

Additive–multiplicative rates model for recurrent events

Yanyan Liu · Yuanshan Wu · Jianwen Cai ·
Haibo Zhou

Received: 20 November 2008 / Accepted: 26 February 2010 / Published online: 14 March 2010
© Springer Science+Business Media, LLC 2010

Abstract Recurrent events are frequently encountered in biomedical studies. Evaluating the covariates effects on the marginal recurrent event rate is of practical interest. There are mainly two types of rate models for the recurrent event data: the multiplicative rates model and the additive rates model. We consider a more flexible additive–multiplicative rates model for analysis of recurrent event data, wherein some covariate effects are additive while others are multiplicative. We formulate estimating equations for estimating the regression parameters. The estimators for these regression parameters are shown to be consistent and asymptotically normally distributed under appropriate regularity conditions. Moreover, the estimator of the baseline mean function is proposed and its large sample properties are investigated. We also conduct simulation studies to evaluate the finite sample behavior of the proposed estimators. A medical study of patients with cystic fibrosis suffered from recurrent pulmonary exacerbations is provided for illustration of the proposed method.

Keywords Recurrent events · Rate regression · Additive–multiplicative rates model · Counting process · Empirical process

Y. Liu · Y. Wu
School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China

Y. Wu · J. Cai (✉) · H. Zhou
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill,
NC 27599-7420, USA
e-mail: cai@bios.unc.edu

1 Introduction

Recurrent event data are common in biomedical studies. For example, patients with cystic fibrosis may suffer from repeated pulmonary exacerbations of respiratory symptoms (Therneau and Grambsch 2000); HIV patients may experience recurrent opportunistic infections (Li and Lagakos 1997). Other examples include myocardial infarctions, tumor metastases etc.

Modeling the occurrence of recurrent events has been a much discussed topic in the last few years and recurrent event data can be viewed as a special case of multivariate failure time data since the different event times within the same subject are ordered and thus correlated. Therefore, these data can be analyzed by well-established marginal intensity model approaches (e.g., Wei et al. 1989; Lee et al. 1992) and conditional intensity model approaches (e.g., Prentice et al. 1981; Andersen and Gill 1982; Chang and Wang 1999). These models are commonly used due to the convention and availability of statistical software. However, when the true underlying covariate effects may add to, rather than multiply, the baseline event intensity, a plausible alternative would be the additive hazards model (e.g., Aalen 1980; Breslow and Day 1980; Cox and Oakes 1984; Buckley 1984; Breslow and Day 1987; Aalen 1989; Huffer and McKeague 1991; Andersen et al. 1993). Most of these approaches are fully parametric except the nonparametric methods of Aalen (1980, 1989) and Huffer and McKeague (1991). Lin and Ying (1994) proposed a semiparametric additive model and gave an explicit estimator of the model parameters through the techniques analogous to partial-likelihood-based method (Andersen and Gill 1982). The additive and multiplicative hazards models postulate two rather different relationship between the covariate and the hazard function. However, these two types of effects could exist for different covariates. Lin and Ying (1995) proposed a general additive–multiplicative intensity model which allows some covariate effects to be additive while others to be multiplicative, and studied the asymptotic properties of their proposed estimates.

In the context of recurrent event data, because the mean number of events is more interpretable than the event intensity, some authors have proposed to model the mean function under the assumption that the covariates act multiplicatively on the unspecified baseline rate function (e.g., Pepe and Cai 1993; Lawless and Nadeau 1995; Lin et al. 2000; Cai and Schaubel 2004). Schaubel et al. (2006) developed a semiparametric additive model for the marginal recurrent event rate, wherein they assume the covariates effects are added to the unspecified baseline rate function. In this article, we consider a general additive–multiplicative rates model in the spirit of the work of Lin and Ying (1995) which includes the multiplicative model and the additive model as special cases. The model assumes that some of the covariates act additively while others multiplicatively on the unspecified baseline rate function.

The remainder of this article is organized as follows. In Sect. 2, we introduce the proposed models and derive a class of estimating equations for the estimates of regression parameters and the baseline rate function. The asymptotic properties of the proposed estimators are investigated in Sect. 3 with the proofs deferred to the Appendix. Simulation studies are conducted in Sect. 4 to evaluate the finite-sample behavior of the asymptotic approximation. The proposed methods are applied to a cystic fibrosis clinical trial in Sect. 5. Some discussions are given in Sect. 6.

2 Model and method

In this section, we describe the inference procedure for the proposed model, beginning with the establishment of some notation.

2.1 Notation and model

Let $N_i^*(t)$ ($i = 1, \dots, n$) be the number of events that occur over the interval $[0, t]$ for the i th subject under study. The corresponding right censoring time is denoted by C_i . Due to the censoring, the number of observed events at time t is $N_i(t) = N_i^*(t \wedge C_i)$, where $a \wedge b$ equals the minimum of a and b . Let $Y_i(t) = I(C_i \geq t)$ denote the at-risk process, where $I(\cdot)$ is the indicator function. Associated with each $N_i(t)$ is a p -dimensional covariate process $W_i(t) = \{Z_i^T(t), X_i^T(t)\}^T$, with any time-dependent elements to be external as defined by Kalbfleisch and Prentice (2002); i.e., to have paths which are known at time $t = 0$, where a^T is the transpose of a vector or matrix a . Furthermore, the censoring mechanism is assumed to be independent in the sense that

$$E\{dN_i^*(t)|W_i(t), C_i \geq t\} = E\{dN_i^*(t)|W_i(t)\}.$$

Assume that the follow-up interval is $[0, \tau]$, where τ is prespecified constant typically meaning the terminal time of study. Thus, the observable data consist of $\{N_i(t), Y_i(t), W_i(t); t \in [0, \tau]\}$ ($i = 1, \dots, n$). Subjects are assumed to be statistically independent, although events within the same subject are ordered and thus naturally expected to be correlated.

We consider the following additive–multiplicative rates model:

$$E\{dN_i^*(t)|W_i(t)\} = g\{\gamma_0^T Z_i(t)\}dt + h\{\beta_0^T X_i(t)\}d\mu_0(t), \quad (1)$$

where $\theta_0 = (\gamma_0^T, \beta_0^T)^T$ is p -vector of unknown regression parameters, g and h are known link functions and $\mu_0(\cdot)$ is an unspecified continuous baseline mean function for subjects with covariates $z_i(t)$ and $x_i(t)$ such that $g\{\gamma_0^T z_i(t)\} = 0$ and $h\{\beta_0^T x_i(t)\} = 1$. Let $\gamma_0 = (\gamma_{01}, \dots, \gamma_{0p_1})^T$ and $\beta_0 = (\beta_{01}, \dots, \beta_{0p_2})^T$, then γ_{0j} can be interpreted as the rate difference for one unit change in Z_{ji} while holding X_i and other components in Z_i the same when $g(x) = x$. As suggested by the associate editor, by holding Z_i and other components in X_i the same, β_{0j} can be interpreted as the logarithm of the ratio of residual rate of risk for one unit change in X_{ji} when $h(x) = e^x$, where the residual rate of risk is defined as the overall rate $E\{dN_i^*(t)|W_i(t)\}$ minus the additive rate of risk. Apparently, model (1) defines a very rich family of models through the link functions g and h , which contains the additive rates model and the multiplicative rates model as its special cases. Selection of these appropriate link functions may be based on prior data or the resulting interpretation of the regression parameters. If the rate of interest is conditioning on the time-independent covariates, i.e., $E\{dN_i^*(t)|W_i\}$, then by integration or summing one can obtain $E\{N_i^*(t)|W_i\}$, to which we refer as means model. Since the means model is a special case of rates model, we focus on the rates model (1) hereafter.

2.2 Inference procedure

We now describe the estimating procedure under model (1), beginning by defining the process:

$$M_i(t, \theta) = N_i(t) - \int_0^t Y_i(u)\{g\{\gamma^T Z_i(u)\}du + h\{\beta^T X_i(u)\}d\mu_0(u)\},$$

where $\theta = (\gamma^T, \beta^T)^T$ is included in the parameter space denoted by Θ , which contains the true value θ_0 as its interior point and is assumed to be compact for technical proof. Abbreviate $M_i(t, \theta_0)$ as $M_i(t)$. Since $E\{dM_i(t)|W_i(t)\} = 0$ under model (1) and independent censoring, an approach analogous to generalized estimating equations (Liang and Zeger 1986) suggests the following estimating equations for $\mu_0(t)$ and θ_0 , respectively:

$$\sum_{i=0}^n \int_0^t dM_i(u) = 0, \tag{2}$$

$$\sum_{i=1}^n \int_0^\tau Q_i(u, \theta)dM_i(u, \theta) = 0, \tag{3}$$

where $Q_i(t, \theta)$ is a smooth p -vector-value function of $W_i(t)$ and θ not involving $\mu_0(t)$.

Solving (2), a class of the baseline mean function estimators for $\mu_0(t)$ (with given θ_0) is directly obtained as follows:

$$\widehat{\mu}_0(t, \theta_0) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)g\{\gamma_0^T Z_i(u)\}du\}}{\sum_{i=1}^n Y_i(u)h\{\beta_0^T X_i(u)\}}.$$

Substituting $\widehat{\mu}_0(t, \theta)$ for $\mu_0(t)$ into (3), followed by some algebraic manipulation, yields an estimating function $U(\tau, \theta)$ for θ_0 which is free of $\mu_0(t)$, where

$$U(t, \theta) = \sum_{i=1}^n \int_0^t \{Q_i(u, \theta) - \overline{Q}(u, \theta)\}dM_i(u, \theta), \tag{4}$$

and

$$\overline{Q}(t, \theta) = \frac{\sum_{i=1}^n Y_i(t)Q_i(t, \theta)h\{\beta^T X_i(t)\}}{\sum_{i=1}^n Y_i(t)h\{\beta^T X_i(t)\}}.$$

Hence, we obtain an estimator, denoted by $\widehat{\theta}$, for θ_0 , which solves equation $U(\tau, \theta) = 0$, and the corresponding estimator for $\mu_0(t)$ is then denoted by $\widehat{\mu}_0(t, \widehat{\theta})$.

To ensure the monotonicity of the estimate of the baseline mean function, the estimator for $\mu_0(t)$ is modified to be:

$$\tilde{\mu}_0(t, \hat{\theta}) = \max_{0 \leq u \leq t} \hat{\mu}_0(u, \hat{\theta}).$$

Instead of constraining $\hat{\theta}$ to force the baseline rate estimator to be positive, the baseline mean estimator $\tilde{\mu}_0(t, \hat{\theta})$ is constrained to be monotone non-decreasing, analogous to the approach of [Lin and Ying \(1994\)](#).

Similarly, one possible choice of $Q_i(t, \theta)$ discussed by [Lin and Ying \(1995\)](#) is obviously suitable here to encompass inference procedures arising from multiplicative rates model and additive rates model, respectively, as its special cases. To be specific, let

$$Q_i(t, \theta) = \left[\begin{array}{c} g'\{\gamma^T Z_i(t)\}Z_i(t)/h\{\beta^T X_i(t)\} \\ h'\{\beta^T X_i(t)\}X_i(t)/h\{\beta^T X_i(t)\} \end{array} \right], \tag{5}$$

where and in what follows $f'(x) = df(x)/dx$. Putting (5) into (4), we have the estimating equations $U(\tau, \theta)$ and therefore give the corresponding estimate of θ_0 .

3 Asymptotic properties

In this section, we establish the large sample properties of the proposed estimators, beginning with the following regularity conditions, analogous to those of [Andersen and Gill \(1982, Theorem 4.1\)](#) and assumed to be held throughout our discussion for $i = 1, \dots, n$.

- C1. $\{N_i^*(\cdot), Y_i(\cdot), W_i(\cdot), Q_i(\cdot, \theta)\}$ are independent and identically distributed;
- C2. $P(C_i \geq \tau) > 0$;
- C3. $N_i(\tau)$ is bounded by a constant;
- C4. $W_i(\cdot)$ has bounded total variation, i.e., for the j -th element of $W_i(\cdot)$, $|W_{ji}(0)| + \int_0^\tau |dW_{ji}(u)| < C^*$ almost surely, where C^* is a constant; moreover, $Q_i(\cdot, \theta)$ has bounded total variation, uniformly in $\theta \in \Theta$;
- C5. The following defined matrix A is nonsingular:

$$A = E \left\{ \int_0^\tau Y_1(u) \{Q_1(u, \theta_0) - \bar{q}(u, \theta_0)\} \left[\begin{array}{c} g'\{\gamma_0^T Z_1(u)\}Z_1(u)du \\ h'\{\beta_0^T X_1(u)\}X_1(u)d\mu_0(u) \end{array} \right]^T \right\},$$

where $\bar{q}(t, \theta) = \frac{E[Y_1(t)Q_1(t, \theta)h\{\beta^T X_1(t)\}]}{E[Y_1(t)h\{\beta^T X_1(t)\}]}$;

- C6. g is nonnegative and $h\{\beta^T X_i(t)\}$ is locally bounded away from 0 for β in a small neighborhood of β_0 ; g and h are continuous differentiable and

$$\begin{aligned} & \{\partial Q_i(t, \theta) / \partial \theta, h\{\beta^T X_i(t)\}, g'\{\gamma^T Z_i(t)\} Z_i(t), \\ & h'\{\beta^T X_i(t)\} X_i(t) : t \in [0, \tau], i = 1, \dots, n \} \end{aligned}$$

are equicontinuous and bounded uniformly in parameter space Θ .

Conditions C1, C3, C4, and C6 simplify our derivation of the asymptotic results but do not impose practical limitations. Condition C2 can be enforced by choosing τ to be not greater than the maximum observation time. Condition C5 is a technique assumption. Note that $\bar{q}(t, \theta)$ is well defined since $E[Y_1(t)h\{\beta^T X_1(t)\}]$ is bounded away from zero whenever $t \in [0, \tau]$ under conditions C2 and C6. We describe the asymptotic behavior of estimates of the regression parameters in the following theorem.

Theorem 1 *Under conditions C1 to C6, $\hat{\theta}$ converges almost surely to θ_0 , while $\sqrt{n}(\hat{\theta} - \theta_0)$ converges weakly to a zero-mean normal distribution with covariance $A^{-1}\Sigma(\tau, \tau)(A^T)^{-1}$, where*

$$\Sigma(s, t) = E \left[\int_0^s \{Q_1(u, \theta_0) - \bar{q}(u, \theta_0)\} dM_1(u) \int_0^t \{Q_1(v, \theta_0) - \bar{q}(v, \theta_0)\}^T dM_1(v) \right]$$

for s and t in $[0, \tau]$.

Theorem 1 can be proved by combining the multivariate central limit theorem and some results from empirical processes theory (Pollard 1990; Biliias et al. 1997) and a lemma from Lin et al. (2000). We present the outline of the proof in the Appendix.

Furthermore, using the Strong Law of Large Numbers (SLLN) and Lemma 1 from Lin et al. (2000) and combining with the results of Theorems 1 and 2 followed below imply that the asymptotic covariance can be consistently estimated by $\hat{A}^{-1}\hat{\Sigma}(\tau, \tau)(\hat{A}^T)^{-1}$, where

$$\begin{aligned} \hat{A} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(u) \{Q_i(u, \hat{\theta}) - \bar{Q}(u, \hat{\theta})\} \begin{bmatrix} g'\{\hat{\gamma}^T Z_i(u)\} Z_i(u) du \\ h'\{\hat{\beta}^T X_i(u)\} X_i(u) d\hat{\mu}_0(u, \hat{\theta}) \end{bmatrix}^T, \\ \hat{\Sigma}(s, t) &= \frac{1}{n} \sum_{i=1}^n \int_0^s \{Q_i(u, \hat{\theta}) - \bar{Q}(u, \hat{\theta})\} d\hat{M}_i(u) \int_0^t \{Q_i(v, \hat{\theta}) - \bar{Q}(v, \hat{\theta})\}^T d\hat{M}_i(v), \end{aligned}$$

with

$$d\hat{M}_i(t) = dN_i(t) - Y_i(t)\{g\{\hat{\gamma}^T Z_i(t)\}dt + h\{\hat{\beta}^T X_i(t)\}d\hat{\mu}_0(t, \hat{\theta})\}. \tag{6}$$

The essential asymptotic results for the baseline mean function estimator are summarized by the following theorem. We firstly introduce some notation:

$$s_{g'}(t, \theta) = E[Y_1(t)g'\{\gamma^T Z_1(t)\}Z_1(t)], s_{h'}(t, \theta) = E[Y_1(t)h'\{\beta^T X_1(t)\}X_1(t)],$$

$$s_h(t, \theta) = E[Y_1(t)h\{\beta^T X_1(t)\}], \tilde{s}(t, \theta) = \begin{bmatrix} \int_0^t \frac{s_{g'}(u, \theta)}{s_h(u, \theta)} du \\ \int_0^t \frac{s_{h'}(u, \theta)}{s_h(u, \theta)} d\mu_0(u) \end{bmatrix}.$$

Theorem 2 Under conditions C1 to C6, $\{\tilde{\mu}_0(t, \hat{\theta}) - \mu_0(t)\}$ converges almost surely to 0, uniformly in $t \in [0, \tau]$, while $\sqrt{n}\{\tilde{\mu}_0(t, \hat{\theta}) - \mu_0(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function $\xi(s, t) = E[\Psi_1(s)\Psi_1(t)]$, where

$$\Psi_i(t) = \int_0^t \frac{1}{s_h(u, \theta_0)} dM_i(u) - \{\tilde{s}(t, \theta_0)\}^T A^{-1} \int_0^\tau \{Q_i(u, \theta_0) - \bar{q}(u, \theta_0)\} dM_i(u).$$

Using arguments analogous to those of Lin and Ying (1994), we can prove $\sqrt{n}\{\tilde{\mu}_0(t, \hat{\theta}) - \hat{\mu}_0(t, \hat{\theta})\} = o_p(1)$, where $o_p(1)$ is uniformly in $t \in [0, \tau]$. Therefore we work with $\hat{\mu}_0(t, \hat{\theta})$ instead of $\tilde{\mu}_0(t, \hat{\theta})$ in the proof of Theorem 2. We then apply the central limit theorem and results from empirical processes theory to the decomposition, $\{\hat{\mu}_0(t, \hat{\theta}) - \mu_0(t)\} = \{\hat{\mu}_0(t, \hat{\theta}) - \hat{\mu}_0(t, \theta_0)\} + \{\hat{\mu}_0(t, \theta_0) - \mu_0(t)\}$, to finish the proof of Theorem 2. An outline of the proof is also provided in the Appendix.

The covariance function $\xi(s, t)$ can be consistently estimated by replacing limiting quantities in $\Psi_i(t)$ with their respectively empirical counterparts. To be specific, $\hat{\xi}(s, t) = \frac{1}{n} \sum_{i=1}^n \hat{\Psi}_i(s)\hat{\Psi}_i(t)$, where

$$\hat{\Psi}_i(t) = \int_0^t \frac{1}{S_h(u, \hat{\theta})} d\hat{M}_i(u) - \{\tilde{S}(t, \hat{\theta})\}^T \hat{A}^{-1} \int_0^\tau \{Q_i(u, \hat{\theta}) - \bar{Q}(u, \hat{\theta})\} d\hat{M}_i(u)$$

with $S_h(t, \theta)$ and $\tilde{S}(t, \theta)$ listed in the Appendix.

4 Simulation studies

Simulation studies were conducted to assess the adequacy of the proposed large-sample approximations for practical sample size. Specially, event times were generated from the following mixed effects rate model:

$$E\{dN_i^*(t)|Z_i(t), X_i(t), \eta_i\} = \eta_i[g\{\gamma_0^T Z_i(t)\}dt + h\{\beta_0^T X_i(t)\}d\mu_0(t)], \quad i = 1, \dots, n,$$

where the frailty variable η_i induces positive correlation among the within-subject events and follows a gamma distribution with unit-mean and variance σ_η . Note that larger σ_η corresponds to higher positive correlation among intra-subject event times and $\sigma_\eta = 0$, i.e., $\eta_i = 1$ for all i , indicates that within-subject event times are independent. We considered the simple cases: $g\{\gamma_0^T Z_i(t)\} = \gamma_0 Z_i$, $h\{\beta_0^T X_i(t)\} = \exp(\beta_0 X_i)$ and the baseline rate $m_0 \equiv d\mu_0(t)/dt$ is a constant. The $(j+1)$ th event time for the i th subject was generated by $T_{i,j+1} = T_{i,j} - \{\eta_i[\gamma_0 Z_i + \exp(\beta_0 X_i)m_0]\}^{-1} \log(U_{i,j+1})$, where $U_{i,j}$ were generated from uniform (0, 1) distribution independently, and $T_{i,0} = 0$. The estimating function $U(\tau, \theta)$ was derived by choosing $Q_i(t, \theta)$ according to (5).

Covariates Z_i 's were independently generated from uniform (0, 1). We generated X_i 's independently from Bernoulli distribution with success probability 0.5. The censoring times C_i 's were generated from uniform (0, 3). The baseline rate $d\mu_0(t)/dt$ varied from 0.125 to 0.25. The expected number of observed events varied from 0.42 to 1.67 and 0.57 to 1.98 at $(\gamma_0, \beta_0) = (0, 0.2)$ and $(\gamma_0, \beta_0) = (0.2, 0.2)$, respectively. We set sample size $n = 50, 100, 200$ and each simulation was repeated 2000 times. The sample mean and sample standard deviation of the 2000 estimates are given in the Mean and SD columns, respectively. The SE columns give the average of the estimated standard errors and CP columns give the coverage probability of the nominal 95% confidence interval for the true parameter using the estimated standard error.

Table 1 displays the simulation results. From Table 1, we make the following observations: (i) the estimators proposed are approximately unbiased when $\gamma_0 = 0$ and slightly underestimated when $\gamma_0 = 0.2$; (ii) the average asymptotic standard error estimators (SE) are approximately equal to the empirical standard deviations (SD); (iii) the corresponding 95% confidence intervals based on the estimated standard errors provide reasonable coverage probabilities; (iv) for fixed n and μ_0 , both SD and SE increase as the intra-subject event times correlation σ_η increases; (v) as expected, as the sample size increases, SE is decreasing and closer to SD.

As suggested by the referees, additional simulation studies were also conducted to investigate the effect of time-dependent baseline rate function $d\mu_0(t)/dt$ on the performances of the proposed estimator. Two cases $d\mu_0(t)/dt = 0.25t$ and $d\mu_0(t)/dt = 0.5t$ are considered with other parameter values not changed. We then use the thinning algorithm (Ross 2006, p. 83) to generate the recurrent event times when the baseline rate function $d\mu_0(t)/dt$ is time-dependent. To be specific, let $m_0(t) = d\mu_0(t)/dt$ and $m^i(t) = \eta_i[\gamma_0 Z_i + \exp(\beta_0 X_i)m_0(t)]$, then the recurrent event times $\{T_{i,j}\}$ for the i th subject over $[0, \tau]$ can be generated as follows:

- Step 1: $t = 0, j = 0$, and $T_{i,j} = 0$.
- Step 2: Generate a random number U from uniform (0, 1) distribution.
- Step 3: $t = t - \frac{1}{m^i(\tau)} \log U$. If $t > \tau$, stop.
- Step 4: Generate a random number V from uniform (0, 1) distribution.
- Step 5: If $V \leq \frac{m^i(t)}{m^i(\tau)}$, set $j = j + 1$ and $T_{i,j} = t$.
- Step 6: Go to Step 2.

The related results are summarized in Table 2. It is evident that the same conclusions can be drawn as that from Table 1.

Table 1 Summary statistics for the simulation studies based on 2000 replicated data set when the baseline rate function $d\mu_0(t)/dt$ is time-independent

n	$\mu_0(t)$	σ_η	$\gamma_0 = 0$				$\gamma_0 = 0.2$			
			Mean	SD	SE	CP	Mean	SD	SE	CP
50	0.125t	0	0.002	0.149	0.153	0.954	0.204	0.191	0.198	0.954
		0.25	-0.001	0.158	0.159	0.950	0.190	0.206	0.206	0.947
	0.25t	0	0.008	0.206	0.216	0.957	0.208	0.232	0.246	0.960
		0.25	-0.003	0.225	0.227	0.940	0.178	0.257	0.258	0.947
100	0.125t	0	0.000	0.108	0.106	0.944	0.196	0.139	0.137	0.949
		0.25	-0.001	0.107	0.108	0.950	0.192	0.143	0.142	0.950
	0.25t	0	-0.002	0.149	0.149	0.949	0.188	0.170	0.169	0.948
		0.25	0.000	0.154	0.154	0.950	0.184	0.176	0.177	0.954
200	0.125t	0	0.001	0.075	0.074	0.947	0.198	0.098	0.096	0.945
		0.25	-0.001	0.076	0.076	0.952	0.190	0.101	0.100	0.949
	0.25t	0	0.000	0.106	0.104	0.949	0.193	0.120	0.118	0.951
		0.25	-0.002	0.108	0.109	0.953	0.178	0.125	0.124	0.949

The additive–multiplicative means model $E(N^*(t)|Z, X) = \gamma_0 Zt + \exp(0.2X)\mu_0(t)$ is considered with $\gamma_0 = 0$ or 0.2, where $\mu_0(t)$ is the baseline mean function, Z is generated from uniform (0, 1), and X is a Bernoulli variable with success probability 0.5. σ_η describes the positive correlation within the recurrent events and n is the number of subjects under study. Mean is the sample mean of the estimator $\hat{\gamma}$, SD is the sampling standard deviation of $\hat{\gamma}$, SE is the sampling mean of the standard error estimator, and CP is the coverage probability of the 95% confidence interval for $\hat{\gamma}$

Table 2 Summary statistics for the simulation studies based on 2000 replicated data set when the baseline rate function $d\mu_0(t)/dt$ is time-dependent

n	$\mu_0(t)$	σ_η	$\gamma_0 = 0$				$\gamma_0 = 0.2$			
			Mean	SD	SE	CP	Mean	SD	SE	CP
50	0.125t ²	0	-0.004	0.149	0.150	0.949	0.199	0.195	0.192	0.950
		0.25	0.001	0.154	0.155	0.947	0.201	0.209	0.201	0.951
	0.25t ²	0	0.001	0.204	0.210	0.951	0.184	0.232	0.240	0.947
		0.25	-0.009	0.213	0.218	0.953	0.182	0.248	0.250	0.953
100	0.125t ²	0	0.000	0.101	0.105	0.951	0.199	0.134	0.137	0.949
		0.25	-0.001	0.106	0.109	0.953	0.185	0.141	0.142	0.948
	0.25t ²	0	0.000	0.145	0.147	0.949	0.192	0.160	0.166	0.953
		0.25	-0.002	0.147	0.153	0.952	0.178	0.169	0.173	0.953
200	0.125t ²	0	0.002	0.073	0.075	0.943	0.192	0.093	0.095	0.956
		0.25	-0.001	0.074	0.077	0.957	0.194	0.102	0.100	0.942
	0.25t ²	0	0.003	0.099	0.103	0.949	0.193	0.113	0.116	0.947
		0.25	0.000	0.109	0.108	0.943	0.179	0.121	0.121	0.950

Note: See Table 1

5 A real example

We now apply the methods developed in Sect. 2 to data from a randomized clinical trial which was conducted to assess the efficacy of treatment rhDNase (TRT), a highly purified recombinant enzyme, in reducing pulmonary exacerbations of respiratory symptoms for patients with cystic fibrosis (Therneau and Grambsch 2000). In the study, 325 patients in the placebo arm and 322 patients in the rhDNase arm were monitored for pulmonary exacerbations, along with measuring of baseline level of forced expiratory volume per second (FEV). Most patients were followed for about 170 days. Table 3 shows the frequency of the number of exacerbations. The endpoint of this study is to evaluate the effects of rhDNase (possibly adjusted by covariate FEV) on the recurrences of pulmonary exacerbations over times.

We will fit the additive rates (AR) model:

$$E\{dN^*(t)|TRT, FEV\} = (\beta_T TRT + \beta_F FEV)dt + d\mu_0(t),$$

the multiplicative rates (MR) model:

$$E\{dN^*(t)|TRT, FEV\} = \exp(\beta_T TRT + \beta_F FEV)d\mu_0(t),$$

and the additive–multiplicative rates models:

$$E\{dN^*(t)|TRT, FEV\} = (\beta_F FEV)dt + \exp(\beta_T TRT)d\mu_0(t)$$

and

$$E\{dN^*(t)|TRT, FEV\} = (\beta_T TRT)dt + \exp(\beta_F FEV)d\mu_0(t),$$

which are denoted by model AMR1 and model AMR2, respectively.

Mimicking Lin et al. (2001), we use mean-square-type distance between the observed and the expected recurrences, defined as

$$D^* = \sum_{l=1}^L \sum_{i=1}^n [\widehat{M}_i(t_l)]^2 / \sum_{l=1}^L \sum_{i=1}^n Y_i(t_l),$$

Table 3 Frequency distribution of pulmonary exacerbations in the rhDNase trial

Group	Patients	Number of exacerbations					
		0	1	2	3	4	5
rhDNase	322	218	65	30	6	3	0
placebo	325	186	97	24	13	4	1

to measure the overall lack of fit in the above four rates models, where $\widehat{M}_i(t)$ is defined by (6) and $0 \equiv t_0 < t_1 < \dots < t_L < t_{L+1} \equiv \tau$ is the set that consists of all the jump points of $\{N_i(t), Y_i(t); t \in [0, \tau)\} (i = 1, \dots, n)$ in addition to 0 and τ .

In order to examine graphically which model fit the rhDNase data better, we replace the continuous covariate FEV with its discretized version, denoted by $dFEV = I(FEV > 57.6)$, where 57.6 is the median of FEV, in the models AR, MR, AMR1, and AMR2, and thus we partition the subjects under study into four strata. The bold and thin solid curves in Fig. 1 display the Nelson-Aalen-type nonparametric and the model-based estimators for the mean numbers of pulmonary exacerbations, respectively, in each stratum under various models. Figure 2 presents the difference of nonparametric estimator for the mean numbers of pulmonary exacerbations in the stratum ($TRT = 0, dFEV = 0$) over that in the stratum ($TRT = 1, dFEV = 0$) indicated by bold solid curve and the corresponding difference of the model-based estimators indicated by the dash curve (for model AR), the dotted-dashed curve (for model AMR1), and the thin solid curve (for model AMR2). Likewise, the difference of the nonparametric and model-based estimators for the mean numbers of pulmonary exacerbations in the stratum ($TRT = 0, dFEV = 1$) over that in the stratum ($TRT = 1, dFEV = 1$) are displayed in Fig. 3. As indicated in these figures, the model AR is unlikely to fit the rhDNase data very well since the difference of the nonparametric means functions are not straight lines, and the model AMR2 may provide a reasonable description for the data set.

For further comparisons, we consider the models AR, MR, AMR1, and AMR2 using FEV instead of $dFEV$, and we also use D^* as a criterion to select model. The analysis results are summarized in Table 4. First, under the measure of D^* , the model AMR2 is the best among the four models AR, MR, AMR1, and AMR2, which has been suggested graphically as reflected in Fig. 1. This indicates that the effect of FEV on the recurrence of pulmonary exacerbations is more likely to be multiplicative while the effect of treatment rhDNase is more likely to be additive. Second, the treatment rhDNase is effective in reducing the recurrence of pulmonary exacerbations. Higher FEV level is associated with lower frequency of the recurrences. Figure 4 displays the mean frequency estimators for the rhDNase and placebo patients with two different levels of FEV under the selected model AMR2.

6 Discussions

We propose a semiparametric regression method for analyzing recurrent event data and consider an unified model in the sense that it allows for both multiplicative and additive covariate effects. The resulting regression parameter estimator is shown to be consistent and asymptotically normally distributed and a baseline mean estimator is proposed, which is also shown to be uniformly consistent and weakly convergent. Numerical results indicate that the proposed method performs well in finite sample. A medical study is provided as an illustration.

As the associate editor pointed out, one potential advantage of the additive–multiplicative rates model (1) is that the additive term is appropriate for detecting large difference in absolute risk, while the multiplicative term is sensitive for detecting large

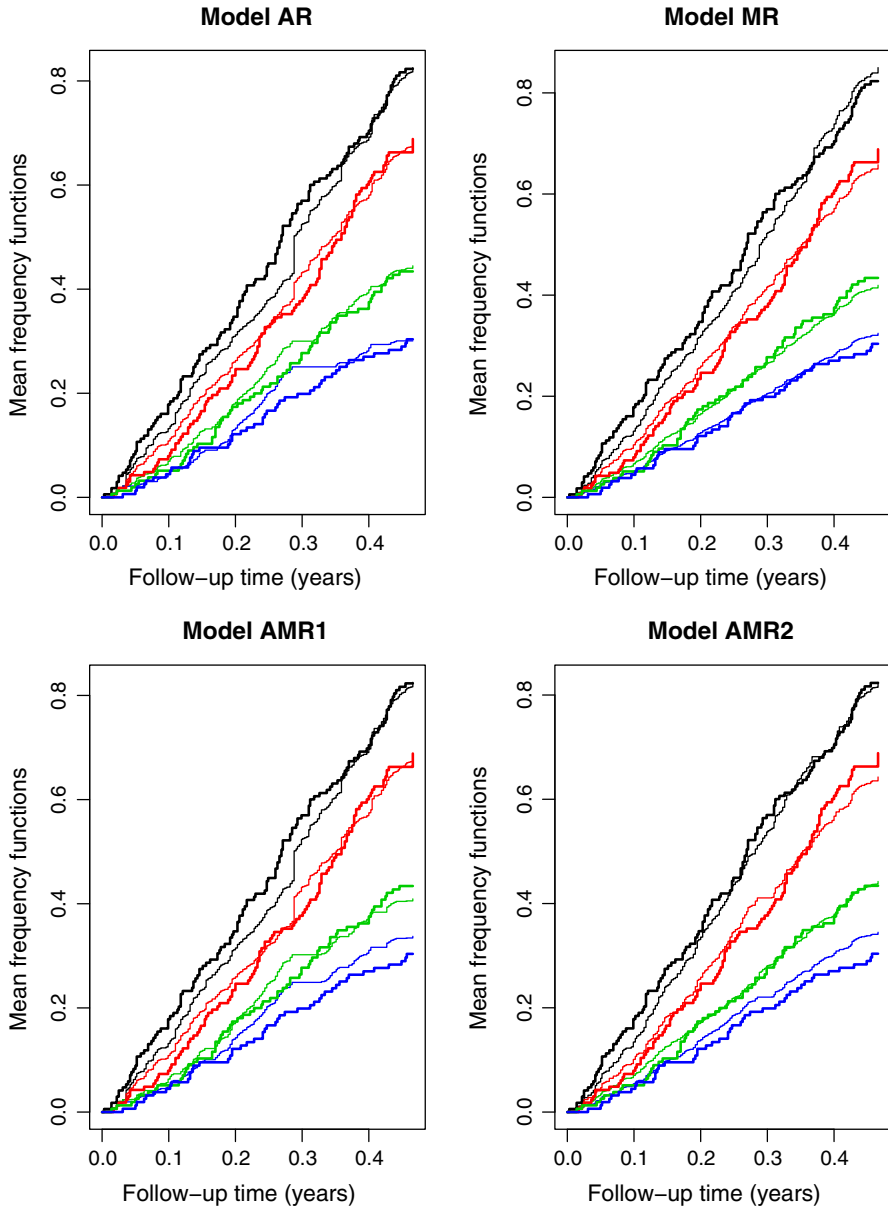


Fig. 1 Mean frequency functions of pulmonary exacerbations for each stratum in the rhDNase trial. The Nelson-Aalen-type nonparametric estimators for the mean numbers of pulmonary exacerbations are shown by *bold solid* curves in the strata (TRT = 0, dFEV = 0), (TRT = 1, dFEV = 0), (TRT = 0, dFEV = 1), and (TRT = 1, dFEV = 1), from top to bottom, respectively. The corresponding model-based estimators are pertained by *thin solid* curves

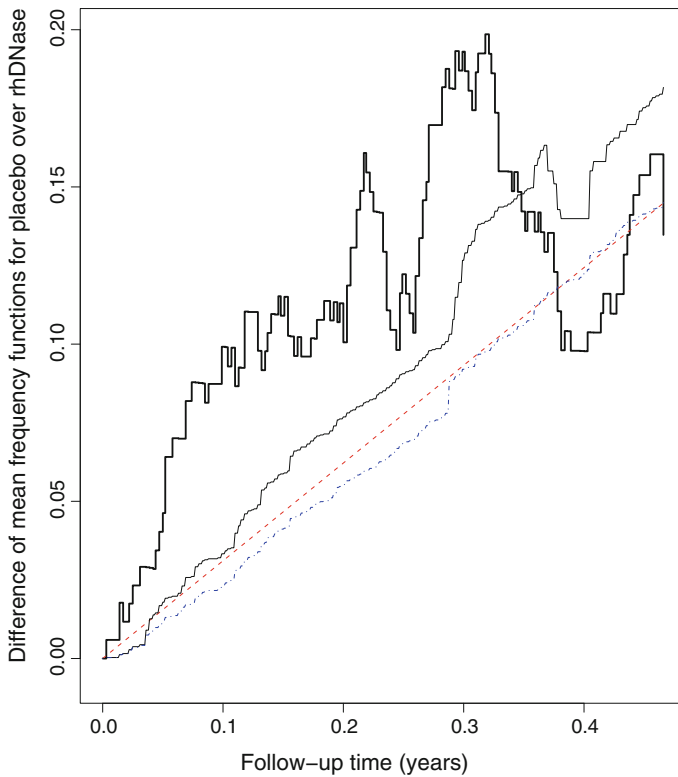


Fig. 2 Difference of the mean frequency functions for the stratum ($TRT = 0$, $dFEV = 0$) versus ($TRT = 1$, $dFEV = 0$). The **bold solid** curve pertains to the nonparametric estimator; the *dashed* curve pertains to the model AR based estimator; the *dotted-dashed* curve pertains to the model AMR1 based estimator; the *thin solid* curve pertains to the model AMR2 based estimator

difference in relative risk (i.e., rate ratio). Thus, in some situations where both types of covariates exist at the same time, one can consider model (1) as a candidate regression model for recurrent event data. Several considerations can be used as guidances in the initial stage. The first factor in deciding the classification is the objectives of the investigator and the underlying biological process. If the investigator is interested in studying the risk difference of some risk factors, then those variables should be added to the additive part. If the investigator is interested in studying the risk ratio of some other risk factors, then those variables should be added to the multiplicative part. Based on the biological process, the covariates anticipated to have a large impact in absolute risks should be included in the additive part and those which could have a large impact in risk ratios should be included in the multiplicative part. When the biological process is not clear, some data-driven methods could be used for the classification of covariates. For example, when the number of covariates is small, all the possible models from different combinations of covariates in the additive and the multiplicative parts can be considered. To facilitate the selection of the models, we have adopted a mean-square-type distance measure D^* between the observed and

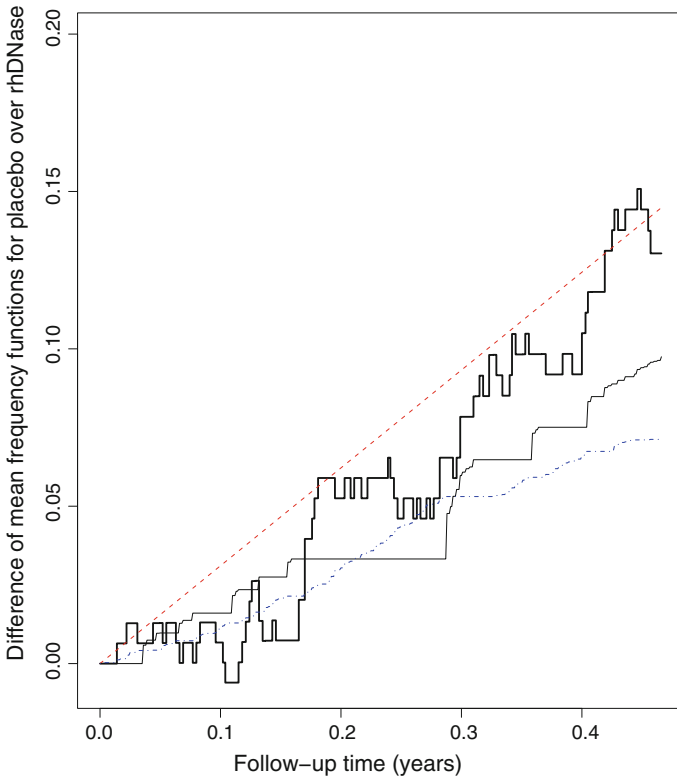


Fig. 3 Difference of the mean frequency functions for the stratum (TRT=0, dFEV=1) versus (TRT=1, dFEV=1). The *bold solid* curve pertains to the nonparametric estimator; the *dashed* curve pertains to the model AR based estimator; the *dotted-dashed* curve pertains to the model AMR1 based estimator; the *thin solid* curve pertains to the model AMR2 based estimator

Table 4 Analysis results for the rhDNase trial

Model	$\hat{\beta}_T$	SE ($\hat{\beta}_T$)	<i>p</i> -Value	$\hat{\beta}_F$	SE ($\hat{\beta}_F$)	<i>p</i> -Value	<i>D</i> *
AR	-0.311	0.145	0.032	-0.0179	0.0027	<0.001	0.8907
MR	-0.256	0.121	0.034	-0.0162	0.0028	<0.001	0.8889
AMR1	-0.135	0.065	0.037	-0.0178	0.0027	<0.001	0.8901
AMR2	-0.313	0.140	0.026	-0.0142	0.0027	<0.001	0.8887

the expected recurrences as model selection criterion. The assignment of the covariates to the additive part or the multiplicative part will follow the model with the smallest *D**. When the number of covariates is moderately large, this comprehensive method could be time-consuming. We recommend to use a two-step approach. First, one can plot the nonparametric estimators for the risk difference and the risk ratio for every covariate. The covariate which has constant risk difference over time should be included in the additive part and the covariate which has constant risk ratio should be

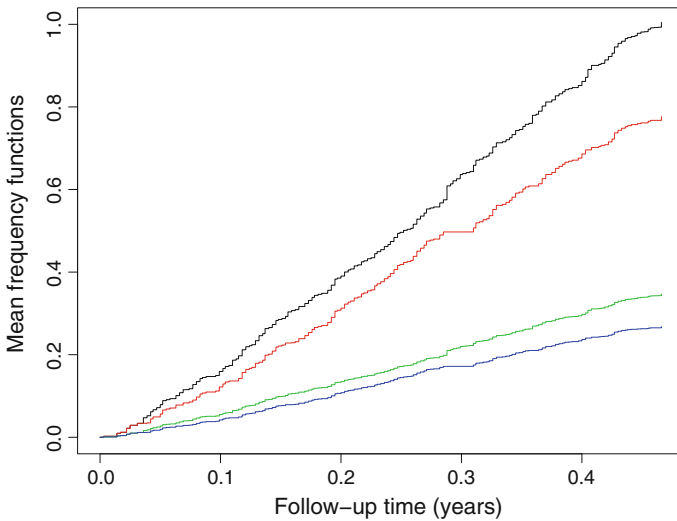


Fig. 4 Estimated mean frequency functions of pulmonary exacerbations in the rhDNase trial under the model AMR2. The curves from top to bottom pertain to the patients with placebo and FEV = 25, the patients with rhDNase and FEV = 25, the patients with placebo and FEV = 100, and the patients with rhDNase and FEV = 100, respectively

included in the multiplicative part. Second, for those covariates which are not clear whether they have constant risk difference or risk ratio, we can consider models from different combinations of these covariates in the additive and the multiplicative parts. The assignment of the covariates to the additive part or the multiplicative part will follow the model with the smallest D^* .

Obviously, a suitable choice of $Q_i(t, \theta)$ would lead to appropriate interpretation of the regression parameters and one form of $Q_i(t, \theta)$ is proposed, which includes multiplicative and additive covariate effects as its special cases. The estimating equations are proposed to allow arbitrary dependence structures within recurrent events. This is analogous to the use of a generalized estimating equation with an independence working assumption for longitudinal data (Liang and Zeger 1986). However, such approach may not necessarily be efficient. One possible way to improve the efficiency of estimate for the regression parameter is to follow the idea in Cai and Prentice (1995) and consider the weighted estimating function

$$U_\omega(\tau, \theta) = \sum_{i=1}^n \int_0^\tau \omega_i(u, \theta) \{Q_i(u, \theta) - \bar{Q}(u, \theta)\} dM_i(u, \theta),$$

where $\omega = (\omega_i; i = 1, \dots, n)$ is a weight. For more discussions on the weight and its choice refer to Cai and Prentice (1995). One can also follow arguments in Lin and Ying (1995) to derive the information bound I_0^{-1} for estimating θ_0 through defining an appropriate parametric submodel and then construct the efficient estimating function $S_{\text{opt}}(\theta)$ using sample-splitting technique such that $-\frac{1}{n} \frac{\partial S_{\text{opt}}(\theta)}{\partial \theta} \rightarrow_p I_0$ and

$\frac{1}{\sqrt{n}}S_{\text{opt}}(\theta) \rightarrow_d N(0, I_0)$. The zero root of such $S_{\text{opt}}(\theta)$ is expected to achieve the information bound at \sqrt{n} rate. An alternative useful approach to accommodating the dependence of recurrent event times within the same subject is to incorporate a random effect or frailty η into model (1) as indicated in Sect. 4:

$$E\{dN^*(t)|W(t), \eta\} = \eta[g\{\gamma_0^T Z(t)\}dt + h\{\beta_0^T X(t)\}d\mu_0(t)], \quad (7)$$

where η follows a specified distribution with some unknown parameters, representing the intra-class correlation. Although there is some literature (e.g., Nielsen et al. 1992; Oakes 1992; Zeng and Lin 2007) on multiplicative intensity-based models with random effect of the form:

$$E\{dN^*(t)|\mathcal{F}_{t-}, \eta\} = \eta \exp\{\theta_0^T W(t)\}\lambda_0(t)dt,$$

where \mathcal{F}_t is the σ -field generated by $\{N^*(u), W(u+) : 0 \leq u \leq t\}$ with $W(t+) = \lim_{s \downarrow t} W(s)$, and $\lambda_0(t)$ is the baseline hazard function, methodology development on model (7) is lacking, although would be valuable, particularly in settings where the association between events within the same subject is of interest in addition to the additive and multiplicative covariate effects.

In the presence of a terminal event, such as death, which precludes further recurrent events, several methods have been proposed for analyzing such recurrent event data (e.g., Cook and Lawless 1997; Ghosh and Lin 2002; Huang and Wang 2004; Liu et al. 2004; Ye et al. 2007). As indicated in Lin et al. (2000) and Schaubel et al. (2006), our proposed model and method could potentially be modified to accommodate such case by substituting $D \wedge C$ for C in the above inference procedures, where D is the time to the terminal event, and model the conditional rate function as follows:

$$E\{dN^*(t)|W(t), D \geq t\} = g\{\gamma_0^T Z(t)\}dt + h\{\beta_0^T X(t)\}d\mu_0(t), \quad (8)$$

and we can show that the basic results in this paper still hold provided that independent censoring is redefined in the sense that

$$E\{dN^*(t)|W(t), D \wedge C \geq t\} = E\{dN^*(t)|W(t), D \geq t\}.$$

However, this approach pertains to the cause-specific rate function, which is analogous to the cause-specific hazard function (Kalbfleisch and Prentice 2002, p. 251). If the terminal event is independent of the recurrent event process, then the cause-specific rate function is the same as the marginal rate function for the recurrences. Otherwise, an alternative method is expected to jointly model the recurrent event process and terminal event process since these two processes are usually highly correlated in applications so that it is unrealistic to take them as separate matters and model them separately (e.g., Cook and Lawless 1997). It may be the combination of censoring and death that poses considerable challenges in statistical inference and would be worthwhile to develop methodology for such data.

In some applications, censoring may depend on the underlying recurrent event process even after conditioning on the covariates in the model (1) or (8). Under the

informative censoring together with multiplicative rates model without terminal event, Wang et al. (2001) described the correlation structure between the recurrent event process and dependent censoring time through latent variables and provided an estimator for cumulative rate function by treating the distributions of the censoring and latent variables as nuisance parameters; Ghosh and Lin (2003) proposed a joint model that formulates the marginal distributions of the recurrent event process and dependent censoring time through scale-change models; Miloslavsky et al. (2004) proposed to adjust for dependent censoring through the inverse probability of censoring weighting. Generalization of this approach beyond the multiplicative rates model would be useful.

Acknowledgements The authors thank the editor, the associate editor, and two anonymous referees for their valuable comments and suggestions. This work was partly supported by National Nature Science Fund of China grant 10771163 (Liu) and U.S. NIH grants R01 CA 79949 (Wu and Zhou) and R01 HL 57444 (Cai).

Appendix

To begin, we state two useful lemmas which are the bases for asymptotic approximation.

Lemma 1 *Under conditions C1 to C6, the processes $\frac{1}{\sqrt{n}}U(t, \theta_0)$ converges weakly to a zero-mean Gaussian process with covariance $\Sigma(s, t)$.*

Lemma 2 *Under conditions C1 to C6,*

$$\widehat{A}(\theta_n^+) \rightarrow_{a.s.} A, \quad \text{for any estimator } \theta_n^+ \rightarrow_{a.s.} \theta_0,$$

$$\text{where } \widehat{A}(\theta) = -\frac{1}{n} \frac{\partial U(\tau, \theta)}{\partial \theta}.$$

The proofs for Lemmas 1 and 2 follow from Lin et al. (2000) by employing various empirical process results and the detailed proofs omitted here are available from the authors.

Proof of Theorem 1

Consistency

Define $\|a\|$ as the max norm for a vector or matrix a . In view of condition C5, let $d = 1/(4\|A^{-1}\|)$ and $d_n = 1/(4\|\widehat{A}(\theta_0)^{-1}\|)$ whenever $\widehat{A}(\theta_0)$ is nonsingular. Select δ sufficiently small such that $\|\widehat{A}(\theta) - \widehat{A}(\theta_0)\| < d$ whenever $\|\theta - \theta_0\| < \delta$, for all n . Since d_n almost surely converges to d under Lemma 2, we can conclude that $\|\widehat{A}(\theta) - \widehat{A}(\theta_0)\| < 2d_n$ for large n , where n does not depend on θ , i.e., one can find a commonly large n for all θ under condition C6.

Write $O_\delta = \{\theta : \|\theta - \theta_0\| < \delta\}$ and it follows from the Inverse Function Theorem (Rudin 1964; Foutz 1977) that $\frac{1}{n}U(\tau, \cdot)$ is a one-to-one mapping from O_δ onto $\frac{1}{n}U(\tau, O_\delta)$ and the image set $\frac{1}{n}U(\tau, O_\delta)$ contains the open neighborhood $\frac{1}{n}U(\tau, \theta_0)$

with radius $d_n\delta$. Hence, when n is taken sufficiently large, image set $\frac{1}{n}U(\tau, O_\delta)$ contains the open neighborhood $\frac{1}{n}U(\tau, \theta_0)$ with radius $d\delta/2$. On the other hand, the convergence of $\frac{1}{n}U(\tau, \theta_0)$ to zero can be derived obviously from Lemma 1. Therefore, $\hat{\theta}$ exists and is unique in O_δ and $\hat{\theta}$ converges to θ_0 almost surely since δ can be taken arbitrarily small. In addition, the argument of Jacobsen (1989) can be used to show the global uniqueness of $\hat{\theta}$ for large n .

Asymptotic normality

Clearly, it follows from the Taylor expansion, Lemmas 1 and 2 that

$$\sqrt{n}(\hat{\theta} - \theta_0) = A^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \{Q_i(u, \theta_0) - \bar{q}(u, \theta_0)\} dM_i(u) + o_p(1).$$

Consequently, $\sqrt{n}(\hat{\theta} - \theta_0)$ behaves asymptotically as a scaled average of independent and identically distributed random vectors, which converges to a normal distribution with mean 0 and covariance $A^{-1}\Sigma(\tau, \tau)(A^{-1})^T$.

Proof of Theorem 2

Consistency

For simplicity, we introduce the following notation:

$$S_{g'}(t, \theta) = \frac{1}{n} \sum_{i=1}^n Y_i(t)g'\{\gamma^T Z_i(t)\}Z_i(t), \quad S_{h'}(t, \theta) = \frac{1}{n} \sum_{i=1}^n Y_i(t)h'\{\beta^T X_i(t)\}X_i(t),$$

$$S_h(t, \theta) = \frac{1}{n} \sum_{i=1}^n Y_i(t)h\{\beta^T X_i(t)\}, \quad \tilde{S}(t, \theta) = \begin{bmatrix} \int_0^t \frac{S_{g'}(u, \theta)}{S_h(u, \theta)} du \\ \int_0^t \frac{S_{h'}(u, \theta)}{S_h(u, \theta)} d\hat{\mu}_0(u, \theta) \end{bmatrix}.$$

Using the Uniform Strong Law of Large Numbers (Pollard 1990, p. 41), we can obtain that as $n \rightarrow \infty$, $\sup_{t \in [0, \tau]} \sup_{\theta \in \Theta} \|S_{g'}(t, \theta) - s_{g'}(t, \theta)\| \rightarrow_{\text{a.s.}} 0$, so are the convergences of $S_{h'}(t, \theta)$ and $S_h(t, \theta)$ to $s_{h'}(t, \theta)$ and $s_h(t, \theta)$, respectively.

We make a simple decomposition:

$$\{\hat{\mu}_0(t, \hat{\theta}) - \mu_0(t)\} = \{\hat{\mu}_0(t, \theta_0) - \mu_0(t)\} + \{\hat{\mu}_0(t, \hat{\theta}) - \hat{\mu}_0(t, \theta_0)\}$$

and then some simple calculations entail that

$$\widehat{\mu}_0(t, \theta_0) - \mu_0(t) = \int_0^t \frac{1}{S_h(u, \theta_0)} d\frac{1}{n} \sum_{i=1}^n M_i(u),$$

which converges almost surely to zero, uniformly in t .

On the other hand, by the Taylor expansion,

$$\{\widehat{\mu}_0(t, \widehat{\theta}) - \widehat{\mu}_0(t, \theta_0)\} = \left\{ \frac{\partial \widehat{\mu}_0(t, \theta_0)}{\partial \theta} \right\}^T (\widehat{\theta} - \theta_0) + o_p(\|\widehat{\theta} - \theta_0\|),$$

where

$$\frac{\partial \widehat{\mu}_0(t, \theta)}{\partial \theta} = - \left[\begin{array}{l} \int_0^t \frac{S_{g'}(u, \theta)}{S_h(u, \theta)} du \\ \int_0^t \frac{S_{h'}(u, \theta)}{S_h(u, \theta)} d\mu_0(u) + \int_0^t \frac{S_{h'}(u, \theta)}{[S_h(u, \theta)]^2} d\frac{1}{n} \sum_{i=1}^n M_i(u, \theta) \end{array} \right].$$

It can be shown that $\frac{\partial \widehat{\mu}_0(t, \theta_0)}{\partial \theta}$ converges almost surely to $-\widetilde{s}(t, \theta_0)$, uniformly in t . Since $\widehat{\theta} \rightarrow_{\text{a.s.}} \theta_0$, the convergence of $\widehat{\mu}_0(t, \widehat{\theta})$ to $\widehat{\mu}_0(t, \theta_0)$ is almost sure, uniformly in t . Hence, it is obvious that $\widehat{\mu}_0(t, \widehat{\theta})$ converges almost surely to $\mu_0(t)$, uniformly in t .

Weak convergence

With respect to the weak convergence of $\widehat{\mu}_0(t, \widehat{\theta})$, we display the following asymptotic approximation:

$$\begin{aligned} \sqrt{n}\{\widehat{\mu}_0(t, \widehat{\theta}) - \mu_0(t)\} &= \int_0^t \frac{1}{S_h(u, \theta_0)} d\frac{1}{\sqrt{n}} \sum_{i=1}^n M_i(u) \\ &\quad + \left\{ \frac{\partial \widehat{\mu}_0(t, \theta_0)}{\partial \theta} \right\}^T \sqrt{n}(\widehat{\theta} - \theta_0) + o_p(\sqrt{n} \|\widehat{\theta} - \theta_0\|) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_i(t) + o_p(1), \end{aligned}$$

where $o_p(1)$ is uniformly in t . Thus, the convergence of $\sqrt{n}\{\widehat{\mu}_0(t, \widehat{\theta}) - \mu_0(t)\}$ in finite dimensional distributions follows from the multivariate central limit theorem. Tightness can be shown by demonstrating the manageability of the components of $\Psi_i(t)$. Hence, we prove that $\sqrt{n}\{\widehat{\mu}_0(t, \widehat{\theta}) - \mu_0(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function $\xi(s, t)$. Then the required weak convergence of $\sqrt{n}\{\widehat{\mu}_0(t, \widehat{\theta}) - \mu_0(t)\}$ follows directly. Furthermore, the consistency of the

covariance function estimator follows from the consistency of $\hat{\theta}$ and $\hat{\mu}_0(t, \hat{\theta})$, and repeated applications of the Uniform SLLN.

References

- Aalen OO (1980) A model for nonparametric regression analysis of counting processes. In: Klonecki N, Kosek A, Rosinski J (eds) *Lecture notes in statistics*, vol 2. Springer, New York, pp 1–25
- Aalen OO (1989) A linear regression model for the analysis of the life times. *Stat Med* 8:907–925
- Andersen PK, Gill RD (1982) Cox's regression model for counting processes: a large sample study. *Ann Stat* 10:1100–1120
- Andersen PK, Borgan O, Gill RD, Keiding N (1993) *Statistical models based on counting processes*. Springer-Verlag, New York
- Biliyas Y, Gu M, Ying Z (1997) Towards a general asymptotic theory for Cox model with staggered entry. *Ann Stat* 25:662–682
- Breslow NE, Day NE (1980) *Statistical methods in cancer research 1. The design and analysis of case-control studies*. IARC, Lyon
- Breslow NE, Day NE (1987) *Statistical methods in cancer research 2. The design and analysis of cohort studies*. IARC, Lyon
- Buckley JD (1984) Additive and multiplicative models for relative survival rates. *Biometrics* 40:51–62
- Cai J, Prentice RL (1995) Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* 82:151–164
- Cai J, Schaubel DE (2004) Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Anal* 10:121–138
- Chang SH, Wang MC (1999) Conditional regression analysis for recurrence time data. *J Am Stat Assoc* 94:1221–1230
- Cook RJ, Lawless JF (1997) Marginal analysis of recurrent events and a terminating event. *Stat Med* 16:911–924
- Cox DR, Oakes D (1984) *Analysis of survival data*. Chapman and Hall, London
- Foutz RV (1977) On the unique consistent solution to the likelihood equations. *J Am Stat Assoc* 72:147–148
- Ghosh D, Lin DY (2002) Marginal regression methods for recurrent and terminal events. *Stat Sin* 12:663–688
- Ghosh D, Lin DY (2003) Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* 59:877–885
- Huang CY, Wang MC (2004) Joint modeling and estimation of recurrent event processes and failure time data. *J Am Stat Assoc* 99:1153–1165
- Huffer FW, McKeague IW (1991) Weighted least squares estimation for Aalen's additive risk model. *J Am Stat Assoc* 86:114–129
- Jacobsen M (1989) Existence and unicity of MLEs in discrete exponential family distributions. *Scand J Stat* 16:335–349
- Kalbfleisch JD, Prentice RL (2002) *The statistical analysis of failure time data*. Wiley, New York
- Lawless JF, Nadeau C (1995) Some simple robust methods for the analysis of recurrent events. *Technometrics* 37:158–168
- Lee EW, Wei LJ, Amato DA (1992) Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein JP, Goel PK (eds) *Survival analysis: state of the art*. Kluwer Academic, Dordrecht, pp 237–247
- Li Q, Lagakos S (1997) Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Stat Med* 16:925–940
- Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22
- Lin DY, Wei LJ, Yang I, Ying Z (2000) Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc B* 62:711–730
- Lin DY, Wei LJ, Ying Z (2001) Semiparametric transformation models for point processes. *J Am Stat Assoc* 96:620–628
- Lin DY, Ying Z (1994) Semiparametric analysis of the additive risk model. *Biometrika* 81:61–71
- Lin DY, Ying Z (1995) Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Ann Stat* 23:1712–1734

- Liu L, Wolfe RA, Huang X (2004) Shared frailty models for recurrent events and a terminal event. *Biometrics* 60:747–756
- Miloslavsky M, Keles S, van der Laan MJ, Butler S (2004) Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *J R Stat Soc B* 66:239–257
- Nielsen GG, Gill RD, Andersen PK, Sorensen TIA (1992) A counting process approach to maximum likelihood estimation in frailty models. *Scand J Stat* 19:25–44
- Oakes D (1992) Frailty models for multiple event times. In: Klein JP, Goel PK (eds) *Survival analysis: state of the art*. Kluwer Academic, Dordrecht, pp 371–379
- Pepe MS, Cai J (1993) Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *J Am Stat Assoc* 88:811–820
- Pollard D (1990) *Empirical processes: theory and applications*. Institute of Mathematical Statistics, Hayward
- Prentice RL, Williams BJ, Peterson AV (1981) On the regression analysis of multivariate failure time data. *Biometrika* 68:373–379
- Ross SM (2006) *Simulation*, 4th edn. Academic Press, New York
- Rudin W (1964) *Principles of mathematical analysis*. McGraw-Hill, New York
- Schaubel DE, Zeng D, Cai J (2006) A semiparametric additive rates model for recurrent event data. *Lifetime Data Anal* 12:389–406
- Therneau TM, Grambsch PM (2000) *Modeling survival data: extending the Cox model*. Springer, New York
- Wang MC, Qin J, Chiang CT (2001) Analyzing recurrent event data with informative censoring. *J Am Stat Assoc* 96:1057–1065
- Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 84:1065–1073
- Ye Y, Kalbfleisch JD, Schaubel DE (2007) Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 63:78–87
- Zeng D, Lin DY (2007) Semiparametric transformation models with random effects for recurrent events. *J Am Stat Assoc* 102:167–180