Published in final edited form as:

Stat Biosci. 2011 December 1; 3(2): 187–207. doi:10.1007/s12561-011-9043-4.

Semiparametric Transformation Rate Model for Recurrent Event Data

Donglin Zeng,

Department of Biostatistics, CB# 7420, University of North Carolina, Chapel Hill, NC 27599-7420

Douglas E. Schaubel, and

Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109-2029

Jianwen Cai

Department of Biostatistics, CB# 7420, University of North Carolina, Chapel Hill, NC 27599-7420,
Tel.: 919-966-7788, Fax: 919-966-3804, cai@bios.unc.edu

Abstract

In this article, we propose a class of semiparametric transformation rate models for recurrent event data subject to right-censoring and potentially stopped by a terminating event (e.g., death). These transformation models include both additive rates model and proportional rates model as special cases. Respecting the property that no recurrent events can occur after the terminating event, we model the conditional recurrent event rate given survival. Weighted estimating equations are constructed to estimate the regression coefficients and baseline rate function. In particular, the baseline rate function is approximated by wavelet function. Asymptotic properties of the proposed estimators are derived and a data-dependent criterion is proposed for selecting the most suitable transformation. Simulation studies show that the proposed estimators perform well for practical sample sizes. The proposed methods are used in two real-data examples: a randomized trial of rhDNase and a community trial of Vitamin A.

Keywords

recurrent event data; transformation model; additive rates model; proportional rates model; terminating event; wavelet approximation

1 Introduction

In many medical studies, subjects may experience the event of interest multiple times. For example, patients with superficial bladder cancer may suffer the occurrence of new tumors; patients who underwent kidney transplantation may have multiple rejection episodes. In a clinical trial conducted to assess the efficacy of rhDNase (a highly purified enzyme) in treating patients with cystic fibrosis, many patients experienced multiple episodes of exacerbations. In studies such as those just described, regression modeling is often of interest in order to simultaneously evaluate the effect of treatment and covariates; the covariates may be of intrinsic interest, or may be modeled to adjust for chance imbalances resulting from imperfect randomization.

Let $N(t)$ be the number of events experienced by time t and let $Z(t)$ be possibly time-dependent covariates. Many statistical methods have been proposed to model the relationship between $N(t)$ and $Z(t)$. Such methods generally fall into one of two broad categories; those that model the conditional mean/rate function of $N(t)$ given $Z(t)$, and those which model the conditional intensity function of $N(t)$ given $Z(t)$ and the event history $\{N(s);$

$s \in (0, t)$. The first category includes the proportional marginal mean/rate models proposed by Pepe and Cai (1993), Lawless and Nadeau (1995), Chen, Wei and Ying (1995), Cook and Lawless (1997), and Lin, Wei, Yang, and Ying (2000), among others. Some work in this category also considers the additive rates model (Schaubel, Zeng and Cai 2006). In the second category, a popular model is the proportional intensity process of Andersen and Gill (1982).

To provide a better fit to the data, Lin, Wei and Ying (2001) proposed a class of more flexible models to describe the relationship between the covariates and recurrent event mean function. In their work, $E[N(t)|Z]$ is modeled as

$$E[N(t)|Z] = g(\mu(t)e^{\beta^T Z}),$$

where $g(\cdot)$ is a known function from a class of monotone transformations. A typical example of $g(\cdot)$ would be the Box-Cox transformation. The best transformation can be selected using an empirical goodness-of-fit criteria. The models in Lin et al (2001) can be considered as extensions of the usual transformation model to recurrent event data. However, the transformations considered in Lin et al (2001) do not include the log-transformation (which induces an additive model). Moreover, only time-independent covariates are allowed in the inference.

In the recurrent event setting, it is often the case that subjects may experience a terminating event which precludes further occurrences of the recurrent event. Referring back to the afore-listed examples, bladder cancer patients may die and hence no longer be subject to further tumor recurrence; similarly, kidney transplant recipients may die and therefore cease to be at risk for subsequent rejection episodes. Several approaches have been suggested to model recurrent/terminal event data. Various authors have proposed estimating the marginal recurrent event mean; e.g., Ghosh and Lin (2000, 2002); Pan and Schaubel (2009); Schaubel and Zhang (2010). Alternatively, one could jointly model the recurrent and terminal events through a shared frailty; e.g., Liu, Wolfe and Huang (2004); Huang and Wang (2004); Ye, Kalbfleisch and Schaubel (2007). In settings where the terminal event is of little or secondary interest, one may choose to model the conditional event rate given survival. If the terminal event is denoted by T , such a model is expressed as $E[dN(t)|Z(t), T \geq t]$.

In this paper, we propose methods for estimating a transformation model suitable for recurrent event data in the presence of a terminating event and covariates which may be time-dependent. In particular, we propose the following transformation model for the conditional rate function given survival

$$E[dN(t)|T \geq t, Z(t)] = I(T \geq t)g(r(t) + \beta^T Z(t))dt, \quad (1)$$

where $Z(t)$ may consist of 'external' time-dependent covariates (Kalbfleisch and Prentice, 2002), $r(t)$ is the baseline rate function, and $g(\cdot)$ is a pre-specified increasing transformation. Specifically, by considering different $g(\cdot)$, the proposed transformation models include some popular rate models including proportional rates model (Lin et al, 2000) with $g(x) = \exp(x)$ and additive rates models (Schaubel et al, 2006) with $g(x) = x$. An even more flexible class of transformations is the Box-Cox family where $g(x) = (1 + sx)^{1/s}$ for $s \geq 0$, where $s = 0$ corresponds to $g(x) = \exp\{x\}$.

Note also that we propose an estimator of the baseline rate, as opposed to mean. Generally, the rate function is defined as the derivative of the mean function. In the context of a conditional event rate model (given survival), the rate has a much clearer interpretation than

the cumulative rate. Very little existing literature has discussed the above transformation rate model. The main difficulty is that, contrast to the transformation model for the mean event function given by Lin et al (2001), the baseline function in the rate model cannot be estimated at the parametric rate. Thus, one would not be able to estimate the baseline function using the empirical estimate, such as that used in Lin et al (2001). However, our current proposal involves circumventing this difficulty through a series of wavelet functions to approximate the baseline rate function and constructing estimating equations in a sieve space. Additionally, our obtained estimate for rate function is smooth so will be useful for visualizing the trend of recurrent events over time. We note that using smoothing estimates for rate/hazard function has been used previously for right-censored or interval censored data (Cai, Hyndman and Wand, 2002; Cai and Betensky, 2003) but there is no work of using smoothing estimates for recurrent events especially with transformations.

The remainder of the article is structured as follows. Section 2 describes the proposed inference procedures and Section 3 presents the asymptotic properties of the proposed estimators. Simulation studies, conducted for various transformation models, are presented in Section 4. In Section 5, we apply our methods to analyze two real-data examples featuring recurrent events. Proofs of all theorems are provided in the Appendix.

2 Inference Procedure

Suppose we observe a recurrent event process, subject to right-censoring, from n independent and identically distributed subjects. We denote the process pertaining to subject i by $N_i(t)$, $0 \leq t \leq T_i \wedge C_i \wedge \tau$ where T_i is the mortality time, C_i is the right-censoring time and τ equals the maximum follow-up time across all subjects. For each subject i , a possibly vector-valued covariate process $Z_i(t) : 0 \leq t \leq T_i \wedge C_i \wedge \tau$ is also observed, with any time-dependent elements assumed to be external (Kalbfleisch and Prentice, 2002). We assume that C_i is independent of the underlying recurrent events, death time and covariates.

We consider the semiparametric transformation rate model defined in (1). Examples of $g(x)$ include the Box-Cox transformation $g(x) = (1 + sx)^{1/s}$ for some $s > 0$ or $g(x) = \exp(x)$.

Denote the at-risk process by $Y_i(t) = I(T_i \wedge C_i \wedge \tau \geq t)$. Our estimates for β and $r(t)$ are proposed as follows. First, we construct the following sieve space: we choose a function $\phi(t)$ satisfying the following conditions:

- i. $\phi(t)$ has a support $[0, \tau]$ and $\phi \in W^{q,2}[0, \tau]$, the Sobolev space consisting of functions whose q th derivative is L_2 -integrable in $[0, \tau]$.
- ii. $\{\phi(t - k) : k \in \mathcal{Z}\}$ is an orthonormal system in $L_2(-\infty, \infty)$, where $\mathcal{Z}$ consists of all integers.
- iii. Let $V_j = \{\sum_k c_k \phi(2^j x - k) : 2^j \sum_k |c_k|^2 < \infty\}$ for any $j \geq 0$. Then $V_0 \subset V_1 \subset \dots$ and $\bigcup_{j \geq 0} V_j$ is dense in $L_2(-\infty, \infty)$.

In the context of wavelet analysis, the sequence V_j is called multi-resolution analysis of $L_2(-\infty, \infty)$ and ϕ is referred to as a father wavelet. Although wavelets are often used in image denoise and signal processing, they have also been successfully applied in statistical estimation for density estimation and nonparametric regression (Hardle et al., 1998). One choice of ϕ is $\psi((x - a)/(b - a)\tau)$, where $\psi(x)$ is the Daubechies 2N-wavelet with support $[a, b]$. Such a wavelet leads to smoother estimate compared to the usual Haar basis and larger N leads to more smoothness. In our subsequent analysis, we find that Daubechies 10-wavelet (see Figure 1) usually works well.

We can approximate the function $r(t)$, $t \in [0, \tau]$, using the Daubechies 2N-wavelet in K_n -level resolution, V_{K_n} . Among the possible bases of V_{K_n} , only $\phi(2^{K_n} t - k + 1)$, $1 \geq k \geq -2^{K_n}\tau$

$+ \tau - 1$ has support in $[0, \tau]$, with the remaining bases having support outside $[0, \tau]$. This motivates the approximation of $r(t)$ by a linear combination of such basis functions which are denoted by $B_1(t), \dots, B_{m_n}(t)$.

Our estimators for β and $r(t)$ are given by $\hat{\beta}$ and $\hat{r}(t) = \sum_{k=1}^{m_n} \hat{\gamma}_k B_k(t)$, respectively, where $\hat{\beta}$ and $\{\hat{\gamma}_k\}$ solve the following generalized estimating equations:

$$\sum_{i=1}^n \int_0^\tau Y_i(t) Z_i(t) g' \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) \times \left\{ dN_i(t) - g \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) dt \right\} = 0 \quad (2)$$

$$\sum_{i=1}^n \int_0^\tau Y_i(t) B_j(t) g' \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) \times \left\{ dN_i(t) - g \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) dt \right\} = 0, j=1, \dots, m_n$$

To solve these estimating questions, we propose the following procedure: we replace the parameters in the weights $g' \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right)$ by previous estimates then iteratively solve equations (2) and (3). Since $g(t)$ is strictly increasing, this guarantees that each step has a unique solution. Furthermore, the final estimators at convergence exactly solve (2) and (3).

When a class of models are used to fit the data, it is natural to ask which model is the best model under some criteria. There are many criteria functions and we consider the QIC as proposed in Pan (2001). In particular, our estimation procedure can be interpreted as maximizing the following quasi-likelihood function:

$$L_n(r, \beta) = -\frac{1}{2} \sum_{i=1}^n \sum_{t \in \mathcal{T}} \{ \delta N_i(t) - g(r(t) + \beta^T Z_i(t)) \delta t \}^2,$$

where $\mathcal{T}$ collects all possible event times, $\delta N_i(t)$ denotes the jump size of N_i at t , and δt is the distance from t to its most recent previous event time. Therefore, according to Pan (2001), a well-performed model selection criterion is to select a $g(\cdot)$ and a K_n which minimize $-2L_n(\hat{r}, \hat{\beta}) + 2\text{trace}(\hat{\Omega} \hat{V})$, where $\hat{\Omega}$ is the negative Hessian matrix of L_n with respect to all parameters and $\hat{V}$ is the sandwiched covariance for parameter estimator. Using this criterion function, one can select both the resolution level, K_n , and the transformation, $g(\cdot)$, most appropriate for the real-data application at hand.

As a remark, for some transformation g such as $g(x) = x$, the obtained estimators for $r(t)$ and β may not necessarily yield non-negative predicted rate. In this case, we suggest to use $\max\{0, g(\hat{r}(t) + \hat{\beta}^T Z(t))\}$ for the predicted rate. We also note that such modification is not necessary for other transformations, for example $g(x) = \exp\{x\}$.

3 Asymptotic Properties

We provide asymptotic results for $(\hat{\beta}, \hat{r})$ in this section. First, we need the following assumptions: Assume that β_0 is the true value for β and $r_0(t)$ is the true baseline rate function for $r(t)$.

- A. The link function, $g(x)$, is continuously twice-differentiable and increasing function. Moreover, there exists a positive constant C_0 such that $\min$

$$\min_{t \in [0, \tau]} g'(x) \Big|_{x=r_0(t)+\beta_0^T Z(t)} \geq C_0 a.s. \quad (4)$$

- B.** If there exist a constant vector a and a deterministic function $b(t)$ such that $P(a^T Z(t) = b(t), \forall t \in [0, \tau]) > 0$, then $a = 0$ and $b(t) = 0$. Moreover, $Z(t)$ is a bounded external process.
- C.** The baseline function, $r_0(t)$, is a continuously q th-differentiable function in $[0, \tau]$ where $q \geq 3$.
- D.** $P(T \wedge C > \tau | Z(t), t \in [0, \tau]) > 0$.
- E.** The number of wavelets, m_n , satisfies

$$m_n \rightarrow \infty, m_n^3 n^{-1/2} \rightarrow 0, \sqrt{n}/m_n^{(2q-2)} \rightarrow 0.$$

Obviously, many link functions satisfy condition (A), including the additive link, $g(x) = x$, and the multiplicative link, $g(x) = \exp(x)$. Condition (B) essentially assumes that elements of the covariate vector $Z(t)$ are linearly independent.

The smoothing condition in (C) is necessary to ensure that the wavelet approximation is reasonable. Condition (D) is a standard identifiability condition in the censored-data setting. In Condition (E), the number of wavelets, m_n , is equal to $[2^{K_n} \tau - \tau + 1]$, where $[x]$ denotes the maximal integer less than or equal to x . Thus, one choice of K_n to satisfy condition (E) is $K_n = \alpha \log_2 n$, where α is some constant in $(1/(4q - 4), 1/6)$.

Under these conditions, the following theorem claims that equations (2) and (3) have a solution and $(\hat{\beta}, \hat{r}(t))$ is consistent.

Theorem 1. Under conditions (A) to (E), there exists a solution $(\hat{\beta}, \{\hat{\gamma}_k\})$ to equations (2) and (3). Moreover,

$$|\hat{\beta} - \beta_0|^2 + m_n \|\hat{r}(t) - r_0(t)\|_{L_2}^2 = O_p \left(\frac{m_n^3}{n} + \frac{1}{m_n^{2q-2}} \right),$$

$$\text{and } \sup_{t \in [0, \tau]} |\hat{r}'(t) - r_0'(t)| \rightarrow 0.$$

The key ideas in the proof of Theorem 1 are as follows. We first construct a function in the sieve space to approximate the true $r_0(t)$ in a suitable norm. In particular, if we define $\{\tilde{\gamma}_k\}$ as the wavelet coefficients of the constructed function, then we delicately find a neighborhood around the $(\beta_0, \{\gamma_k\})$ and show there is a $(\hat{\beta}, \{\hat{\gamma}_k\})$ in this neighborhood solving equations (2) and (3). Specifically, the existence of a solution to (2) and (3) comes from the well-known Brouwer's Fixed Theorem, which can be intuitively understood as follows: if a continuous vector field has its directions pointing to the outside of a ball, then this vector field must have a zero-vector within the ball. In the proof, the rate of m_n (equivalently, the multi-resolution level K_n) as given in (E) ensures that neither bias due to wavelet approximation nor variability due to undersmoothing is dominating. The next theorem provides the asymptotic distribution of the estimators of the regression coefficients.

Theorem 2. In addition to conditions (A)–(E), if we suppose that

$$Q(t) \equiv \frac{E[Y(t)Z(t)g'(r_0(t)+\beta'_0 Z(t))^2]}{E[Y(t)g'(r_0(t)+\beta'_0 Z(t))^2]}$$

is q -th ($q \geq 3$) continuously differentiable in $(0, \tau)$, then $\sqrt{n}(\widehat{\beta} - \beta_0)$ has an asymptotic normal distribution with mean zero and variance equal to

$$\Sigma = \left\{ \int_0^\tau \Psi_1(t) dt \right\}^{-1} \left\{ \int_0^\tau \int_0^\tau \Psi_2(t, s) dt ds \right\} \left\{ \int_0^\tau \Psi_1(t) dt \right\}^{-1},$$

where $\Psi_1(t)$ and $\Psi_2(t, s)$ are matrix functions given by

$$\begin{aligned} \Psi_1(t) &= E[Y(t)(Z(t) - Q(t))(Z(t) - Q(t))^T g'(r_0(t) + \beta_0^T Z(t))^2], \\ \Psi_2(t, s) dt ds &= E[Y(t)Y(s)(Z(t) - Q(t))(Z(s) - Q(s))^T g'(r_0(t) + \beta_0^T Z(t)) \times g'(r_0(s) + \beta_0^T Z(s)) \{dN(t) - g(r_0(t) + \beta_0^T Z(t))dt\} \times \{dN(s) - g(r_0(s) + \beta_0^T Z(s))ds\}]. \end{aligned}$$

The derivation of the asymptotic distribution relies on the linearization of the estimating equations and the existence of a particular function which can be used to remove the variation from $\widehat{r}(t)$.

A consistent estimator for Σ is given by

$$\widehat{\Sigma} = \left\{ \int_0^\tau \widehat{\Psi}_1(t) dt \right\}^{-1} \left\{ \int_0^\tau \int_0^\tau \widehat{\Psi}_2(t, s) dt ds \right\} \left\{ \int_0^\tau \widehat{\Psi}_1(t) dt \right\}^{-1},$$

where we set

$$\begin{aligned} \widehat{Q}(t) &= \frac{\sum_{i=1}^n Y_i(t) Z_i(t) g'(\widehat{r}(t) + \widehat{\beta}^T Z_i(t))^2}{\sum_{i=1}^n Y_i(t) g'(\widehat{r}(t) + \widehat{\beta}^T Z_i(t))^2}, \\ \widehat{\Psi}_1(t) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) (Z_i(t) - \widehat{Q}(t))(Z_i(t) - \widehat{Q}(t))^T g'(\widehat{r}(t) + \widehat{\beta}^T Z_i(t))^2 \end{aligned}$$

and

$$\widehat{\Psi}_2(t, s) = \frac{1}{n} \sum_{i=1}^n \left[Y_i(t) Y_i(s) (Z_i(t) - \widehat{Q}(t))(Z_i(s) - \widehat{Q}(s))^T g'(\widehat{r}(t) + \widehat{\beta}^T Z_i(t)) \times g'(\widehat{r}(s) + \widehat{\beta}^T Z_i(s)) \{dN_i(t) - g(\widehat{r}(t) + \widehat{\beta}^T Z_i(t))dt\} \times \{dN_i(s) - g(\widehat{r}(s) + \widehat{\beta}^T Z_i(s))ds\} \right]$$

A confidence interval for β_0 can be constructed according to the normal approximation.

4 Simulation Studies

We performed simulation studies to examine the finite sample performance of our methods. Simulated recurrent event data were generated from one of the following transformation rate models:

$$\begin{aligned}
E[dN(t)|Z_1, Z_2, T \geq t, \xi] &= I(T \geq t)(1 + 3t + Z_1 - Z_2)\xi dt, \\
E[dN(t)|Z_1, Z_2, T \geq t, \xi] &= I(T \geq t)\{(3t + Z_1 - Z_2)/2 + 1\}^{1/2}\xi dt, \\
E[dN(t)|Z_1, Z_2, T \geq t, \xi] &= I(T \geq t) \exp\{t + Z_1 - Z_2\}\xi dt,
\end{aligned}$$

where Z_1 is a covariate from the uniform distribution in $[0, 1]$, Z_2 follows a Bernoulli distribution with probability 0.5, and ξ is a random effect from the exponential distribution with mean 1. Moreover, the study ending time is set to be $\tau = 1$ and the survival time for each subject is generated from a proportional hazards model with hazards rate $\exp\{-Z_1 + Z_2\}$. In addition, a random right-censoring time C is generated from a uniform distribution between $(\tau/2, 3\tau/2)$. Thus, the recurrent events of each subject are right-censored at $\min(C, T, \tau)$.

After integrating out the random effect, it is clear that each of the above three models satisfies semiparametric transformed rate models with the following corresponding transformations

$$g(x) = 1 + x, g(x) = (1 + x/2)^{1/2}, \text{ and } g(x) = e^x.$$

The first model is an additive model, the second model is the power transformation model with $s = 1/2$, and the third model is a multiplicative model. The two regression coefficients are in turn $\beta_1 = 1$, $\beta_2 = -1$.

For each simulated data set, we applied the proposed estimating equations to find the estimates for the regression coefficients as well as the estimates for the baseline rate function $r_0(t)$. In constructing the sieve function for $r_0(t)$, we chose the father wavelet as one of the Daubechies wavelets; in particular, the DB10 wavelet. This choice ensures the bounded support and the required smoothness of the basis function. In the estimation, the multi-resolution level is selected as 4; however, only small differences are seen when the multi-resolution level was set to 3 or 5. The asymptotic variance for the regression coefficients was estimated using the formulae in Theorem 2.

When data are generated from the additive model, the average number of events experienced by each subject is 2.5. When data are generated from the Box-Cox transformation model with $s = 1/2$, the average number of events experienced by each subject is 3.9. This number becomes 1.7 when data are from the multiplicative model. Table 1 summarizes the results of the simulation studies from 1,000 repetitions. In Table 1, column ‘‘Ave. Bias’’ is the average of the biases for the regression coefficients; column ‘‘SEE’’ is the standard deviation of the 1,000 estimates; column ‘‘Ave. SE’’ is the average of the estimated standard errors; ‘‘95% CP’’ gives the coverage proportion of the 95-percent confidence intervals for the regression coefficients. Table 1 shows that the proposed method produces small-biased estimates; that the estimated standard errors agree reasonably well with the empirical standard deviations and although they tend to underestimate the true variability, the performance improves with increasing sample sizes; and that the confidence intervals have reasonably good coverage.

We also calculated the mean-square errors between the estimated $r_0(t)$ and the true $r_0(t)$ over a grid of points with equal space (of 0.005) in $[0, \tau]$. Such distances for each of the three models are in turn 0.714, 0.444, 0.193 for $n = 400$ and decreases to 0.365, 0.191, 0.087 for $n = 800$. Figure 2 plots one example of the estimated baseline function versus the true $r_0(t)$ within the interval $(0, \tau)$, from which we also notice that the estimates are overall close to the truth but possess a fair amount of variability.

5 Application

We first consider a randomized clinical trial to assess the efficacy of rhDNase, a highly purified recombinant enzyme, in reducing exacerbations of respiratory symptoms for patients with cystic fibrosis (Therneau and Hamilton 1997). A total of 321 patients were assigned to rhDNase and 324 to placebo. Most patients were followed for approximately 170 days. By the end of follow-up, 65 patients in the rhDNase group experienced only 1 exacerbation and 39 experienced ≥ 2 exacerbations; in the placebo group, 97 patients experienced 1 exacerbation and 42 experienced ≥ 2 exacerbations.

We fit the transformation rate models with the family of logarithmic transformations (time is measured in years). The models include the treatment indicator and baseline level of forced expiratory volume per second (FEV_1) as the covariates. We also consider the use of different resolution levels: $K_n = 2, \dots, 8$. To avoid sparse data points in the supports of some wavelet bases, we only use those wavelet bases whose supports contain at least 20% of observed times in the regression. The QIC criterion (as described in Section 2) is used to select the resolution level and the appropriate transformation from the Box-Cox family with $s = 0, 0.1, \dots, 1$.

It turned out that the best transformation corresponded to the additive rates model, with $g(x) = 1+x$, and that the best resolution level was $K_n = 3$. Table 2 summarizes the estimated covariate effects. From the additive model results in Table 2, it is concluded that the treatment significantly reduced the recurrence of the exacerbation and the patients with higher FEV_1 had significantly less exacerbations.

The first plot of Figure 3 depicts the estimated baseline rate function, $\hat{r}(t)$, which indicates a bell-shape. The second plot of Figure 3 gives the estimated cumulative baseline function,

$\int_0^t \hat{r}(s)ds$, using the proposed approach (solid curve) along with the estimated cumulative baseline function in the additive model using the method in Schaubel et al (2006) (dashed line). This plot shows that both methods provide similar cumulative baseline rate estimates, especially at the beginning of the follow-up. Note that the methods of Schaubel et al (2006) do not produce a baseline rate estimator, only its cumulative counterpart.

As a second example, we apply the proposed models to analyze the data from the Vitamin A Community Trial (VACT) conducted in Brazil (Barreto et al, 1994). The study was a randomized community trial to examine the effect of the supplementation of vitamin A on diarrhea morbidity and diarrhea severity in children living in areas where its intake was inadequate. Some previous information showed that the vitamin A supplement could reduce child mortality by 23 to 34% in populations where vitamin deficiency was endemic. However, little was known of the effect vitamin A supplementation on morbidity, before the VACT. The study consisted of 1240 children, aged 6–48 months at baseline, who were assigned to receive either vitamin A or placebo every 4 months for 1 year in a small city in the Northeast Brazil. During the study, a number of children had one or more diarrheal episodes, where an episode of diarrhea was defined as a sequence of days with diarrhea and the episode was considered ended if there were 3 or more days without diarrhea. Besides the covariate for treatment assignment, we included gender in the regression. The average number of days of follow-up is 331 days, with 83.7% of the children having daily continuous information for one year. A total of 1063 children had at least one diarrheal episode during their follow-up period in the study and the average number of episodes was 5.9. Table 3 summarizes this information in both treatment group and placebo group.

We modeled the data set using the proposed methods, with a variety of transformations. The QIC criterion was used to choose the best transformation and the most appropriate resolution level. It turns out that the best model was still an additive rate model and the best resolution level was found to be $K_n = 3$. The estimated covariate effects are given in Table 4. It shows that treatment effect was not significant, although the treatment tended to reduce the risk of recurrence. The estimated baseline rate function is plotted in Figure 4, which shows a non-monotone pattern.

6 Concluding Remarks

We have proposed a class of transformation models for modeling the rates of recurrent events. The particular class of transformations we considered include both the additive and multiplicative rates models; thus, it provides a feasible way of choosing between these two models. We have introduced the wavelet based sieve estimation. However, the choice of the wavelets is not essential; other approximation methods such as smoothing splines or B-splines could also be used in the estimation.

One way to obtain more efficient estimation for β_0 is to consider the following weighted estimating equations:

$$\begin{aligned} \sum_{i=1}^n \int_0^\tau w(t) Y_i(t) Z_i(t) g' \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) \times \left\{ dN_i(t) - g \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) dt \right\} &= 0, \\ \sum_{i=1}^n \int_0^\tau w(t) Y_i(t) B_k(t) g' \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) \times \left\{ dN_i(t) - g \left(\sum_{k=1}^{m_n} \gamma_k B_k(t) + \beta^T Z_i(t) \right) dt \right\} &= 0, \end{aligned}$$

where $w(t)$ is a weight function. Some particular choice of the weight function will theoretically improve the efficiency of $\hat{\beta}$. Following a similar proof of Theorem 2, the asymptotic covariance of $\sqrt{n}(\hat{\beta} - \beta_0)$, where $\hat{\beta}$ solves the above equation, is equal to

$$\left\{ \int_0^\tau w(t) \Psi_1(t) dt \right\}^{-1} \left\{ \int_0^\tau \int_0^\tau w(t) w(s) \Psi_2(t, s) dt ds \right\} \left\{ \int_0^\tau w(t) \Psi_1(t) dt \right\}^{-1}.$$

Thus, we could select some weight to minimize this variance. For example, when β 's dimension is one, finding an optimal weight function becomes to find a $w(t) > 0$ such that the ratio

$$\frac{\int_0^\tau \int_0^\tau w(t) w(s) \Psi_2(t, s) dt ds}{\left\{ \int_0^\tau w(t) \Psi_1(t) dt \right\}^2}$$

attains the minimum. Suppose that $w(t)$ is a piece-wise constant function with constants on each piece being $w_1, \dots, w_N$; that is,

$$w(t) = \sum_{l=1}^N w_l I(t \in [t_l, t_{l+1})),$$

where $\{t_1 = 0, \dots, t_N = \tau\}$ is a partition of $[0, \tau)$. Denote the matrix $\Psi_2 = (\Psi_{2,kl})$ and the vector $\Psi_1 = \{\Psi_{1,l}\}$, where $\Psi_{1,l} = \int_{t_l}^{t_{l+1}} \Psi_1(t) dt$ and $\Psi_{2,lk} = \int_{t_l}^{t_{l+1}} \int_{t_k}^{t_{k+1}} \Psi_2(t, s) ds dt$. Then the above minimization is equivalent to finding $w_1, \dots, w_N > 0$ such that $(w_1, \dots, w_N) \Psi_2 (w_1, \dots, w_N)^T$ attains the minimum among those $\{w_k\}$'s satisfying that $\sum_k w_k \Psi_{1,k} = 1$. Then the Lagrange-Multiplier calculation shows that

$$\begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} = \frac{1}{N \Psi_1^T \Psi_2^{-1} \Psi_1} \Psi_2^{-1} \Psi_1.$$

In practice, we estimate $\Psi_{1,l}$ and $\Psi_{2,lk}$ using $\int_{t_l}^{t_{l+1}} \widehat{\Psi}_1(t) dt$ and $\int_{t_l}^{t_{l+1}} \int_{t_k}^{t_{k+1}} \widehat{\Psi}_2(t, s) ds dt$, respectively. When β 's dimension is more than one and suppose $w(t)$ is a piece-wise constant function, derivation of the optimal weight function $w(t)$ does not have an explicit solution of $\{w_l\}$. However, we can numerically search for the minimum by solving the following constrained optimization problem:

$$\min_{w_1, \dots, w_N} \text{Tr} \{ \Sigma(w_1, \dots, w_N) \}$$

subject to constraints

$$w_1, \dots, w_N \geq 0, \sum_{l=1}^N w_l = 1.$$

In practical implementation, $\Sigma(w_1, \dots, w_N)$ can be estimated using the observed data, where

both $\Psi_{1,l}$ and $\Psi_{2,lk}$ are estimated using $\int_{t_l}^{t_{l+1}} \widehat{\Psi}_1(t) dt$ and $\int_{t_l}^{t_{l+1}} \int_{t_k}^{t_{k+1}} \widehat{\Psi}_2(t, s) ds dt$ respectively.

The proposed method can be also generalized to model the intensity functions of recurrent events. For the latter, we assume that, given some latent frailty ξ_i , the intensity of $N_i(t)$ given the past history is $I(T_i \geq t)g(r(t) + \beta^T Z(t))\xi_i dt$. Similarly, the wavelet approximation can be used to estimate $r(t)$; but the parameter estimation will be based on maximizing the observed likelihood function. Both the inference and asymptotic properties require quite a bit of work.

We have adopted the QIC to choose between the transformation models. However, this may not be the best choice in terms of model prediction. An alternative criterion could be based on cross-validation, in which case each subject is randomly assigned to either a training or validation data set. The model based on the training data set is used to predict the event rates in the validation data; discrepancies between such predictions and their corresponding observed counts are used to judge the appropriateness of a particular model. The model yielding the smallest discrepancy would be deemed the most appropriate choice. The validity of such a method will be developed in our future work.

Acknowledgments

The authors would like to thank Dr. Barreto and colleagues for providing the Vitamin A data. This work was partially supported by National Institutes of Health grants R01-HL57444, P01-CA142538 and R01-DK070869.

APPENDIX

A.1 Proof of Theorem 1

First, we construct a function $\tilde{r}(t)$ in the linear span of $\{B_1(t), \dots, B_{m_n}(t)\}$ and $\tilde{r}(t)$ approximates $r_0(t)$ in a suitable norm. Since $r_0(t) \in C^2[0, \tau]$, from the wavelet analysis, we can find a $\tilde{r}(t)$ in the linear span of $\{B_1(t), \dots, B_{m_n}(t)\}$ (cf. Chapter 8, Hardle et al., 1998) such that

$$\|\tilde{r}(t) - r_0(t)\|_\infty = O\left(\frac{1}{m_n^q}\right), \|\tilde{r}'(t) - r_0'(t)\|_\infty = O\left(\frac{1}{m_n^{q-1}}\right),$$

where $\|\cdot\|_\infty$ is the supreme norm in $[0, \tau]$. We denote $\tilde{r}(t) = \sum_{k=1}^{m_n} \tilde{\gamma}_k B_k(t)$ and let $\gamma = (\gamma_1, \dots, \gamma_{m_n})$.

The next step is also the main step of our proof. Denote $l_\beta(\beta, \gamma)$ as the left-hand side of equation (2) and denote $l_{\gamma_k}(\beta, \gamma)$ as the left-hand side of the equation (3) for $k = 1, \dots, m_n$. We will show that when

$$|\beta - \beta_0|^2 + \sum_{k=1}^{m_n} |\gamma_k - \tilde{\gamma}_k|^2 = \varepsilon_n^2$$

for some ε_n , it holds that

$$l_\beta(\beta, \gamma)^T (\beta - \beta_0) + \sum_{k=1}^{m_n} l_{\gamma_k}(\beta, \gamma) (\gamma_k - \tilde{\gamma}_k) < 0. \quad (\text{A.1})$$

Then, according to the Brouwer fixed point theorem, there exists a solution $(\hat{\beta}, \hat{\gamma})$ satisfying equations (2) and (3) and

$$|\hat{\beta} - \beta_0|^2 + \sum_{k=1}^{m_n} |\hat{\gamma}_k - \tilde{\gamma}_k|^2 \leq \varepsilon_n^2.$$

Correspondingly, since $|B_k(t)| \leq 1$ and $|B_k'(t)| = O(m_n)$ we have $|\hat{\beta} - \beta_0| \leq \varepsilon_n$ and

$$\begin{aligned} |\hat{r}(t) - \tilde{r}(t)| &\leq \sum_{k=1}^{m_n} |\hat{\gamma}_k - \tilde{\gamma}_k| \leq \sqrt{m_n} \sqrt{\sum_{k=1}^{m_n} |\hat{\gamma}_k - \tilde{\gamma}_k|^2} \leq \sqrt{m_n} \varepsilon_n, \\ |\hat{r}'(t) - \tilde{r}'(t)| &= O_p(m_n) \sum_{k=1}^{m_n} |\hat{\gamma}_k - \tilde{\gamma}_k| = O_p(m_n \sqrt{m_n}) \varepsilon_n. \end{aligned}$$

Thus,

$$|\hat{\beta} - \beta_0| + \|\hat{r}(t) - r_0(t)\|_{W^{1,\infty}[0,\tau]} = O_p\left(\frac{1}{m_n} + m_n \sqrt{m_n} \varepsilon_n\right).$$

Thus, the consistency holds if we choose ε_n such that $m_n \sqrt{m_n} \varepsilon_n \rightarrow 0$. It remains to show that (A.1) holds for the choice of ε_n .

We use the notation $\mathbf{P}_n$ to denote the empirical measure based on these n i.i.d observations and denote its expectation as $\mathbf{P}$. Then

$$l_\beta(\beta, \gamma)^T (\beta - \beta_0) + \sum_{k=1}^{m_n} l_{\gamma_k}(\beta, \gamma)(\gamma_k - \tilde{\gamma}_k) = \mathbf{P}_n \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \times \{dN(t) - g(r^*(t) + \beta^T Z(t))dt\} \right],$$

where $r^*(t) = \sum_{k=1}^{m_n} \gamma_k B_k(t)$. The latter can be written as the sum of the two terms (I) and (II) such that

$$(I) = (\mathbf{P}_n - \mathbf{P}) \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \times \{dN(t) - g(r^*(t) + \beta^T Z(t))dt\} \right],$$

$$(II) = \mathbf{P} \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \times \{dN(t) - g(r^*(t) + \beta^T Z(t))dt\} \right].$$

Note that $\sqrt{n} \times (I)$ belongs to an empirical process indexed by the class of functions

$$\left\{ \int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \{dN(t) - g(r^*(t) + \beta^T Z(t))dt\} : |\beta - \beta_0| + \|r^*(t) - \tilde{r}(t)\|_{W^{L,\infty}[0,\tau]} \leq L \right\}.$$

This class has finite bracket entropy integral and moreover, such a class has an envelop function

$$\int_0^\tau Y(t) \sqrt{m_n} \varepsilon_n \{dN(t) + dt\}.$$

By Lemma 19.38 (van der Vaart 1998), $\sqrt{n} \times (I) = O_p(\sqrt{m_n} \varepsilon_n)$. For the term (II), since

$$0 = \mathbf{P} \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \times \{dN(t) - g(r_0(t) + \beta_0^T Z(t))dt\} \right],$$

we have that

$$\begin{aligned}
(II) &= \mathbf{P} \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \times \{g(r_0(t) + \beta_0^T Z(t)) - g(r^*(t) + \beta^T Z(t))\} dt \right] \\
&= \mathbf{P} \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \times \{g(\tilde{r}(t) + \beta_0^T Z(t)) - g(r^*(t) + \beta^T Z(t))\} dt \right] + \mathbf{P} \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\} g'(r^*(t) + \beta^T Z(t)) \times \{g(r_0(t) + \beta_0^T Z(t)) - g(\tilde{r}(t) + \beta_0^T Z(t))\} dt \right] \\
&\leq -\mathbf{P} \left[\int_0^\tau Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\}^2 g'(r_0(t) + \beta_0^T Z(t))^2 \times \{1 + O(|\tilde{r}(t) - r^*(t) + (\beta - \beta_0)^T Z(t)|)\} dt \right] + \int_0^\tau E[Y(t) |\tilde{r}(t) - r^*(t) + (\beta - \beta_0)^T Z(t)|] dt \\
&\leq -C_1 \int_0^\tau E[Y(t) \{(\beta - \beta_0)^T Z(t) + (r^*(t) - \tilde{r}(t))\}^2] dt + \|r_0(t) - \tilde{r}(t)\|_\infty \int_0^\tau E[Y(t) |\tilde{r}(t) - r^*(t) + (\beta - \beta_0)^T Z(t)|] dt.
\end{aligned}$$

On the other hand, we note that

$$E[Y(t) |\tilde{r}(t) - r^*(t) + (\beta - \beta_0)^T Z(t)] = O_p(|\beta - \beta_0|^2 + \sum_{k=1}^{m_n} |\gamma_k - \tilde{\gamma}_k|^2)^{1/2} = O_p(\varepsilon_n)$$

and according to assumption (B)

$$E[Y(t) \{ \tilde{r}(t) - r^*(t) + (\beta - \beta_0)^T Z(t) \}^2] \geq \tilde{C}_1 \{ E[Y(t)^2 (\beta - \beta_0)^T Z(t) Z(t)^T (\beta - \beta_0)] + E[Y(t)^2] (\tilde{r}(t) - r^*(t))^2 \}.$$

Thus,

$$(II) \leq -C_2 \left[\int_0^\tau \{ \tilde{r}(t) - r^*(t) \}^2 dt + |\beta - \beta_0|^2 \right] + O\left(\frac{\varepsilon_n}{m_n^q}\right) = -C_2 (|\beta - \beta_0|^2 + \sum_{k=1}^{m_n} |\gamma_k - \tilde{\gamma}_k|^2) m_n^{-l} + O\left(\frac{\varepsilon_n}{m_n^q}\right).$$

Conclusively, we obtain that

$$l_\beta(\beta, \gamma)^T (\beta - \beta_0) + \sum_{k=1}^{m_n} l_{\gamma_k}(\beta, \gamma) (\gamma_k - \tilde{\gamma}_k) \leq -C m_n^{-1} \varepsilon_n^2 + O\left(\frac{\varepsilon_n}{m_n^q}\right) + \frac{\sqrt{m_n}}{\sqrt{n}} \varepsilon_n.$$

This is negative when ε_n is chosen in the order of $\sqrt{m_n m_n} / \sqrt{n+1} / m_n^{q-1}$. Thus, there exists a solution to equations (2) and (3).

Furthermore, since

$$\left| \hat{\beta} - \beta_0 \right| + \left\| \tilde{r}(t) - r_0(t) \right\|_{W^{1,\infty}[0,\tau]} = O_p\left(\frac{1}{m_n} + m_n \sqrt{m_n} \varepsilon_n\right),$$

the second half of Theorem 1 holds.

A.2 Proof of Theorem 2

Equations (2) and (3) can be written as

$$\begin{aligned} \mathbf{P}_n \left[\int_0^\tau Y(t)Z(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right] &= 0, \\ \mathbf{P}_n \left[\int_0^\tau Y(t)Q_n(t)\{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right] &= 0, \end{aligned}$$

where $Q_n(t)$ is any linear combination of $B_1(t), \dots, B_{m_n}(t)$. Equivalently, it holds that

$$\begin{aligned} & \mathbf{G}_n \left[\int_0^\tau Y(t)Z(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right] \\ &= -\sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Z(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right], \\ & \mathbf{G}_n \left[\int_0^\tau Y(t)Q_n(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right] \\ &= -\sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Q_n(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right]. \end{aligned}$$

Here, $\mathbf{G}_n = \sqrt{n}(\mathbf{P}_n - \mathbf{P})$. We perform Taylor expansion on the right-hand sides of the above equations at (β_0, r_0) and note $E[\int_0^\tau \xi(Y(t), Z(t))\{dN(t) - g(r_0(t)+\beta_0^T Z(t))dt\}] = 0$. Then we obtain that

$$\begin{aligned} & \mathbf{G}_n \left[\int_0^\tau Y(t)Z(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right] \\ &= \sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Z(t)Z(t)^T g'(r_0(t)+\beta_0^T Z(t))^2 dt \right] (\widehat{\beta} - \beta_0) + \sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Z(t)g'(r_0(t)+\beta_0^T Z(t))^2 (\widehat{r}(t) - r_0(t))dt \right] \\ & \quad + \sqrt{n}O_p(|\widehat{\beta} - \beta_0|^2 + \|\widehat{r}(t) - r_0(t)\|) \end{aligned} \quad (\text{A.2})$$

and

$$\begin{aligned} & \mathbf{G}_n \left[\int_0^\tau Y(t)Q_n(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right] \\ &= \sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Q_n(t)Z(t)^T g'(r_0(t)+\beta_0^T Z(t))^2 dt \right] (\widehat{\beta} - \beta_0) + \sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Q_n(t)g'(r_0(t)+\beta_0^T Z(t))^2 (\widehat{r}(t) - r_0(t))dt \right] \\ & \quad + \sqrt{n}O_p(|\widehat{\beta} - \beta_0|^2 + \|\widehat{r}(t) - r_0(t)\|) \end{aligned} \quad (\text{A.3})$$

According to the condition, we can find $Q_n(t)$ as a linear combination of $B_1(t), \dots, B_{m_n}(t)$ such that

$$\|Q_n(t) - Q(t)\|_{L_2}^2 = O\left(\frac{1}{m_n^q}\right), \|Q_n(t) - Q(t)\|_{w^{1,2}} \rightarrow 0.$$

Thus, equation (A.3) becomes

$$\begin{aligned} & \mathbf{G}_n \left[\int_0^\tau Y(t)Q_n(t)g'(\widehat{r}(t)+\widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t)+\widehat{\beta}^T Z(t))dt\} \right] \\ &= \sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Q(t)Z(t)^T g'(r_0(t)+\beta_0^T Z(t))^2 dt \right] (\widehat{\beta} - \beta_0) + \sqrt{n}\mathbf{P} \left[\int_0^\tau Y(t)Q(t)g'(r_0(t)+\beta_0^T Z(t))^2 (\widehat{r}(t) - r_0(t))dt \right] \\ & \quad + \sqrt{n}O_p(|\widehat{\beta} - \beta_0|^2 + \|\widehat{r}(t) - r_0(t)\|) \end{aligned} \quad (\text{A.4})$$

After taking difference between (A.2) and (A.4), we obtain

$$\mathbf{G}_n \left[\int_0^\tau Y(t)(Z(t) - Q_n(t))g'(\widehat{r}(t) + \widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t) + \widehat{\beta}^T Z(t))dt\} \right] \\ = \sqrt{n} \mathbf{P} \left[\int_0^\tau Y(t)(Z(t) - Q(t))Z(t)^T g'(r_0(t) + \beta_0^T Z(t))^2 dt \right] (\widehat{\beta} - \beta_0) + \sqrt{n} O_p(|\widehat{\beta} - \beta_0|^2 + \|\widehat{r}(t) - r_0(t)\|_{L_2}^2 + \|Q_n(t) - Q(t)\|_{L_2}^2).$$

Note that on the left-hand side, the function

$$\int_0^\tau Y(t)(Z(t) - Q_n(t))g'(\widehat{r}(t) + \widehat{\beta}^T Z(t)) \{dN(t) - g(\widehat{r}(t) + \widehat{\beta}^T Z(t))dt\}$$

belongs to a P-Donsker class and it converges in $L_2(P)$ to

$$\int_0^\tau Y(t)(Z(t) - Q(t))g'(r_0(t) + \beta_0^T Z(t)) \{dN(t) - g(r_0(t) + \beta_0^T Z(t))dt\}.$$

From the results of Theorem 1,

$$\sqrt{n} O_p(|\widehat{\beta} - \beta_0|^2 + \|\widehat{r}(t) - r_0(t)\|_{L_2}^2 + \|Q_n(t) - Q(t)\|_{L_2}^2) = O_p\left(\frac{m_n^3}{\sqrt{n}} + \frac{\sqrt{n}}{m_n^{2q-2}}\right) = o_p(1).$$

We thus obtain that

$$\sqrt{n} \mathbf{P} \left[\int_0^\tau Y(t)(Z(t) - Q(t)) + Z(t)^T g'(r_0(t) + \beta_0^T Z(t))^2 dt \right] (\widehat{\beta} - \beta_0) \\ = \mathbf{G}_n \left[\int_0^\tau Y(t)(Z(t) - Q(t))g'(r_0(t) + \beta_0^T Z(t)) \{dN(t) - g(r_0(t) + \beta_0^T Z(t))dt\} \right] + o_p(1).$$

Furthermore, since the matrix

$$\Sigma_1 \equiv \mathbf{P} \left[\int_0^\tau Y(t)(Z(t) - Q(t))^{\otimes 2} g'(r_0(t) + \beta_0^T Z(t))^2 dt \right]$$

is non-singular, $\sqrt{n}(\widehat{\beta} - \beta_0)$ has an asymptotic normal distribution with mean zero and variance $\Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}$ where

$$\Sigma_2 = E \left\{ \left[\int_0^\tau Y(t)(Z(t) - Q(t))g'(r_0(t) + \beta_0^T Z(t)) \{dN(t) - g(r_0(t) + \beta_0^T Z(t))dt\} \right]^{\otimes 2} \right\}.$$

References

1. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann. Stat.* 1982; 10:1100–1120.

2. Barreto ML, Santos LMP, Assis AMO, Araujo MPN, Farenzena GG, Santos PAB, Fiaccone BL. Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. *Lancet*. 1994; 344:228–231. [PubMed: 7913157]
3. Cai T, Betensky RA. Hazard regression for interval-censored data with penalized spline. *Biometrics*. 2003; 59:570–579. [PubMed: 14601758]
4. Cai T, Hyndman R, Wand MP. Mixed model-based hazard estimation. *Journal of Computational and Graphical Statistics*. 2002; 11:784–798.
5. Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82:835–845.
6. Cook RJ, Lawless JF. Marginal analysis of recurrent events and a termination event. *Statistics in Medicine*. 1997; 16:911–924. [PubMed: 9160488]
7. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics*. 56:554–562. [PubMed: 10877316]
8. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. *Statistica Sinica*. 2002; 12:663–688.
9. Hardle, W.; Kerkycharian, G.; Picard, D.; Tsybakov, A. *Lecture Notes in Statistics*. Vol. 129. Springer-Verlag; 1998. Wavelets, Approximation, and Statistical Applications.
10. Huang CY, Wang MC. Joint modeling and estimation of recurrent event processes and failure time. *Journal of the American Statistical Association*. 2004; 99(468):1153–1165.
11. Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2nd Edition. New York: Wiley; 2002.
12. Lawless JF, Nadeau C. Some simple robust methods for analysis of recurrent events. *Technometrics*. 1995; 37:158–168.
13. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate function of recurrent events. *JRSS(B)*. 2000:711–730.
14. Lin DY, Wei LJ, Ying Z. Semiparametric transformation models for point processes. *JASA*. 2001; 96:620–628.
15. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004; 60:747–756. [PubMed: 15339298]
16. Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics*. 2001; 57:120–125. [PubMed: 11252586]
17. Pan Q, Schaubel DE. Flexible estimation of differences in treatment-specific recurrent event means in the presence of a terminating event. *Biometrics*. 2009; 65(3):753–761. [PubMed: 19053997]
18. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *JASA*. 1993; 88:811–820.
19. Schaubel DE, Zeng D, Cai J. A Semiparametric Additive Rates Model for Recurrent Event Data. *Lifetime Data Analysis*. 2006; 12:386–406.
20. Therneau TM, Hamilton SA. rhDNase as an example of recurrent event analysis. *Statistics in Medicine*. 1997; 16:2029–2047. [PubMed: 9308130]
21. van der Vaart, AW. *Asymptotic Statistics*. Cambridge University Press; 1998.
22. Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics*. 2007; 63:78–87. [PubMed: 17447932]
23. Schaubel DE, Zhang M. Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event. *Lifetime Data Analysis*. 2010; 16:451–477. [PubMed: 20063183]

Fig. 1.
Father wavelet $\phi(t)$ from Daubechies 10-wavelets

Fig. 2. Estimated baseline rate function from the simulation study $s = 1/2$ and $n = 400$. Note. The solid line is the true baseline function; the dashed line is the average of the estimated baseline function; the top and bottom curves are the respective 95- and 5-percentiles from the empirical estimates.

Fig. 3. Estimated baseline rate function and cumulative baseline rate function in the analysis of rhDNase data. The solid line in (b) is based on the proposed method and the dashed line is based on the method in Schaubel et al (2006).

Fig. 4.
Estimated baseline rate function in the analysis of VACT data

Table 1

Summary from Simulation Study with 1,000 Repetitions

Model	n	Coefficient	Ave. Bias	SEE	Ave. SE	95% CP
Additive Model ($s = 1$)	400	$\beta_1 = 1$	-0.030	0.879	0.848	0.94
		$\beta_2 = -1$	0.001	0.475	0.466	0.94
	800	$\beta_1 = 1$	-0.021	0.625	0.605	0.95
		$\beta_2 = -1$	0.039	0.329	0.332	0.95
Transformation ($s = 1/2$)	400	$\beta_1 = 1$	-0.058	0.763	0.732	0.94
		$\beta_2 = -1$	0.010	0.407	0.403	0.94
	800	$\beta_1 = 1$	-0.047	0.531	0.518	0.95
		$\beta_2 = -1$	0.052	0.285	0.286	0.95
Multiplicative Model ($s = 0$)	400	$\beta_1 = 1$	0.068	0.399	0.355	0.94
		$\beta_2 = -1$	-0.053	0.205	0.193	0.93
	800	$\beta_1 = 1$	0.019	0.265	0.250	0.94
		$\beta_2 = -1$	-0.020	0.141	0.137	0.94

Table 2Analysis of rhDNase data from additive rate model with $K_n = 3$

covariates	Est	SE	P-value
Treatment	-0.327	0.147	0.027
FEV	-0.018	0.003	< 0.001

Table 3

Summary of diarrheal episodes in VACT data

Recurrent diarrheal episodes		
Number of episodes	Vitamin A group	Placebo group
0–3	271	256
4–7	157	153
8–11	89	94
12–15	49	54
≥ 16	39	45

Table 4Analysis of VACT data from additive rate model with $K_n = 3$

covariates	Est	SE	P-value
Treatment	-0.422	0.322	0.189
Gender	0.344	0.321	0.286