Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring

Xingqiu Zhao*[,1], **Li Liu**[2], **Yanyan Liu**[2], and **Wei Xu**[3]

[1] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong
[2] School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, China
[3] Dalla Lana School of Public Health, University of Toronto, Toronto, ON, M5G 2M9, Canada

Received 16 September 2011; revised 26 May 2012; accepted 8 June 2012

Multivariate recurrent event data are usually encountered in many clinical and longitudinal studies in which each study subject may experience multiple recurrent events. For the analysis of such data, most existing approaches have been proposed under the assumption that the censoring times are noninformative, which may not be true especially when the observation of recurrent events is terminated by a failure event. In this article, we consider regression analysis of multivariate recurrent event data with both time-dependent and time-independent covariates where the censoring times and the recurrent event process are allowed to be correlated via a frailty. The proposed joint model is flexible where both the distributions of censoring and frailty variables are left unspecified. We propose a pairwise pseudolikelihood approach and an estimating equation-based approach for estimating coefficients of time-dependent and time-independent covariates, respectively. The large sample properties of the proposed estimates are established, while the finite-sample properties are demonstrated by simulation studies. The proposed methods are applied to the analysis of a set of bivariate recurrent event data from a study of platelet transfusion reactions.

Keywords: Frailty; Informative censoring; Marginal model; Multivariate recurrent event data; Pairwise pseudolikelihood.

Supporting Information for this article is available from the author or on the WWW under http://dx.doi.org/10.1002/bimj.201100194

1 Introduction

Recurrent event data arise in epidemiologic studies, reliability experiments, and longitudinal studies in which the event of interest may occur repeatedly. Examples include repeated hospitalizations, repeated tumor metastases, and repeated acute myocardial infarctions. The most important feature of recurrent event data is that the recurrent event times within a subject are ordered and correlated. Hence, recurrent event data can be viewed as an ordered multivariate failure time data. Many authors have discussed the analysis of recurrent event data (e.g., Andersen and Gill, 1982; Pepe and Cai, 1993; Lin et al., 2000). Cook and Lawless (2007) also provided a detailed review for the analysis of recurrent events.

A subject may experience several types of correlated recurrent events, resulting in multivariate recurrent event data. For example, hematology/oncology patients may experience different febrile nonhemolytic transfusion reactions (FNHTRs), pulmonary exacerbations may be differentiated by severity, transient ischemic attacks may be classified according to location in cardiovascular trials, and infections in bone marrow transplantation may be subtyped as of bacterial, fungal, and viral origin.

*Corresponding author: e-mail: xingqiu.zhao@polyu.edu.hk, Phone: +852-2766-6921, Fax: +852-2362-9045

To analyze the multivariate recurrent event data, some inference procedures have been developed, for example, Abu-Libdeh et al. (1990), Spiekerman and Lin (1998), Cai and Schaubel (2004), Schaubel and Cai (2005), and Sun et al. (2009). Many of these methods are based on the marginal model in which the information of the correlation structure of the recurrent events is not incorporated. An alternative approach is based on the conditional model (e.g., the frailty model). However, the existing methods usually require dependence structures specified in conditional models. For example, Ye et al. (2007) studied joint semiparametric models and specified frailty as a gamma distribution. This parameterizing assumption seems overrestricted because it is hard to check distributional assumptions about unobservable frailty random variables. In this article, we will use the frailty model while leaving distributions of frailty variables to be arbitrary.

Independent censoring is another stringent assumption in the aforementioned studies. In many applications, however, censoring can be informative about the recurrent event process, such as death, and it is not sensible to assume independence between the censoring mechanism and the recurrent event process. To resolve the informative censoring problem, joint marginal models and shared-frailty models have been developed for the univariate recurrent time data in the literature. Lancaster and Intrator (1998) presented an inference procedure through parametric modeling of recurrent event and survival data, where the dependency between the two outcomes was induced by sharing an unobserved frailty in the intensity model of the recurrent event process and the hazard model of the failure event. Rondeau et al. (2007) proposed joint semiparametric models for recurrent events and terminal events with a specified frailty distribution by using maximum penalized likelihood estimation. Wang et al. (2001) and Huang and Wang (2004) relaxed the stringent distributional assumption on the frailty in Rondeau et al. (2007) by extending the work of Lancaster and Intrator (1998) to a regression model where covariates were assumed to be time-independent and the distributions of both the censoring and frailty variables were treated as nuisance parameters. Zhu et al. (2010) extended the estimation procedure in Wang et al. (2001) to multivariate recurrent event data with time-invariant covariates. In practice, covariate information is often collected longitudinally in many studies. In other words, some covariates may vary with time. For example, in the FNHTRs studies the platelet product donor type may be collected at several time points and moreover some treatments such as premedication may change over time. In this article, we propose an extension of the model in Huang et al. (2010) to handle multivariate recurrent data with time-dependent covariates and informative censoring. For inference, we develop a pairwise likelihood method for estimating the coefficient parameters of time-dependent covariates and construct an estimating equation to estimate the coefficient parameters of time-independent covariates.

The remainder of the article is organized as follows. Section 2 describes a marginal subject-specific nonstationary Poisson process model taking the informative censoring into consideration via an unobservable frailty. Section 3 presents an estimation procedure and the asymptotic properties of the resulting estimators. Results from a series of simulation studies conducted for evaluating the finite-sample properties of the proposed estimates are reported in Section 4. Section 5 illustrates the proposed methodology by using a set of bivariate recurrent event data from a platelet transfusion reaction study. Some concluding remarks are made in Section 6.

2 Joint modeling of multivariate recurrent events

Consider n independent subjects who experience J different types of recurrent events in observation time period $[0, \tau]$. For subject i, let $N_{ij}^*(t)$ be the number of events of type j that occur over the interval $[0, t]$, $0 \le t \le \tau$, $i = 1, \ldots, n$, $j = 1, \ldots, J$. Moreover, suppose that there exist two types of covariates $X_{ij}(\cdot)$ and Z_{ij}, where $X_{ij}(\cdot)$ is a bounded p-dimensional vector of time-dependent covariates and Z_{ij} is a q-dimensional vector of time-independent covariates. Denote by $\mathcal{X}_{ij}(t) = \{X_{ij}(u), 0 \le u \le t\}$ the covariate history of X_{ij} up to time t. In this article, the covariate histories $\mathcal{X}_{ij}$'s are assumed to be left-continuous and observed. Let C_{ij} be the censoring time for subject i with recurrent event type j.

Define $N_{ij}(t) = N^*_{ij}(t \wedge C_{ij})$ and $\Delta_{ijk} = I(T_{ijk} \le C_{ij})$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ denotes the indicator function. Denote by m_{ij} the number of recurrent events that occurred before time C_{ij}, and $T_{ij1} < \cdots < T_{ijm_{ij}}$ the j-th type recurrent event time points. Thus, the observed data for the i-th subject consist of $D_i = \{D_{ij}, j = 1, \ldots, J\}$ with $D_{ij} = \{C_{ij}, \mathcal{X}_{ij}(C_{ij}), Z_{ij}, m_{ij}, (T_{ij1}, \ldots, T_{ijm_{ij}})\}$. Assume that $\{(C_{ij}, X_{ij}, Z_{ij}, N_{ij}), i = 1, \ldots, n\}$ is a random sample of (C_j, X_j, Z_j, N_j).

Let β and γ be vectors of unknown parameters, λ_{0j} an unspecified continuous baseline intensity function with $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(u)du$, and ξ_{i0} a nonnegative valued subject-specific frailty variable. Conditioning on $(\xi_{i0}, \mathcal{X}_{ij}, Z_{ij})$, the occurrence of recurrent events is modeled by a nonhomogeneous Poisson process with intensity function

$$\lambda_{ij}(t) = \xi_{i0}\lambda_{0j}(t)\exp\{X_{ij}(t)'\beta + Z'_{ij}\gamma\}, \quad t \in [0, \tau]. \tag{1}$$

For an identifiability reason, we assume $E\{\xi_{i0}|Z_{ij}, \mathcal{X}_{ij}(\tau)\} = 1$. Here, the subject-specific frailty variable ξ_{i0} characterizes the correlation of the within-subject multivariate recurrent events.

Because of the memoryless property of a Poisson process, conditional on ξ_{i0}, the rate function equals the intensity function of the recurrent event process. Thus, the marginal rate function of event occurrence is given by $\lambda_{0j}(t)\exp\{X_{ij}(t)'\beta + Z'_{ij}\gamma\}$, which has been discussed by Schaubel and Cai (2005) and Lin et al. (2000) where the censoring time was assumed to be independent of the recurrent event process. In this article, we assume that conditional on $(\xi_{i0}, \mathcal{X}_{ij}, Z_{ij})$, C_{ij} is independent of $N^*_{ij}(\cdot)$. The unobservable frailty ξ_{i0} inflates/deflates the intensity and this assumption allows the censoring time C_{ij} to depend on the frailty ξ_{i0}, which substantially relaxes the usual requirement of independent censoring models. Note that the censoring time can handle a composite censoring event, that is, $C_{ij} = \min(C^{(0)}_{ij}, C^{(1)}_{ij})$, where $C^{(0)}_{ij}$ represents a noninformative censoring time independent of $N^*_{ij}(t)$ and $C^{(1)}_{ij}$ represents an informative censoring time correlated with $N^*_{ij}(t)$. Furthermore, we assume that the regression coefficients are the same for different types of event without loss of generality. In fact, the methodology developed below still applies by defining a new vector of covariates if they are type-specific. For example, for different β_j and γ_j, $X_{ij}(t)'\beta_j + Z'_{ij}\gamma_j$ can be rewritten as $X^*_{ij}(t)'\beta^* + Z^{*\prime}_{ij}\gamma^*$, where $X^*_{ij}(t) = (0, \ldots, X_{ij}(t)', \ldots, 0)'$, $Z^*_{ij} = (0, \ldots, Z'_{ij}, \ldots, 0)'$, $\beta^* = (\beta'_1, \ldots, \beta'_J)'$, and $\gamma^* = (\gamma'_1, \ldots, \gamma'_J)'$.

3 Estimation procedure

We first estimate the regression parameter β. Conditional on $(C_{ij}, \mathcal{X}(C_{ij}), Z_{ij}, m_{ij}, \xi_{i0})$, the pseudo-likelihood of $(T_{ij1}, \ldots, T_{ijm_{ij}})$ is proportional to

$$\prod_{i=1}^{n}\prod_{j=1}^{J}\prod_{k=1}^{m_{ij}} \frac{\lambda_{0j}(T_{ijk})\exp\{X_{ij}(T_{ijk})'\beta\}}{\int_0^{C_{ij}} \lambda_{0j}(u)\exp\{X_{ij}(u)'\beta\}du} = \prod_{i=1}^{n}\prod_{j=1}^{J}\prod_{k=1}^{m_{ij}} \frac{p_{ij}(T_{ijk})}{P_{ij}(C_{ij})}, \tag{2}$$

where

$$p_{ij}(t) = \frac{\lambda_{0j}(t)\exp\{X_{ij}(t)'\beta\}}{\int_0^{\tau} \lambda_{0j}(u)\exp\{X_{ij}(u)'\beta\}du} I(0 \le t \le \tau)$$

and

$$P_{ij}(t) = \int_0^t p_{ij}(s)ds.$$

Since $X_{ij}(\cdot)$ is time-dependent, the integral in the denominator of the conditional likelihood does not have a closed form unless λ_{0j} is specified. It is challenging to maximize the conditional likelihood function. A natural idea is to eliminate the nonparametric component λ_{0j} in the likelihood.

Let $\delta_{ijk,ljs} = 1$ if $\max\{T_{ijk}, T_{ljs}\} \le C_{ij} \wedge C_{lj}$, and 0 otherwise. When $\delta_{ijk,ljs} = 1$, T_{ijk} and T_{ljs} are called a comparable pair for risk type j. By following the idea of Huang et al. (2010), the pairwise pseudolikelihood of (T_{ijk}, T_{ljs}), $i < l$, given the order statistics of (T_{ijk}, T_{ljs}) and $\delta_{ijk,ljs} = 1$, can be expressed as

$$\frac{p_{ij}(T_{ijk})p_{lj}(T_{ljs})}{p_{ij}(T_{ijk})p_{lj}(T_{ljs}) + p_{ij}(T_{ljs})p_{lj}(T_{ijk})} = \frac{1}{1+\exp\{\rho_{ij,lj}(T_{ijk}, T_{ljs})'\beta\}},$$

where

$$\rho_{ij,lj}(t,u) = X_{ij}(u) + X_{lj}(t) - X_{ij}(t) - X_{lj}(u).$$

Note that as nuisance parameters, both frailty ξ_{i0} and baseline intensity function $\lambda_{0j}(t)$ are eliminated from the likelihood. Hence, β can be estimated by maximizing the pairwise pseudolikelihood

$$\prod_{i<l}\prod_{j=1}^{J}\prod_{k=1}^{m_{ij}}\prod_{s=1}^{m_{lj}}\left(\frac{1}{1+\exp\{\rho_{ij,lj}(T_{ijk}, T_{ljs})'\beta\}}\right)^{\delta_{ijk,ljs}}.$$

The score function is given by

$$l(\beta) = \frac{1}{\binom{n}{2}}\sum_{i<l} H(D_i, D_l; \beta),$$

where

$$H(D_i, D_l; \beta) = \sum_{j=1}^{J}\int_0^{C_{ij}\wedge C_{lj}}\int_0^{C_{ij}\wedge C_{lj}} h_{ij,lj}(t,u;\beta)dN_{ij}(t)dN_{lj}(u)$$

and

$$h_{ij,lj}(t,u;\beta) = \frac{-\exp\{\rho_{ij,lj}(t,u)'\beta\}}{1+\exp\{\rho_{ij,lj}(t,u)'\beta\}}\rho_{ij,lj}(t,u).$$

Let $\hat{\beta}$ be the solution of $l(\beta) = 0$ and β_0 be the true value of β. Under some regularity conditions, $\hat{\beta}$ is a consistent estimator of β_0 and $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean 0 and a covariance matrix that can be consistently estimated by $\hat{V}$ as given in Theorem 1 of Appendix A1.

To estimate $\Lambda_{0j}(t)$, we first consider estimation of $F_{0j}(t) = \Lambda_{0j}(t)/\Lambda_{0j}(\tau)$, which is a proper cumulative distribution function. Note that for given β, maximizing the pseudolikelihood in (2) is equivalent to maximizing

$$L_j(\lambda_{0j}|\beta) = \prod_{i=1}^{n}\prod_{k=1}^{m_{ij}}\frac{\lambda_{0j}(T_{ijk})\exp\{X_{ij}(T_{ijk})'\beta\}}{\int_0^{C_{ij}}\lambda_{0j}(u)\exp\{X_{ij}(u)'\beta\}du} = \prod_{i=1}^{n}\prod_{k=1}^{m_{ij}}\frac{p_{ij}(T_{ijk})}{P_{ij}(C_{ij})},\ \ j = 1,\ldots,J.$$

Thus, following Huang et al. (2010), the L_j can be viewed as the likelihood of a biased sample from the distribution function $F_{0j}(t)$, where the observations are sampled with sampling weights proportional to $\exp\{X_{ij}(t)'\beta\}$ and are right truncated by C_{ij}. Thus, the estimator $\hat{F}_{0j}(t)$ of $F_{0j}(t)$ can be obtained by modifying the truncation product-limit estimator (Wang et al., 1986) and using the inverse probability weighting technique, where $\hat{F}_{0j}(t)$ and the asymptotic results are given in Appendix A2, and the proofs can be found in Supplementary Information on the journal's website (`http://www.biometrical-journal.com`).

For estimation of $\Lambda_{0j}(\tau)$ and γ, we can use an estimating equation approach. The estimating equation-based estimators $\hat{\Lambda}_{0j}(\tau)$ and $\hat{\gamma}$ as well as an alternative estimator $\tilde{\gamma}$ are given in Appendix A2. The asymptotic results are also given in this Appendix with proofs available in Supplementary Information on the journal's website (`http://www.biometrical-journal.com`).

4 Simulation study

In this section, the finite-sample properties of the proposed estimators are evaluated through simulation studies. In the study, we considered the situation where there are $J = 2$ types of recurrent events of interest and two covariates with $p = q = 1$. The time-independent covariate, Z_{ij}, was generated from Bernoulli distribution with success probability 0.5, and the time-dependent covariate $X_{ij}(t)$ takes the form $X_{ij}\log(t)$, where X_{ij} has a uniform $[0, 1]$ distribution. For subject i, we generated a frailty, ξ_{i0}, inducing positive correlation among the within-subject event times, from a gamma distribution with unit mean and variance σ^2. We took $\sigma^2 = 1, 0.7, 0.5, 0.3$ and $\lambda_{01}(t) = 1/2, \lambda_{02}(t) = 1/4$. We generated recurrent event times from model (1), where the subject's underlying recurrent event process is a nonhomogeneous Poisson process with intensity function $\xi_{i0}\lambda_{0j}(t)\exp\{X_{ij}(t)'\beta_0 + Z_{ij}'\gamma_0\}$, $i = 1, \ldots, n$, $j = 1, 2$. The censoring times were generated as $C_{ij} = \min(C_{ij}^{(0)}, C_{ij}^{(1)})$, where $C_{ij}^{(0)} = 10$ is the time of end of the study and $C_{ij}^{(1)}$ is from an exponential distribution with mean 10 for subjects in the treatment arm ($Z_{ij} = 1$) or an exponential distribution with mean $6\xi_{i0} + 4$ for subjects in the control arm ($Z_{ij} = 0$). Thus, the censoring time is correlated with $N_{ij}(\cdot)$ through $(\xi_{i0}, \mathcal{X}_{ij}, Z_{ij})$. Set $(\beta_0, \gamma_0) = (0, 0.5), (0.2, 0.5), (0.2, -0.3)$, representing the different effects of the covariates X_{ij} and Z_{ij} on the recurrent events. For each setting, we considered two sample sizes, $n = 100$ and 200, respectively. All the results reported here are based on 1000 Monte Carlo replications using R software.

Table 1 presents the simulation results on estimation of β and γ for different situations. It can be seen from the estimated bias given by the average of proposed estimates minus the true value that $\hat{\beta}$ and $\hat{\gamma}$ are approximately unbiased for all the data configurations considered here. To evaluate the validity of the estimators, the 95% bootstrap confidence intervals for $\hat{\gamma}$ and $\tilde{\gamma}$ were produced based on 200 bootstrap replications, while the 95% confidence intervals for $\hat{\beta}$ were computed according to Theorem 1. The average of estimated standard error, denoted by ESE, closely approximates the sample standard error, SSE, with empirical coverage probabilities (CP) close to the nominal value, 0.95. These results indicate that the proposed variance estimation procedure provides reasonable estimates and the normal approximation seems to be appropriate. For $n = 100$, slight undercoverage occurs, which reduced progressively with increasing sample size. Furthermore, the efficiency gains for $\tilde{\gamma}$ over $\hat{\gamma}$ are reduced by decreasing standard deviations of frailty variables. This reduction is reasonable because the correlation between censoring times and frailty variables and the correlation among the within-subject event times become weak when the standard deviations of frailty variables decrease. In the simulation, we investigated the effect of the size of bootstrap samples on variance estimation and the size of 200 used here seems reasonable.

In practice, a natural question that one may ask is if one could simply apply the inference procedures developed under the independent censoring assumption to informative censoring situations. To answer this, we also investigated the sample bias and SSE of the estimates proposed by Schaubel and Cai (2005)

Table 1 Simulation results for proposed estimates $\hat{\beta}$, $\hat{\gamma}$, and $\tilde{\gamma}$.

n	β_0	γ_0	σ	$\hat{\beta}$				$\hat{\gamma}$				$\tilde{\gamma}$				
				Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	Ref
100	0	0.5	1	0.000	0.156	0.134	0.903	−0.030	0.200	0.189	0.931	−0.025	0.170	0.150	0.920	1.59
			0.7	−0.003	0.152	0.138	0.926	−0.024	0.169	0.158	0.939	−0.021	0.155	0.139	0.928	1.30
			0.5	−0.004	0.157	0.140	0.925	−0.018	0.150	0.139	0.934	−0.017	0.144	0.131	0.930	1.12
			0.3	0.005	0.153	0.141	0.927	−0.002	0.137	0.126	0.941	−0.003	0.138	0.128	0.940	0.96
	0.2	0.5	1	−0.002	0.165	0.142	0.913	−0.033	0.208	0.190	0.928	−0.029	0.179	0.153	0.919	1.54
			0.7	−0.001	0.157	0.143	0.918	−0.026	0.168	0.159	0.935	−0.023	0.158	0.141	0.927	1.28
			0.5	0.000	0.166	0.146	0.919	−0.021	0.159	0.141	0.926	−0.021	0.157	0.134	0.923	1.10
			0.3	0.006	0.158	0.147	0.934	−0.003	0.139	0.128	0.939	−0.003	0.142	0.130	0.936	0.98
	0.2	−0.3	1	−0.001	0.201	0.171	0.896	−0.021	0.225	0.206	0.935	−0.018	0.201	0.173	0.919	1.41
			0.7	0.001	0.197	0.177	0.922	−0.022	0.187	0.175	0.934	−0.022	0.178	0.160	0.928	1.19
			0.5	0.001	0.204	0.179	0.920	−0.016	0.171	0.158	0.932	−0.019	0.171	0.154	0.933	1.06
			0.3	0.008	0.201	0.180	0.919	−0.003	0.156	0.146	0.929	−0.005	0.156	0.150	0.945	0.94
200	0	0.5	1	−0.002	0.107	0.099	0.926	−0.029	0.147	0.136	0.928	−0.032	0.123	0.108	0.919	1.59
			0.7	0.007	0.107	0.100	0.939	−0.023	0.124	0.116	0.941	−0.024	0.117	0.102	0.936	1.30
			0.5	0.000	0.106	0.102	0.942	−0.016	0.106	0.102	0.942	−0.013	0.105	0.097	0.945	1.10
			0.3	−0.002	0.113	0.102	0.922	0.001	0.105	0.093	0.934	0.000	0.110	0.096	0.932	0.94
	0.2	0.5	1	−0.010	0.114	0.106	0.922	−0.035	0.152	0.138	0.920	−0.038	0.128	0.110	0.913	1.56
			0.7	0.007	0.113	0.104	0.928	−0.024	0.125	0.117	0.929	−0.025	0.120	0.104	0.927	1.28
			0.5	0.002	0.109	0.106	0.942	−0.018	0.113	0.104	0.940	−0.015	0.111	0.099	0.939	1.10
			0.3	−0.001	0.117	0.106	0.932	0.001	0.106	0.094	0.929	−0.002	0.112	0.097	0.929	0.94
	0.2	−0.3	1	−0.004	0.136	0.125	0.919	−0.026	0.164	0.149	0.920	−0.030	0.146	0.126	0.905	1.39
			0.7	0.008	0.133	0.128	0.941	−0.022	0.138	0.129	0.937	−0.025	0.130	0.118	0.937	1.19
			0.5	0.007	0.138	0.131	0.928	−0.014	0.130	0.119	0.927	−0.014	0.132	0.116	0.921	1.06
			0.3	0.003	0.141	0.131	0.934	−0.004	0.121	0.108	0.930	−0.005	0.129	0.112	0.919	0.92

SSE represents the sample standard error of estimates, ESE represents the mean of the estimated standard errors, CP represents the empirical 95% coverage probability, and Ref represents the relative efficiency of $\tilde{\gamma}$ vs. $\hat{\gamma}$.

Table 2 Simulation results based on Schaubel and Cai's method and the proposed method with $n = 100$, where censoring time $C_{ij} = \min(10, C_{ij}^{1})$ with $C_{ij}^{1} \sim E(1/10)$ if $Z_{ij} = 1$, $C_{ij}^{1}|(\xi_{i0}, \mathcal{X}_{ij}, Z_{ij}) \sim E(1/(6\xi_{i0} + 4))$ if $Z_{ij} = 0$.

β_0	γ_0	σ	$\bar{\beta}$		$\bar{\gamma}$		$\hat{\beta}$		$\hat{\gamma}$	
			Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
0	0.5	1	−0.001	0.179	−0.193	0.189	0.000	0.156	−0.030	0.200
		0.7	−0.003	0.140	−0.106	0.145	−0.003	0.152	−0.024	0.169
		0.5	0.004	0.123	−0.057	0.120	−0.004	0.157	−0.018	0.150
		0.3	0.005	0.103	−0.022	0.094	0.005	0.153	−0.002	0.137
0.2	0.5	1	−0.005	0.182	−0.207	0.187	−0.002	0.165	−0.033	0.208
		0.7	−0.004	0.143	−0.113	0.144	−0.001	0.157	−0.026	0.168
		0.5	0.006	0.126	−0.061	0.118	0.000	0.166	−0.021	0.159
		0.3	0.003	0.100	−0.023	0.091	0.006	0.158	−0.003	0.139
0.2	−0.3	1	−0.008	0.198	−0.199	0.200	−0.001	0.201	−0.021	0.225
		0.7	−0.002	0.157	−0.113	0.154	0.001	0.197	−0.022	0.187
		0.5	0.006	0.138	−0.060	0.129	0.001	0.204	−0.016	0.171
		0.3	0.003	0.118	−0.024	0.106	0.008	0.201	−0.003	0.156

$\bar{\beta}$ and $\bar{\gamma}$ represent the estimates of regression parameters β and γ in Schaubel and Cai (2005). SSE represents the sample standard error of estimates.

based on 1000 samples of the same simulated data as those considered in Table 1. Table 2 shows that $\bar{\gamma}$, the estimate obtained by using Schaubel and Cai's method, seems biased, which results from the restriction of independent censoring required in Schaubel and Cai (2005). In fact, higher frailty values tend to have a longer observation period and the corresponding risk sets are more likely to consist of sicker subjects at later time points. As a result, Schaubel and Cai's method underestimates the difference between treatment group and control group under the specified conditions in our simulations. Table 2 shows that our proposed method provides reasonable estimates. It deserves to note that $\bar{\beta}$ seems unbiased in Table 2, which may be caused by the data setting that C_{ij} is independent of the covariate $X_{ij}(t)$. Table 3 presents the estimated bias and SSE of the estimates based on 1000 samples of the same data configurations as those considered in Table 1 except that censoring times were generated as $C_{ij} = \min(C_{ij}^{(0)}, C_{ij}^{(1)})$, where $C_{ij}^{(0)} = 10$ and $C_{ij}^{(1)}$ follows an exponential distribution with mean $6\xi_{i0} + 4$ for subjects satisfying $Z_{ij} = 1$ or $X_{ij} \leq 0.5$, and an exponential distribution with mean 10 otherwise. Table 3 also shows that under the informative censoring mechanism $\bar{\beta}$ and $\bar{\gamma}$, the estimates by using Schaubel and Cai's method have larger absolute biases than our estimates, while their SSE is slightly smaller than ours. These simulation results indicate that the proposed methods perform well in all the situations considered here.

In addition, we conducted the simulation studies under the same setups as those in Tables 2 and 3 for the sample size of 200 and obtained similar results as shown in Supplementary Information on the journal's website (`http://www.biometrical-journal.com`).

5 Example: a platelet transfusion reaction study

We apply the proposed method to a set of bivariate recurrent event data on FNHTRs characterized by fever, chills, rigors, hives, and others arising within 4–6 h of transfusion. The data were collected among hematology/oncology patients at five university teaching hospitals in Toronto coded A–E over three

Table 3 Simulation results based on Schaubel and Cai's method and the proposed method with $n = 100$, where censoring time $C_{ij} = \min(10, C_{ij}^1)$ with $C_{ij}^1|(\xi_{i0}, \mathcal{X}_{ij}, Z_{ij}) \sim E(1/(6\xi_{i0} + 4))$ if $Z_{ij} = 1$ or $X_{ij} \leq 0.5$, $C_{ij}^1 \sim E(1/10)$ otherwise.

β_0	γ_0	σ	$\bar{\beta}$		$\bar{\gamma}$		$\hat{\beta}$		$\hat{\gamma}$	
			Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
0	0.5	1	−0.077	0.177	0.082	0.183	−0.004	0.156	0.005	0.196
		0.7	−0.046	0.143	0.047	0.144	−0.003	0.155	0.005	0.167
		0.5	−0.019	0.120	0.031	0.119	−0.002	0.151	0.003	0.149
		0.3	−0.008	0.100	0.013	0.100	0.001	0.152	0.007	0.138
0.2	0.5	1	−0.087	0.180	0.090	0.181	−0.009	0.164	0.002	0.199
		0.7	−0.048	0.143	0.054	0.143	−0.001	0.159	0.004	0.167
		0.5	−0.020	0.122	0.034	0.116	0.001	0.156	0.001	0.154
		0.3	−0.009	0.097	0.013	0.096	0.002	0.157	0.004	0.143
0.2	−0.3	1	−0.122	0.196	0.099	0.191	0.007	0.194	0.017	0.221
		0.7	−0.070	0.154	0.053	0.153	0.019	0.191	−0.004	0.175
		0.5	−0.033	0.135	0.033	0.124	−0.004	0.204	0.011	0.175
		0.3	−0.008	0.121	0.013	0.107	−0.003	0.188	−0.005	0.167

$\bar{\beta}$ and $\bar{\gamma}$ represent the estimates of regression parameters β and γ in Schaubel and Cai (2005). SSE represents the sample standard error of estimates.

consecutive summers from 1996 to 1998, which were conducted by Patterson et al. (2000). During the first summer, an overview regarding platelet transfusion practices, premedication, and reaction rate was obtained. In the second summer, platelet transfusion premedications were standardized and the effect on reaction rate analyzed. The third summer addressed the effect of prestorage platelet leukoreduction on the reaction rate. The occurrence of FNHTRs is temporary and it is natural to treat a reaction as a recurrent event. We will focus on two types of causes for the occurrence of a reaction in the following analysis: Type I = "fever" and Type II = "all other reactions", where fever is defined as the presence of a temperature increase of $\geq 1^{\circ}$C within 6 h post transfusion.

The data available here are restricted to those collected during the 1997 summer where eligible patients include all hematology/oncology patients who experienced at least one FNHTR reaction. There are 254 patients with a total of 1395 transfusions. The missing values of the FNHTR data during the 1997 summer concentrate on five types of transfusion reactions defined based on a temperature increase ≥ 1 (fever), chills, rigors, hives, and other symptoms. Among 1395 transfusions, the missing frequencies of fever, chill, rigors, hives, and other symptoms are 195, 65, 65, 65, and 66, respectively. These missing values are imputed by carrying forward the last observed value. The covariates of interest in our analysis include platelet product donor type ($X_{ij}^{(1)}(t) = 1$ if random donor platelet at time t, 0 if single donor platelet or HLA-matched platelet), premedication ($X_{ij}^{(2)}(t) = 0$ if yes at time t, 1 if no), hospital center ($Z_{ij}^{(1)} = 1$ if hospital A,B, or C, 0 if hospital D or E), gender ($Z_{ij}^{(2)} = 1$ if female, 0 if male) and age of patient at start of study ($Z_{ij}^{(3)} = 1$ if age in (0, 27], 2 if age in (27, 42], 3 if age in (42, 55], 4 if age greater than 55). Censoring time C_{ij} is defined as the day of the last visit at the study and τ denotes the maximum time of C_{ij}'s. Since the time-dependent covariates $X_{ij}(t) = (X_{ij}^{(1)}(t), X_{ij}^{(2)}(t))$ were observed only at observation times and missing at nonobservation times, following the ideas of Huang et al. (2010), we approximate the time-dependent covariates $X_{ij}(t)$ in the interval between two observation times by the measurement of $X_{ij}(t)$ at the time point nearest to t. We suppose that the

Table 4 Estimation results of the covariate effects on the transfusion reactions with the corresponding *p*-values in parentheses.

Type	$\hat{\beta}$		$\hat{\gamma}$		
	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\gamma}^{(1)}$	$\hat{\gamma}^{(2)}$	$\hat{\gamma}^{(3)}$
I	2.129 (<0.01)	0.367 (0.143)	−0.219 (0.597)	0.038 (0.916)	0.104 (0.521)
II	1.436 (<0.01)	1.133 (< 0.01)	−0.765 (0.032)	0.163 (0.651)	0.337 (0.022)

Type I = "fever"; Type II = "all other reactions."

covariates are type-specific. The goal is to investigate the effect of the covariates on the risk of two types of transfusion reactions.

To analyze the data, we denote $N_{i1}^*(t)$ and $N_{i2}^*(t)$ to be the numbers of fever and all other reactions over interval $[0, t]$ with patient i, respectively, with risk intensity function $\lambda_{ij}(t)$, $j = 1, 2$ satisfying (1). The numbers of recurrence for two events range from 0 to 7 and 0 to 8 with the mean values of 0.76 and 1.01, respectively. Let $\beta_{0j}^{(k)}$, $j, k = 1, 2$ be the regression coefficients of $X_{ij}^{(k)}(t)$, and let $\gamma_{0j}^{(k)}$, $j = 1, 2, k = 1, 2, 3$ be the coefficients of $Z_{ij}^{(k)}$. The application of the proposed estimation procedure in Section 3 gave the estimates of regression parameters with *p*-values in Table 4. These results show that random donor platelets significantly increase the rate of both types of FNHTRs, while a priori with medication significantly decreases the rate of Type II platelet transfusion reactions and those patients in hospital A, B, and C and younger patients experienced less Type II platelet transfusion reactions. On the other hand, neither FNHTR rates seem to be related to the gender of the patients. Figure1 presents the estimated $\Lambda_{0j}(t)$, $j = 1, 2$ with the pointwise 95% bootstrap confidence intervals. By comparing Fig. 1 (a) and (b), it can be observed that FNHTR accompanied with fever has a higher risk than other reactions and the risk of the former increases over time.

We also investigated the effects of the covariates $(X_{ij}^{(1)}(t), X_{ij}^{(2)}(t), Z_{ij}^{(1)}, Z_{ij}^{(2)}, Z_{ij}^{(3)})$ on the risk of four types of transfusion reactions, that is, fever, chill, rigor, and all other reactions and obtained similar results as shown in Supplementary Information on the journal's website (`http://www.biometrical-journal.com`).

6 Concluding remarks

In this article we have propose a joint semiparametric frailty-based proportional intensity model for regression analysis of multivariate recurrent event data. The model allows for both time-dependent and time-independent covariates and informative censoring. Through the use of frailties, the proposed model relaxes the noninformative censoring condition for the recurrent event process. Simulation results show that the proposed method performed well and is more robust. The proposed model is flexible in the sense that the frailty distribution is treated as a nuisance parameter and no parametric assumptions are imposed. However, the lack of involvement of the correlation structure in the parameter estimation can cause to lose efficiency. To enhance the efficiency of the estimates, a borrow-strength procedure is developed. Simulation studies indicate that the efficiency gain is reduced by less correlation between different risk types.

One limitation is that each recurrent event process has been assumed to be a mixed Poisson process and it would be useful to generalize the approach to situations where the recurrent event process satisfies a proportional mean/rate model. In addition, one may investigate other models such as additive and additive-multiplicative intensity models as well as mean/rate models. Another direction for future research is to develop model-checking techniques.

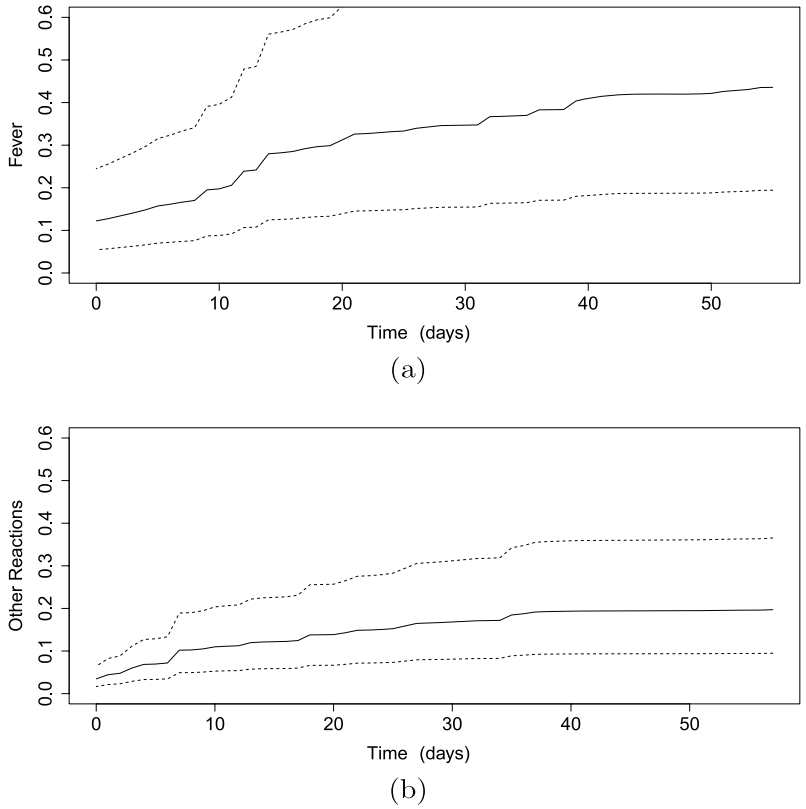

Figure 1 Plots of $\hat{\Lambda}_{0j}(t)$, $j = 1, 2$, the estimated mean numbers of recurrence of fever, and all other reactions in the FNHTR data, with pointwise bootstrap 95% confidence intervals.

In the application, the missing values for transfusion reactions are imputed by carrying forward the last observed. This approach takes all unobserved measurements as the last observed measurement, resulting in either underestimating or overestimating the treatment effects. Furthermore, it may lead to an underestimate of the standard deviation and inflation of the Type I error rate just like other single imputation methods, for example, baseline observation carried forward or mean imputation. On the other hand, multiple imputation methods, such as regression method, propensity score method, and Markov chain Monte Carlo (MCMC) method, can be taken into consideration. As the latter replaces each missing value with two or more plausible values that represent the uncertainty about the true value, it may overcome the drawbacks of single imputation methods. It deserves for us to study the effects of different imputation methods on FNHTR data in our future research.

Acknowledgments The authors are grateful to the editor, Dr Lutz Edler, an associate editor, and the two referees for their many helpful comments and suggestions that greatly improved this article. Zhao's work is partly supported by the Research Grant Council of Hong Kong (PolyU 5040/11P), Li Liu's work is partly supported by the Natural Science Foundation of China (nos. 11101315, 11171262), and Yanyan Liu's work is partly supported by the Natural Science Foundation of China (no. 11171263). The authors thank Dr. Bruce J. Patterson for providing us with example data.

Conflict of interest
The authors have declared no conflict of interest.

Appendix

For the sake of convenience, let $\sum_{i<l}$ stand for $\sum_{i=1}^{n}\sum_{l=i+1}^{n}$ and $\sum_{i<l_1<l_2}$ stand for $\sum_{i=1}^{n}\sum_{l_1=i+1}^{n}\sum_{l_2=l_1+1}^{n}$.

A.1 Asymptotic results of $\hat{\boldsymbol{\beta}}$

Theorem 1. *Assume that $X_j(t)$'s be bounded by M and $E[\overline{N}(\tau)]] < \infty$, where $\overline{N}(\tau) = \sum_{j=1}^{J} N_j(\tau)$. Then $\hat{\beta}$ is a consistent estimator of β_0. Furthermore, $\sqrt{n}(\hat{\beta} - \beta_0)$ converges weakly to a normal distribution with mean 0 and covariance $V(\beta_0) = V_2(\beta_0)^{-1}V_1(\beta_0)V_2(\beta_0)^{-1}$, where $V_1(\beta_0) = 4E[H(D_1, D_2; \beta_0)H(D_1, D_3; \beta_0)']$ and $V_2(\beta_0) = -E[\partial H(D_1, D_2; \beta_0)/\partial\beta_0]$. The covariance matrix V can be estimated by $\hat{V}_2^{-1}\hat{V}_1\hat{V}_2^{-1}$, where*

$$\hat{V}_1 = \frac{4}{n-2}\sum_{i=1}^{n-2}\frac{1}{\binom{n-i}{2}}\sum_{i<j<l} H(D_i, D_j; \hat{\beta})H(D_i, D_l; \hat{\beta})',$$

and

$$\hat{V}_2 = -\frac{1}{\binom{n}{2}}\sum_{i<l}\frac{\partial H(D_i, D_l; \hat{\beta})}{\partial\beta}.$$

A.2 Asymptotic results of the estimators of $\boldsymbol{\Lambda_{0j}(t)}$ and $\boldsymbol{\gamma}$

For each j, let $s_{j(l)}$'s be the ordered and distinct values of the j-th type event times T_{ijk}'s,

$$d_{j(l)}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m_{ij}} I(T_{ijk} = s_{j(l)})\exp\{-X_{ij}(T_{ijk})'\beta\},$$

and

$$R_{j(l)}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m_{ij}} I(T_{ijk} \le s_{j(l)} \le C_{ij})\exp\{-X_{ij}(T_{ijk})'\beta\}.$$

Then $F_{0j}(t)$ can be estimated by

$$\hat{F}_{0j}(t) = \hat{F}_{0j}(t, \hat{\beta}) = \prod_{s_{j(l)}>t}\left(1 - \frac{d_{j(l)}(\hat{\beta})}{R_{j(l)}(\hat{\beta})}\right).$$

To state the asymptotic distribution of $(\hat{F}_{0j}(t), j = 1, \ldots, J)'$, for any $u \in [0, \tau]$ and $d = 0, 1$, define

$$Q_{0j}^d(u) = \int_0^u G_{0j}(v) d\Lambda_{0j}(v) = \int_0^u E[\xi_{10} I(C_{1j} \geq v) \exp(Z_{1j}'\gamma_0) X_{1j}(v)^d] d\Lambda_{0j}(v),$$

$$R_{0j}^d(u) = \int_0^u E[\xi_{10} I(C_{1j} \geq u) \exp(Z_{1j}'\gamma_0) X_{1j}(v)^d] d\Lambda_{0j}(v),$$

$$\zeta_{0j}(D_i, D_l; t, \beta) = \left(- \int_t^\tau \frac{dQ_{0j}^1(u)}{R_{0j}^0(u)} + \int_t^\tau R_{0j}^1(u) \frac{dQ_{0j}^0(u)}{R_{0j}^0(u)^2} \right) V_2^{-1}(\beta) H(D_i, D_l; \beta),$$

$$\psi_{0j}(D_i; t, \beta) = \sum_{k=1}^{m_{ij}} \frac{I(t < T_{ijk} \leq \tau) \exp(-X_{ij}(T_{ijk})'\beta)}{R_{0j}^0(T_{ijk})} - \sum_{k=1}^{m_{ij}} \int_t^\tau I(t < T_{ijk} \leq C_{ij}) \exp(-X_{ij}(T_{ijk})'\beta) \frac{dQ_{0j}^0(u)}{R_{0j}^0(u)^2},$$

and

$$\kappa_{0j}(D_i, D_l; t, \beta) = \zeta_{0j}(D_i, D_l; t, \beta) + \{\psi_{0j}(D_i; t, \beta) + \psi_{0j}(D_l; t, \beta)\}/2.$$

Lemma 1. *Assume that (a)* $0 < \lambda_{0j}(\tau) < \infty$ *for each* $j = 1, \ldots, J$, *(b)* $P(C_{ij} > \tau, \xi_{i0} > 0) > 0$, *and (c)* $G_{0j}(u) = E\{\xi_{i0} I(C_{ij} \geq u) \exp(Z_{ij}'\gamma_0)\}$ *is a continuous function for* $u \in [0, \tau]$. *Let* $\tau_0 = \sup_{1 \leq j \leq J} \inf\{y : \Lambda_{0j}(y) > 0\}$. *Then for each* $t \in (\tau_0, \tau]$, $\sqrt{n}(\hat{F}_{0j}(t) - F_{0j}(t), j = 1, \ldots, J)'$ *converges weakly to a normal distribution with mean 0 and variance* $\Sigma_\kappa(t)$, *where* $\Sigma_\kappa(t) = 4E\{\kappa(D_1, D_2; t, \beta_0)\kappa(D_1, D_3; t, \beta_0)'\}$ *with* $\kappa(D_i, D_l; t, \beta) = (\kappa_{0j}(D_i, D_l; t, \beta) F_{0j}(t), j = 1, \ldots, J)'$.

For estimation of $\Lambda_{0j}(\tau)$ and γ, noting that

$$E\{m_{ij} | \xi_{i0}, C_{ij}, Z_{ij}, \mathcal{X}_{ij}(C_{ij})\} = \xi_{i0} \Lambda_{0j}(\tau) \exp(Z_{ij}'\gamma) \int_0^{C_{ij}} \exp\{X_{ij}(u)'\beta\} dF_{0j}(u), \tag{A1}$$

we have

$$E\left[\frac{m_{ij}}{\int_0^{C_{ij}} \exp(X_{ij}(u)'\beta) dF_{0j}(u)} \middle| C_{ij}, Z_{ij}, \mathcal{X}_{ij}(C_{ij}) \right] = \Lambda_{0j}(\tau) \exp(Z_{ij}'\gamma).$$

Let e_j be a J-dimensional vector with the j-th entry being 1 and other entries being 0, and let $\eta = (\log \Lambda_{01}(\tau), \ldots, \log \Lambda_{0J}(\tau), \gamma')'$, $\overline{Z}_{ij} = (e_j', Z_{ij}')'$, $Z_i^* = (\overline{Z}_{i1}, \ldots, \overline{Z}_{iJ})$,

$$M_{ij}(\beta, F_{0j}) = \frac{m_{ij}}{\int_0^{C_{ij}} \exp\{X_{ij}(u)'\beta\} dF_{0j}(u)},$$

and

$$M_i(\eta;\beta,F)=\left(M_{i1}(\beta,F_{01})-\exp(\overline{Z}_{i1}{}'\eta),\ldots,M_{iJ}(\beta,F_{0J})-\exp(\overline{Z}_{iJ}{}'\eta)\right)',$$

where $F=(F_{01},\ldots,F_{0J})$. Thus, η can be estimated by solving the following equation:

$$\frac{1}{n}\sum_{i=1}^{n}Z_i^*M_i(\eta;\hat{\beta},\hat{F})=0, \tag{A2}$$

where $F=(F_{01},\ldots,F_{0J})$. Let $\hat{\eta}=(\hat{\eta}_1,\ldots,\hat{\eta}_J,\hat{\gamma}')'$ be the root of the estimating equation (A2). So, $(\hat{\Lambda}_{0j}(t),\ j=1,\ldots,J)'$ can be estimated by $(\hat{F}_{0j}(t)\exp(\hat{\eta}_j),\ j=1,\ldots,J)'$. In the following, let γ_0 and η_0 be the true values of γ and η.

Theorem 2. *Under the conditions of Theorem 1 and Lemma 1, $\sqrt{n}(\hat{\gamma}-\gamma_0)$ converges weakly to a normal distribution with mean 0 and covariance Σ_γ, which is the $q\times q$ submatrix consisting of the last q row and the last q column of $\Sigma_\eta=\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}$ with $\Sigma_1=4E[\iota(D_1,D_2)\iota(D_1,D_3)']$ and $\Sigma_2=-E[\partial\iota(D_1,D_2)/\partial\eta_0]$, where*

$$\begin{aligned}\iota(D_i,D_l)=&\int z^*(\mathcal{M}_j(D_i,D_l;\beta_0),\ j=1,\ldots,J)'d\mathcal{V}(z^*,m,x,c)\\&+\frac{1}{2}Z_i^*M_i(\eta_0;\beta_0,F)+\frac{1}{2}Z_l^*M_l(\eta_0;\beta_0,F)\end{aligned}$$

with $\mathcal{V}$ being the joint probability measure of

$$(Z^*,m,X,C)=(Z^*,(m_1,\ldots,m_J),(X_1,\ldots,X_J),(C_1,\ldots,C_J)),$$

and

$$\begin{aligned}\mathcal{M}_j(D_i,D_l;\beta_0)=&\frac{-m_j}{\{\int_0^{c_j}\exp\{x_j(u)\beta_0\}dF_{0j}(u)\}^2}\\&\times\left[V_2^{-1}(\beta_0)H(D_i,D_l;\beta_0)\int_0^{c_j}\exp\{x_j(u)'\beta_0\}x_j(u)'dF_{0j}(u)\right.\\&\left.-\int_0^{c_j}\exp\{x_j(u)'\beta_0\}d(\kappa_{0j}(D_i,D_l;u,\beta_0)F_{0j}(u))\right].\end{aligned}$$

Moreover, for fixed $t\in(\tau_0,\tau]$, $\sqrt{n}(\hat{\Lambda}_{0j}(t)-\Lambda_{0j}(t),\ j=1,\ldots,J)$ converges weakly to a normal distribution with mean 0 and covariance matrix $\Sigma_\Lambda(t)=4E[f(D_1,D_2;t,\beta_0)f(D_1,D_3;t,\beta_0)']$ with $f(D_i,D_l;t,\beta_0)=(F_{0j}(t)\exp(\eta_j)(f_{0j}(D_i,D_l)+\kappa_{0j}(D_i,D_l;t,\beta_0)),\ j=1,\ldots,J)'$, where $f_{0j}(D_i,D_l)$ is the j-th entry of the vector function $\Sigma_2^{-1}\iota(D_i,D_l)$.

To enhance the efficiency of the estimator of γ, we propose an alternative estimating equation where the effect of the shared frailty is incorporated.

$$\frac{1}{n}\sum_{i=1}^{n}Z_i^*\left(M_{i1}(\beta,F_{01})-\xi_{i0}\exp(\overline{Z}_{i1}'\eta),\ldots,M_{i1}(\beta,F_{01})-\xi_{i0}\exp(\overline{Z}_{i1}'\eta)\right)'=0,$$

where frailty ξ_{i0} is unobservable. From (A1), given η, a natural estimator for ξ_{i0} is

$$\tilde{\xi}_{i0} = \frac{1}{J}\sum_{k=1}^{J} M_{ik}(\hat{\beta}, \hat{F}_{0k}) \exp(-\overline{Z}_{ik}'\eta)$$

by using the borrow strength ideas (Huang and Wang, 2004). Let

$$\tilde{M}_i(\eta; \beta, F) = \left(M_{ij}(\beta, F_{0j}) - \exp(\overline{Z}_{ij}'\eta)\frac{1}{J}\sum_{k=1}^{J} M_{ik}(\beta, F_{0k}) \exp(-\overline{Z}_{ik}'\eta),\ j = 1, \ldots, J \right)'.$$

Then η can be estimated by the solution to the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} Z_i^* \tilde{M}_i(\eta; \hat{\beta}, \hat{F}) = 0,$$

denoted by $\tilde{\eta} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_J, \tilde{\gamma}')'$, and $\Lambda_{0j}(t)$ can be estimated by $\tilde{\Lambda}_{0j}(t) = \hat{F}_{0j}(t) \exp(\tilde{\eta}_j)$.

Theorem 3. *Under the conditions of Theorem 1 and Lemma 1, $\sqrt{n}(\tilde{\gamma} - \gamma_0)$ converges weakly to a normal distribution with mean 0 and covariance $\tilde{\Sigma}_\gamma$, which is the $q \times q$ submatrix consisting of the last q row and the last q column of $\tilde{\Sigma}_\eta = \tilde{\Sigma}_2^{-1}\tilde{\Sigma}_1\tilde{\Sigma}_2^{-1}$ where $\tilde{\Sigma}_1 = 4E[\tilde{\iota}(D_1, D_2)\tilde{\iota}(D_1, D_3)']$, $\tilde{\Sigma}_2 = -E[\partial\tilde{\iota}(D_1, D_2)/\partial\eta_0]$, and*

$$\tilde{\iota}(D_i, D_l) = \int z^* \left(\mathcal{M}_j(D_i, D_l; \beta_0) - \exp(\overline{z}_j'\eta_0)\frac{1}{J}\sum_{k=1}^{J} \frac{\mathcal{M}_k(D_i, D_l; \beta_0)}{\exp(\overline{z}_k'\eta_0)},\ j = 1, \ldots, J \right)' \times d\mathcal{V}(z^*, m, x, c) + \frac{1}{2} Z_i^* \tilde{M}_i(\eta_0; \beta_0, F) + \frac{1}{2} Z_l^* \tilde{M}_l(\eta_0; \beta_0, F).$$

Moreover, for fixed $t \in (\tau_0, \tau]$, $\sqrt{n}(\tilde{\Lambda}_{0j}(t) - \Lambda_{0j}(t),\ j = 1, \ldots, J)$ converges weakly to a normal distribution with mean 0 and covariance $\tilde{\Sigma}_\Lambda(t) = 4E[\tilde{f}(D_1, D_2; t, \beta_0)\tilde{f}(D_1, D_3; t, \beta_0)']$ where

$$\tilde{f}(D_i, D_l; t, \beta_0) = \left(F_{0j}(t) \exp(\eta_j)(\tilde{f}_{01}(D_i, D_l) + \kappa_{0j}(D_i, D_l; t, \beta_0)),\ j = 1, \ldots, J \right)',$$

and $\tilde{f}_{0j}(D_i, D_l)$ is the j-th entry of the vector function $\tilde{\Sigma}_2^{-1}\tilde{\iota}(D_i, D_l)$.

References

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

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics* **10**, 1100–1120.

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

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. Springer-Verlag, New York.

Huang, C. Y., Qin, J. and Wang, M. C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics* **66**, 39–49.

Huang, C. Y. and Wang, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99**, 1153–1165.

Lancaster, T. and Intrator, O. (1998). Panel data with survival: hospitalization of HIV-positive patients. *Journal of the American Statistical Association* **93**, 46–53.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B* **62**, 711–730.

Patterson, B. J., Freedman, J., Blanchette, V., Sher, G., Pinkerton, P., Hannach, B., Meharchand, J., Lau, W., Boyce, N., Pinchefsky, E., Tasev, T., Pinchefsky, J., Poon, S., Shudman, L., Mack, P., Thomas, K., Blanchette, N., Greenspan, D., and Panzarella, T. (2000). Effect of premedication guidelines and leukoreduction on the rate of febrile nonhaemolytic platelet transfusion. *Transfusion Medicine* **10**, 199–206.

Pepe, M. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88**, 811–820.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. and Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: Application on cancer events. *Biostatistics* **8**, 708–721.

Schaubel, D. E. and Cai, J. (2005). Semiparametric methods for clustered recurrent event data. *Lifetime Data Analysis* **11**, 405–425.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93**, 1164–1175.

Sun, L., Zhu L. and Sun, J. (2009). Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Journal of Multivariate Analysis* **100**, 2214–2223.

Wang, M. C., Jewell, N. P. and Tsai, W. Y. (1986). Asymtotic properties of the product limit estimate under random truncation. *Annals of Statistics* **14**, 1597–1650.

Wang, M. C., Qin, J. and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96**, 1057–1065.

Ye, Y., Kalbfleisch, J. D. and Schaubel, E. S. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63**, 78–87.

Zhu, L., Sun, J. and Tong, X. (2010). Regression analysis of multivariate recurrent event data with a dependent terminal event. *Lifetime Data Analysis* **16**, 478–490.

