Lifetime Data Anal
DOI 10.1007/s10985-012-9226-4

Semiparametric additive marginal regression models for multiple type recurrent events

Xiaolin Chen · Qihua Wang · Jianwen Cai · Viswanathan Shankar

Received: 2 April 2011 / Accepted: 1 August 2012

Abstract Recurrent event data are often encountered in biomedical research, for example, recurrent infections or recurrent hospitalizations for patients after renal transplant. In many studies, there are more than one type of events of interest. Cai and Schaube (Lifetime Data Anal 10:121–138, 2004) advocated a proportional marginal rate model for multiple type recurrent event data. In this paper, we propose a general additive marginal rate regression model. Estimating equations approach is used to obtain the estimators of regression coefficients and baseline rate function. We prove the consistency and asymptotic normality of the proposed estimators. The finite sample properties of our estimators are demonstrated by simulations. The proposed methods are applied to the India renal transplant study to examine risk factors for bacterial, fungal and viral infections.

Keywords Additive model · Empirical process · Multiple type recurrent events · Recurrent events

Mathematical Subject Classification (2000) 62N02 · 62N03

X. Chen · Q. Wang (✉)
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
e-mail: qhwang@amss.ac.cn

J. Cai · V. Shankar
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA

Published online: 17 August 2012

1 Introduction

In biomedical research, the events of interest can occur multiple times on the same subject. This kind of events are termed as recurrent events in the literature. Examples of recurrent events include repeated opportunistic infections among HIV patients, recurrent seizures in epileptic patients, multiple hospitalizations among organ transplant recipients and so on.

Developing statistical methods for the analysis of recurrent events has received much attention during the last two decades. Methods for the regression analysis of recurrent events can be classified into two categories: the conditional methods and the marginal methods. The conditional methods are based on modeling the intensity or hazard functions (Prentice et al. 1981; Anderson and Gill 1982), while the marginal methods model the mean or rate function (Pepe and Cai 1993; Lawless and Nadeau 1995; Lin et al. 2000; Schaubel et al. 2006). Compared with the conditional methods, the marginal methods have the advantage that the mean number of the recurrent events is easier to interpret than the event intensity.

For the recurrent event data, Lin et al. (2000) established the rigorous theory for both the parameter estimation and model checking for the multiplicative rate model. As noted by Schaubel et al. (2006), the popularity of multiplicative rate model derives not only from their utility and wide applicability, but also from convention and the availability of statistical software. The additive rate model is an useful alternative to the multiplicative rate model when absolute rate differences is of interest. Schaubel et al. (2006) proposed a semiparametric additive rate model to analyze the recurrent event data. Liu et al. (2010) generalized this model to allow for the additive effects and the multiplicative effects simultaneously. Zeng and Cai (2010) proposed an additive rate model for the recurrent events with an informative terminal event.

All the aforementioned methods are for the analysis of recurrent events of a single type. Frequently, in many studies, the recurrent events of interest are of multiple types. Our motivating problem concerns end-stage renal disease patients from India. End-stage renal disease patients receive different types of immunosuppressions and being so, they are more susceptible to opportunistic infections. In our study, infections are ascribed to one of the three types: bacterial, fungal and viral origin. It has been established that exposure to infective agents and net state of immunosuppression are important determinants of infection risk in transplantation. Though in the developing countries, the immunosuppresive protocols are similar to that of developed countries, pandemicity of infective agents and other environmental factors increase the risk of infections in these patients to many folds. Hence, the main objective of the study is to investigate the risk factors for recurrent infections and to examine the absolute rate of bacterial, fungal and viral infections. Of primary interest is to study the role of immunosuppression on different types of infections which is very useful information to nephrologist in care for the renal transplant patients.

Statistical methods for handling multiple type recurrent events are relatively limited. Prentice et al. (1981) firstly suggested extending their conditional intensity model for recurrent event data to model multiple type infections classified as bacterial, fungal and viral origin. Abu-Libdeh et al. (1990) studied the relation of the nutritional supplement of selenium and the recurrence of several kinds of skin cancers. A joint regression model for cumulative mean functions arising from bivariate point process was studied by Ng and Cook (1999). Lawless et al. (2001) studying failures of surgically implanted shunts in children with hydrocephalus under gap times models discussed the possibility of extension to multiple type recurrent events and the problem of terminal events. Cai and Schaubel (2004) modeled physician office visitis and hospitalizations attributable to asthma using proportional marginal rate model. Chen et al. (2005) developed joint models for multiple type recurrent events under interval censored data setup and described Gibbs sampling algorithms for fitting mixed Poisson models with piecewise constant baselines and multivariate lognormal random effects. Schaubel and Cai (2006a,b) considered the analysis of multiple type recurrent event data with missing event types using estimating equations and multiple imputation approaches. Chen and Cook (2009) described a likelihood approach based on the joint models for the multiple type recurrent events with partially missing event types. Sun et al. (2009) investigated the multiplicative rate model with the time-varying covariate effects for the multiple type recurrent events. In this article, we propose a general additive marginal rate model for the multiple type recurrent events. We consider the estimating equations approach (Liang and Zeger 1986) for obtaining the estimators of regression coefficients and baseline rate function.

The remainder of this article is organized as follows. In Sect. 2, we set up the necessary notations and describe the proposed semiparametric additive marginal regression model for the multiple type recurrent events along with the inference procedure. The asymptotic properties of the proposed estimators are stated in Sect. 3, the proofs of which are provided in the Appendix. In Sect. 4, we give a goodness-of-fit test. We report the simulation studies in Sect. 5 and illustrate the method by applying it to the India renal transplant infection data in Sect. 6. Finally, we provide some concluding remarks in Sect. 7.

2 Model and inference method

Assume that there are K different types of events of interest, i.e., every subject in the population may experience K different types of events. Let $N^*_{i,k}(t)$ denote the number of events of type k during the time $[0, t]$ for subject i. Similarly, let $C_{i,k}$ be the censoring time for the recurrent events of type k for subject i. Usually the censoring times for the K different recurrent events are equal, i.e., $C_{i,k} = C_i$. Let $Z_{i,k}(t)$ denote the $p \times 1$ covariate vector for the recurrent events of type k. The covariate $Z_{i,k}(t)$ is possibly time-dependent and external (Kalbfleisch and Prentice 2002). Let the at risk process be $Y_{i,k}(t) = I(C_{i,k} \geq t)$. Subject to right censoring, the kth observed recurrent event process is

$$N_{i,k}(t) = \int_0^t I(C_{i,k} \geq s) dN^*_{i,k}(s).$$

The rate function for the kth recurrent event process is

$$d\mu_k(t|Z_{i,k}(t)) = E\left\{dN^*_{i,k}(t)|Z_{i,k}(t)\right\}.$$

In this article, we consider the following semiparametric additive rate model:

$$d\mu_k(t|Z_{i,k}(t)) = d\mu_{0,k}(t) + g\left(\beta_0^T Z_{i,k}(t)\right) dt, \tag{1}$$

where $\mu_{0,k}(t)$ is true baseline mean function, β_0 is the true regression coefficient and g is a prespecified link function satisfying some regularity conditions given in Sect. 3. For the censoring mechanism, we assume that

$$E\left\{dN_{i,k}^*(t)|Z_{i,k}(t), C_{i,k} \geq t\right\} = E\left\{dN_{i,k}^*(t)|Z_{i,k}(t)\right\},$$

which is referred to as independent censoring in the literature.

Model (1) specifies a very large class of models. For example, it can accommodate event-type specific effects. Different regression coefficients for different event types can be obtained by introducing event-type specific covariates in the covariate $Z_{i,k}$. Event-type specific effects can be obtained by defining $\beta = \left(\beta_1^T, \ldots, \beta_k^T, \ldots, \beta_K^T\right)^T$ and $Z_{ik}(t) = \left[\mathbf{0}_{i1}^T, \mathbf{0}_{i2}^T, \ldots, \mathbf{0}_{i(k-1)}^T, \left[Z_{ik}(t)\right]^T, \mathbf{0}_{i(k+1)}^T, \ldots, \mathbf{0}_{iK}^T\right]^T$, where $\mathbf{0}$s are zero vectors. This is illustrated in Sect. 5. Note that the baseline rate function is explicitly event specific.

Define $M_{i,k}(t,\beta) = N_{i,k}(t) - \int_0^t Y_{i,k}(s)\{d\mu_{0,k}(s) + g(\beta^T Z_{i,k}(s))ds\}$. It is easy to see that $M_{i,k}(t,\beta_0)$ is a zero-mean stochastic process under the proposed model and the assumption of independent censoring. Throughout the paper, we use $M_{i,k}(t)$ to denote $M_{i,k}(t,\beta_0)$. Following the idea of generalized estimating equations (Liang and Zeger 1986), we specify the following estimating equations:

$$\sum_{i=1}^{n}\left[dN_{i,k}(t) - Y_{i,k}(t)\left\{d\mu_k(t) + g\left(\beta^T Z_{i,k}(t)\right)\right\}\right] = 0,\ 0 \leq t \leq \tau, 1 \leq k \leq K, \tag{2}$$

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau Z_{i,k}(t) g^{(1)}\left(\beta^T Z_{i,k}(t)\right)\left[dN_{i,k}(t) - Y_{i,k}(t)\left\{d\mu_k(t) + g\left(\beta^T Z_{i,k}(t)\right)dt\right\}\right] = 0, \tag{3}$$

where $g^{(1)}(\cdot)$ is the first order derivative of $g(\cdot)$ and τ is a prespecified study ending time satisfying $P(Y_{i,k}(\tau) = 1) > 0$ for $k = 1, \ldots, K$ and $i = 1, \ldots, n$. For fixed β, based on (2), after some algebraic manipulations, we can obtain that

$$\hat{\mu}_{0,k}(t,\beta) = \sum_{i=1}^{n}\int_0^t \frac{Y_{i,k}(s)\left\{dN_{i,k}^*(s) - g\left(\beta^T Z_{i,k}(s)\right)ds\right\}}{\sum_{j=1}^{n} Y_{j,k}(s)} \tag{4}$$

for $k = 1, \ldots, K$. Substituting (4) into (3), again followed by some algebraic manipulations, leads to the following estimating equation:

$$
\begin{aligned}
U_n(\tau,\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} &\left\{Z_{i,k}(t)g^{(1)}\left(\beta^T Z_{i,k}(t)\right)\right. \\
&\left. - \frac{\sum_{j=1}^{n} Y_{j,k}(t)Z_{j,k}(t)g^{(1)}(\beta^T Z_{j,k}(t))}{\sum_{l=1}^{n} Y_{l,k}(t)}\right\} \\
&\times \left\{dN_{i,k}(t) - Y_{i,k}(t)g\left(\beta^T Z_{i,k}(t)\right)dt\right\} = 0.
\end{aligned}
$$

For the simplicity of expression, we introduce some notations. Let $S_k^{(d)}(t,\beta) = \frac{1}{n}\sum_{i=1}^{n} Y_{i,k}(t)Z_{i,k}(t)^{\otimes d} g^{(d)}(\beta^T Z_{i,k}(t))$ for $d = 0, 1, 2$, where $g^{(0)}(x) = 1$, $g^{(d)}(x)$ is the dth-order derivative of $g(x)$ for $d = 1, 2$, and for a vector a, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, $a^{\otimes 2} = aa^T$. Define $E_k(t,\beta) = S_k^{(1)}(t,\beta)/S_k^{(0)}(t,\beta)$. The limits of $S_k^{(d)}(t,\beta)$ and $E_k(t,\beta)$ are $s_k^{(d)}(t,\beta)$ and $e_k(t,\beta)$ respectively for $k = 1,\ldots,K$ and $d = 0, 1, 2$.

Using the above notations, we get the following estimating equation:

$$
\begin{aligned}
U_n(\tau,\beta) &= \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}(t)g^{(1)}\left(\beta^T Z_{i,k}(t)\right) - E_k(t,\beta)\right\} \\
&\quad\times\left\{dN_{i,k}(t) - Y_{i,k}(t)g\left(\beta^T Z_{i,k}(t)\right)dt\right\} \\
&= 0, \qquad (5)
\end{aligned}
$$

which can also be expressed as

$$
U_n(\tau,\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}(t)g^{(1)}\left(\beta^T Z_{i,k}(t)\right) - E_k(t,\beta)\right\}dM_{i,k}(t,\beta) = 0. \qquad (6)
$$

Different methods, such as Newton–Raphson algorithm, can be used to solve the above estimating equation when the equation does not have an explicit solution. As a specifical case, for function $g(x) = x$, we can obtain an explicit expression of the solution for (5),

$$
\begin{aligned}
\hat{\beta} = &\left[\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} Y_{i,k}(t)\left\{Z_{i,k}(t) - \frac{\sum_{j=1}^{n} Y_{j,k}(t)Z_{j,k}(t)}{\sum_{l=1}^{n} Y_{l,k}(t)}\right\}^{\otimes 2} dt\right]^{-1} \\
&\times\left[\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}(t) - \frac{\sum_{j=1}^{n} Y_{j,k}(t)Z_{j,k}(t)}{\sum_{l=1}^{n} Y_{l,k}(t)}\right\}dN_{i,k}(t)\right].
\end{aligned}
$$

Once we get a solution of the above estimating equation, denoted by $\hat{\beta}$, we can estimate the baseline mean function of the kth recurrent event by the Breslow–Aalen type estimator

$$\hat{\mu}_{0,k}(t) = \hat{\mu}_{0,k}(t, \hat{\beta}) = \sum_{i=1}^{n} \int_{0}^{t} \frac{Y_{i,k}(s) \left\{ dN_{i,k}^{*}(s) - g\left(\hat{\beta}^T Z_{i,k}(s)\right) ds \right\}}{n S_k^{(0)}(s, \hat{\beta})} \tag{7}$$

for $1 \leq k \leq K$.

3 Asymptotic results

In this section, we present the asymptotic results for the proposed estimators. Similar to Cai and Schaubel (2004) and Liu et al. (2010), we assume the following regularity conditions:

(a) $\{N_{i,k}^{*}(\cdot), C_{i,k}, Z_{i,k}(\cdot)\}_{k=1}^{K}$, $i = 1, \ldots, n$, are independent and identically distributed.
(b) $P(C_{i,k} \geq \tau) > 0$, for $k = 1, \ldots, K, i = 1, \ldots, n$.
(c) $N_{i,k}(\tau)$ are bounded by a constant for $k = 1, \ldots, K, i = 1, \ldots, n$.
(d) $Z_{i,k}(\cdot), i = 1, \ldots, n$, have bounded total variations, i.e., $|Z_{j,i,k}(0)| + \int_0^\tau |dZ_{j,i,k}(t)| \leq H$ for all $j = 1, \ldots, p, k = 1, \ldots, K, i = 1, \ldots, n$, where $Z_{j,i,k}(t)$ is the jth component of $Z_{i,k}(t)$ and $0 < H < \infty$. $g^{(1)}\left(\beta^T Z_{i,k}(t)\right) Z_{i,k}(t)$ has bounded total variation, uniformly in β.
(e) The matrix A is nonsingular, where

$$A = E\left[\sum_{k=1}^{K} \int_{0}^{\tau} Y_{1,k}(t) \left\{ Z_{1,k}(t) g^{(1)}\left(\beta_0^T Z_{1,k}(t)\right) - e_k(t, \beta_0) \right\} g^{(1)}\left(\beta_0^T Z_{1,k}(t)\right) Z_{1,k}(t)^T dt \right].$$

(f) g is twice continuous differentiable; $\{g(\cdot), g^{(1)}(\cdot), g^{(2)}(\cdot)\}$ are uniformly continuous and bounded.

In the Appendix, we have derive the asymptotic properties of the estimated regression coefficients. Special results are listed in the following theorem.

Theorem 1 *Under the regularity conditions (a) to (f), $\hat{\beta}$ converges almost surely to β_0. Furthermore, $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ converges to a normal distribution with mean zero and covariance matrix $\Sigma_0 = A^{-1}\Sigma_1(A^{-1})^T$, where A is defined in Condition (e) and $\Sigma_1 = E[[\sum_{k=1}^{K} \int_0^\tau \{Z_{1,k}(t) g^{(1)}(\beta_0^T Z_{1,k}(t)) - e_k(t, \beta_0)\} dM_{1,k}(t)]^{\otimes 2}]$.*

The covariance matrix can be consistently estimated by $\hat{\Sigma}_0 = \hat{A}^{-1}\hat{\Sigma}_1(\hat{A}^{-1})^T$, where

$$\begin{aligned} \hat{A} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} Y_{i,k}(t) & \left\{ Z_{i,k}(t) g^{(1)}\left(\hat{\beta}^T Z_{i,k}(t)\right) \right. \\ & \left. - E_k(t, \hat{\beta}) \right\} g^{(1)}\left(\hat{\beta}^T Z_{i,k}(t)\right) Z_{i,k}(t)^T dt \end{aligned} \tag{8}$$

and

$$\hat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}(t)g^{(1)}\left(\hat{\beta}^T Z_{i,k}(t)\right) - E_k(t,\hat{\beta})\right\}d\hat{M}_{i,k}(t,\hat{\beta})\right]^{\otimes 2}$$

with

$$d\hat{M}_{i,k}(t,\hat{\beta}) = dN_{i,k}(t) - Y_{i,k}(t)\left\{d\hat{\mu}_{0,k}(t,\hat{\beta}) + g\left(\hat{\beta}^T Z_{i,k}(t)\right)dt\right\}. \tag{9}$$

To study the asymptotic properties of estimators for the K different baseline mean functions, we have to define a new metric space (Spiekerman and Lin 1998; Cai and Schaubel 2004). Let $D[0,\tau]$ be the space of cadlag functions on $[0,\tau]$. Set $f(t) = [f_1(t),\ldots,f_K(t)]^T$, $f_k(t) \in D[0,\tau]$ for $k = 1,\ldots,K$. Define $D[0,\tau]^K$ to be the space consisting of such functions with the metric $\rho(f,g) = \max_k\{\sup_{t\in[0,\tau]}|f_k(t) - g_k(t)|\}$ for any $f, g \in D[0,\tau]^K$. Define $W_n(t) = [W_{1,n}(t),\ldots,W_{K,n}(t)]^T$ where $W_{k,n}(t) = n^{\frac{1}{2}}(\hat{\mu}_{0,k}(t) - \mu_{0,k}(t))$ for $k = 1,\ldots,K$. The asymptotic results for the estimated baseline mean functions are given in the following theorem.

Theorem 2 *Under the regularity conditions (a) to (f), $\hat{\mu}_{0,k}(t)$ converges almost surely to $\mu_{0,k}(t)$, uniformly in $t \in [0,\tau]$. Furthermore, $W_n(t)$ converges to a zero-mean Gaussian field in $D[0,\tau]^K$, with covariance function $\psi_{k,l}(s,t) = E[\phi_{1,k}(s)\phi_{1,l}(t)]$ where*

$$\phi_{1,k}(t) = \int_0^t \frac{dM_{1,k}(s)}{s_k^{(0)}(s)} - h_k(t)^T A^{-1}\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{1,k}(t)g^{(1)}\left(\beta_0^T Z_{1,k}(t)\right) - e_k(t,\beta_0)\right\}dM_{1,k}(t)$$

with $h_k(t) = \int_0^t e_k(s,\beta_0)ds$.

The covariance function can be consistently estimated by $\hat{\psi}_{k,l}(s,t) = n^{-1}\sum_{i=1}^{n}\hat{\phi}_{i,k}(s)\hat{\phi}_{i,l}(t)$, where

$$\hat{\phi}_{i,k}(t) = \int_0^t \frac{d\hat{M}_{i,k}(s,\hat{\beta})}{S_k^{(0)}(s)} - \int_0^t \frac{S_k^{(1)}(s,\hat{\beta})}{S_k^{(0)}(s,\hat{\beta})}ds\hat{A}^{-1}\hat{U}_i$$

with $\hat{A}$ and $d\hat{M}_{i,k}(t,\hat{\beta})$ as defined in (8), (9) and

$$\hat{U}_i = \sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}(t)g^{(1)}\left(\hat{\beta}^T Z_{i,k}(t)\right) - E_k(t,\hat{\beta})\right\}d\hat{M}_{i,k}(t,\hat{\beta}).$$

Pointwise confidence interval for $\mu_{0,k}(t)$ can be constructed according to Theorem 2. To construct simultaneous confidence band for $\mu_{0,k}(t)$ over a time interval, we need to obtain the distribution of the supremum of the process $n^{\frac{1}{2}}(\hat{\mu}_{0,k}(t) - \mu_{0,k}(t))$, which is difficult analytically. Alternatively, we adapt the idea in Lin et al. (2000) to our situation and summarize the result in the following theorem.

Theorem 3 *Assume that $G_1, \ldots, G_n$ are independent standard normal variables which are independent of $\{N_{i,k}^*(\cdot), C_{i,k}, Z_{i,k}(\cdot)\}$, $i = 1, \ldots, n$. Then under the conditions (a) to (f),*

$$\tilde{W}_{k,n}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \hat{\phi}_{i,k}(t) G_i$$

has the same limit distribution as $W_{k,n}(t)$.

The proof of Theorem 3 is similar to that of Lin et al. (2000) and we will not repeat it here. Using the result in Theorem 3, to approximate the distribution of $W_{k,n}(t)$, we can obtain a large number of realizations from $\tilde{W}_{k,n}(t)$ by repeatedly generating the standard normal random sample $G_1, \ldots, G_n$ while fixing the data $\{N_{i,k}^*(\cdot), C_{i,k}, Z_{i,k}(\cdot)\}$, $i = 1, \ldots, n$. We then can have the approximation of the distribution of the supremum of the process of $W_{k,n}(t)$ over the interval of interest and construct the simultaneous confidence bands. It is noted that the confidence intervals or bands constructed by the above method can not guarantee the confidence intervals or bands to be positive. As Lin et al. (1994, 1998) to avoid this issue, the log transformation can be used in the constructed of the confidence intervals or bands. Also this method can also improve the coverage probability in small samples. We omit the details about the special procedure. Interested readers can consult Lin et al. (1994, 1998).

4 Goodness-of-fit test

In order to examine whether the fitted model is adequate, we consider a goodness-of-fit test statistics. We consider this problem under the situation when the covariates are time-independent. Define $\hat{M}_{i,k}(t) = \hat{M}_{i,k}(t, \hat{\beta})$. Following the idea of Lin et al. (1993, 2000), we define the following test statistics:

$$V(t, z) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} I(Z_{i,k} \leq z) \hat{M}_{i,k}(t),$$

where $(Z_{i,k} \leq z)$ denotes the situation that every component of $Z_{i,k}$ is no larger than that of z. For the null distribution of $V(t, z)$, we have the following theorem.

Theorem 4 *Under the regularity conditions listed in Sect. 3, the null distribution of $V(t, z)$ converges weakly to a zero-mean Gaussian process with covariance function $E\{\zeta_i(t_1, z_1)\zeta_i(t_2, z_2)\}$ at (t_1, z_1) and (t_2, z_2), where*

$$\begin{aligned}\zeta_i(t, z) = & \sum_{k=1}^{K} \int_0^t \left\{ I(Z_{i,k} \leq z) - \frac{f_k(s, z)}{s_k^{(0)}(s)} \right\} dM_{i,k}(s) \\ & - \sum_{k=1}^{K} (h(t, z) + l(t, z)) A^{-1} \int_0^\tau \left\{ Z_{i,k} g^{(1)} \left(\beta_0^T Z_{i,k} \right) - e_k(t, \beta_0) \right\} dM_{i,k}(t)\end{aligned}$$

with

$$f_k(t,z) = E\{I(Z_{i,k} \le z)Y_{i,k}(t)\},$$
$$h(t,z) = \sum_{k=1}^{K} \int_0^t f_k(s,z)\,dh_k(s)^T,$$
$$h_k(t) = \int_0^t e_k(s,\beta_0)\,ds$$

and

$$l(t,z) = \sum_{k=1}^{K} E\left\{\int_0^t I(Z_{i,k} \le z)Y_{i,k}(s)g^{(1)}\left(\beta_0^T Z_{i,k}\right) Z_{i,k}^T ds\right\}.$$

From the proof of Theorem 4, we can see that the null distribution of $V(t,z)$ can be approximated by the zero-mean Gaussian process

$$\begin{aligned}\tilde{V}(t,z) &= n^{-\frac{1}{2}} \sum_{i=1}^{n}\sum_{k=1}^{K} \int_0^t \left\{I(Z_{i,k} \le z) - \frac{\hat{f}_k(s,z)}{S_k^{(0)}(s)}\right\} d\hat{M}_{i,k}(s) \\ &\quad - n^{-\frac{1}{2}} \sum_{i=1}^{n}\sum_{k=1}^{K} \left\{\hat{h}(t,z) + \hat{l}(t,z)\right\} \hat{A}^{-1} \\ &\quad \times \int_0^\tau \left\{Z_{i,k}g^{(1)}\left(\beta_0^T Z_{i,k}\right) - E_k(t,\hat{\beta})\right\} d\hat{M}_{i,k}(t), \end{aligned} \tag{10}$$

where

$$\hat{f}_k(t,z) = \frac{1}{n}\sum_{i=1}^{n} I(Z_{i,k} \le z)Y_{i,k}(t),$$
$$\hat{h}(t,z) = \sum_{k=1}^{K} \int_0^t \hat{f}_k(s,z)\,d\hat{h}_k(s)^T,$$
$$\hat{h}_k(t) = \int_0^t E_k(s,\hat{\beta})\,ds$$

and

$$\hat{l}(t,z) = \sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} \int_0^t I(Z_{i,k} \le z)Y_{i,k}(s)g^{(1)}\left(\hat{\beta}^T Z_{i,k}\right) Z_{i,k}^T ds.$$

It is difficult to obtain the analytical expression of the covariance function. Thus we can utilize the method of resampling to get the asymptotic null distribution of $V(t, z)$. Specifically, define

$$\hat{V}(t,z) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \Bigg[\sum_{k=1}^{K} \int_0^t \left\{ I(Z_{i,k} \le z) - \frac{\hat{f}_k(s,z)}{S_k^{(0)}(s)} \right\} d\hat{M}_{i,k}(s) - \sum_{k=1}^{K} \left\{ \hat{h}(t,z) + \hat{l}(t,z) \right\} \hat{A}^{-1} \times \int_0^\tau \left\{ Z_{i,k} g^{(1)} \left(\beta_0^T Z_{i,k} \right) - E_k(t,\hat{\beta}) \right\} d\hat{M}_{i,k}(t) \Bigg] G_i, \tag{11}$$

where $G_i, i = 1, \ldots, n$ are i.i.d. standard normal distribution random variates, which are independent of $\{N_{i,k}^*(\cdot), C_{i,k}, Z_{i,k}\}_{k=1}^K, i = 1, \ldots, n$. It is easy to show that the null distribution of $V(t, z)$ can be approximated by the conditional distribution of $\hat{V}(t, z)$. Hence to obtain the asymptotic null distribution of $V(t, z)$, we can get a large number of realizations of $\hat{V}(t, z)$ by repeatedly generating random numbers $G_i, i = 1, \ldots, n$, from standard normal distribution, while fixing the $\{N_{i,k}^*(\cdot), C_{i,k}, Z_{i,k}\}_{k=1}^K$, $i = 1, \ldots, n$, at their observed values. To assess the goodness-of-fit of the proposed model, we could plot a few number of realizations of $\hat{V}(t, z)$ and compare them with $V(t, z)$ to see if there are some unusual patterns. But it is noted that plotting $\hat{V}(t, z)$ versus t and z involves high-dimensional graphics. And this is difficult now. Alternatively, we can use the supremum test $\sup_{t,z}|V(t, z)|$ to obtain the p-value of the test, which could be used to judge the adequacy of the model. The p-value can be obtained by comparing the observed value of $\sup_{t,z}|V(t, z)|$ to a large number of realizations from $\sup_{t,z}|\hat{V}(t, z)|$.

5 Simulation studies

We conducted extensive simulation studies to examine the finite sample properties of our proposed estimators in this section.

Specially, we simulated two different types of recurrent events with the sample size being $n = 50$, 100 and 200. Firstly, we considered the model with the same regression coefficients and covariates for the two types of recurrent events. The dimension of the covariates Z_i is 2. The first component of Z_i follows the uniform distribution on (0, 1), while the second follows Bernoulli distribution with success probability being 0.5. Let $\beta_0 = (0, 0.5)^T$. For the same subject, we introduced a frailty variable to induce positive correlation for the within-subject events. We assume that the frailty variable, $Q_{i,k} \equiv Q_i$, followed a Gamma distribution with expectation 0.25 and the variance σ^2 with $\sigma^2 = 0, 0.25, 0.5$ and 1. To avoid too large frailty variable, which will induce too many recurrent events for one subject, we took $Q_i^* = \min(Q_i, 1.5)$. For the two different baseline mean functions, we considered $\mu_{0,1}(t) = \gamma_{0,1} \cdot t = 0.25t$ and $\mu_{0,2}(t) = \gamma_{0,2} \cdot t = 0.5t$ respectively. Then the kth type recurrent events were generated from the Poisson process with intensity function

$$\lambda_{i,k}(t) = Q_i^* + \gamma_{0,k} + \beta_0^T Z_i.$$

Under the above setups, the marginal additive rate model was

$$E\left\{dN_{i,k}^*(t)|Z_i\right\} = E(Q_i^*)dt + d\mu_{0,k}(t) + \beta_0^T Z_i dt = \left\{E(Q_i^*) + \gamma_{0,k} + \beta_0^T Z_i\right\} dt,$$

for $k = 1, 2$. The censoring time was generated from $U(0, \tau)$ with $\tau = 5$. Under the above settings, the average number of events for the first type of recurrent events is approximatively 1.8 per subject, while that of the second type of recurrent events is approximatively 2.5 per subject.

As advised by one reviewer, secondly, we consider the model with different regression coefficients for different types of recurrent events. The covariate $Z_{i,1}$ and $Z_{i,2}$ are independent and identically distributed as the Bernoulli distribution with the success probability 0.5, where $Z_{i,1}$ is the covariate for the first type of the recurrent events and $Z_{i,2}$ is the covariate for the second type of the recurrent events. The regression coefficient for $Z_{i,1}$ was set to be 0.5, while that for $Z_{i,2}$ was 0.3. Following the descriptions in Sect. 2, we defined $\beta_0 = (0.5, 0.3)^T$, $Z_{i,1}^* = (Z_{i,1}, 0)^T$ and $Z_{i,2}^* = (0, Z_{i,2})^T$. The other design elements were the same as the model with the same regression coefficients and the same covariates. Under these design elements, the kth type recurrent events were generated from the Poisson process with intensity function

$$\lambda_{i,k}(t) = Q_i^* + \gamma_{0,k} + \beta_0^T Z_{i,k}^*.$$

Then the marginal additive rate model was

$$E\left\{dN_{i,k}^*(t)|Z_{i,k}^*\right\} = E(Q_i^*)dt + d\mu_{0,k}(t) + \beta_0^T Z_{i,k}^* dt = \left\{E(Q_i^*) + \gamma_{0,k} + \beta_0^T Z_{i,k}^*\right\} dt,$$

for $k = 1, 2$. In this situation, the average number of events for the first type of recurrent events is about 1.8 per subject, while that of the second type of recurrent events is about 2 per subject.

The summaries of our simulation results are presented in Tables 1 and 2. All the results are based on 1000 replicates. It can be seen from the results in Tables 1 and 2 that our methods work well. The biases of the estimated regression coefficients are approximately 0s. The asymptotic standard errors (ASEs) are very close to the empirical standard errors (ESEs). The empirical coverage probabilities are very close to their nominal level, 0.95. As the sample size n increases, the biases, ASEs and ESEs decrease while the CPs are closer to the nominal level in general. For the fixed sample size, the estimated regression coefficients with $\sigma^2 = 0$ have smallest biases, ASEs and ESEs in general. But because of the truncation for the frailty variable, the ASEs and ESEs of the estimated regression coefficients become larger firstly and then smaller as the σ^2 increases.

Table 1 Summaries of the simulation results for the estimated regression coefficients for the model with the same regression coefficient for different types of recurrent events

n	σ^2	$\beta_{10} = 0$				$\beta_{20} = 0.5$			
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP
50	0	0.009	0.202	0.216	0.925	0.003	0.117	0.125	0.931
	0.25	−0.004	0.276	0.305	0.921	−0.004	0.163	0.174	0.929
	0.5	0.005	0.278	0.315	0.924	0.001	0.164	0.176	0.936
	1	0.009	0.265	0.292	0.917	0.003	0.156	0.171	0.922
100	0	−0.004	0.143	0.147	0.942	−0.002	0.083	0.086	0.940
	0.25	0.007	0.200	0.203	0.955	−0.002	0.117	0.118	0.937
	0.5	0.007	0.202	0.215	0.942	0.001	0.119	0.122	0.935
	1	0.008	0.195	0.207	0.943	0.000	0.113	0.120	0.937
200	0	−0.003	0.101	0.100	0.942	−0.003	0.059	0.060	0.942
	0.25	−0.003	0.144	0.153	0.934	0.000	0.084	0.087	0.943
	0.5	−0.007	0.146	0.151	0.939	0.005	0.084	0.086	0.940
	1	−0.006	0.140	0.147	0.939	−0.002	0.081	0.083	0.948

Table 2 Summaries of the simulation results for the estimated regression coefficients for the model with the different regression coefficients for different types of recurrent events

n	σ^2	$\beta_{10} = 0.5$				$\beta_{20} = 0.3$			
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP
50	0	0.009	0.151	0.162	0.931	−0.017	0.161	0.167	0.936
	0.25	−0.006	0.188	0.204	0.924	−0.011	0.199	0.205	0.924
	0.5	0.005	0.188	0.197	0.945	−0.019	0.192	0.208	0.932
	1	−0.005	0.180	0.185	0.933	−0.008	0.184	0.196	0.936
100	0	0.004	0.108	0.109	0.948	−0.010	0.116	0.119	0.941
	0.25	−0.005	0.136	0.143	0.938	−0.005	0.142	0.143	0.942
	0.5	−0.009	0.135	0.135	0.950	0.008	0.141	0.140	0.947
	1	0.005	0.130	0.129	0.954	−0.002	0.136	0.139	0.946
200	0	0.001	0.078	0.077	0.944	−0.005	0.087	0.083	0.935
	0.25	−0.001	0.096	0.097	0.952	−0.006	0.105	0.102	0.949
	0.5	0.006	0.101	0.096	0.945	−0.005	0.105	0.101	0.940
	1	−0.002	0.096	0.093	0.940	−0.001	0.100	0.098	0.943

6 Analysis of the India renal transplant data

We now apply the proposed methods to a cohort of renal transplant patients who had acquired opportunistic infections during the post transplantation follow-up. Patients who received primary renal transplantation at a tertiary care teaching hospital in Southern India between 1994 and 2007 were studied. We considered a subset of 991 transplant patients for this illustration. Patients were followed to the end of 2008. The median follow-up time was 75.2 months (range 0–179.5 months).

Table 3 Recurrent infections by type in the India renal transplant patients

Infections type	Recurrent infections						
	0	1	2	3	4	5	6
Bacteria	708	165	67	38	8	3	2
Systemic mycoses	937	47	7	0	0	0	0
Virus	865	187	45	5	1	1	0

Data were collected prospectively on the transplant patients, which included date of transplantation, information on the subsequent infections that are ascribed to one of the three organism types: bacterial, systemic mycoses (fungal) and viral. A total of 829 infection episodes were observed among these patients. The average number of infections per patient observed was 2 (sd = 1.2). Nearly 43 % had at least one infection and 22 % ($n = 219$) had multiple infections of either the same type or of different types. Of those who had multiple infections, 40 % acquired infections of the same type, while 52 % had two out of three infection types and 8 % acquired all three infection types. Table 3 summarizes the distribution of infections across patients and types of infections. The primary clinical objective of the investigation is to find the risk factors associated with the multiple type infections especially the role of immunosuppression. Risk factors of interest include type of immunosuppression, role of pre-transplant (pre Tx DM) or post-transplant diabetes (post Tx DM) mellitus, acute rejection, age of patient and gender of patient. Patients received different combination of primary immunosuppression and we grouped them into three major regimens: (i) a combination of prednisolone (Pred), azathiaprine (AZA) and calcineurin inhibitor (CNI) (PAC), (ii) combination of prednisolone, CNI and mycophenolate mofetil (MMF) or mycophenolate sodium (MPA) (PCM), and (iii) a group consisting of drugs that are non-CNI based regimens. We centered the age of the patient to 34 years (Age-34).

We considered the following model:

$$E\left[dN_k^*(t)|Z_k(t)\right] = d\mu_{0,k}(t) + \beta_0^T Z_k(t)dt,$$

with event-specific effect for immunosuppression and common effect for other covariates with

$$\begin{aligned}\beta^T \mathbf{Z}(t) = {}& \beta_{1,B}\mathrm{IMM}_{\mathrm{PAC}} * I(\mathrm{Type}=1) + \beta_{2,B}\mathrm{IMM}_{\mathrm{PCM}} * I(\mathrm{Type}=1)\\ &+\beta_{3,F}\mathrm{IMM}_{\mathrm{PAC}} * I(\mathrm{Type}=2) + \beta_{4,F}\mathrm{IMM}_{\mathrm{PCM}} * I(\mathrm{Type}=2)\\ &+\beta_{5,V}\mathrm{IMM}_{\mathrm{PAC}} * I(\mathrm{Type}=3) + \beta_{6,V}\mathrm{IMM}_{\mathrm{PCM}} * I(\mathrm{Type}=3) + \beta_7\mathrm{Age}\\ &+\beta_8\mathrm{Sex} + \beta_9\mathrm{PreTxDM} + \beta_{10}\mathrm{PostTxDM} + \beta_{11}\mathrm{Rejection},\end{aligned}$$

where β_B, β_F and β_V denote the bacterial, fungal and viral infection specific parameters, respectively.

Table 4 presents the parameter estimates, standard error of the estimates, and the associated Wald test and p values. The values presented in the table are 10^3 times the original values. Interestingly, the effect of immunosuppression is different for

Table 4 Regression analysis of multiple type infections in the India renal transplant study

Covariates	Estimate ($\times 10^3$)	SE ($\times 10^3$)	Wald test $(\beta/\text{SE})^2$	P-value
Immunosuppression				
Bacteria				
PAC	0.096	0.074	1.686	0.194
PCM	7.173	2.591	7.661	0.006
Non CNI	Ref			
Fungus				
PAC	−0.101	0.079	1.642	0.200
PCM	−1.168	0.392	8.896	0.003
Non CNI	Ref			
Virus				
PAC	−0.116	0.112	1.063	0.303
PCM	−0.788	1.360	0.336	0.562
Non CNI	Ref			
Age (centered 34 years)	0.001	0.016	0.005	0.942
Gender (male = 1, female = 0)	−0.684	0.501	1.867	0.172
Diabetes mellitus (DM)				
Pre Tx DM	1.170	0.888	1.738	0.187
Post Tx DM	0.952	0.511	3.480	0.062
No DM	Ref			
Acute rejection (yes = 1, no = 0)	1.263	0.340	13.767	<0.001

different type of infections. For bacterial infections, patients in the PCM group have 7.2 more infections per 1,000 patients per month compared to those in the non-CNI regimen. While, for fungal infection, patients in the PCM group have 1 less infection per 1,000 patients per month than those in the non-CNI immunosuppression group. Similarly decreased effect were observed for viral infection in the PAC group and the PCM group compared to the non-CNI group, however, it was not statistically significant.

Post transplant diabetes mellitus (risk difference (RD) = 0.95) and post transplant episodes of acute rejections (RD = 1.26) are associated with increased rate of infections. Though the presence of pre-transplant hyperglycemia showed an elevated rate of infection (RD = 1.17) compared to those without diabetes mellitus, it was not statistically significant. Older patients tend to have more frequent infection while male patients tend to have a decreased infections rate compared to females, although none of these effects were statistically significant.

7 Discussion

In this paper, we proposed a semiparametric additive marginal rate model for the analysis of multiple type recurrent event data, which is an useful alternative model to the semiparametric multiplicative marginal rate model suggested by Cai and Schaubel (2004). The class of the link function g is very rich. Many conventional functions, such as $g(x) = e^x$, belong to this class. If we choose $g(x) = x$, we can obtain an explicit expression of estimators of the true regression parameters. In this article, we used the estimating equations approach to make inference. The consistency and asymptotic normality of the proposed estimator were established. Extensive simulation studies demonstrated the validity of the asymptotic approximation in finite samples.

Schaubel and Cai (2006a,b) considered the analysis of multiple type recurrent event data with possibly missing event type under the semiparametric multiplicative rate model. For the additive rate model, our method can be similarly extended to incorporate such missing event type recurrent event data. Another complication arises when recurrent events are stopped by a terminal event such as death. Methods for a single type of recurrent events with a terminal event have been proposed in the literature, for example, Ghosh and Lin (2000, 2002); Liu et al. (2004); Zeng and Cai (2010), among others. However, terminal event has not been dealt with for the multiple type recurrent events. This is a topic meriting further research.

Appendix: Sketch proofs of the main results

Proof of Theorem 1 Firstly, we prove the strong consistency. Define $\hat{A}_1(\beta) = -\frac{1}{n}\frac{\partial U_n(\tau,\beta)}{\partial \beta}$. After some algebraic computations, we can obtain

$$\begin{aligned}
\hat{A}_1(\beta) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}(t)g^{(1)}\left(\beta^T Z_{i,k}(t)\right) - E_k(t,\beta)\right\} Y_{i,k}(t)g^{(1)}\left(\beta^T Z_{i,k}(t)\right) Z_{i,k}(t)^T dt \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} Z_{i,k}(t)g^{(2)}\left(\beta^T Z_{i,k}(t)\right) Z_{i,k}(t)^T dM_{i,k}(t,\beta) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} \frac{\partial E_k(t,\beta)}{\partial \beta} dM_{i,k}(t,\beta) \\
&\equiv a(\beta) - b(\beta) + c(\beta).
\end{aligned} \tag{A.1}$$

Under Conditions (d) and (f), we can find a $\delta > 0$ such that $\hat{A}_1(\beta)$ is sufficiently close to $\hat{A}_1(\beta_0)$ whenever $\|\beta - \beta_0\| \le \delta$, uniformly in n. On the other hand, we can write

$$\begin{aligned}
a(\beta_0) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}(t)g^{(1)}\left(\beta_0^T Z_{i,k}(t)\right) - e_k(t,\beta_0)\right\} Y_{i,k}(t)g^{(1)}\left(\beta_0^T Z_{i,k}(t)\right) Z_{i,k}(t)^T dt \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\{e_k(t,\beta_0) - E_k(t,\beta_0)\} Y_{i,k}(t)g^{(1)}\left(\beta_0^T Z_{i,k}(t)\right) Z_{i,k}(t)^T dt.
\end{aligned} \tag{A.2}$$

Through repeated applications of the strong law of large numbers (Sen and Singer 1993), the second term of $a(\beta_0)$ can be proven to converge almost surely to 0. By the strong law of large numbers (Sen and Singer 1993), the first term of $a(\beta_0)$ can be shown to converge almost surely to A, which is defined in Condition (e). Similarly, we can prove that $b(\beta_0)$ and $c(\beta_0)$ converges almost surely to 0. Therefore, we get that $\hat{A}_1(\tilde{\beta})$ converges to A almost surely if $\tilde{\beta}$ converges to β_0 almost surely. The consistency of $\hat{\beta}$ can then be proved using similar arguments as in Theorem 1 in Liu et al. (2010).

Secondly, we prove the asymptotic normality of $\hat{\beta}$. By Taylor expansion of $U_n(\tau, \hat{\beta})$ at β_0, we obtain

$$\sqrt{n}(\hat{\beta} - \beta_0) = \hat{A}_1(\hat{\beta^*})^{-1} \frac{1}{\sqrt{n}} U_n(\tau, \beta_0), \tag{A.3}$$

where $\hat{\beta^*}$ is between $\hat{\beta}$ and β_0. Substituting the expression of $U_n(\tau, \beta_0)$ into the above equation, we get

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta_0) &= \hat{A}_1(\hat{\beta^*})^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau \left\{ Z_{i,k}(t) g^{(1)}\left(\beta_0^T Z_{i,k}(t)\right) - E_k(t, \beta_0) \right\} dM_{i,k}(t) \\
&= \hat{A}_1(\hat{\beta^*})^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau \left\{ Z_{i,k}(t) g^{(1)}\left(\beta_0^T Z_{i,k}(t)\right) - e_k(t, \beta_0) \right\} dM_{i,k}(t) \\
&\quad + \hat{A}_1(\hat{\beta^*})^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau \{ e_k(t, \beta_0) - E_k(t, \beta_0) \} dM_{i,k}(t).
\end{aligned} \tag{A.4}$$

Using the empirical process techniques (Pollard 1990; Bilias et al. 1997), it can be shown that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau \{ e_k(t, \beta_0) - E_k(t, \beta_0) \} dM_{i,k}(t) = o_p(1). \tag{A.5}$$

Because $\hat{\beta} \to_{a.s.} \beta$ and $\hat{\beta^*}$ is between $\hat{\beta}$ and β_0, we get $\hat{\beta^*} \to_{a.s.} \beta$ and $\hat{A}_1(\hat{\beta^*}) \to_{a.s.} A$. Hence, we have

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta_0) &= A^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau \left\{ Z_{i,k}(t) g^{(1)}\left(\beta_0^T Z_{i,k}(t)\right) - e_k(t, \beta_0) \right\} dM_{i,k}(t) \\
&\quad + o_p(1).
\end{aligned} \tag{A.6}$$

Then the proof of asymptotic normality is completed by applying the multivariate central limit Theorem (Sen and Singer 1993).

Proof of Theorem 2 We firstly prove the uniform strong consistency. As usual, we make a simple decomposition

$$\hat{\mu}_{0,k}(t) - \mu_{0,k}(t) = \hat{\mu}_{0,k}(t) - \hat{\mu}_{0,k}(t, \beta_0) + \hat{\mu}_{0,k}(t, \beta_0) - \mu_{0,k}(t). \tag{A.7}$$

For the first term of (A.7), using Taylor expansion, we can write

$$\begin{aligned}
\hat{\mu}_{0,k}(t) - \hat{\mu}_{0,k}(t, \beta_0) &= \sum_{i=1}^{n} \int_0^t \frac{Y_{i,k}(s)\left\{dN^*_{i,k}(s) - g\left(\hat{\beta}^T Z_{i,k}(s)\right)ds\right\}}{\sum_{j=1}^n Y_{j,k}(s)} \\
&\quad - \sum_{i=1}^{n} \int_0^t \frac{Y_{i,k}(s)\left\{dN^*_{i,k}(s) - g\left(\beta_0^T Z_{i,k}(s)\right)ds\right\}}{\sum_{j=1}^n Y_{j,k}(s)} \\
&= -\sum_{i=1}^{n} \int_0^t \frac{Y_{i,k}(s)\left\{g(\hat{\beta}^T Z_{i,k}(s)) - g\left(\beta_0^T Z_{i,k}(s)\right)\right\}ds}{\sum_{j=1}^n Y_{j,k}(s)} \\
&= \sum_{i=1}^{n} \int_0^t \frac{Y_{i,k}(s) Z^T_{i,k}(s) g^{(1)}\left(\hat{\beta^*}^T Z_{i,k}(s)\right)ds}{\sum_{j=1}^n Y_{j,k}(s)} (\beta_0 - \hat{\beta}).
\end{aligned} \tag{A.8}$$

Under Conditions (d) and (f), using the strong consistency of $\hat{\beta}$, we can prove that the first term converges almost surely to 0 uniformly in $t \in [0, \tau]$.

For the second term of (A.7), we can write

$$\begin{aligned}
\hat{\mu}_{0,k}(t, \beta_0) - \mu_{0,k}(t) &= \sum_{i=1}^{n} \int_0^t \frac{Y_{i,k}(s)\left\{dN^*_{i,k}(s) - g\left(\beta_0^T Z_{i,k}(s)\right)ds\right\}}{\sum_{j=1}^n Y_{j,k}(s)} - \mu_{0,k}(t) \\
&= \sum_{i=1}^{n} \int_0^t \frac{Y_{i,k}(s)\left\{dN^*_{i,k}(s) - g(\beta_0^T Z_{i,k}(s))ds\right\}}{\sum_{j=1}^n Y_{j,k}(s)} - \int_0^t d\mu_{0,k}(s) \\
&= \sum_{i=1}^{n} \int_0^t \frac{1}{\sum_{j=1}^n Y_{j,k}(s)} Y_{i,k}(s)\left\{dN^*_{i,k}(s) - g\left(\beta_0^T Z_{i,k}(s)\right)ds - d\mu_{0,k}(s)\right\} \\
&= \sum_{i=1}^{n} \int_0^t \frac{1}{\sum_{j=1}^n Y_{j,k}(s)} dM_{i,k}(s).
\end{aligned} \tag{A.9}$$

Using the uniform strong law of large number (Pollard 1990) and Lemma 1 of Lin et al. (2000), the second term converges almost surely to 0 uniformly in $t \in [0, \tau]$.

Consequently, $\hat{\mu}_{0,k}(t)$ converges almost surely to $\mu_{0,k}(t)$ uniformly in $t \in [0, \tau]$ for $k = 1, \ldots, K$.

We now prove the weak convergence of $W_n(t)$. Using the above decomposition, we have

$$\begin{aligned}
W_{k,n}(t) &= \sqrt{n}(\hat{\mu}_{0,k}(t) - \mu_{0,k}(t)) \\
&= \sqrt{n}\{\hat{\mu}_{0,k}(t) - \hat{\mu}_{0,k}(t, \beta_0)\} + \sqrt{n}\{\hat{\mu}_{0,k}(t, \beta_0) - \mu_{0,k}(t)\}.
\end{aligned} \tag{A.10}$$

For the first term of (A.10), using Taylor expansion, it is easy to obtain

$$
\begin{aligned}
\sqrt{n}\{\hat{\mu}_{0,k}(t)-\hat{\mu}_{0,k}(t,\beta_0)\} &= \sum_{i=1}^{n}\int_0^t \frac{Y_{i,k}(s)Z_{i,k}^T(s)g^{(1)}\left(\hat{\beta^*}^T Z_{i,k}(s)\right)ds}{\sum_{j=1}^n Y_{j,k}(s)}\sqrt{n}(\beta_0-\hat{\beta}) \\
&= \sum_{i=1}^{n}\int_0^t \frac{Y_{i,k}(s)Z_{i,k}^T(s)g^{(1)}\left(\beta_0^T Z_{i,k}(s)\right)ds}{\sum_{j=1}^n Y_{j,k}(s)}\sqrt{n}(\beta_0-\hat{\beta}) \\
&\quad+\sum_{i=1}^{n}\int_0^t \frac{Y_{i,k}(s)Z_{i,k}^T(s)\{g^{(1)}\left(\hat{\beta}^{*T} Z_{i,k}(s)\right)-g^{(1)}\left(\beta_0^T Z_{i,k}(s)\right)\}ds}{\sum_{j=1}^n Y_{j,k}(s)} \\
&\quad\times\sqrt{n}\left(\beta_0-\hat{\beta}\right) \\
&\equiv \eta_1(t)+\eta_2(t).
\end{aligned}
\tag{A.11}
$$

It is easy to see that

$$
\begin{aligned}
\eta_1(t) &= \int_0^t E_k(s,\hat{\beta}^*)ds\sqrt{n}(\beta_0-\hat{\beta}) \\
&= \int_0^t e_k(s,\beta_0)ds\sqrt{n}(\beta_0-\hat{\beta})+\int_0^t\{E_k(s,\hat{\beta}^*)-e_k(s,\beta_0)\}ds \\
&\quad\times\sqrt{n}(\beta_0-\hat{\beta}).
\end{aligned}
\tag{A.12}
$$

From Theorem 1, we have $\sqrt{n}(\beta_0-\hat{\beta})=O_p(1)$. It can be shown that

$$
sup_{t\in[0,\tau]}\left|\int_0^t\{E_k(s,\hat{\beta}^*)-e_k(s,\beta_0)\}ds\right|=o_p(1). \tag{A.13}
$$

Hence we have

$$
\eta_1(t)=\int_0^t e_k(s,\beta_0)ds\sqrt{n}(\beta_0-\hat{\beta})+o_p(1). \tag{A.14}
$$

Under Conditions (d) and (f), it can be shown that

$$
sup_{t\in[0,\tau]}\left|\sum_{i=1}^{n}\int_0^t \frac{Y_{i,k}(s)Z_{i,k}^T(s)\{g^{(1)}(\hat{\beta^*}^T Z_{i,k}(s))-g^{(1)}(\beta_0^T Z_{i,k}(s)\}ds}{\sum_{j=1}^n Y_{j,k}(s)}\right|=o_p(1). \tag{A.15}
$$

Hence we have $\eta_2(t) = o_p(1)$ uniformly in $t \in [0, \tau]$. Then

$$\sqrt{n}\{\hat{\mu}_{0,k}(t) - \hat{\mu}_{0,k}(t, \beta_0)\} = \int_0^t e_k(s, \beta_0)ds\sqrt{n}(\beta_0 - \hat{\beta}) + o_p(1)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t e_k(s, \beta_0)ds \sum_{k=1}^K \int_0^\tau \Big\{Z_{i,k}(t)g^{(1)}(\beta_0^T Z_{i,k}(t))$$
$$-e_k(t, \beta_0)\Big\}\, dM_{i,k}(t) + o_p(1). \tag{A.16}$$

For the second term of (A.10), we have

$$\sqrt{n}\{\hat{\mu}_{0,k}(t, \beta_0) - \mu_{0,k}(t)\} = \sqrt{n}\sum_{i=1}^n \int_0^t \frac{1}{\sum_{j=1}^n Y_{j,k}(s)} dM_{i,k}(s)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{1}{S_k^{(0)}(s)} dM_{i,k}(s)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{1}{s_k^{(0)}(s)} dM_{i,k}(s)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \left\{\frac{1}{S_k^{(0)}(s)} - \frac{1}{s_k^{(0)}(s)}\right\} dM_{i,k}(s). \tag{A.17}$$

Using the uniform strong law of large number (Pollard 1990) and Lemma 1 of Lin et al. (2000), it can be shown that

$$sup_{t\in[0,\tau]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \left\{\frac{1}{S_k^{(0)}(s)} - \frac{1}{s_k^{(0)}(s)}\right\} dM_{i,k}(s)\right| = o_p(1). \tag{A.18}$$

Hence we have

$$\sqrt{n}\{\hat{\mu}_{0,k}(t, \beta_0) - \mu_{0,k}(t)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{1}{s_k^{(0)}(s)} dM_{i,k}(s) + o_p(1). \tag{A.19}$$

Summarizing the above results, we arrive at

$$
\begin{aligned}
\sqrt{n}(\hat{\mu}_{0,k}(t) - \mu_{0,k}(t)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \frac{1}{s_k^{(0)}(s)} dM_{i,k}(s) \\
&\quad + h_k(t) A^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau \Big\{ Z_{i,k}(t) g^{(1)}(\beta_0^T Z_{i,k}(t)) \\
&\quad - e_k(t, \beta_0) \Big\} dM_{i,k}(t) \\
&\quad + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{i,k}(t) + o_p(1), \qquad \text{(A.20)}
\end{aligned}
$$

where $h_k(t) = \int_0^t e_k(s, \beta_0) ds$.

From the above expression, the convergence in finite-dimensional distribution of W_n follows from the Multivariate Central Limit theorem.

To complete the proof, we also need the tightness of W_n. Because $\phi_{i,k}(t)$ consists of monotone functions, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{i,k}(t)$ is tight (Van der Vaart and Wellner, 1996, p.215). This completes the proof.

Proof of Theorem 4 It is easy to see that

$$
\begin{aligned}
V(t,z) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} I(Z_{i,k} \le z) \hat{M}_{i,k}(t) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^t I(Z_{i,k} \le z) d\hat{M}_{i,k}(s) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^t I(Z_{i,k} \le z) dM_{i,k}(s) \\
&\quad - n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^t I(Z_{i,k} \le z) Y_{i,k}(s) d\{\hat{\mu}_{0,k}(s) - \mu_{0,k}(s)\} \\
&\quad - n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^t I(Z_{i,k} \le z) Y_{i,k}(s) \left\{ g\left(\hat{\beta}^T Z_{i,k}\right) - g\left(\beta_0^T Z_{i,k}\right) \right\} ds \\
&\equiv \omega_1(t,z) - \omega_2(t,z) - \omega_3(t,z). \qquad \text{(A.21)}
\end{aligned}
$$

Note that $\omega_2(t, z)$ can be rewritten as

$$\omega_2(t, z) = \sum_{k=1}^{K} \int_0^t \frac{1}{n} \sum_{i=1}^{n} I(Z_{i,k} \le z) Y_{i,k}(s) dn^{\frac{1}{2}} \{\hat{\mu}_{0,k}(s) - \mu_{0,k}(s)\}. \quad \text{(A.22)}$$

From the proof of Theorem 2, we can see that

$$\sqrt{n}\{\hat{\mu}_{0,k}(t) - \mu_{0,k}(t)\} = -\int_0^t \frac{\sum_{i=1}^n Y_{i,k}(s) Z_{i,k}^T g^{(1)} \left(\hat{\beta^*}^T Z_{i,k}\right)}{\sum_{j=1}^n Y_{j,k}(s)} ds \sqrt{n}(\hat{\beta} - \beta_0)$$
$$+ n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t \frac{1}{\frac{1}{n} \sum_{j=1}^n Y_{j,k}(s)} dM_{i,k}(s). \quad \text{(A.23)}$$

So $\omega_2(t, z)$ can be written as

$$\omega_2(t, z) = -\sum_{k=1}^{K} \int_0^t \left(\frac{1}{n} \sum_{i=1}^{n} I(Z_{i,k} \le z) Y_{i,k}(s) \right) \frac{\sum_{j=1}^n Y_{j,k}(s) Z_{j,k}^T g^{(1)} \left(\hat{\beta^*}^T Z_{j,k}\right)}{\sum_{l=1}^n Y_{l,k}(s)} ds$$
$$\times \sqrt{n}(\hat{\beta} - \beta_0) + n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^t \frac{\frac{1}{n} \sum_{j=1}^n I(Z_{j,k} \le z) Y_{j,k}(s)}{\frac{1}{n} \sum_{l=1}^n Y_{l,k}(s)} dM_{i,k}(s).$$
(A.24)

By the Uniform SLLN (Pollard 1990), $\frac{\sum_{j=1}^n Y_{j,k}(t) Z_{j,k} g^{(1)} (\hat{\beta^*}^T Z_{j,k})}{\sum_{l=1}^n Y_{l,k}(t)}$ converges to $e_k(t, \beta_0)$ uniform in t. Similarly, it can be proven that $\frac{1}{n} \sum_{j=1}^n I(Z_{j,k} \le z) Y_{j,k}(t)$ converges to $f_k(t, z)$ uniformly in t and z. Thus, the first term of (A.24) equals

$$-\sum_{k=1}^{K} \int_0^t f_k(s, z) e_k^T(s, \beta_0) ds \sqrt{n}(\hat{\beta} - \beta_0) + o_p(1) \quad \text{(A.25)}$$

uniformly in t and z. By the Uniform SLLN and Lemma 1 of Lin et al. (2000), the second term of (A.24) is

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^t \frac{f_k(s, z)}{s_k^{(0)}(s)} dM_{i,k}(s) + o_p(1) \quad \text{(A.26)}$$

uniformly in t and z. From the proof of Theorem 1, it has been shown that

$$\sqrt{n}(\hat{\beta}-\beta_0) = A^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{Z_{i,k}g^{(1)}\left(\beta_0^T Z_{i,k}\right)-e_k(t,\beta_0)\right\}dM_{i,k}(t)$$
$$+o_p(1). \tag{A.27}$$

Combing the (A.24, A.25, A.26) and (A.27), we arrive at

$$\omega_2(t,z) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{t}\frac{f_k(s,z)}{s_k^{(0)}(s)}dM_{i,k}(s)$$
$$+n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K}h(t,z)A^{-1}\int_0^{\tau}\left\{Z_{i,k}g^{(1)}\left(\beta_0^T Z_{i,k}\right)-e_k(t,\beta_0)\right\}dM_{i,k}(t)$$
$$+o_p(1). \tag{A.28}$$

By Taylor expansion, we have

$$\begin{aligned}\omega_3(t,z) &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{t}I(Z_{i,k}\le z)Y_{i,k}(s)\left\{g\left(\hat{\beta}^T Z_{i,k}\right)-g\left(\beta_0^T Z_{i,k}\right)\right\}ds\\ &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{t}I(Z_{i,k}\le z)Y_{i,k}(s)g^{(1)}\left(\hat{\beta}^{*T} Z_{i,k}\right)Z_{i,k}^T ds\sqrt{n}\{\hat{\beta}-\beta_0\},\end{aligned} \tag{A.29}$$

where $\hat{\beta}^*$ lies between β_0 and $\hat{\beta}$. It can be shown that $n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t I(Z_{i,k}\le z)Y_{i,k}(s)g^{(1)}(\hat{\beta}^{*T}Z_{i,k})Z_{i,k}^T ds$ converges to $l(t,z)$ almost surely uniformly in t and z. So we have

$$\omega_3(t,z) = l(t,z)\sqrt{n}\{\hat{\beta}-\beta_0\}+o_p(1). \tag{A.30}$$

Using the results from the proof of Theorem 1, we get

$$\omega_3(t,z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{K}l(t,z)A^{-1}\int_0^{\tau}\left\{Z_{i,k}g^{(1)}\left(\beta_0^T Z_{i,k}\right)-e_k(t,\beta_0)\right\}dM_{i,k}(t)$$
$$+o_p(1). \tag{A.31}$$

Based on the above results, we can arrive at

$$
\begin{aligned}
V(t,z) = n^{-\frac{1}{2}} \sum_{i=1}^{n}\sum_{k=1}^{K} \int_{0}^{t} \left\{ I(Z_{i,k} \le z) - \frac{f_k(s,z)}{s_k^{(0)}(s)} \right\} dM_{i,k}(s) \\
-n^{-\frac{1}{2}} \sum_{i=1}^{n}\sum_{k=1}^{K} \{h(t,z) + l(t,z)\} A^{-1} \int_{0}^{\tau} \left\{ Z_{i,k} g^{(1)}(\beta_0^T Z_{i,k}) \right. \\
\left. -e_k(t,\beta_0) \right\} dM_{i,k}(t) \\
+o_p(1). \qquad \text{(A.32)}
\end{aligned}
$$

Since $V(t,z)$ is represented asymptotically as a sum of independent identical distribution random vectors, the convergence of finite-dimensional distributions can be proven by the multivariate CLT. Using the results from modern empirical process, the tightness can also be obtained. Consequently the weak convergence is proven. This completes the proof.

References

Abu-Libdeh H, Turnbull B, Clark L (1990) Analysis of multi-type recurrent events in longitudinal studies: Application to a skin cancer prevention trial. Biometrics 46:1017–1034

Anderson P, Gill R (1982) Cox's regression model for counting processes: a large sample study. Ann Stat 10:1100–1120

Bilias Y, Gu M, Ying Z (1997) Towards a general asymptotic theory for the cox model with ataggered entry. Ann Stat 25:662–682

Cai J, Schaubel D (2004) Marginal means/rates models for multiple type recurrent event data. Lifetime Data Anal 10:121–138

Chen B, Cook R, Lawless F, Zhan M (2005) Statistical methods for multivariate interval-censored recurrent events. Stat Med 24:671–691

Chen B, Cook R (2009) The analysis of multivariate recurrent events with partially missing event types. Lifetime Data Anal 15:41–58

Ghosh D, Lin D (2000) Nonparametric analysis of recurrent events and death. Biometrics 56:554–562

Ghosh D, Lin D (2002) Marginal regression models for recurrent and terminal events. Stat Sinica 12:88–663

Kalbfleisch J, Prentice R (2002) The statistical analysis of failure time data. Wiley, New York

Lawless J, Nadeau C (1995) Some simple robust methods for the analysis of recurrent events. Technometrics 37:68–158

Lawless J, Wigg M, Tuli S, Drake J, Lamberti-Pasculli M (2001) Analysis of repeated failures or durations, with application to shunt failures for patients with paediatric hydrocephalus. Appl Stat 50:449–465

Liang K, Zeger S (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22

Lin D, Wei L, Ying Z (1993) Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika 80:72–557

Lin D, Fleming T, Wei L (1994) Confidence bands for survival curves under the proportional hazards model. Biometrika 81:73–81

Lin D, Wei L, Ying Z (1998) Accelerated failure time models for counting processes. Biometrika 85:605–618

Lin D, Wei L, Ying Z (2000) Semiparametric regression for the mean and rate functions of recurrent events. J R Stat Soc B 62:30–711

Liu L, Wolfe R, Huang X (2004) Shared frailty models for recurrent events and a terminal event. Biometrics 60:747–756

Liu Y, Wu Y, Cai J, Zhou H (2010) Additive-multiplicative rates model for recurrent events. Lifetime Data Anal 16:353–373

Ng E, Cook R (1999) Robust inference for bivariate point processes. Can J Stat 27:509–524

Pepe M, Cai J (1993) Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. J Am Stat Assoc 88:20–811

Pollard D (1990) Empirical processes:theory and application. Institute of Mathematical Statistics, Hyward

Prentice R, Williams B, Peterson A (1981) On the regression analysis of multivariate failure time data. Biometrika 68:79–373

Schaubel D, Zeng D, Cai J (2006) A semiparametric additive rates model for recurrent event data. Lifetime Data Anal 12:389–406

Schaubel D, Cai J (2006a) Rate/mean regression for multiple-sequence recurrent event data with missing event category. Scand J Stat 33:191–207

Schaubel D, Cai J (2006b) Multiple imputation methods for recurrent event data with missing event category. Can J Stat 34:677–692

Sen P, Singer J (1993) Large sample methods in statistics. Chapman & Hall, New York

Spiekerman C, Lin D (1998) Marginal regression models for multivariate failure time data. J Am Stat Assoc 93:75–1164

Sun L, Zhu L, Sun J (2009) Regression analysis of multivariate recurrent event data with time-varying covariate effects. J Multivar Anal 100:2214–2223

Van der Vaart A, Wellner J (1996) Weak convergence and empirical processes. Springer, New York

Zeng D, Cai J (2010) A semiparametric additive rate model for recurrent events with an informative terminal event. Biometrika 97:699–712