta)$.

A profile weighted least squares type estimator for β can be obtained by minimizing

$$l(\beta) = \sum_{i=1}^n \int_{t_1}^{t_2} W_i(s) \left(Y_i(s) - \tilde{\mu}_0(s; \beta) \exp\{\beta^T Z_i(s)\} \right)^2 dN_i(s), \quad (4)$$

where $W_i(t)$ are weight processes and $[t_1, t_2]$ is a subinterval of $[0, \tau]$. The trimming of the boundary points is to avoid the complicated boundary problems involved in the proofs of the asymptotic properties. Our simulations have shown that t_1 and t_2 can be taken to be very close to 0 and τ , respectively, usually less than one bandwidth in difference. Note that $\partial\mu_0(t; \beta)/\partial\beta = -\mu_0(t; \beta) Z(t; \beta)$. By taking the partial derivatives of $l(\beta)$ with respect to β , we have

$$\partial l(\beta) / \partial \beta = \sum_{i=1}^n \int_{t_1}^{t_2} W_i(s) \tilde{\mu}_0(s; \beta) \exp\{\beta^T Z_i(s)\} \left(Z_i(s) - \tilde{Z}(s; \beta) \right) \left(Y_i(s) - \tilde{\mu}_0(s; \beta) \exp\{\beta^T Z_i(s)\} \right) dN_i(s). \quad (5)$$

Let $W_i(t) = W(t)(\tilde{\mu}_0(t; \beta) \exp\{\beta^T Z_i(t)\})^{-1}$ where $W(t)$ is some weight process not depending on β and i . We obtain the following estimation function:

$$U(\beta) = \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s; \beta) \right) \left(Y_i(s) - \tilde{\mu}_0(s; \beta) \exp\{\beta^T Z_i(s)\} \right) dN_i(s). \quad (6)$$

Let $\hat{\beta}$ be the estimator of β such that $U(\hat{\beta}) = 0$. We show that $\hat{\beta}$ is asymptotically consistent and has an asymptotic normal distribution. A estimator for the baseline mean function is given by $\hat{\mu}_0(t) = \mu_0(t; \hat{\beta})$.

As we shall see later in Sect. 2.3, the proposed estimator gains efficiency over the estimator of Hu et al. (2003) by centering $Y_i(\cdot)$ around its estimated mean, thus reducing variance, while Hu et al. (2003) avoided dealing with the baseline function $\mu_0(\cdot)$ in their estimation of parametric components.

Let $\tilde{\mu}_1(t; \beta) = \partial\tilde{\mu}_0(t; \beta)/\partial\beta$. Taking partial derivative of $U(\beta)$ with respect to β , we have

$$\begin{aligned} \frac{\partial U(\beta)}{\partial \beta^T} = & - \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s; \beta) \right)^{\otimes 2} \tilde{\mu}_0(s; \beta) \exp\{\beta^T Z_i(s)\} dN_i(s) \\ & - \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(\frac{\tilde{S}^{(2)}(s; \beta)}{\tilde{S}^{(0)}(s; \beta)} - \tilde{Z}(s; \beta)^{\otimes 2} \right) \left(Y_i(s) - \tilde{\mu}_0(s; \beta) \exp\{\beta^T Z_i(s)\} \right) dN_i(s). \end{aligned} \quad (7)$$

By Lemma 1 given in the Appendix, the second term above is at the order of $o_p(n)$. Let β_0 be the true value of β under model (1). By the first order Taylor expansion, we have

$$n^{1/2}(\hat{\beta} - \beta_0) = \left[-n^{-1} \frac{\partial U(\beta^*)}{\partial \beta^T} \right]^{-1} n^{-1/2} U(\beta_0), \quad (8)$$

where β^* is on the line segment between $\hat{\beta}$ and β_0 .

Let $\xi_i(t) = I(C_i \geq t)$, $s_Y(t) = E(\xi_i(t) \alpha_i(t) Y_i(t))$ and $s^{(k)}(t) = E[\xi_i(t) \alpha_i(t) Z_i^{\otimes k}(t) \exp\{\beta^T Z_i(t)\}]$,

for $k = 0, 1, 2$. Let $w(t)$ be a deterministic nonnegative function such that $W(t) \xrightarrow{P} w(t)$ uniformly in t . Define $\bar{z}(t) = S^{(1)}(t)/S^{(0)}(t)$,

$$A = \int_{t_1}^{t_2} w(s) \mu_0(s) s^{(0)}(u) \left(s^{(2)}(u) / s^{(0)}(u) - \bar{z}(u)^{\otimes 2} \right) du, \text{ and}$$

$$\Sigma = E \left[\left(\int_{t_1}^{t_2} W(s) \left(Z_i(s) - \bar{Z}(s) \right) \left(Y_i(s) - \mu_0(s) \exp \{ \beta^T Z_i(s) \} \right) dN_i(s) \right)^{\otimes 2} \right]. \quad (9)$$

The following theorem presents the asymptotic consistency of the estimator $\hat{\beta}$. By (8), the asymptotic normality of $\hat{\beta}$ follows from the asymptotic consistency of $\hat{\beta}$, uniform convergence of $n^{-1} \partial U(\beta) / \partial \beta^T$ in a neighborhood of β_0 and asymptotic normality of $n^{-1/2} U(\beta_0)$.

Theorem 1—Under Condition A given in the Appendix,

$\hat{\beta} \xrightarrow{P} \beta_0, n^{1/2} (\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N(0, A^{-1} \Sigma A^{-1})$ and $\hat{\mu}_0(t) \xrightarrow{P} \mu_0(t)$ uniformly in $t \in [t_1, t_2]$. Consistent estimators for $\hat{A} = -n^{-1} \partial U(\hat{\beta}) / \partial \beta^T$ and

$$\hat{\Sigma} = n^{-1} \sum_{i=1}^n \left(\int_{t_1}^{t_2} W(s) \left(Z_i(s) - \bar{Z}(s; \hat{\beta}) \right) \hat{\epsilon}_i(s) dN_i(s) \right)^{\otimes 2},$$

where $\hat{\epsilon}_i(s) = Y_i(s) - \hat{\mu}_0(s) \exp \{ \hat{\beta}^T Z_i(s) \}$.

From (9), we see that the weight process $W(\cdot)$ assigns the weight $W(t_{ij})$ to the difference $Y_i(t_{ij}) - \mu_0(t_{ij}) \exp \{ \beta^T Z_i(t_{ij}) \}$ at the sampling time t_{ij} . Taking $W(\cdot) = 1$ implies that the longitudinal observations at all the times are equally weighted for the estimation of β , which is what we recommend for most applications. On the other hand, one can choose to emphasize the early or late observations by selecting $W(\cdot)$ to be decreasing or increasing. The optimal choice of $W(\cdot)$ for the observed data such that Σ defined in (9) is minimized is a challenge problem and needs further exploration.

In practice, the appropriate bandwidth can be selected using a leave-one-subject-out cross validation approach suggested by Rice and Silverman (1991). In particular, Let

$$PE(b) = n^{-1} \sum_{i=1}^n \left[\int_0^\tau \left(Y_i(t) - \hat{\mu}_{0(i)}(t) \exp \{ \hat{\beta}_{(i)}^T Z_i(t) \} \right) dN_i(t) \right]^2, \quad (10)$$

where, for a given bandwidth b , $\hat{\beta}_{(i)}$ and $\hat{\mu}_{0(i)}(t)$ are the estimators of β and $\mu_0(t)$ based on the data without subject i . The data-driven bandwidth selection method is to choose the bandwidth b that minimizes the mean squares of fitted residuals $PE(b)$. This data-driven bandwidth selection method is used in Sect. 4 for analyzing the bladder cancer data.

Next, we present an asymptotic result for the estimator $\hat{\mu}_0(t)$ of the baseline function. The result is useful for constructing confidence intervals for the mean response curve given the covariates. Let $\mu_2 = \int u^2 K(u) du$ and $\eta(t) = (s^{(0)}(t))^{-1} [(s_y(t))'' - \mu_0(t)(s^{(0)}(t))'']$, where $(f(t))''$ denotes the second derivative of $f(t)$ with respect to t .

Theorem 2—Under Conditions A and B given in the Appendix,

$$(nb)^{1/2} \left\{ \hat{\mu}_0(t) - \mu_0(t) - \frac{1}{2} b^2 \mu_2 \eta(t) \right\} \xrightarrow{\mathcal{D}} N \left\{ 0, (s^{(0)}(t))^{-2} \sigma^2(t) \right\}$$

for $t \in [t_1, t_2]$. The asymptotic variance can be estimated consistently by $(\tilde{S}^{(0)}(t))^{-2} \hat{\sigma}^2(t)$, where $\hat{\sigma}^2(t) = bn^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) \hat{\epsilon}_i(s) dN_i(s) \right]^2$.

2.2 Constructing confidence intervals for baseline function

The pointwise confidence intervals for the baseline function $\mu_0(t)$ can be constructed based on the asymptotic normality given in Theorem 2. However, since $\hat{\mu}_0(t)$ has a slow convergence rate of $(nb)^{-1/2}$, the resampling-based method yields more accurate coverage probability. Here we present a method by perturbing the estimation equations. This method has been studied and shown to have good empirical properties by Jin et al. (2001) and by other authors. Let $\zeta_1, \dots, \zeta_n$ be iid random variables with mean 1 and variance 1, say, exponential random variables with mean 1. The perturbed estimating equation for $\mu_0(t)$ for fixed β is

$$\sum_{i=1}^n \zeta_i \left[Y_i(t) dN_i(t) - \mu_0(t) \exp \left\{ \beta^T Z_i(t) \right\} dN_i(t) \right] = 0. \quad (11)$$

Similar to (3), we obtain the following smoothed estimator of $\mu_0(t)$ for given β :

$$\mu_0^*(t; \beta) = \frac{\sum_{i=1}^n \zeta_i \int_0^\tau K_b(t-s) Y_i(s) dN_i(s)}{\sum_{i=1}^n \zeta_i \int_0^\tau K_b(t-s) \exp \left\{ \beta^T Z_i(s) \right\} dN_i(s)}. \quad (12)$$

Perturbing the estimation Eq. 6 yields

$$U^*(\beta) = \sum_{i=1}^n \zeta_i \int_{t_1}^{t_2} W(s) \left(Y_i(s) - \mu_0^*(s; \beta) \exp \left\{ \beta^T Z_i(s) \right\} \right) dN_i(s). \quad (13)$$

Let $\hat{\beta}^*$ be the estimator of β such that $U^*(\hat{\beta}^*) = 0$. Let $\hat{\mu}_0^*(t) = \mu_0^*(t; \hat{\beta}^*)$. Similar to Jin et al. (2001), the distributions of $n^{1/2}(\hat{\beta} - \beta)$ and $n^{1/2}(\hat{\mu}_0(t) - \mu_0(t))$ can be approximated by the distributions of $n^{1/2}(\hat{\beta}^* - \beta)$ and $n^{1/2}(\hat{\mu}_0^*(t) - \mu_0(t))$. A $100(1 - \alpha)\%$ confidence interval for $\mu_0(t)$, $0 \leq t \leq \tau$, can be constructed by $(\hat{\mu}_0(t) - q_{1-\alpha/2}(t), \hat{\mu}_0(t) + q_{\alpha/2}(t))$, where $q_{\alpha/2}(t)$, $q_{1-\alpha/2}(t)$ are the $\alpha/2$ and $1 - \alpha/2$ quantiles of $\{\hat{\mu}_0^{*k}(t) - \mu_0(t), k=1, \dots, B\}$ based on B sets of perturbed estimation equations. Alternatively, the confidence interval for $\mu_0(t)$ can be obtained by $\hat{\mu}_0(t) \pm z_{\alpha/2} SE^*(t)$, where $SE^*(t)$ is the standard deviation of the estimators $\{\hat{\mu}_0^{*k}(t), k=1, \dots, B\}$ based on B sets of perturbed estimation equations. Our simulations show that the two approaches have similar performances, only the first is presented in Sect. 3.

2.3 Asymptotic efficiency considerations

Cheng and Wei (2000) studied model (1) by assuming that $\alpha_i(t) = \lambda_0(t)$ does not depend on $Z_i(t)$. Their estimator of β has the asymptotic variance of $A^{-1} \Sigma_{CW} A^{-1}$, where

$$\Sigma_{CW} = E \left\{ \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \bar{z}(s) \right) \left[Y_i(s) dN_i(s) - \mu_0(s) \lambda_0(s) \exp \left\{ \beta^T Z_i(s) \right\} \xi_i(s) ds \right] \right\}^{\otimes 2}.$$

Under noninformative censoring and the independence between $Y_i(\cdot)$ and $N_i^*(\cdot)$ given $Z_i(\cdot)$,

$$\Sigma_{CW} = \Sigma + E \left\{ \int_{t_1}^{t_2} w(s) \left(Z_i(s) - \bar{z}(s) \right) \mu_0(s) \exp \left\{ \beta^T Z_i(s) \right\} (dN_i(s) - \xi_i(s) \lambda_0(s) ds) \right\}^{\otimes 2},$$

where Σ is defined in (9). It is clear that our estimator $\hat{\beta}$ has smaller asymptotic variance, thus more efficient than the estimator of Cheng and Wei (2000).

Hu et al. (2003) studied model (1) under the assumption that the observation process $N_i^*(t)$ follows the conditional proportional mean rate model

$$\alpha_i(t) = \lambda_0(t) \exp \left\{ \alpha^T Z_i \right\} \quad (14)$$

with $\lambda_0(t)$ an unspecified baseline mean function, and that covariate Z_i is time independent. Hu et al. (2003, Sect. 2.3) obtained a joint estimator $(\hat{\beta}^M, \hat{\alpha}^M)$ for (β, α) . Here we show that the HSW procedure is also applicable to time-dependent covariate and that our robust estimator is more efficient than the HSW estimator. We use a different parametrization, expressing the estimation equations as the functions of (β, α) instead of $(\hat{\beta}, \alpha)$ as in Hu et al. (2003), where $\hat{\beta} = \beta + \alpha$. Of course, the different ways of parametrization do not change the estimator for β or its asymptotic variance. Our parametrization only makes it easier to show that the proposed estimator $\hat{\beta}$ is more efficient than the HSW estimator $\hat{\beta}^M$.

Let

$$S_M^{(j)}(t; \beta, \alpha) = n^{-1} \sum_{i=1}^n \xi_i(t) Z_i(t)^{\otimes j} \exp \left\{ (\beta + \alpha)^T Z_i(t) \right\}, \quad \text{for } j=0, 1, 2$$

and $\bar{Z}_M(t; \beta, \alpha) = S_M^{(1)}(t; \beta, \alpha) / S_M^{(0)}(t; \beta, \alpha)$. The joint estimation equations of Hu et al. (2003) for (β, α) using data on $[t_1, t_2] \subset [0, \tau]$ are

$$\begin{aligned} U_1^M(\beta, \alpha) &= \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \bar{Z}_M(s; \beta, \alpha) \right) Y_i(s) dN_i(s) \\ U_2^M(\alpha) &= \sum_{i=1}^n \int_{t_1}^{t_2} W_O(s) \left(Z_i(s) - \bar{Z}_M(s; 0, \alpha) \right) dN_i(s), \end{aligned}$$

where $W(t)$ and $W_O(t)$ are the weight processes. Hu et al. (2003) have focused the investigation on the unit weights of $W(t) = 1$ and $W_O(t) = 1$. Since $W_O(t) = 1$ is the optimal weight for estimating α , we let $W_O(t) = 1$.

Let $\bar{Z}_M(t; \beta, \alpha)$ be the limit of $\bar{Z}_M(t; \beta, \alpha)$ in probability. Let

$$\begin{aligned} M_i^O(t) &= \int_{t_1}^t \xi_i(s) \left[dN_i(s) - \exp \left(\alpha^T Z_i(s) \right) \lambda_0(s) ds \right], \\ M_i^M(t) &= \int_{t_1}^t \xi_i(s) \left[Y_i(s) dN_i(s) - \mu_0(s) \exp \left((\beta + \alpha)^T Z_i(s) \right) \lambda_0(s) ds \right] \text{ and} \\ \Delta &= E \left\{ \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) \mu_0(s) \exp \left\{ \beta^T Z_i(s) \right\} dM_i^O(s) \right\}^{\otimes 2}. \end{aligned}$$

Let

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = E \left\{ \begin{bmatrix} \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) dM_i^M(s) \\ \int_{t_1}^{t_2} \xi_i(s) \left(Z_i(s) - \bar{z}_M(s; 0, \alpha) \right) dM_i^O(s) \end{bmatrix}^{\otimes 2} \right\} \quad (15)$$

where B_{11} is the variance matrix of the top term in the square bracket, B_{12} is the covariance matrix the top term and the bottom term, $B_{21} = B_{12}^T$ and B_{22} is the variance matrix of the bottom term in the square bracket.

The following Theorem shows that our estimator $\hat{\beta}$ is more efficient than the HSW estimator $\hat{\beta}^M$ for β .

Theorem 3

Under the assumption that the observation process $N_i^*(t)$ has the conditional proportional mean rate model (14), the asymptotic variance of HSW estimator $\hat{\beta}^M$ for β is $A^{-1}\Sigma A^{-1} + D$,

where $D = A^{-1}\Delta A^{-1} - A^{-1}B_{12}A_{\alpha}^{-1} - (A^{-1}B_{12}A_{\alpha}^{-1})^T + A_{\alpha}^{-1}B_{22}A_{\alpha}^{-1}$ is semi-positive definite and A_{α} is the usual information matrix associated with the proportional mean rate model defined in Condition (e) of Lin et al. (2000), which can be obtained by letting $w(t) = 1$, $\mu_0(t) = 1$ and $\beta = 0$ in A .

Hu et al. (2003) have taken $[t_1, t_2]$ to be $[0, \tau]$ while we take $[t_1, t_2]$ to be a subinterval of $[0, \tau]$ to avoid dealing with boundary problems. This small[trim] of data will not result in much efficiency loss in analyzing the data. In fact, t_1 and t_2 can be chosen arbitrarily close to 0 and τ , respectively.

When the observation process is a counting process, $M_i^O(t)$ is a martingale. It follows that $B_{22} = A_{\alpha}$ and

$$\begin{aligned} B_{12} &= E \left\{ \left[\int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) \mu_0(s) \exp(\beta^T Z_i(s)) dM_i^O(s) \right] \times \left[\int_{t_1}^{t_2} \xi_i(s) \left(Z_i(s) - \bar{z}_M(s; 0, \alpha) \right) dM_i^O(s) \right]^T \right\} \\ &= E \left\{ \left[\int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) \left(Z_i(s) - \bar{z}_M(s; 0, \alpha) \right)^T \times \mu_0(s) \lambda_0(s) \exp((\beta + \alpha)^T Z_i(s)) ds \right] \right\} \\ &= E \left\{ \left[\int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right)^{\otimes 2} \mu_0(s) \lambda_0(s) \exp((\beta + \alpha)^T Z_i(s)) ds \right] \right\} + E \left\{ \left[\int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) \left(\bar{z}_M(s; \beta, \alpha) - \bar{z}_M(s; 0, \alpha) \right) ds \right] \right\} \\ &= E \left\{ \left[\int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right)^{\otimes 2} \mu_0(s) \lambda_0(s) \exp((\beta + \alpha)^T Z_i(s)) ds \right] \right\} \\ &= A. \end{aligned}$$

By Theorem 3, the asymptotic variance of the HSW estimator $\hat{\beta}^M$ is $A^{-1}(\Sigma + \Delta)A^{-1} - A_{\alpha}^{-1}$. One can also see from the proof of Theorem 3 that $\hat{\beta}^M$ and $\hat{\alpha}^M$ are asymptotically independent. In this case, the matrix $A^{-1}\Delta A^{-1} - A_{\alpha}^{-1}$ is semi-positive definite. This expression shows that the variance for the estimator $\hat{\beta}^M$ is reduced by modeling the observation process $N_i^*(t)$ using the proportional mean rate model as expected. However, the proposed estimator gains more efficiency by subtracting the mean of $Y_i(t)$ in the estimation function $U(\beta)$ in (6).

3 A simulation study

Some numerical simulation results are presented in this section to illustrate the feasibility and validity of the proposed methods. The responses are generated from the following model

$$Y_i(t) = \mu_0(t) \exp(\beta_1 Z_{1i} + \beta_2 Z_{2i}) + \epsilon_i(t), \quad i=1, \dots, n, \quad (16)$$

over the time interval $[0, \tau]$ where $\beta_1 = 1, \beta_2 = -0.5$ and (Z_{1i}, Z_{2i}) are independent identically distributed. We take $\tau = 24$. The Z_{1i} has a Bernoulli distribution with $P(Z_{1i} = 1) = 0.4$ and Z_{2i} is uniformly distributed on $(0, 1)$. We consider two baseline functions $\mu_0(t) = 0.2 + 0.5t$ and $\mu_0(t) = 0.05t^2$. The $\epsilon_i(t)$ has a normal distribution conditional on the i th subject with mean ϕ_i and variance $\sigma_e^2 = 1$, and ϕ_i is normal with mean zero and variance $\sigma_\phi^2 = 1$. The counting process $N_i^*(t)$ is set to be a Poisson process with intensity rate of $\alpha_i(t)$ over the interval $[0, 24]$. We consider two models for $\alpha_i(t)$. One is the proportional mean rate model with $\alpha_i(t) = 0.4 \exp(0.3Z_{1i} + 0.6Z_{2i})$. The other is the additive mean rate model with $\alpha_i(t) = 0.48Z_{1i} + 0.96Z_{2i}$. We generate the censoring random variable C_i from the uniform distribution on $(0, 60)$. With this censoring distribution, about 36% of subjects have censored observations under both the proportional mean rate model and the additive mean rate model. The average uncensored observations per subject is about 12 under the proportional mean rate model and 13 under the additive mean rate model. Trimming away a little bit of boundary points, we take $[t_1, t_2] = [1, 23]$. For simplicity, the weight function $W(t) = 1$ and use the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$.

Table 1 summarizes some simulation results from the estimators of β_1 and β_2 for $n = 100, 200, 300$ under the proportional mean rate sampling model $\alpha_i(t) = 0.4 \exp(0.3Z_{1i} + 0.6Z_{2i})$ and for two different baseline functions. The bandwidths $b = 1.5, 2.0$ and 2.5 are used for the kernel smoothing. The simulation results based on the method of Hu et al. (2003) are indicated with HSW under the bandwidth column. The method of Hu et al. (2003) is developed under the proportional mean rate sampling model. In Table 2 we also list its simulation results under the misspecified additive mean rate sampling model $\alpha_i(t) = 0.48Z_{1i} + 0.96Z_{2i}$, to show its sensitivity to the model specification and to demonstrate the robust property of the proposed estimator. Each entry in Table 1 and 2 is based on 1000 repetitions (samples), where, for $k = 1, 2$, under $b = 1.5, 2.0$ and 2.5 , $\text{Bias}(\beta_k)$ is the average of the estimation bias of β_k ; $\text{SSE}(\beta_k)$ is the sampling standard error of β_k ; $\text{ESE}(\beta_k)$ is the estimated standard error of β_k ; $\text{CP}(\beta_k)$ is the coverage probability of 95% confidence interval for β_k using $\hat{\beta}_k$, and under $b = \text{HSW}$, they are corresponding summary statistics for the estimators of Hu et al. (2003).

The biases for the proposed estimators are generally small and the coverage probabilities are close to 95% indicating appropriateness of the proposed estimation procedures for β . Under the proportional mean rate model, our estimator outperforms that of Hu et al. (2003) in terms of bias and standard error; see Table 1. When $\mu_0(t) = 0.2 + 0.5t$, the ratios of the standard errors of the proposed estimator to those of the HSW estimator are around 0.72 for β_1 and 0.53 for β_2 . When $\mu_0(t) = 0.05t^2$, the ratios of the standard errors of the proposed estimator to those of the HSW estimator are around 0.42 for β_1 and 0.30 for β_2 . Our estimators have similar performances under the additive mean rate model while the estimators of Hu et al. (2003) fall apart, see Table 2. These numerical results are consistent with the large sample results derived in Sect. 2.

The biases for the baseline function estimator $\hat{\mu}_0(t)$, and the lengths and coverage probabilities of the 95% confidence intervals for $\mu_0(t)$ for $n = 100$ under the proportional

mean rate sampling model $\alpha_i(t) = 0.4 \exp(0.3Z_{1i} + 0.6Z_{2i})$ and the additive mean rate sampling model $\alpha_i(t) = 0.48Z_{1i} + 0.96Z_{2i}$ at a number of the grid points are given Table 3. The perturbation method described in Sect. 2.2 is used to obtain the lengths and the coverage probabilities. The first numbers in parenthesis are biases, the second and third numbers are coverage probabilities and lengths of 95% confidence intervals for $\mu_0(t)$. Each entry in Table 3 is evaluated using 500 repetitions and 500 perturbation samples. We find in a simulation not presented here that using the perturbation method produces better results than using the bootstrap resampling at the subject level. Perhaps this is due to the variability in the number of observations across subjects. Table 3 shows that the biases are generally small, the coverage probabilities are close to 95% nominal level. The lengths of the 95% confidence intervals increase with t due to the increased variations in the estimator $\hat{\mu}_0(t)$.

4 An Application

In this section, we apply the proposed method to analyze a data set from the bladder cancer study, conducted by the Veterans Administration Cooperative Urological Research Group of USA over four years of period. All the 121 subjects entered trial had superficial bladder tumors (Byar 1980). These tumors were removed transurethrally and then subjects were randomly allocated to one of the three treatments, placebo, thiotepa and pyridoxine. There were 47 subjects in the placebo group, 38 in the thiotepa group and the rest in the pyridoxine. Identical tablets were given daily by mouth to the subjects in the placebo and pyridoxine groups, while for the subjects in the thiotepa group, thiotepa was instilled into the bladder for 2 hours once a week for 4 weeks and once a month thereafter. Many subjects had multiple new tumors during the study. The new tumors were removed at the clinical visits of the subjects. One of the study objectives was to evaluate the effectiveness of thiotepa by comparing the placebo and the thiotepa groups in tumor accumulation. Thus the number of subjects of interest is $n = 85$ and the follow-up time is $\tau = 48$ months.

Let $Y_i(t)$ be the number of the accumulated new tumors for subject i by time t , $t \in [0, \tau]$. The process $Y_i(\cdot)$ is only observable at the subject's finite number of clinical visits. The times of clinical visits varied among individuals and it has been noticed that the thiotepa group tended to visit the clinics more frequently than the placebo group. Let $N_i^*(\cdot)$ be the counting process of the visiting times for subject i over the time period $[0, \tau]$. Let Z_i be a 3-D vector with the first component indicating whether the subject was in thiotepa group, and the second and third component being the number of tumors observed at the beginning of the study and the size of the largest initial tumors, respectively. The average number of tumors observed at the beginning of the study is 1.936 for the placebo group and 2.316 for the thiotepa group. The average size of the largest initial tumors is 2.085 for the placebo group and 1.921 for the thiotepa group. The exact censoring time C_i for subject i is not available and is taken to be the subject's last visit time. The bladder cancer data set is published in Hu et al. (2003). The plots of the new tumor accumulations against entry time are given in Fig. 1(a) for subjects in the placebo group and in Fig. 1(b) for subjects in the thiotepa group.

The bladder cancer data has been analyzed by Hu et al. (2003) under the models (1) and (14) where they obtained an estimate of $\hat{\beta}^M = (-1.482, 0.285, -0.083)$ for β and the standard error of (0.329, 0.062, 0.105). To apply the proposed robust method, we first choose a bandwidth using the leave-one-subject-out cross validation method described in Sect. 2. The plot of the mean squares of fitted residuals $PE(b)$ defined in (10) against different bandwidths are given in Fig. 2. The minimal value of $PE(b)$ is found at the bandwidth $b = 9.0$.

We trim a small part of boundary points and take $[t_1, t_2] = [1, 47]$. With $b = 9.0$, our method yields $\hat{\beta} = (-1.310, 0.248, -0.067)$ and the standard error of (0.315, 0.062, 0.098). For this

example, Hu et al. (2003) have shown that the proportional mean rate model (14) fits the clinical visit times well. The two methods produce similar estimates. We have tried different bandwidths ranging from 3.0 to 14.0, the results on the estimation are very similar, differ only on the thousand-th decimal place. We also estimate the baseline function in the model (1) and its 95% pointwise confidence intervals. The plot is given in Fig. 3(a). The plots of mean accumulated new tumors for the placebo group and the thiotepa group at the average number of tumors at the beginning of the study and the average size of largest initial tumors are given in Fig. 3(b), showing that the thiotepa group has much less accumulated new tumors on average. The residual plots of $\{\hat{\epsilon}_i(t), i = 1, \dots, 85\}$ versus entry time are given in Fig. 3(c).

The plot shows that the distribution of $\hat{\epsilon}_i(t)$ for a given t is skewed to the right and is not symmetric around zero. The mean value of $\hat{\epsilon}_i(t)$ seems to be around zero. This is reasonable since the response process $Y_i(t)$ here is the accumulated new tumor counts by time t , whose distribution exhibits similar pattern as that of a Poisson distribution.

5 Discussion

The procedure proposed here has several advantages over the existing methods. It is more flexible. The covariate processes need not to be observed at all the time, which is practically a useful feature when some of the covariates may change with time. It is robust to the assumptions of the sampling models since we do not need to specify a particular form for the conditional mean rate $\alpha(t, Z_i(t))$. This is an important improvement and generalization of the existing methods. One does not need to know whether $\alpha(t, Z_i(t))$ is of the proportional model (Lin et al. 2000), or the additive model (Scheike 2002), or the transformation model (Lin and Ying 2001). The proposed estimator is more efficient than the estimator of Cheng and Wei (2000) which assumes $\alpha(t, Z_i(t))$ not depending on $Z_i(t)$. It is also more efficient than the estimator of Hu et al. (2003) which assumes that $\alpha(t, Z_i(t))$ is proportional.

In the situations where the response process is the number of recurrent events that have occurred by time t , it is expected that the baseline function $\mu_0(t)$ increases with time when covariate Z_i is time-independent. Although the proposed estimator $\hat{\mu}_0(t)$ is still a valid estimator for $\mu_0(t)$, it is desirable to have an estimator that increases with time. Sun and Kalbfleisch (1995) proposed an estimator for the mean function of point processes of recurrent events based on isotonic regression, which is shown to be a pseudo-maximum likelihood estimator by Wellner and Zhang (2000). It would be interesting to construct an isotonic regression type estimator for $\mu_0(t)$.

6 Appendix

Large sample theory, such as presented by Van der Vaart (1998), together with an elegant technical lemma by Lin and Ying (2001), will be applied to prove the asymptotic results. We assume the following conditions throughout the paper:

Condition A. The censoring time C_i is noninformative in the sense that

$E\{dN_i^*(t)|Z_i(t), C_i \geq t\} = E\{dN_i^*(t)|Z_i(t)\}$ and $E\{Y_i(t)|Z_i(t), C_i \geq t\} = E\{Y_i(t)|Z_i(t)\}$; the censoring time C_i is allowed to depend on the covariate process $Z_i(\cdot)$; the process $N_i^*(\cdot)$ is independent of $Y_i(\cdot)$ given $Z_i(\cdot)$; the processes $Y_i(t)$, $Z_i(t)$ and $\alpha_i(t)$, $0 \leq t \leq \tau$, are bounded and their total variations are bounded by a constant; $E|N_i(t_2) - N_i(t_1)|^2 \leq L(t_2 - t_1)$ for $0 \leq t_1 \leq t_2 \leq \tau$, where $L > 0$ is a constant; the weight function $W(t)$ can be written as a difference of two monotone functions, each of which converges in probability to a deterministic function, such that $W(t) \xrightarrow{P} w(t)$; the kernel function $K(\cdot)$ is symmetric with compact support on $[-1,$

1] and bounded variation; $n \rightarrow \infty$, $nb^2 \rightarrow \infty$ and $nb^4 \rightarrow 0$; $s_y(t)$ and $s^{(k)}(t)$, $k = 0, 1, 2$, are twice differentiable; $(s^{(0)}(t))^{-1}$ are bounded over $0 \leq t \leq \tau$; and that A and Σ are positive definite.

Condition B. $E|N_i(t+b) - N_i(t-b)|^{2+v} = O(b)$, for some $v > 0$; the limit

$\sigma^2(t) = \lim_{n \rightarrow \infty} bE \left[\int_0^\tau K_b(t-s) \epsilon_i(s) dN_i(s) \right]^2$ exists and is finite, where $\epsilon_i(s) = Y_i(s) - \mu_0(s) \exp\{\beta^T Z_i(s)\}$.

We remark that under the conditional independence of $N_i^*(\cdot)$ and $Y_i(\cdot)$ given $Z_i(\cdot)$ and assuming noninformative censoring C_i , Condition B holds if $N_i^*(\cdot)$ is Poisson process and if $E\{(\epsilon_i(s))^2 \xi_i(s) \alpha_i(s)\}$ is continuous in $s \in (t-b, t+b)$.

6.1 Technical lemmas

In the rest of the section, we drop β in the terms $\tilde{S}^{(k)}(t; \beta)$ and $\tilde{Z}(t; \beta)$ for ease of notation. Let $S^{(k)}(t) = n^{-1} \sum_{i=1}^n \xi_i(t) \alpha_i(t) Z_i^{(k)}(t) \exp\{\beta^T Z_i(t)\}$, $k=0, 1, 2$. Let $\tilde{Z}(t) = S^{(1)}(t)/S^{(0)}(t)$. Let $M_i(t) = N_i(t) - \int_0^t \xi_i(s) \alpha_i(s) ds$. The process $M_i(t)$ is a mean-zero process.

Lemma 1

Under Condition A, $\tilde{S}^{(k)}(t) \xrightarrow{P} s^{(k)}(t)$ for $k=0, 1, 2$, $\tilde{Z}(t) \xrightarrow{P} \tilde{z}(t)$ and $\tilde{\mu}_0(t; \beta) \xrightarrow{P} \mu_0(t)$, uniformly in $t \in [t_1, t_2]$ and $\beta \in \mathcal{N}(\beta_0)$, a neighborhood of β_0 .

Proof Similar to the proof given in the Appendix of Gilbert and Sun (2005), under Condition A, by Theorem 19.4 and 19.5 of Van der Vaart (1998), we have

$n^{-1} \sum_{i=1}^n \int_0^t Z_i^{\otimes 2}(u) \exp\{\beta^T Z_i(u)\} dN_i(u) \xrightarrow{P} s^{(k)}(t)$ for $k = 0, 1, 2$, and

$n^{-1} \sum_{i=1}^n \int_0^t Y_i(u) dN_i(u) \xrightarrow{P} s_y(t)$, uniformly in $t \times \beta \in [t_1, t_2] \times \mathcal{N}(\beta_0)$. Since $K(\cdot)$ has

bounded variation, $nb^2 \rightarrow \infty$ and $nb^4 \rightarrow 0$, it follows $\tilde{S}^{(k)}(t) \xrightarrow{P} s^{(k)}(t)$ and $\tilde{S}_y(t) \xrightarrow{P} s_y(t)$,

uniformly in $t \times \beta \in [t_1, t_2] \times \mathcal{N}(\beta_0)$, by integrations by parts. Thus, $\tilde{Z}(t) \xrightarrow{P} \tilde{z}(t)$ and

$\tilde{\mu}_0(t; \beta) \xrightarrow{P} \mu_0(t)$, uniformly in $t \times \beta \in [t_1, t_2] \times \mathcal{N}(\beta_0)$. $\square$

Lemma 2

Let $\eta(t) = (s^{(0)}(t))^{-1}[(s_y(t))'' - \mu_0(t)(s^{(0)}(t))'']$.

(a) Under Condition A and B,

$$(nb)^{1/2} \left(\tilde{\mu}_0(t; \beta) - \mu_0(t) - \frac{1}{2} b^2 \mu_2 \eta(t) \right) \xrightarrow{\mathcal{D}} N \left(0, (s^{(0)}(t))^{-2} \sigma^2(t) \right),$$

(b) Under Condition A, $\int_{t_1}^t n^{1/2} \left(\tilde{\mu}_0(s; \beta) - \mu_0(s) \right) ds$ converges weakly to a mean-zero

Gaussian process on $[t_1, t_2]$, and $n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s) \right) \exp\{\beta^T Z_i(s)\} dM_i(s)$ converges weakly to mean-zero normal distribution.

Proof We begin with the proof of the assertion (a). Since $\mu_0(t) = s_y(t)/s^{(0)}(t)$, we have

$$\begin{aligned}
n^{1/2} \left(\tilde{\mu}_0(t; \beta) - \mu_0(t) \right) &= n^{1/2} \left(\frac{\tilde{S}_y(t)}{\tilde{S}^{(0)}(t)} - \frac{s_y(t)}{s^{(0)}(t)} \right) \\
&= \left(\tilde{S}^{(0)}(t) \right)^{-1} n^{1/2} \left(\tilde{S}_y(t) - s_y(t) \right) - s_y(t) \left(\tilde{S}^{(0)}(t) s^{(0)}(t) \right)^{-1} n^{1/2} \left(\tilde{S}^{(0)}(t) - s^{(0)}(t) \right).
\end{aligned} \quad (17)$$

Note that $\int_0^\tau K_b(t-s) s_y(s) ds = s_y(t) + \frac{1}{2} b^2 \mu_2(s_y(t))'' + o(b^2)$ and

$\int_0^\tau K_b(t-s) s^{(0)}(s) ds = s^{(0)}(t) + \frac{1}{2} b^2 \mu_2(s^{(0)}(t))'' + o(b^2)$, uniformly in $t \in [t_1, t_2]$. Let

$S_y(t) = n^{-1} \sum_{i=1}^n \xi_i(t) \alpha_i(t) Y_i(t)$. We have

$$n^{1/2} \left(\tilde{S}_y(t) - s_y(t) \right) = \int_0^\tau K_b(t-s) n^{1/2} \left(S_y(s) - s_y(s) \right) ds + \int_0^\tau K_b(t-s) n^{-1/2} \sum_{i=1}^n Y_i(s) dM_i(s) + \frac{1}{2} n^{1/2} b^2 \mu_2(s_y(t))'' + o(n^{1/2} b^2), \quad (18)$$

and

$$\begin{aligned}
n^{1/2} \left(\tilde{S}^{(0)}(t) - s^{(0)}(t) \right) &= \int_0^\tau K_b(t-s) n^{1/2} \left(S^{(0)}(s) - s^{(0)}(s) \right) ds \\
&+ \int_0^\tau K_b(t-s) n^{-1/2} \sum_{i=1}^n \exp \{ \beta^T Z_i(s) \} dM_i(s) \\
&+ \frac{1}{2} n^{1/2} b^2 \mu_2(s^{(0)}(t))'' \\
&+ o(n^{1/2} b^2).
\end{aligned} \quad (19)$$

By (17), (18), (19), the convergence of $(\tilde{S}^{(0)}(t))^{-1}$ in probability, the weak convergence of the processes $n^{1/2}(S_y(t) - s_y(t))$ and $n^{1/2}(\tilde{S}^{(0)}(t) - s^{(0)}(t))$ on the interval $[t_1, t_2]$, and Lemma 1 of Sun and Wu (2005), we have

$$\begin{aligned}
(nb)^{1/2} \left(\tilde{\mu}_0(t; \beta) - \mu_0(t) \right) &= n^{-1/2} b^{1/2} \left(s^{(0)}(t) \right)^{-1} \int_0^\tau K_b(t-s) \sum_{i=1}^n \xi_i(s) \alpha_i(s) \epsilon_i(s) ds \\
&+ n^{-1/2} b^{1/2} \left(s^{(0)}(t) \right)^{-1} \int_0^\tau K_b(t-s) \sum_{i=1}^n Y_i(s) dM_i(s) \\
&- n^{-1/2} b^{1/2} s_y(t) \left(s^{(0)}(t) \right)^{-2} \int_0^\tau K_b(t-s) \sum_{i=1}^n \exp \{ \beta^T Z_i(s) \} dM_i(s) \\
&+ \frac{1}{2} n^{1/2} b^{5/2} \mu_2(s^{(0)}(t))^{-1} \left[(s_y(t))'' - \mu_0(t) (s^{(0)}(t))'' \right] \\
&+ o_p(b + n^{1/2} b^{5/2}).
\end{aligned} \quad (20)$$

Let $\phi_i(t) = b^{1/2} \int_0^\tau K_b(t-s) \epsilon_i(s) dN_i(s)$. It follows that

$$(nb)^{1/2} \left(\tilde{\mu}_0(t; \beta) - \mu_0(t) - \frac{1}{2} b^2 \mu_2 \eta(t) \right) = (s^{(0)}(t))^{-1} n^{-1/2} \sum_{i=1}^n \phi_i(t) + o_p(1). \quad (21)$$

Note that $E(\phi_i(t)) = 0$. Let $B_n^2 = n\sigma_n^2(t)$, where

$\sigma_n^2(t) = \text{Var}(\phi_i(t)) = bE \left[\int_0^\tau K_b(t-s) \epsilon_i(s) dN_i(s) \right]^2 \rightarrow \sigma^2(t)$. Under Condition A and B, for $v > 0$,

$$\begin{aligned} \sum_{i=1}^n E|\phi_i(t)|^{2+v} &= nb^{-(1+v/2)} E \left| \int_0^\tau K_b((t-s)/b) \epsilon_i(s) dN_i(s) \right|^{2+v} \\ &= O(1) nb^{-(1+v/2)} b = O(1) nb^{-v/2}. \end{aligned}$$

Hence, $\sum_{i=1}^n E|\phi_i(t)|^{2+v} / B_n^{2+v} = O(1) (nb)^{-v/2} = o(1)$. It follows that

$n^{-1/2} \sum_{i=1}^n \phi_i(t) \xrightarrow{\mathcal{D}} N(0, \sigma^2(t))$ by applying the Lindeberg-Feller central limit theorem for double arrays of random variables (cf. Serfling (1980), Corollary, page 32). Consequently,

$$(nb)^{1/2} \left(\tilde{\mu}_0(t; \beta) - \mu_0(t) - \frac{1}{2} b^2 \mu_2 \eta(t) \right) \xrightarrow{\mathcal{D}} N(0, (s^{(0)}(t))^{-2} \sigma^2(t)). \quad (22)$$

Next, we prove the assertion (b). By Lemma 1, $(\tilde{S}^{(0)}(t))^{-1}$ converges in probability uniformly on the interval $[t_1, t_2]$. By Lemma 1 of Sun and Wu (2005), the processes $n^{1/2}(S_y(t) - s_y(t))$, $n^{-1/2} \sum_{i=1}^n \int_{t_1}^t Y_i(s) dM_i(s) n^{1/2}(S^{(0)}(t) - s^{(0)}(t))$ and $n^{-1/2} \sum_{i=1}^n \int_{t_1}^t \exp\{\beta^T Z_i(s)\} dM_i(s)$ jointly converge in distribution to some mean-zero Gaussian processes on $[t_1, t_2]$. By (17), (18) and (19), applying Lemma 2 of Sun and Wu (2005) and Lemma A.1 of Lin and Ying (2001), we have

$$\begin{aligned} & n^{1/2} \left(\int_{t_1}^t \left(\tilde{\mu}_0(s, \beta) - \mu_0(s) \right) ds \right) \\ &= n^{-1/2} \sum_{i=1}^n \int_{t_1}^t (s^{(0)}(s))^{-1} \int_0^\tau K_b(s-u) (Y_i(u) dN_i(u) - s_y(u) du) ds - n^{-1/2} \sum_{i=1}^n \int_{t_1}^t s_y(s) (s^{(0)}(s))^{-2} \int_0^\tau K_b(s-u) (\exp\{\beta^T Z_i(u)\} dN_i(u) - s_0(u) du) ds \\ &= n^{-1/2} \sum_{i=1}^n \int_{t_1}^t (s^{(0)}(s))^{-1} (Y_i(s) dN_i(s) - s_y(s) ds) - n^{-1/2} \sum_{i=1}^n \int_{t_1}^t s_y(s) (s^{(0)}(s))^{-2} (\exp\{\beta^T Z_i(s)\} dN_i(s) - s_0(s) ds) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n \int_{t_1}^t (s^{(0)}(s))^{-1} (Y_i(s) - \mu_0(s) \exp\{\beta^T Z_i(s)\}) dN_i(s) + o_p(1), \end{aligned} \quad (23)$$

which converges weakly to mean-zero Gaussian process by Lemma 1 of Sun and Wu (2005).

The weak convergence of $n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s) \right) \exp \left\{ \beta^T Z_i(s) \right\} dM_i(s)$, follows from the uniform convergence in probability of $W(t)$ and $Z(t)$, the weak convergence of $n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} Z_i(s) \exp \left\{ \beta^T Z_i(s) \right\} dM_i(s)$ and $n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} \exp \left\{ \beta^T Z_i(s) \right\} dM_i(s)$, and applications of Lemma A.1 of Lin and Ying (2001).

6.2 Proof of Theorem 1

Proof of the asymptotic consistency of $\hat{\beta}$

We use a similar argument to the Appendix A.1 of Lin et al. (2001). By Lemma 1 and Theorem 19.4 of Van der Vaart (1998), we have $n^{-1} U(\beta_0) \xrightarrow{P} 0$ as $n \rightarrow \infty$. Further, by (7), $-n^{-1} \partial U(\beta) / \partial \beta \xrightarrow{P} A$ uniformly in a neighborhood of β_0 . Note that A is continuous in β and is nonsingular in a neighborhood of β_0 . Then, for any $\delta > 0$, there exists an event J with $P(J) < \delta$ and a small neighborhood of β_0 inside of which the eigenvalues of $-n^{-1} \partial U(\beta) / \partial \beta$ are bounded away from zero for all large n on J^c , the complement of J . Thus, by the inverse function theorem (Goffman 1965, p.92), we can find a small neighborhood of β_0 , inside of which there exists a unique solution $\hat{\beta}$ to $U(\beta) = 0$ for every sufficiently large n on J^c . Since δ can be arbitrarily small, it follows that unique $\hat{\beta}$ exists in a neighborhood of β_0 with probability 1. The nonnegative definiteness of $-n^{-1} \partial U(\beta) / \partial \beta$ in the entire domain of β implies the global uniqueness of $\hat{\beta}$.

Considering an ϵ neighborhood of β_0 in the preceding arguments for any $\epsilon > 0$, we see that $\hat{\beta}$ converges to β_0 on J^c . This follows by $P(\{\hat{\beta} \in \mathcal{N}_{\epsilon}(\beta_0)\} \cap J^c) \rightarrow 1$ as $n \rightarrow \infty$. Thus, $P(\hat{\beta} \in \mathcal{N}_{\epsilon}(\beta_0)) \rightarrow 1$ since δ can be arbitrarily small, where $\mathcal{N}_{\epsilon}(\beta_0)$ is an ϵ neighborhood of β_0 . This proves the consistency of $\hat{\beta}$.

Proof of the asymptotic normality of $\hat{\beta}$

Taking partial derivative of $U(\beta)$ with respect to β and applying Lemma 1, we have

$$\begin{aligned} \frac{\partial U(\beta)}{\partial \beta^T} &= - \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left[\mu_0(s; \beta) \exp \left\{ \beta^T Z_i(s) \right\} Z_i(s) + \exp \left\{ \beta^T Z_i(s) \right\} \mu_1(s; \beta) \right] \left(Z_i(s) - \tilde{Z}(s) \right)^T dN_i(s) + o_p(n) \\ &= - \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s) \right)^{\otimes 2} \mu_0(s; \beta) \exp \left\{ \beta^T Z_i(s) \right\} dN_i(s) + o_p(n) \\ &= - \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s) \right)^{\otimes 2} \mu_0(s) \exp \left\{ \beta^T Z_i(s) \right\} dN_i(s) + o_p(n), \end{aligned}$$

It follows that

$$n^{-1} \frac{\partial U(\beta)}{\partial \beta^T} \xrightarrow{P} -A. \quad (24)$$

Now, consider

$$\begin{aligned} n^{-1/2} U(\beta) &= n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s) \right) \left(Y_i(s) - \mu_0(s) \exp \left\{ \beta^T Z_i(s) \right\} \right) dN_i(s) \\ &\quad - n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s) \right) \left(\mu_0(s; \beta) - \mu_0(s) \right) \exp \left\{ \beta^T Z_i(s) \right\} dN_i(s). \end{aligned} \quad (25)$$

By Lemma 1, 2 and repeatedly applying Lemma A.1 of Lin and Ying (2001), the second term of (25) equals to

$$\begin{aligned} & n^{1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(\bar{Z}(s) - \tilde{Z}(s) \right) \left(\tilde{\mu}_0(s; \beta) - \mu_0(s) \right) \tilde{S}^{(0)}(s) ds \\ & + n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \tilde{Z}(s) \right) \left(\tilde{\mu}_0(s; \beta) - \mu_0(s) \right) \times \exp \{ \beta^T Z_i(s) \} dM_i(s) = o_p(1). \end{aligned}$$

Hence,

$$n^{-1/2} U(\beta) = n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(Z_i(s) - \bar{z}(s) \right) \left(Y_i(s) - \mu_0(s) \exp \{ \beta^T Z_i(s) \} \right) dN_i(s) + o_p(1) \quad (26)$$

which converges in distribution to a normal random variable with variance equal to Σ .

By (8), (24) and (26), we have

$$n^{1/2} (\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N(0, A^{-1} \Sigma A^{-1}). \quad (27)$$

Proof of the uniform consistency of $\hat{\mu}_0(t)$

The uniform consistency $\hat{\mu}_0(t) \xrightarrow{P} \mu_0(t)$ for $t \in [t_1, t_2]$ follows by $\hat{\beta} \xrightarrow{P} \beta$ and the uniform convergence of $\tilde{\mu}_0(t; \beta) \xrightarrow{P} \mu_0(t)$ on $t \in [t_1, t_2]$ and $\beta \in \mathcal{N}(\beta_0)$ in Lemma 1.

The consistency of $\hat{A}$ and $\hat{\Sigma}$ follows from the consistency of $\hat{\beta}$, the uniform consistency of $\hat{\mu}_0(t)$ and by Lemma 1.

6.3 Proof of Theorem 2

Under Condition A, for $t \in [t_1, t_2]$, $(nb)^{1/2} \left(\tilde{S}^{(0)}(t, \hat{\beta}) - \tilde{S}^{(0)}(t, \beta) \right) \xrightarrow{P} 0$. We have $(nb)^{1/2} \left(\tilde{\mu}_0(t, \hat{\beta}) - \tilde{\mu}_0(t, \beta) \right) \xrightarrow{P} 0$. Thus, by Lemma 2, for $t \in [t_1, t_2]$,

$$(nb)^{1/2} \left(\tilde{\mu}_0(t, \hat{\beta}) - \mu_0(t) - \frac{1}{2} b^2 \mu_2 \eta(t) \right) \xrightarrow{\mathcal{D}} N \left(0, \left(s^{(0)}(t) \right)^{-2} \sigma^2(t) \right).$$

Since $\tilde{S}^{(0)}(t) \xrightarrow{P} s^{(0)}(t)$ by Lemma 1, to prove the consistency of the variance estimator, it suffices to show

$$\hat{\sigma}^2(t) = b n^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) \hat{\epsilon}_i(s) dN_i(s) \right]^2 \xrightarrow{P} \sigma^2(t). \quad (28)$$

The left hand side of (28) equals to

$$bn^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) \epsilon_i(s) dN_i(s) + \int_0^\tau K_b(t-s) (\widehat{\epsilon}_i(s) - \epsilon_i(s)) dN_i(s) \right]^2.$$

The (28) is implied by

$$bn^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) \epsilon_i(s) dN_i(s) \right]^2 \xrightarrow{P} \sigma^2(t) \quad (29)$$

$$bn^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) (\widehat{\epsilon}_i(s) - \epsilon_i(s)) dN_i(s) \right]^2 \xrightarrow{P} 0 \quad (30)$$

Let $\phi_i(t) = b^{1/2} \int_0^\tau K_b(t-s) \epsilon_i(s) dN_i(s)$. The limit (29) holds by using Chebychev's inequality because $E\phi_i(t)^2 \rightarrow \sigma^2(t)$ and

$\text{Var} \{n^{-1} \sum_{i=1}^n (\phi_i^2(t) - E\phi_i^2(t))\} = n^{-1} \{E\phi_i^4(t) - (E\phi_i^2(t))^2\} \leq n^{-1} E\phi_i^4(t) = O((nb^2)^{-1})$, under Condition A and B. To show (30), note that

$$\begin{aligned} & bn^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) (\widehat{\epsilon}_i(s) - \epsilon_i(s)) dN_i(s) \right]^2 \\ & \leq \max_{1 \leq i \leq n} \sup_{s \in [t-b, t+b]} |\widehat{\epsilon}_i(s) - \epsilon_i(s)|^2 bn^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) dN_i(s) \right]^2. \end{aligned}$$

Under Condition A and B, $bE \left[\int_0^\tau K_b(t-s) dN_i(s) \right]^2 = O(1)$, we have

$bn^{-1} \sum_{i=1}^n \left[\int_0^\tau K_b(t-s) dN_i(s) \right]^2 = O_p(1)$ by the Markov inequality. By the consistency of $\hat{\beta}$ and the uniform consistency of $\hat{\mu}_0(t)$ in Theorem 1, we have

$\max_{1 \leq i \leq n} \sup_{s \in [t-b, t+b]} |\widehat{\epsilon}_i(s) - \epsilon_i(s)| \xrightarrow{P} 0$. Hence the limit (30) holds.

6.4 Proof of Theorem 3

By Hu et al. (2003),

$$U_1^M(\beta, \alpha) = \sum_{i=1}^n \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) dM_i^M(s) + o_p(n^{1/2}) \quad (31)$$

$$U_2^M(\alpha) = \sum_{i=1}^n \int_{t_1}^{t_2} \xi_i(s) \left(Z_i(s) - \bar{z}_M(s; 0, \alpha) \right) dM_i^O(s) + o_p(n^{1/2}),$$

where $\bar{z}_M(t; \beta, \alpha)$ is the limit of $\bar{Z}_M(t\beta, \alpha)$ in probability.

Taking the partial derivatives of $U_1^M(\beta, \alpha)$ and $U_2^M(\alpha)$ with respect to β and α , Hu et al. (2003) showed that

$$\begin{aligned}\frac{\partial U_1^M(\beta, \alpha)}{\partial \beta^T} &= - \sum_{i=1}^n \int_{t_1}^{t_2} W(s) \left(\frac{S_M^{(2)}(s; \beta, \alpha)}{S_M^{(0)}(s; \beta, \alpha)} - \bar{Z}_M(s; \beta, \alpha)^{\otimes 2} \right) \times Y_i(s) dN_i(s) \xrightarrow{P} -A \\ \frac{\partial U_2^M(\alpha)}{\partial \alpha^T} &= - \sum_{i=1}^n \int_{t_1}^{t_2} \left(\frac{S_M^{(2)}(s; 0, \alpha)}{S_M^{(0)}(s; 0, \alpha)} - \bar{Z}_M(s; 0, \alpha)^{\otimes 2} \right) dN_i(s) \xrightarrow{P} -A_\alpha,\end{aligned}\quad (32)$$

$$\text{and that } \frac{\partial U_1^M(\beta, \alpha)}{\partial \alpha^T} = \frac{\partial U_1^M(\beta, \alpha)}{\partial \beta^T} \xrightarrow{P} -A \text{ and } \frac{\partial U_2^M(\alpha)}{\partial \beta^T} = 0.$$

By (15), (31) and (32), the asymptotic covariance matrix of (β^M, α^M) is

$$\begin{aligned}\Sigma_M &= \begin{bmatrix} A & A \\ 0 & A_\alpha \end{bmatrix}^{-1} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} A & 0 \\ A & A_\alpha \end{bmatrix}^{-1} \\ &= \begin{bmatrix} A^{-1} & -A_\alpha^{-1} \\ 0 & A_\alpha^{-1} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ -A_\alpha^{-1} & A_\alpha^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_M^{11} & \Sigma_M^{12} \\ \Sigma_M^{21} & \Sigma_M^{22} \end{bmatrix},\end{aligned}\quad (33)$$

where

$$\Sigma_M^{11} = A^{-1} B_{11} A^{-1} - A^{-1} B_{12} A_\alpha^{-1} - (A^{-1} B_{12} A_\alpha^{-1})^T + A_\alpha^{-1} B_{22} A_\alpha^{-1} \quad (34)$$

is the asymptotic variance matrix of β^M , $\Sigma_M^{21} = A_\alpha^{-1} B_{21} A^{-1} - A_\alpha^{-1} B_{22} A_\alpha^{-1}$ is the asymptotic covariance matrix of α^M and β^M , $\Sigma_M^{12} = (\Sigma_M^{21})^T$ and $\Sigma_M^{22} = A_\alpha^{-1} B_{22} A_\alpha^{-1}$ is the asymptotic variance matrix of α^M .

Note that

$$M_i^M(t) = \int_{t_1}^t \xi_i(s) \left[Y_i(s) - \mu_0(s) \exp(\beta^T Z_i(s)) \right] dN_i(s) + \int_{t_1}^t \xi_i(s) \mu_0(s) \exp(\beta^T Z_i(s)) dM_i^O(s). \quad (35)$$

We have the following decomposition for the top term in the square bracket in (15):

$$\begin{aligned}& \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{Z}_M(s; \beta, \alpha) \right) dM_i^M(s) \\ &= \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{Z}_M(s; \beta, \alpha) \right) \left[Y_i(s) - \mu_0(s) \exp(\beta^T Z_i(s)) \right] dN_i(s) \\ &+ \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{Z}_M(s; \beta, \alpha) \right) \mu_0(s) \exp(\beta^T Z_i(s)) dM_i^O(s).\end{aligned}\quad (36)$$

Since the two terms in (36) are uncorrelated, we have

$$\begin{aligned}
B_{11} &= \Sigma + E \left\{ \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) \mu_0(s) \exp \left\{ \beta^T Z_i(s) \right\} dM_i^O(s) \right\}^{\otimes 2} \\
&= \Sigma + \Delta.
\end{aligned} \tag{37}$$

Hence we can write

$$\begin{aligned}
\Sigma_M^{11} &= A^{-1} \Sigma A^{-1} + A^{-1} \Delta A^{-1} - A^{-1} B_{12} A_\alpha^{-1} - \left(A^{-1} B_{12} A_\alpha^{-1} \right)^T + A_\alpha^{-1} B_{22} A_\alpha^{-1} \\
&= A^{-1} \Sigma A^{-1} + D.
\end{aligned} \tag{38}$$

Note that the first term is asymptotic variance matrix of our estimator $\hat{\beta}$. It remain to show that the matrix D is semi-positive definite.

Because the first term in (36) is uncorrelated to the second term in (36) and is also uncorrelated to the bottom term in in the square bracket in (15), replacing the top term in (15) by the second term of its decomposition in (36), we have

$$\begin{aligned}
E & \left\{ \left[\begin{array}{c} \int_{t_1}^{t_2} \xi_i(s) w(s) \left(Z_i(s) - \bar{z}_M(s; \beta, \alpha) \right) \mu_0(s) \exp \left(\beta^T Z_i(s) \right) dM_i^O(s) \\ \int_{t_1}^{t_2} \xi_i(s) \left(Z_i(s) - \bar{z}_M(s; 0, \alpha) \right) dM_i^O(s) \end{array} \right]^{\otimes 2} \right\} \\
&= \begin{bmatrix} \Delta & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.
\end{aligned} \tag{39}$$

Now replacing the middle matrix of (33) by the above matrix, we have that the matrix

$$\begin{bmatrix} A & A \\ 0 & A_\alpha \end{bmatrix}^{-1} \begin{bmatrix} \Delta & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} A & 0 \\ A & A_\alpha \end{bmatrix}^{-1}$$

is semi-positive definite. Hence, its first block matrix

$$D = A^{-1} \Delta A^{-1} - A^{-1} B_{12} A_\alpha^{-1} - \left(A^{-1} B_{12} A_\alpha^{-1} \right)^T + A_\alpha^{-1} B_{22} A_\alpha^{-1}$$

is also semi-positive definite.

References

- Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Annal Stat* 1982;10:1100–1120.
- Byar, DP. The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparison of placebo, pyridoxine, and topical thiotepa. In: Pavone-Macaluso, M.; Smith, PH.; Edsmyrn, F., editors. *Bladder tumors and other topics in urological oncology*. Plenum; New York: 1980. p. 363-370.
- Cheng SC, Wei LJ. Inferences for a semiparametric model with panel data. *Biometrika* 2000;87:89–97.
- Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J Am Stat Assoc* 2004;99:710–723.
- Gilbert PB, Sun Y. Failure time analysis of HIV vaccine effects on viral load and treatment initiation. *Biostatistics* 2005;6:374–394. [PubMed: 15831584]
- Goffman, C. *Calculus of several variables*. Harper and Row; New York: 1965.
- Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 1998;85:809–822.

- Hu XJ, Lagakos SW, Lockhart RA. Generalized least squares estimation of the mean function of a counting process based on panel counts. *Stat Sinica* 2009;19:561–580.
- Hu XJ, Sun J, Wei LJ. Regression parameter estimation from panel counts. *Scand J Stat* 2003;30:25–43.
- Jin Z, Ying Z, Wei LJ. A simple resampling method by perturbing the minimand. *Biometrika* 2001;88:381–390.
- Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995;37:158–168.
- Liang H, Wu H, Carroll RJ. The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient semiparametric models with measurement error. *Biostatistics* 2003;4:297–312. [PubMed: 12925523]
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J Royal Stat Soc B* 2000;62:711–730.
- Lin DY, Wei LJ, Ying Z. Semiparametric transformation models for point processes. *J Am Stat Assoc* 2001;96:620–628.
- Lin DY, Ying Z. Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *J Am Stat Assoc* 2001;96:103–113.
- Lu M, Zhang Y, Huang J. Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* 2007;94:705–718.
- Martinussen T, Scheike TH. A semiparametric additive regression model for longitudinal data. *Biometrika* 1999;86:691–702.
- Martinussen T, Scheike TH. A nonparametric dynamic additive regression model for longitudinal data. *Annal Stat* 2000;28:1000–1025.
- Martinussen T, Scheike TH. Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scand J Stat* 2001;28:303–323.
- Moyeed RA, Diggle PJ. Rates of convergence in semiparametric modelling of longitudinal data. *Aust J Stat* 1994;36:75–93.
- Rice JA, Silverman B. Estimating the mean and covariance structure nonparametrically when the data are curves. *J Royal Stat Soc B* 1991;53:233–243.
- Scheike TH. The additive nonparametric and semiparametric Aalen model as the rate function for a counting process. *Lifetime Data Anal* 2002;8:247–262. [PubMed: 12182121]
- Serfling, RJ. Approximation theorems of mathematical statistics. Wiley; New York: 1980.
- Sun J, Kalbfleisch JD. Estimation of the mean function of point process based on panel count data. *Stat Sinica* 1995;5:279–289.
- Sun Y, Wu H. Semiparametric time-varying coefficients regression model for longitudinal data. *Scand J Stat* 2005;32:21–47.
- Sun J, Wei LJ. Regression analysis of panel count data with covariate-dependent observation and censoring times. *J Royal Stat Soc B* 2000;62:293–302.
- Van der Vaart, AW. Asymptotic statistics. Cambridge University Press; Cambridge: 1998.
- Wellner JA, Zhang Y. Two estimators of the mean of a counting process with panel count data. *Annal Stat* 2000;28:779–814.
- Wu CO, Chiang CT, Hoover D. Asymptotic confidence regions for kernel smoothing of a time-varying coefficient model with longitudinal data. *J Am Stat Assoc* 1998;88:1388–1402.
- Wu H, Liang H. Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scand J Stat* 2004;31:3–19.
- Wu H, Zhang JT. Local polynomial mixed-effects models for longitudinal data. *J Am Stat Assoc* 2002;97:883–897.
- Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* 1994;55:452–459.
- Zhang Y. A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* 2002;89:39–48.

Fig. 1.
The plots of the new tumor accumulations against entry time; **(a)** for the placebo group and **(b)** for the thiotepa group

Fig. 2.
The plot of the mean squares of fitted residuals against different bandwidths

Fig. 3.

(a) The plot of the baseline function estimate $\hat{\mu}_0(t)$ and 95% pointwise confidence intervals with the bandwidth $b = 9.0$ months; (b) The plots of mean accumulated new tumors for the placebo group and the thiotepa group at the average number of tumors (2.106) at the beginning of the study and the average largest initial tumors size (2.012); (c) The residual plots of $\{\hat{\epsilon}_i(t), i = 1, \dots, 85\}$

Table 1

Summary statistics for the estimators $\hat{\beta}$ and $\hat{\beta}_{HSW}$ under the proportional mean sampling model $\alpha_i(t) = 0.4 \exp(0.3Z_{1i} + 0.6Z_{2i})$

$\mu_0(t)$	n	b	$\text{Bias}(\beta_1)$	$\text{Bias}(\beta_2)$	$\text{SSE}(\beta_1)$	$\text{SSE}(\beta_2)$	$\text{ESE}(\beta_1)$	$\text{ESE}(\beta_2)$	$\text{CP}(\beta_1)$	$\text{CP}(\beta_2)$		
0.2 + 0.5t	100	1.5	-0.0001	-0.0008	0.0391	0.0539	0.0370	0.0513	0.933	0.937		
		2.0	-0.0003	-0.0000	0.0369	0.0532	0.0369	0.0510	0.947	0.935		
		2.5	0.0025	0.0011	0.0377	0.0542	0.0369	0.0507	0.944	0.923		
		HSW	-0.0010	0.0018	0.0548	0.1026	0.0522	0.0976	0.940	0.920		
		200	1.5	-0.0004	-0.0003	0.0264	0.0366	0.0263	0.0364	0.946	0.947	
			2.0	0.0013	0.0009	0.0268	0.0379	0.0263	0.0361	0.936	0.932	
	300	2.5	0.0010	0.0003	0.0264	0.0372	0.0263	0.0360	0.949	0.935		
		HSW	-0.0009	0.0015	0.0377	0.0713	0.0369	0.0694	0.942	0.930		
		0.05t ²	100	1.5	0.0004	0.0005	0.0222	0.0306	0.0216	0.0297	0.940	0.952
				2.0	0.0007	0.0002	0.0210	0.0301	0.0216	0.0296	0.960	0.946
				2.5	0.0002	-0.0001	0.0215	0.0294	0.0216	0.0296	0.948	0.952
				HSW	-0.0001	0.0027	0.0309	0.0570	0.0301	0.0568	0.943	0.952
200	1.5			-0.0001	-0.0005	0.0276	0.0381	0.0262	0.0364	0.932	0.939	
	2.0			-0.0003	-0.0000	0.0261	0.0378	0.0261	0.0362	0.947	0.936	
0.05t ²	200	2.5	0.0018	0.0008	0.0267	0.0385	0.0261	0.0360	0.942	0.925		
		HSW	-0.0015	0.0032	0.0662	0.1284	0.0631	0.1230	0.939	0.928		
		300	1.5	-0.0003	-0.0002	0.0187	0.0258	0.0186	0.0257	0.945	0.948	
			2.0	0.0009	0.0006	0.0189	0.0268	0.0186	0.0256	0.938	0.933	
		2.5	0.0006	0.0002	0.0187	0.0264	0.0186	0.0255	0.949	0.932		
			HSW	-0.0011	0.0025	0.0462	0.0904	0.0447	0.0880	0.938	0.934	
	300	1.5	0.0003	0.0003	0.0156	0.0216	0.0152	0.0209	0.940	0.951		
			2.0	0.0004	0.0001	0.0149	0.0213	0.0152	0.0209	0.960	0.946	
			2.5	0.0001	-0.0001	0.0151	0.0208	0.0152	0.0209	0.949	0.950	
			HSW	-0.0006	-0.0036	0.0375	0.0708	0.0365	0.0723	0.938	0.947	
			2.5	0.0001	-0.0001	0.0151	0.0208	0.0152	0.0209	0.949	0.950	
				HSW	-0.0006	-0.0036	0.0375	0.0708	0.0365	0.0723	0.938	0.947

Table 2

Summary statistics for the estimators $\hat{\beta}$ and $\hat{\beta}_{HSW}$ under the additive mean sampling model with $\alpha_i(t) = 0.48Z_{1i} + 0.96Z_{2i}$.

$\mu_0(t)$	n	b	Bias(β_1)	Bias(β_2)	SSE(β_1)	SSE(β_2)	ESE(β_1)	ESE(β_2)	CP(β_1)	CP(β_2)		
0.2 + 0.5t	100	1.5	0.0012	-0.0009	0.0446	0.0526	0.0431	0.0509	0.938	0.933		
		2.0	0.0005	-0.0016	0.0443	0.0544	0.0428	0.0517	0.940	0.929		
		2.5	0.0009	-0.0027	0.0433	0.0520	0.0427	0.0512	0.940	0.947		
		HSW	0.048	-0.1956	0.0615	0.0984	0.0592	0.0955	0.861	0.490		
		200	1.5	0.0006	-0.0008	0.0308	0.0379	0.0307	0.0366	0.938	0.938	
			2.0	0.0007	-0.0014	0.0304	0.0366	0.0305	0.0366	0.951	0.946	
	300	2.5	0.0002	0.0018	0.0304	0.0382	0.0305	0.0366	0.948	0.932		
		HSW	0.0459	-0.1962	0.0414	0.0676	0.0418	0.0683	0.828	0.177		
		0.05t ²	100	1.5	0.0004	-0.0017	0.0248	0.0295	0.0251	0.0300	0.954	0.957
				2.0	0.0003	0.0013	0.0243	0.0308	0.0250	0.0300	0.946	0.936
				2.5	-0.0001	0.0006	0.0251	0.0307	0.0251	0.0298	0.952	0.929
				HSW	0.0459	-0.1959	0.0331	0.0549	0.0341	0.0559	0.750	0.055
200	1.5			0.0008	-0.0004	0.0316	0.0374	0.0305	0.0361	0.937	0.934	
	2.0			0.0003	-0.0010	0.0313	0.0387	0.0302	0.0367	0.940	0.932	
0.05t ²	200	2.5	0.0005	-0.0016	0.0306	0.0370	0.0301	0.0364	0.942	0.951		
		HSW	0.0485	-0.1939	0.0692	0.1233	0.0668	0.1201	0.884	0.640		
		300	1.5	0.0005	-0.0005	0.0217	0.0268	0.0217	0.0259	0.939	0.938	
			2.0	0.0005	-0.0009	0.0214	0.0259	0.0215	0.0259	0.952	0.949	
		300	2.5	0.0002	0.0014	0.0215	0.0271	0.0216	0.0259	0.946	0.934	
			HSW	0.0463	-0.1951	0.0466	0.0856	0.0471	0.0860	0.854	0.392	
	0.05t ²	300	1.5	0.0003	-0.0011	0.0176	0.0208	0.0177	0.0212	0.954	0.959	
			2.0	0.0003	0.0011	0.0172	0.0218	0.0177	0.0212	0.947	0.937	
			2.5	-0.0001	0.0005	0.0177	0.0216	0.0177	0.0210	0.952	0.931	
			HSW	0.0465	-0.1933	0.0380	0.0706	0.0384	0.0706	0.773	0.223	

Table 3

Summary statistics for the estimator $\hat{\mu}_0(t)$ under the proportional mean sampling model (Cox) $\alpha_i(t) = 0.4 \exp(0.3Z_{1i} + 0.6Z_{2i})$ and the additive mean sampling model (Aalen) $\alpha_i(t) = 0.48Z_{1i} + 0.96Z_{2i}$ for $n = 100$. The first number in parenthesis is the bias, the second and third numbers the are coverage probability and length of a 95% confidence interval for $\mu_0(t)$

$\alpha_i(t)$	b	$t = 7$	$t = 10$	$t = 13$	$t = 16$	$t = 19$
$\mu_i(t) = (0.2 + 0.5t) \exp(Z_{1i} - 0.5Z_{2i})$						
Cox	1.5	(-0.0009, 92.4, 0.683)	(0.0020, 94.4, 0.856)	(0.0016, 94.0, 1.037)	(-0.0013, 94.2, 1.228)	(0.0016, 93.4, 1.425)
	2.0	(-0.0036, 92.6, 0.668)	(-0.0012, 94.4, 0.843)	(-0.0026, 93.2, 1.026)	(-0.0054, 94.4, 1.217)	(-0.0037, 93.8, 1.416)
	2.5	(-0.0135, 93.2, 0.657)	(-0.0008, 95.0, 0.836)	(-0.0099, 94.8, 1.021)	(-0.0048, 95.8, 1.220)	(-0.0159, 94.8, 1.414)
Aalen	1.5	(-0.0097, 95.0, 0.748)	(-0.0070, 94.4, 0.979)	(-0.0058, 94.6, 1.217)	(-0.0051, 92.2, 1.465)	(-0.0069, 93.6, 1.713)
	2.0	(-0.0120, 94.2, 0.736)	(-0.0103, 94.0, 0.970)	(-0.0079, 94.8, 1.209)	(-0.0095, 92.2, 1.459)	(-0.0104, 93.8, 1.711)
	2.5	(-0.0092, 92.8, 0.734)	(-0.0113, 92.4, 0.962)	(-0.0191, 93.2, 1.199)	(-0.0211, 92.0, 1.446)	(-0.0212, 93.2, 1.700)
$\mu_i(t) = 0.05t^2 \exp(Z_{1i} - 0.5Z_{2i})$						
Cox	1.5	(0.0191, 92.2, 0.526)	(0.0201, 92.4, 0.706)	(0.0158, 93.2, 0.992)	(0.0085, 95.2, 1.387)	(0.0191, 93.8, 1.881)
	2.0	(0.0327, 91.8, 0.510)	(0.0307, 92.8, 0.699)	(0.0232, 92.6, 0.993)	(0.0138, 95.2, 1.395)	(0.0138, 92.8, 1.895)
Aalen	2.5	(0.0427, 93.0, 0.502)	(0.0512, 94.6, 0.704)	(0.0322, 94.4, 1.010)	(0.0261, 95.6, 1.423)	(0.0003, 93.2, 1.927)
	1.5	(0.0115, 93.6, 0.526)	(0.0096, 94.0, 0.763)	(0.0059, 94.2, 1.142)	(0.0008, 92.8, 1.657)	(0.0006, 94.8, 2.290)
	2.0	(0.0252, 94.0, 0.513)	(0.0194, 94.2, 0.761)	(0.0162, 94.2, 1.145)	(0.0039, 94.4, 1.666)	(0.0020, 94.4, 2.305)
	2.5	(0.0503, 93.0, 0.513)	(0.0410, 92.4, 0.763)	(0.0173, 93.6, 1.144)	(0.0056, 93.2, 1.661)	(-0.0131, 93.4, 2.301)