Statistica Sinica **27** (2017), 625-643
doi:http://dx.doi.org/10.5705/ss.202015.0152

SEMIPARAMETRIC ACCELERATED INTENSITY MODELS FOR CORRELATED RECURRENT AND TERMINAL EVENTS

Sangbum Choi[1], Xuelin Huang[2], Hyunsu Ju[2] and Jing Ning[2]

[1]*Korea University and* [2]*The University of Texas*

Abstract: In clinical and epidemiological studies, recurrent events can arise when a subject repeatedly experiences the event of interest. Often, a terminal event such as death may preclude further occurrence of recurrent events in an informative manner such that the terminal event is strongly correlated with the recurrent event process. In this article, we propose a semiparametric joint analysis of correlated recurrent and terminal events. Specifically, we consider an accelerated intensity model for the recurrent events and an accelerated failure time model for the terminal event. We assess the dependency between the two event processes through a commonly used log-normal or gamma shared frailty. To estimate regression parameters and unspecified baseline intensity functions, we develop an EM algorithm with kernel smoothing adapted for both intensity functions, and perform variance estimation via numerical differentiation of the profile likelihoods. We evaluated the finite sample performance of the proposed method via simulation studies for both gamma and log-normal frailty models, and applied our method to the analysis of tumor recurrences and patient survival times in a soft tissue sarcoma study.

Key words and phrases: Accelerated intensity regression, frailty model, informative censoring, kernel smoothing, nonparametric likelihood.

1. Introduction

In many clinical and epidemiological studies, recurrent event data are frequently encountered when a subject repeatedly experiences the event of interest. The recurrent event times for the subject are thus ordered and correlated. Statistical analysis of recurrent event data has been broadly explored in theory and extensively used in practice (Andersen and Gill (1982), Pepe and Cai (1993), Lin et al. (2000), Cai and Schaubel (2004), Liu, Lu and Zhang (2014)). Cook and Lawless (2007) provides a comprehensive review of the analysis of recurrent events.

When analyzing recurrent event data, an independent censoring condition is typically assumed in the construction of statistical methods. However, in many applications, especially in medical research, the timing of the terminating event such as death is likely to correlate with the recurrent event process, so that

the assumption of independent censoring between two event processes can be violated. For this situation, one has to consider not only the dependence among recurrent events, but also the association between recurrent and terminal events; ignorance of either could lead to biased estimates. A particularly important approach to this problem is to consider shared random-effects or frailty models (Wang, Qin and Chiang (2001), Huang and Wang (2004), Liu, Wolfe and Huang (2004), Ye, Kalbfleisch and Schaubel (2007), Liu and Huang (2007), Zeng and Lin (2009), Kalbfleisch et al. (2013)), in which a latent variable is incorporated to characterize the association between the recurrent event process and the failure time by allowing a common frailty variable to appear in the intensity of the recurrent event process and the hazard of the failure time. Similar techniques have been also extensively explored to accommodate an analysis of longitudinal data with informative right censoring (Wulfsohn and Tsiatis (1997), Tsiatis and Davidian (2004), Liu and Ying (2007), among others).

Many frailty-based models have assumed Cox-type multiplicative intensity/hazard functions for modeling both recurrent and terminal event processes. Wang, Qin and Chiang (2001) and Huang and Wang (2004) proposed an estimating equation approach for the shared frailty proportional intensity model that is flexible, in that no assumption for the frailty distribution is required. These methods, however, are statistically inefficient compared to a likelihood-based approach and do not allow for assessment of the degree of association, as well as a multivariate frailty structure. Ye, Kalbfleisch and Schaubel (2007) considered a gamma frailty for the same model and used estimating functions for marginal models along with a closed-form variance estimation, conditioning only on the covariates and the frailty. This approach was further extended by Kalbfleisch et al. (2013) to avoid the strong Poisson-type likelihood assumptions. Based on earlier work by Huang and Wolfe (2002) for informative censoring in clustered data, Liu, Wolfe and Huang (2004) proposed frailty proportional hazards models for the recurrent and terminal event processes, and carried out maximum likelihood estimation and inference via a Monte Carlo expectation-maximization (EM) algorithm. Alternatively, Zeng and Lin (2009) assumed a class of semiparametric transformation models for the intensity functions of both recurrent and terminal event processes. Their formulation allows flexible dependence structures on two event processes, including negative correlations between recurrent event times. Zeng and Lin developed an efficient EM algorithm to calculate the nonparametric maximum likelihood estimators (NPMLEs) and established asymptotic theories for the proposed estimators. Their algorithm was further extended by Zhu et al. (2011) for the situations with multiple types of recurrent events.

As an alternative to the Cox-type regression models, several authors have used semiparametric accelerated failure time (AFT) models for the counting process (Kalbfleisch and Prentice (2002, Chap. 7), Ying (1993), Lin, Wei and Ying

(1998), Jin et al. (2003), Zeng and Lin (2007)), where a known transformation of the failure time is linearly related to the covariates, while the distribution of the residuals is left unspecified. Because of its quite direct physical interpretation, the AFT model is often preferred in applications, but less popularized due to lack of efficient and reliable inferential procedures. Jin et al. (2003) and Jin, Lin and Ying (2006) developed rank regression estimators that can be obtained through linear programming. Zeng and Lin (2007) used a kernel-based profile likelihood maximization to achieve an NPMLE for the AFT model with censored data. Liu, Lu and Zhang (2014) extended their algorithm to facilitate a semiparametric accelerated intensity model for correlated recurrent event survival data.

In this article, we present a joint estimation of the accelerated intensity models for the recurrent/terminal event data and assess their dependence through a shared frailty with gamma and log-normal distribution assumptions. We propose to estimate the model parameters by the NPMLEs and to establish their theoretical properties. In addition, we provide simple and efficient numerical algorithms to implement the proposed inference procedures. The remainder of the article is organized as follows. Section 2 describes joint accelerated intensity frailty models for recurrent/terminal event data, taking potential dependency into consideration via a latent variable. In the same section, an estimation procedure and the asymptotic properties of the resulting estimators are provided. Section 3 presents results from simulation studies conducted to evaluate the finite-sample properties of the proposed estimates. Section 4 illustrates the application of the proposed methodology to a sarcoma cancer study. Section 5 presents some concluding remarks.

2. Statistical Models and Likelihood Inference

2.1. Model and notation

Suppose there are n independent subjects. Let C_i and D_i be the censoring and death times of individual $i = 1, \ldots, n$. Let $X_i = \min(D_i, C_i \wedge \tau)$ be the follow-up time and $\delta_i = I(D_i \leq C_i \wedge \tau)$ the observed terminal event (death) indicator, where τ is the maximum follow-up time, $a \wedge b = \min(a, b)$, and $I(\cdot)$ denotes an indicator function. Then, $N_i^D(t) = I(X_i \leq t, \delta_i = 1)$ is the observed death process. Let $N_i^R(t)$ denote the observed recurrence process that counts the number of observed recurrences up to time $t \in [0, X_i]$. Both N_i^R and N_i^D are the observed parts of the underlying recurrence and death counting processes N_i^{R*} and N_i^{D*}, respectively, which might increase after X_i when $\delta_i = 0$. These two sets of processes are related through $N_i^R(t) = \int_0^t Y_i(s) dN_i^{R*}(s)$ and $N_i^D(t) = \int_0^t Y_i(s) dN_i^{D*}(s)$, where $Y_i(t) = I(X_i \geq t)$ is the at-risk indicator at t. For subject i, we may observe the recurrent events at distinct times, $0 < T_{i1} < \cdots < T_{in_i} \leq$

X_i, where $n_i = N_i^R(X_i)$ represents the total number of recurrences. There may exist a p-vector of possibly time-dependent covariates $\mathbf{z}_i(t)$. Then, the observed data of individual i include $\mathbf{O}_i = \{Y_i(t), N_i^R(t), N_i^D(t), \mathbf{z}_i(t); 0 \le t \le \tau\}$, and $\mathbf{O} = \{\mathbf{O}_i, i = 1, \ldots, n\}$ is all the observed data. For simplicity of presentation, we assume that the covariates are time-independent, but the following argument can be easily generalized to the analysis with time-dependent covariates.

To evaluate the effects of covariates $\mathbf{z}_i$, we specify the intensity function of the recurrent event process $N_i^{R*}(t)$ and the hazard function of the terminal event process $N_i^{D*}(t)$, respectively, as

$$\begin{cases} r_i(t|\nu_i, \mathbf{z}_i) = \nu_i e^{\alpha' \mathbf{z}_i} r_0(te^{\alpha' \mathbf{z}_i}) = \nu_i r_i(t|\mathbf{z}_i), \\ \lambda_i(t|\nu_i, \mathbf{z}_i) = \nu_i e^{\beta' \mathbf{z}_i} \lambda_0(te^{\beta' \mathbf{z}_i}) = \nu_i \lambda_i(t|\mathbf{z}_i), \end{cases} \tag{2.1}$$

where $r_0(\cdot)$ and $\lambda_0(\cdot)$ are the unspecified baseline intensity functions, respectively corresponding to the recurrent and terminal events, α and β are the regression parameters of primary interest, and ν_i's are unobserved i.i.d. realizations of a latent variable $\nu > 0$. It is assumed that, given $(\mathbf{z}, \nu)$, $\{N^{R*}(\cdot), N^{D*}(\cdot), C\}$ are mutually independent. The occurrence of recurrent events is modeled by a subject-specific Poisson process via a latent variable, and conditioning on it, the rate function equals the intensity function. A multiplicative hazard function with the same latent variable but a different baseline function is assumed for the hazard of the failure event.

In model (2.1), dependence between the recurrent and terminal events is characterized by a latent variable ν_i, whose interpretation is analogous to a random effect from a mixed model. Technically, $\nu \sim f_\theta(\nu)$ can follow any distribution with strictly positive support, where the parameter θ indexes the frailty distribution. In the semiparametric setting, some distributional constraint on $f_\theta(\nu)$, such as $E(\nu) = 1$, may be required for model identification. A popular choice is a one-parameter Gamma distribution with mean 1 and variance θ:

$$f_\theta(\nu) = \frac{\nu^{\theta-1} \exp(-\theta\nu)\theta^\theta}{\Gamma(\theta)}. \tag{2.2}$$

In addition to assuming (2.2), we consider a log-normal distribution for $f_\theta(\nu)$. When $\theta \downarrow 0$, the frailty terms ν_i's are identically 1, so that the heterogeneity in both the recurrent and terminal event rates is explained solely by $\mathbf{z}_i$.

The two sub-models in (2.1) are equivalent to a class of AFT models for counting processes (Lin, Wei and Ying (1998), Zhang and Peng (2007)). One major advantage of this approach is that it allows for a direct relationship between event times and covariates, and marginal interpretation within the framework of accelerated regression modeling is possible. To see this, model (2.1) implies

$$E[N_i^{R*}(t)|\nu_i, \mathbf{z}_i] = \exp\{-\nu_i R_0(te^{\alpha' \mathbf{z}_i})\}, \quad E[N_i^{D*}(t)|\nu_i, \mathbf{z}_i] = \exp\{-\nu_i \Lambda_0(te^{\beta' \mathbf{z}_i})\},$$

where $R_0(t) = \int_0^t r_0(u)du$ and $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. Therefore,

$$E[N_i^{R*}(te^{-\alpha'\mathbf{z}_i})|\nu_i] = \exp\{-\nu_i R_0(t)\}, \quad E[N_i^{D*}(te^{-\beta'\mathbf{z}_i})|\nu_i] = \exp\{-\nu_i \Lambda_0(t)\}.$$

According to this formulation, the set of covariates $\mathbf{z}_i$ affects both the frequency of recurrences and the risk of death over time by expanding or contracting the time scale on which the events occur by a multiplicative factor of $e^{\alpha'\mathbf{z}_i}$ and $e^{\beta'\mathbf{z}_i}$, respectively, relative to that of a zero-valued covariate vector. Clearly, a large value of the shared frailty ν_i inflates both the intensity of the recurrent events and the hazard of the failure event. However, this does not alter the direct association between the event times and covariates, and holds even after the latent variable is integrated out.

2.2. Nonparametric maximum likelihood estimation

A likelihood can be constructed with a general approach for counting processes (Kalbfleisch and Prentice (2002, Chap. 6.2)) from the intensity specification in model (2.1). The specification in (2.1) implicitly assumes that, conditioning on the subject-specific random effect ν_i, two processes for recurrence and death are independent. Therefore, given ν_i, the "complete-data" likelihood of the ith subject can be factored into the products of the conditional distribution of N_i^{R*}, the conditional distribution of N_i^{D*}, and the distribution of ν_i. Integrating over ν_i, the joint nonparametric likelihood for $\vartheta = \{\alpha, \beta, \theta, r_0(\cdot), \lambda_0(\cdot)\}$ is

$$\begin{aligned} L(\vartheta) = \prod_{i=1}^{n} \int_{\nu_i} \Big[\Big\{ \prod_{k=1}^{n_i} \nu_i e^{\alpha'\mathbf{z}_i} r_0(t_{ik} e^{\alpha'\mathbf{z}_i}) \Big\} \exp\{-\nu_i R_0(x_i e^{\alpha'\mathbf{z}_i})\} \Big] \\ \times \Big[\{\nu_i e^{\beta'\mathbf{z}_i} \lambda_0(x_i e^{\beta'\mathbf{z}_i})\}^{\delta_i} \exp\{-\nu_i \Lambda_0(x_i e^{\beta'\mathbf{z}_i})\} \Big] f_\theta(\nu_i) d\nu_i. \end{aligned} \tag{2.3}$$

Contrary to the shared gamma frailty models (Nielsen et al. (1992)), the full log-likelihood (2.3) of the joint frailty model does not take a simple form because the integrals do not have a closed form. Also, without a parametric assumption on the baseline hazards on $r_0(\cdot)$ and $\lambda_0(\cdot)$, direct inference on the marginal likelihood is virtually impossible. For that situation, accomplishing a closed form expression for the complete data likelihood corresponding to (2.3) makes it feasible and attractive to use the EM algorithm that treats the ν_i's as missing values.

The complete data log-likelihood, based on the observed data $\mathbf{O}_i$ and a random effect ν_i, consists of three components:

$$l^c(\vartheta|\nu) = l_1^c(\alpha, r_0|\nu) + l_2^c(\beta, \lambda_0|\nu) + l_3^c(\theta|\nu), \tag{2.4}$$

where

$$l_1^c(\alpha, r_0|\nu) = n^{-1} \sum_{i=1}^{n} \Big[\sum_{k=1}^{n_i} \{\alpha'\mathbf{z}_i + \log r_0(e^{\tilde{\eta}_{ik}(\alpha)})\} - \nu_i R_0(e^{\tilde{\varepsilon}_i(\alpha)}) \Big],$$

$$l_2^c(\beta, \lambda_0|\nu) = n^{-1}\sum_{i=1}^{n}\Big[\delta_i\{\beta'\mathbf{z}_i + \log\lambda_0(e^{\tilde{\varepsilon}_i(\beta)})\} - \nu_i\Lambda_0(e^{\tilde{\varepsilon}_i(\beta)})\Big],$$
$$l_3^c(\theta|\nu) = n^{-1}\sum_{i=1}^{n}\left\{(n_i+\delta_i)\log\nu_i + \log f_\theta(\nu_i)\right\},$$

with $\tilde{\eta}_{ik}(\alpha) = \log(t_{ik}) + \alpha'\mathbf{z}_i$, $\tilde{\varepsilon}_i(\alpha) = \log(x_i) + \alpha'\mathbf{z}_i$, and $\tilde{\varepsilon}_i(\beta) = \log(x_i) + \beta'\mathbf{z}_i$.

E-step. In the E-step, we calculate the conditional expectation of $g(\nu_i)$ for some function $g(\cdot)$, given the observed data and the estimate of ϑ in the previous step. By Bayes' rule, the conditional distribution of ν_i, given $(\vartheta, \mathbf{O}_i)$, is

$$f(\nu_i|\vartheta, \mathbf{O}_i) = \frac{\nu_i^{n_i+\delta_i}e^{-\nu_i R_0(e^{\tilde{\varepsilon}_i(\alpha)})}e^{-\nu_i\Lambda_0(e^{\tilde{\varepsilon}_i(\beta)})}f_\theta(\nu_i)}{\int_{\nu_i}\nu_i^{n_i+\delta_i}e^{-\nu_i R_0(e^{\tilde{\varepsilon}_i(\alpha)})}e^{-\nu_i\Lambda_0(e^{\tilde{\varepsilon}_i(\beta)})}f_\theta(\nu_i)d\nu_i}. \tag{2.5}$$

If $\hat{\vartheta}^{(s)} = \{\hat{\alpha}^{(s)}, \hat{\beta}^{(s)}, \hat{\theta}^{(s)}, \hat{r}_0^{(s)}(\cdot), \hat{\lambda}_0^{(s)}(\cdot)\}$ is obtained in the sth iteration of the EM step, the $(s+1)$th iteration, taking the expectation with respect to ν_i in (2.4) given $(\hat{\vartheta}^{(s)}, \mathbf{O}_i)$ reduces to computing

$$\begin{cases} \hat{\nu}_i^{(s)} = E[\nu_i|\hat{\vartheta}^{(s)}, \mathbf{O}_i], \\ \widehat{\log\nu_i}^{(s)} = E[\log\nu_i|\hat{\vartheta}^{(s)}, \mathbf{O}_i], \\ \widehat{\log f_\theta(\nu_i)}^{(s)} = E[\log f_\theta(\nu_i)|\hat{\vartheta}^{(s)}, \mathbf{O}_i]. \end{cases} \tag{2.6}$$

When the frailty follows the gamma distribution, given in (2.2), we have

$$\begin{cases} \hat{\nu}_i^{(s)} = \dfrac{\hat{\theta}^{(s)} + n_i + \delta_i}{\{\hat{\theta}^{(s)} + \hat{R}_0^{(s)}(e^{\tilde{\varepsilon}(\hat{\alpha}^{(s)})}) + \hat{\Lambda}_0^{(s)}(e^{\tilde{\varepsilon}(\hat{\beta}^{(s)})})\}}, \\ \widehat{\log\nu_i}^{(s)} = \Psi(\hat{\theta}^{(s)} + n_i + \delta_i) - \log\{\hat{\theta}^{(s)} + \hat{R}_0^{(s)}(e^{\tilde{\varepsilon}(\hat{\alpha}^{(s)})}) + \hat{\Lambda}_0^{(s)}(e^{\tilde{\varepsilon}(\hat{\beta}^{(s)})})\}, \\ \widehat{\log f_\theta(\nu_i)}^{(s)} = (\theta-1)\widehat{\log\nu_i}^{(s)} - \theta\hat{\nu}_i^{(s)} + \theta\log\theta - \log\Gamma(\theta), \end{cases}$$

where $\Psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function. In general, $E\{g(\nu_i)|\hat{\vartheta}^{(s)}, \mathbf{O}_i\}$ can be approximated by the Monte Carlo simulation of ν_i or the Gaussian-quadrature approximation when ν_i follows a log-normal distribution.

M-step. In the M-step, we need to maximize the conditional expectation of the pseudo-complete data log-likelihood function given the observed data and the estimated quantities for the subject-specific frailty variable. Suppose the quantities (2.6) associated with unknown ν_i's are obtained in the sth iteration of

the E-step. Then we maximize the expectation of the complete data log-likelihood in equation (2.4)

$$l^e(\vartheta|\hat{\nu}^{(s)}) = l_1^e(\alpha, r_0|\hat{\nu}^{(s)}) + l_2^e(\beta, \lambda_0|\hat{\nu}^{(s)}) + l_3^e(\theta|\hat{\nu}^{(s)}), \tag{2.7}$$

where

$$l_1^e(\alpha, r_0|\hat{\nu}^{(s)}) = n^{-1}\sum_{i=1}^{n}\Big[\sum_{k=1}^{n_i}\{\alpha'\mathbf{z}_i + \log r_0(e^{\tilde{\eta}_{ik}(\alpha)})\} - \hat{\nu}_i^{(s)}R_0(e^{\tilde{\varepsilon}_i(\alpha)})\Big], \tag{2.8}$$

$$l_2^e(\beta, \lambda_0|\hat{\nu}^{(s)}) = n^{-1}\sum_{i=1}^{n}\Big[\delta_i\{\beta'\mathbf{z}_i + \log \lambda_0(e^{\tilde{\varepsilon}_i(\beta)})\} - \hat{\nu}_i^{(s)}\Lambda_0(e^{\tilde{\varepsilon}_i(\beta)})\Big], \tag{2.9}$$

$$l_3^e(\theta|\hat{\nu}^{(s)}) = n^{-1}\sum_{i=1}^{n}\left[(n_i+\delta_i)\widehat{\log \nu_i}^{(s)} + \widehat{\log f_\theta(\nu_i)}^{(s)}\right]. \tag{2.10}$$

The maximization of (2.10) for θ is relatively easy with a standard optimization algorithm and we let $\hat{\theta}^{(s+1)}$ denote the maximizer of (2.10). However, maximums of (2.8) and (2.9) do not exist when considering nonparametric estimation of $r_0(\cdot)$ and $\lambda_0(\cdot)$ (Zeng and Lin (2007)). To see this, under model (2.1) and the conditional independent censoring assumption, the intensity functions of the two counting processes, $N_i^R(te^{-\alpha'\mathbf{z}_i})$ and $N_i^D(te^{-\beta'\mathbf{z}_i})$, are given by $\nu_i r_0(t)I(e^{\tilde{\varepsilon}_i(\alpha)} \geq t)$ and $\nu_i\lambda_0(t)I(e^{\tilde{\varepsilon}_i(\beta)} \geq t)$, respectively, conditional on ν_i and $\mathbf{z}_i$. This implies that the nonparametric maximum likelihood estimators of $R_0(t)$ and $\Lambda_0(t)$ are

$$\tilde{R}_0(t;\alpha) = \sum_{i=1}^{n}\int_0^t \Delta\hat{R}_0(u;\alpha)du, \qquad \tilde{\Lambda}_0(t;\beta) = \sum_{i=1}^{n}\int_0^t \Delta\hat{\Lambda}_0(u;\beta)du, \tag{2.11}$$

where

$$\Delta\tilde{R}_0(t;\alpha) = \frac{dN_i^R(te^{-\alpha'\mathbf{z}_i})}{\sum_{j=1}^{n}\hat{\nu}_j^{(s)}I(e^{\tilde{\varepsilon}_j(\alpha)} \geq t)}, \quad \Delta\tilde{\Lambda}_0(t;\beta) = \frac{dN_i^D(te^{-\beta'\mathbf{z}_i})}{\sum_{j=1}^{n}\hat{\nu}_j^{(s)}I(e^{\tilde{\varepsilon}_j(\beta)} \geq t)}.$$

Here $\Delta\tilde{R}_0(t;\alpha)$ and $\Delta\tilde{\Lambda}_0(t;\beta)$ represent the jump size of the step function $\tilde{R}_0(t;\alpha)$ and $\tilde{\Lambda}_0(t;\beta)$ at time t, respectively, for subject i. It follows from the martingale property that $\sum_{i=1}^{n}\hat{\nu}_i^{(s)}\tilde{R}_0\{e^{\tilde{\varepsilon}_i(\alpha)};\alpha\} = \sum_{i=1}^{n} n_i$ and $\sum_{i=1}^{n}\hat{\nu}_i^{(s)}\tilde{\Lambda}_0\{e^{\tilde{\varepsilon}_i(\beta)};\beta\} = \sum_{i=1}^{n}\delta_i$. The profile log-likelihoods corresponding to (2.8) and (2.9) can then be reduced to, ignoring constants independent of the parameters, to

$$l_1^p(\alpha|\hat{\nu}^{(s)}) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{n_i}\left\{\alpha'\mathbf{z}_i + \log\Delta\tilde{R}_0(e^{\tilde{\eta}_{ik}(\alpha)};\alpha)\right\}, \tag{2.12}$$

$$l_2^p(\beta|\hat{\nu}^{(s)}) = n^{-1}\sum_{i=1}^{n}\delta_i\left\{\beta'\mathbf{z}_i + \log\Delta\tilde{\Lambda}_0(e^{\tilde{\varepsilon}_i(\beta)};\beta)\right\}. \tag{2.13}$$

However, maximums of these respective objective functions cannot be achieved with finite values of α and β, essentially because the estimators for $R_0(\cdot)$ and $\Lambda_0(\cdot)$ are very nonsmooth, involving only the ranks of $\tilde{\varepsilon}_i(\alpha)$ and $\tilde{\varepsilon}_i(\beta)$ through the indicator function $I(\cdot)$.

To handle such complexities, we seek smoothed versions of the baseline intensity functions in (2.11) using a symmetric kernel function $K(\cdot)$ with bandwidths a_n and b_n. Specifically, the estimated intensity function $\Delta\tilde{R}_0(t;\alpha)$ for $N_i^R(te^{-\alpha'\mathbf{z}_i})$ can be approximated by its smoothed version

$$\begin{aligned}\hat{r}_0^{(s)}(t;\alpha) &= \frac{(na_n)^{-1}\sum_{i=1}^n \int_0^\infty K\{(u-t)/a_n\}dN_i^R(ue^{-\alpha'\mathbf{z}_i})}{n^{-1}\sum_{i=1}^n \hat{\nu}_i^{(s)} \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\alpha)}-t)/a_n} K(u)du} \\ &= \frac{(na_n)^{-1}\sum_{i=1}^n \sum_{k=1}^{n_i} K\{(e^{\tilde{\eta}_{ik}(\alpha)}-t)/a_n\}}{n^{-1}\sum_{i=1}^n \hat{\nu}_i^{(s)} \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\alpha)}-t)/a_n} K(u)du}, \end{aligned} \tag{2.14}$$

as $a_n \to 0$. Similarly, the intensity function $\Delta\tilde{\Lambda}_0(t;\beta)$ for $N_i^D(te^{-\beta'\mathbf{z}_i})$ can be approximated by

$$\begin{aligned}\hat{\lambda}_0^{(s)}(t;\beta) &= \frac{(nb_n)^{-1}\sum_{i=1}^n \int_0^\infty K\{(u-t)/b_n\}dN_i^D(ue^{-\beta'\mathbf{z}_i})}{n^{-1}\sum_{i=1}^n \hat{\nu}_i^{(s)} \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\beta)}-t)/b_n} K(u)du} \\ &= \frac{(nb_n)^{-1}\sum_{i=1}^n \delta_i K\{(e^{\tilde{\varepsilon}_i(\beta)}-t)/b_n\}}{n^{-1}\sum_{i=1}^n \hat{\nu}_i^{(s)} \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\beta)}-t)/b_n} K(u)du}, \end{aligned} \tag{2.15}$$

as $b_n \to 0$. The corresponding estimators of $R_0(t)$ and $\Lambda_0(t)$ are respectively given by

$$\begin{aligned}\hat{R}_0^{(s)}(t;\alpha) &= \int_{-\infty}^t \frac{(na_n)^{-1}\sum_{i=1}^n \sum_{k=1}^{n_i} K\{(e^{\tilde{\eta}_{ik}(\alpha)}-s)/a_n\}}{n^{-1}\sum_{i=1}^n \hat{\nu}_i^{(s)} \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\alpha)}-s)/a_n} K(u)du} ds, \\ \hat{\Lambda}_0^{(s)}(t;\beta) &= \int_{-\infty}^t \frac{(nb_n)^{-1}\sum_{i=1}^n \delta_i K\{(e^{\tilde{\varepsilon}_i(\beta)}-s)/b_n\}}{n^{-1}\sum_{i=1}^n \hat{\nu}_i^{(s)} \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\beta)}-s)/b_n} K(u)du} ds. \end{aligned}$$

Therefore, the smoothed profile likelihoods for α and β, corresponding to (2.12) and (2.13), can be represented by

$$\begin{aligned} l_1^s(\alpha|\hat{\nu}^{(s)}) = {} & \frac{1}{n}\sum_{i=1}^n n_i(\alpha'\mathbf{z}_i) \\ & + \frac{1}{n}\sum_{i=1}^n \sum_{k=1}^{n_i} \log\Big\{\frac{1}{na_n}\sum_{i'=1}^n \sum_{k'=1}^{n_i} K\Big(\frac{e^{\tilde{\eta}_{i'k'}(\alpha)} - e^{\tilde{\eta}_{ik}(\alpha)}}{a_n}\Big)\Big\} \\ & - \frac{1}{n}\sum_{i=1}^n \sum_{k=1}^{n_i} \log\Big\{\frac{1}{n}\sum_{i'=1}^n \hat{\nu}_{i'}^{(s)} \int_{-\infty}^{(e^{\tilde{\varepsilon}_{i'}(\alpha)} - e^{\tilde{\varepsilon}_i(\alpha)})/a_n} K(u)du\Big\}, \end{aligned} \tag{2.16}$$

$$l_2^s(\beta|\hat{\nu}^{(s)}) = \frac{1}{n}\sum_{i=1}^{n}\delta_i(\beta'\mathbf{z}_i) + \frac{1}{n}\sum_{i=1}^{n}\delta_i \log\Big\{\frac{1}{nb_n}\sum_{j=1}^{n}\delta_j K\Big(\frac{e^{\tilde{\varepsilon}_j(\beta)} - e^{\tilde{\varepsilon}_i(\beta)}}{b_n}\Big)\Big\} - \frac{1}{n}\sum_{i=1}^{n}\delta_i \log\Big\{\frac{1}{n}\sum_{j=1}^{n}\hat{\nu}_j^{(s)}\int_{-\infty}^{(e^{\tilde{\varepsilon}_j(\beta)} - e^{\tilde{\varepsilon}_i(\beta)})/b_n} K(u)du\Big\}. \tag{2.17}$$

In the $(s+1)$th iteration of the M-step, we propose to maximize the objective functions (2.16) and (2.17) over α and β, respectively, to obtain $\hat{\alpha}^{(s+1)}$ and $\hat{\beta}^{(s+1)}$. Because $K(\cdot)$ is a smooth kernel function, the maximum of each function can be achieved using the Newton-Raphson method or any optimization search algorithm. In the implementation, we set the initial values of $\hat{\alpha}^{(0)}$ and $\hat{\beta}^{(0)}$ to the values of the estimators from the naïve Cox analyses that ignore correlations between recurrent and terminal events, and $\hat{\theta}^{(0)} = 1$. Letting $\hat{\nu}_i^{(0)} = 1$, we can also calculate $\hat{r}_0^{(0)}(\cdot)$ from (2.14) and $\hat{\lambda}_0^{(0)}(\cdot)$ from (2.15) that are required to initiate the E-step. Then the process between the E-step and the M-step is iterated until it reaches convergence. In each step, it usually suffices to obtain one-step estimates. Our experience with simulation studies in practical settings reveals that the results do not seem to be sensitive to the choice of initial values and that the maximizers that optimize the smoothed profile likelihood functions (2.16) and (2.17) also increase the actual complete-data log-likelihood function (2.7).

2.3. Variance estimation

To estimate the variance of the parameters of interest, we use the EM-aided numerical differentiation method (Chen and Little (1999)), which numerically computes the empirical information matrix of the observed profile likelihood. Similar inferential procedures were adapted by Liu, Lu and Zhang (2013, 2014) to facilitate clustered and recurrent event time data. To be specific, let $l_i^e(\vartheta|\hat{\nu})$ denote the ith component of the expected complete-data log-likelihood $l^e(\vartheta|\hat{\nu})$ in (2.7). Here, we focus on variance estimation of the finite-dimensional parameter of interest. The variance of $\hat{\alpha}$ is evaluated as follows. By perturbing the jth component $\hat{\alpha}_j$ of $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)'$ by a small amount, ϵ, we obtain the pair of perturbed estimates, denoted by $\hat{\alpha}_{j+} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_j + \epsilon, \ldots, \hat{\alpha}_p)$ and $\hat{\alpha}_{j-} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_j - \epsilon, \ldots, \hat{\alpha}_p)$ for $j = 1, \ldots, p$. We perform the EM algorithm until convergence to obtain a new estimator, $\hat{\vartheta}_{\alpha,j+}$, which consists of updated components of $\{\beta, \theta, r_0(\cdot), \lambda_0(\cdot)\}$ while fixing α at $\hat{\alpha}_{j+}$. We run another EM step to obtain $\hat{\vartheta}_{\alpha,j-}$ while fixing α at $\hat{\alpha}_{j-}$. Then we calculate $\tilde{S}_{\alpha,i} = (S_{\alpha,i1}, \ldots, S_{\alpha,ip})'$

for $i = 1, \ldots, n$, where $S_{\alpha,ij} = \{l_i^e(\hat{\vartheta}_{\alpha,j+}|\hat{\nu}) - l_i^e(\hat{\vartheta}_{\alpha,j-}|\hat{\nu})\}/(2\epsilon)$. We obtain $\tilde{S}_{\beta,i}$ and $\tilde{S}_{\theta,i}$ in a similar manner so as to form $\tilde{\mathcal{S}}_i = (\tilde{S}'_{\alpha,i}, \tilde{S}'_{\beta,i}, \tilde{S}'_{\theta,i})'$. We let $\tilde{\mathcal{I}} = \sum_{i=1}^n \tilde{\mathcal{S}}_i \tilde{\mathcal{S}}'_i$. Finally, we approximate the variance-covariance matrix of $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$ by $\tilde{\mathcal{I}}^{-1}$, the diagonal entities of which may provide desirable variance estimators for the NPMLEs.

2.4. Asymptotic results

In this section, we show that the proposed NPMLEs are consistent, asymptotically normal and semiparametrically efficient. The proofs are sketched in the Web Appendix. We need certain regularity conditions.

(C1) The parameter value $(\alpha_0, \beta_0, \theta_0)$ belongs to the interior of a known compact set Θ in $\mathcal{R}^d$. The covariate matrix $\mathbf{z} \in \mathcal{R}^p$ is bounded and has full rank.

(C2) The true rate function of $r_0(t)$ and the true hazard function of $\lambda_0(t)$ are both positive, at least twice continuously differentiable, and have bounded variations over $t \in [0, \tau]$.

(C3) With probability one, there exists a $\kappa_0 > 0$ such that $P(C \geq \tau|\mathbf{z}) > \kappa_0$. With probability one, $E[N^{R*}(\tau)] < \infty$ and $E[N^{D*}(\tau)] < \infty$.

(C4) The kernel function $K(\cdot)$ is thrice continuously differentiable and the rth derivative $K^{(r)}(\cdot), r = 0, 1, 2, 3$ has bounded variation in $(-\infty, \infty)$.

(C5) The information matrix $\mathcal{I}_0$ is finite and positive definite.

Theorem 1. *If* (C1)−(C4) *hold and, as* $n \to \infty$, $na_n^2 \to \infty$, $na_n^4 \to 0$, $nb_n^2 \to \infty$, *and* $nb_n^4 \to 0$, *then* $\hat{\vartheta}$ *is strongly consistent for* ϑ_0 *as* $n \to \infty$.

Theorem 2. *If* (C1)−(C5) *hold and, as* $n \to \infty$, $na_n^2 \to \infty$, $na_n^4 \to 0$, $nb_n^2 \to \infty$, *and* $nb_n^4 \to 0$, *then* $\sqrt{n}(\hat{\alpha}-\alpha_0)$, $\sqrt{n}(\hat{\beta}-\beta_0)$ *and* $\sqrt{n}(\hat{\theta}-\theta_0)$ *converge in distribution to respective mean-0 normal random vectors, as* $n \to \infty$.

3. Simulation Studies

We conducted a set of simulation studies to examine the performance of the proposed methods in practical settings. We generated recurrent and terminal event times from the frailty models

$$R_i(t|\nu_i) = \nu_i R_0(\xi_1 t e^{\alpha' \mathbf{z}_i}), \qquad \Lambda_i(t|\nu_i) = \nu_i \Lambda_0(\xi_2 t e^{\beta' \mathbf{z}_i}), \tag{3.1}$$

which involve two covariates, $\mathbf{z}_i = (z_{1i}, z_{2i})'$, where z_{1i} is a uniform $(-1, 1)$ variable and z_{2i} is a Bernoulli variable with 0.5 success probability. The frailty variable ν_i follows a gamma distribution or a log-normal distribution. Specifically, we considered (a) $\nu_i \sim \text{Gamma}(1/\theta, 1/\theta)$ and (b) $\log \nu_i \sim N(0, \sigma^2)$. We

let $\alpha = (0.5, 0.5)'$, $\beta = (0.6, 1)'$, $\theta = 1$ and $\sigma^2 = 0.8$. For the given frailty variable, we considered two scenarios where baseline intensity functions for recurrent and terminal event processes follow (i) an exponential ("Exp") distribution with $R_0(x) = \Lambda_0(x) = x$, and (ii) a standard log-logistic ("SL") distribution with $R_0(x) = \Lambda_0(x) = \log(1 + x)$. Of note, the marginal model given the frailty in (i) is equivalent to

$$\log T_{ik} = \log \xi_1 - \alpha' \mathbf{z}_i + \log \nu_i + \eta_{ik}, \qquad \log D_i = \log \xi_2 - \beta' \mathbf{z}_i + \log \nu_i + \varepsilon_i,$$

where the respective error terms $(\eta_{ik}, \varepsilon_i)$ have an extreme-value distribution and the model falls in the class of frailty proportional hazard models. We varied the values of ξ_1 and ξ_2 to retain a 35% censoring rate for the terminal event and to allow for about three recurrent events on average under the noninformative censoring scheme from the uniform (0.5) distribution.

In Table 1, we summarize the results of the proposed NPMLEs, obtained from 1,000 simulation runs with the sample sizes $n = 100$ and 200. Overall, the proposed method performs well in all cases. The estimators for the regression parameters are virtually unbiased, the standard error estimators accurately reflect the true variations and the confidence intervals have reasonable coverage probabilities. The standard errors for the frailty parameter are underestimated, resulting in lower coverage probabilities, but show improved accuracy as the sample size increases. We chose the kernel function $K(\cdot)$ to be the standard normal density for convenience and tractability. For bandwidth, we used the optimal bandwidth (Jones (1990)), given by $(4/n)^{1/3}\tilde{\sigma}_R$ and $(4/n)^{1/3}\tilde{\sigma}_D$ for models involving recurrent and terminal events, respectively, where $\tilde{\sigma}_R$ and $\tilde{\sigma}_D$ represent the sample standard deviation of $\{\log T_{ik} + \alpha^{(0)'}\mathbf{z}_i\}$ and $\{\log X_i + \beta^{(0)'}\mathbf{z}_i\}$. Another choice of bandwidth was considered and our limited experience reveals that the proposed method is not very sensitive to bandwidth selection and generally works well in all scenarios for the gamma and log-normal frailties.

We performed a sensitivity analysis to examine the effect of misspecification of the frailty distribution. With the specification $R_0(x) = x^2$ and $\Lambda_0(x) = x^3$, the simulated data involved one covariate z from the Bernoulli (0.5) distribution. The frailty variable followed (a) gamma and (b) log-normal distributions that have both mean 1 and variance $\theta = 0.8$ or 1.5 to reflect small and moderate variations in the frailty distribution. The variance of the log-normal frailty variable is $\theta = e^{\sigma^2} - 1$. We then applied the gamma-frailty model and the log-normal frailty model, respectively, as a working model to the simulated datasets. Table 2 presents the simulation results from 500 replications with $n = 100$. We observe that the variance estimates are seriously biased when the model is not correctly specified. When the true frailty distribution is gamma, the log-normal working

Table 1. Summary statistics for the simulation studies under (a) gamma frailty distribution and (b) log-normal frailty distribution.

		(a) Gramma frailty					(b) Log-normal frailty				
n	Dist	Parameter	Est	SE	SEE	CP	Parameter	Est	SE	SEE	CP
100	Exp	$\alpha_1 = 0.5$	0.490	0.344	0.328	0.938	$\alpha_1 = 0.5$	0.532	0.287	0.304	0.944
		$\alpha_2 = 0.5$	0.448	0.415	0.392	0.942	$\alpha_2 = 0.5$	0.478	0.356	0.327	0.945
		$\beta_1 = 0.6$	0.626	0.461	0.461	0.954	$\beta_1 = 0.6$	0.637	0.393	0.415	0.961
		$\beta_2 = 1.0$	1.110	0.601	0.585	0.944	$\beta_2 = 1.0$	1.088	0.542	0.530	0.949
		$\theta = 1.0$	1.096	0.266	0.198	0.885	$\sigma^2 = 0.8$	0.738	0.283	0.198	0.892
	SL	$\alpha_1 = 0.5$	0.452	0.306	0.289	0.936	$\alpha_1 = 0.5$	0.485	0.261	0.266	0.939
		$\alpha_2 = 0.5$	0.558	0.304	0.294	0.946	$\alpha_2 = 0.5$	0.506	0.333	0.345	0.956
		$\beta_1 = 0.6$	0.607	0.552	0.548	0.951	$\beta_1 = 0.6$	0.588	0.478	0.481	0.947
		$\beta_2 = 1.0$	1.025	0.660	0.671	0.952	$\beta_2 = 1.0$	1.069	0.606	0.601	0.935
		$\theta = 1.0$	1.104	0.324	0.244	0.896	$\sigma^2 = 0.8$	0.724	0.285	0.213	0.906
200	Exp	$\alpha_1 = 0.5$	0.492	0.185	0.176	0.941	$\alpha_1 = 0.5$	0.498	0.179	0.191	0.956
		$\alpha_2 = 0.5$	0.486	0.229	0.210	0.944	$\alpha_2 = 0.5$	0.514	0.239	0.237	0.948
		$\beta_1 = 0.6$	0.574	0.310	0.311	0.961	$\beta_1 = 0.6$	0.629	0.271	0.281	0.944
		$\beta_2 = 1.0$	1.063	0.400	0.378	0.944	$\beta_2 = 1.0$	1.072	0.359	0.351	0.958
		$\theta = 1.0$	1.072	0.166	0.131	0.911	$\sigma^2 = 0.8$	0.764	0.168	0.133	0.901
	SL	$\alpha_1 = 0.5$	0.533	0.155	0.152	0.937	$\alpha_1 = 0.5$	0.503	0.156	0.158	0.948
		$\alpha_2 = 0.5$	0.531	0.202	0.186	0.943	$\alpha_2 = 0.5$	0.487	0.212	0.211	0.954
		$\beta_1 = 0.6$	0.625	0.363	0.363	0.945	$\beta_1 = 0.6$	0.636	0.329	0.336	0.964
		$\beta_2 = 1.0$	1.103	0.445	0.455	0.955	$\beta_2 = 1.0$	1.031	0.408	0.395	0.944
		$\theta = 1.0$	1.089	0.196	0.176	0.904	$\sigma^2 = 0.8$	0.785	0.194	0.163	0.903

Note: Est and SE are the estimate and standard error of the parameter estimator, SEE is the mean of the standard error estimator, and CP is the coverage probability of the 95% confidence interval. "Exp" and "SL" represent exponential and standard log-logistic distributions, respectively, for the baseline hazard function.

model overestimates the frailty variance. Such biasedness seems amplified to an extent as θ increases. When the frailty is obtained from the log-normal distribution but the gamma model is applied, the frailty variance is underestimated. Nonetheless, the estimators of the regression parameters seem unbiased in nearly all cases and the proposed variance estimation gives proper coverage probabilities. In the supplementary material, we also display true and estimated frailty distributions that are fairly similar under model misspecification. This may indicate that the performance of the NPMLEs is robust to the choice of frailty distribution if the primary focus is assessing the covariate effects and such an approximation will not lessen the value of our method.

In a similar set-up as above, we conducted simulation studies to compare

Table 2. Simulation studies for sensitivity analysis under a misspecified frailty assumption with the underlying model having (a) gamma frailty with mean 1 and variance θ and (b) log-normal frailty with mean 1 and variance $\theta = e^{\sigma^2} - 1$.

Frailty			Working models							
			Gamma				Log-normal			
Distribution	Par	True	Est	SE	SEE	CP	Est	SE	SEE	CP
Gamma	α	-0.2	-0.198	0.108	0.104	0.948	-0.196	0.116	0.111	0.965
	β	-1.0	-0.997	0.088	0.085	0.950	-1.009	0.094	0.089	0.954
	θ	0.8	0.820	0.109	0.114	0.920	2.089	0.600	0.574	0.222
Gamma	α	-0.2	-0.198	0.108	0.104	0.948	-0.196	0.116	0.111	0.965
	β	-1.0	-0.997	0.088	0.085	0.950	-1.009	0.094	0.089	0.924
	θ	1.5	1.488	0.191	0.181	0.907	8.533	4.438	4.200	0.116
Log-normal	α	-0.2	-0.187	0.093	0.089	0.939	-0.192	0.092	0.089	0.940
	β	–1.0	-0.992	0.077	0.076	0.942	-1.006	0.075	0.075	0.952
	θ	0.8	0.465	0.072	0.080	0.020	0.789	0.182	0.192	0.922
Log-normal	α	-0.2	-0.187	0.093	0.089	0.939	-0.192	0.092	0.089	0.940
	β	-1.0	-0.992	0.077	0.076	0.942	-1.006	0.075	0.075	0.952
	θ	1.5	0.689	0.100	0.104	0.000	1.481	0.384	0.384	0.936

Table 3. Simulation studies for sensitivity analysis between joint and reduced models. The joint model and the reduced model are a correct model in scenario (I) and (II), respectively. Gamma frailty distribution is assumed in all cases.

			Joint model				Reduced model				
Scenario	Par	True	Est	SE	SEE	CP	Est	SE	SEE	CP	RE
(I)	α	1.0	0.995	0.215	0.220	0.952	0.974	0.225	0.232	0.942	1.09
	β	1.0	0.991	0.318	0.308	0.940	0.917	0.317	0.306	0.926	1.01
	θ	0.5	0.494	0.111	0.111	0.944	0.531	0.136	0.141	0.971	1.49
	α	1.0	1.035	0.295	0.287	0.965	1.024	0.306	0.300	0.945	1.10
	β	1.0	1.041	0.364	0.366	0.958	0.919	0.366	0.371	0.951	1.05
	θ	1.0	0.994	0.187	0.189	0.934	1.070	0.229	0.233	0.950	1.49
(II)	α	1.0	1.028	0.238	0.183	0.902	1.016	0.236	0.211	0.938	0.98
	β	1.0	1.007	0.242	0.253	0.966	1.004	0.224	0.232	0.962	0.86
	θ	0.5	0.243	0.069	0.071	0.126	0.481	0.145	0.137	0.914	4.38
	α	1.0	1.018	0.259	0.219	0.918	1.008	0.257	0.271	0.958	0.99
	β	1.0	0.974	0.265	0.268	0.946	0.982	0.243	0.231	0.943	0.79
	θ	1.0	0.413	0.088	0.098	0.000	0.991	0.228	0.232	0.940	7.14

our joint model (2.1) with the reduced model

$$r_i(t|\nu_i, z_i) = \nu_i e^{\alpha z_i} r_0(te^{\alpha z_i}), \qquad \lambda_i(t|\nu_i, z_i) = e^{\beta z_i}\lambda_0(te^{\beta z_i}). \tag{3.2}$$

Here the random effect appears only in the recurrent event model that is equivalent to that of Liu, Lu and Zhang (2014), and recurrent and terminal event processes are independent. We considered two scenarios (I) and (II), where the data sets were simulated according to models (2.1) and (3.2), respectively, and these models were used as a working model. The joint and reduced models were correct under scenario (I) and (II), respectively. Gamma frailty distribution was assumed in all cases. This study aimed to check the effect of model misspecification and parameter efficiency. We let $\alpha = \beta = 1$ and assumed a gamma frailty distribution with $\theta \in \{0.5, 1\}$. As seen in Table 3, the frailty parameter estimators appear to be biased to an extent under model misspecification, while the regression parameter estimators are relatively consistent. We also give the "relative efficiency (RE)", the mean square error of the reduced model divided by that of the joint model. The joint model gains more efficiency under case (I) but not under case (II), as expected. Efficiency loss becomes large when θ increases. The degree of sensitivity is more severe when the joint model is applied to the data from the reduced model, as a single frailty variable in the joint model accounts for the association among recurrent event processes as well as that between recurrent and terminal event processes. The joint modeling approach seems ineffective in the independent cases, and may require separate frailty formulations for different types of association.

4. Application to a Soft Tissue Sarcoma Study

We applied our method to the data set from a soft tissue sarcoma study (Cormier et al. (2004), Huang, Cormier and Pisters (2006)), in which patients may experience local recurrence of sarcoma (in the same or nearby part of the body where the primary cancer occurred), distant recurrence (in a different part of the body), and death. A cohort of 679 patients was identified from two major cancer centers. In their initial treatments, all patients received definitive surgical resection of the tumor. The objective of this analysis was to evaluate the impact of chemotherapy and radiation while accounting for known prognostic variables. Among the 679 patients, 228 received adjuvant radiation alone, 109 received adjuvant chemotherapy alone, 207 received both, and 135 received none of these treatments. Of the 316 patients treated with adjuvant chemotherapy, 148 (46.8%) died from sarcomas; and of the 363 patients not treated with adjuvant chemotherapy, 140 (38.6%) died from sarcomas. The maximum number of tumor recurrences in one patient was three. The total number of tumor recurrences was 537, and 350 patients had at least one local or distant sarcoma recurrence. The median follow-up time was 6.87 years.

Table 4. Soft tissue sarcoma study: parameter estimate and inference.

		Cancer recurrence			Death		
Model	Effect	Est	SE	P-value	Est	SE	P-value
Gamma	Chemotherapy	-0.046	0.170	0.786	-0.055	0.125	0.661
	Radiation	-0.326	0.177	0.067	-0.313	0.130	0.016
	Tumor size	0.063	0.016	<0.001	0.054	0.011	<0.001
	θ	3.042	0.243	<0.001			
Log-normal	Chemotherapy	0.128	0.187	0.496	0.031	0.133	0.811
	Radiation	-0.372	0.193	0.055	-0.344	0.138	0.013
	Tumor size	0.072	0.016	<0.001	0.060	0.011	<0.001
	σ^2	3.632	0.385	<0.001			

In the treatment of sarcomas, chemotherapy is used to destroy cancer cells and prevent distant recurrences, while radiation is used to shrink the tumor and prevent local recurrences. Although this effect of radiation is well accepted, the effect of adjuvant chemotherapy remains uncertain. It has been explored in several studies that have compared outcomes for patients who have received adjuvant chemotherapy with those of patients who have not received adjuvant chemotherapy. We are interested in evaluating the effects of adjuvant chemotherapy and radiation on cancer recurrence and survival using the proposed method for the joint semiparametric accelerated intensity model (2.1). In each submodel, we included three covariates: the indicators of receiving chemotherapy and radiation and the maximum tumor size at baseline, which ranges from 5 to 41 cm.

Table 4 summarizes the estimation results under the proposed model with gamma and log-normal frailty distributions. In both models, radiation is significant in reducing the risks of death and shows a moderate effect in decreasing sarcoma recurrence. The regression parameter estimators associated with chemotherapy showed opposite signs, depending on the choice of frailty distribution, but there is no treatment difference with chemotherapy for either disease recurrence or death. Not surprisingly, the patients who had a large tumor size tended to experience disease recurrence more frequently and to die earlier. The estimated variance being significantly greater than zero indicates that after adjusting for treatments and clinical factors, there appears to be a strong association between cancer recurrence and death due to unknown factors. Figure 1 displays estimated cumulative rate functions of sarcoma recurrences for (a) patients who received radiation and (b) patients who did not receive radiation, along with nonparametric curves that are estimated by $\sum_{i=1}^{n} I(X_i \geq t)N_i^R(t)/\sum_{i=1}^{n} I(X_i \geq t)$ for each group. The proposed estimates reasonably follow the nonparametric estimates, supporting the choice of the method.

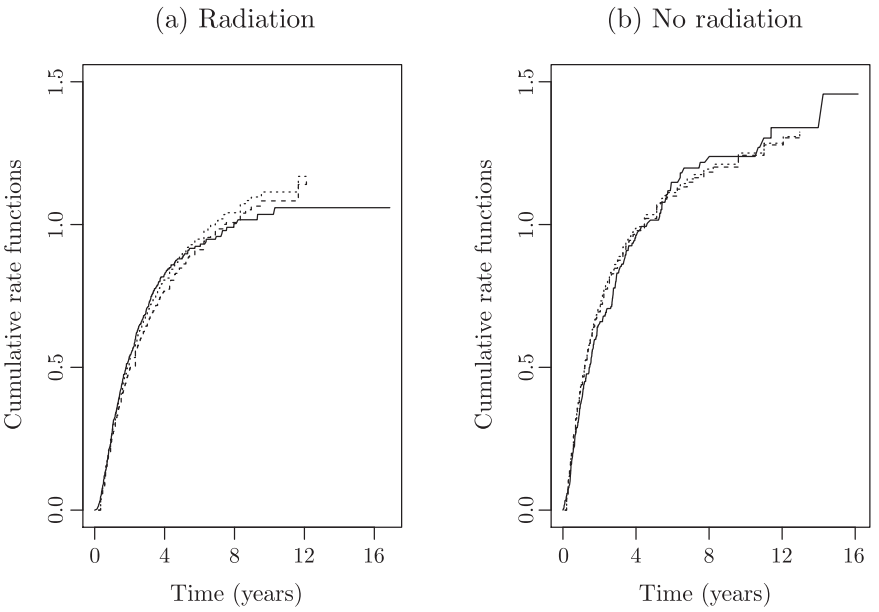

Figure 1. Estimated cumulative rate functions of sarcoma recurrences for (a) patients who received radiation and (b) patients who did not receive radiation. The solid, dashed and dotted lines pertain to the nonparametric and semiparametric estimates from the gamma and log-normal frailty models, respectively.

5. Discussion

In this paper, we propose a joint accelerated intensity model for correlated recurrent and terminal event data. The approach directly deals with the association between recurrent events and terminal events while allowing each event time to have an arbitrary residual error distribution. A latent frailty variable, gamma or log-normally distributed, is assumed to explain the dependency. For the estimation, we developed the EM algorithm that extends the method by Liu, Lu and Zhang (2014) for univariate recurrent event data, in which the regression parameters of interest are estimated by maximizing the kernel-smoothed profile likelihood functions. One advantage of the proposed model is that it preserves a direct relationship between the event time and covariates for both recurrent/terminal events even after the frailty variable is integrated out, and thus regression parameters still have a marginal interpretation as in univariate AFT models. In many public health and biomedical studies, this approach may be preferred for analysis, especially in identifying treatment effects and risk factors. Moreover, compared to the marginal approach (Lin, Wei and Ying (1998)), the joint frailty approach offers the additional ability to quantify the dependence

between different types of event processes and the efficiency gain via likelihood-based estimation.

Simulation studies demonstrated that the proposed method works well and the regression estimators do not seem to be very sensitive to the choice of the frailty distribution. As shown in Table 3, however, the joint model performs unsatisfactorily when the recurrent event processes are correlated within subject but not with the terminal event process. In the situation, we can consider the modification

$$r_i(t|\nu_i, \mathbf{z}_i) = \nu_i r_i(t|\mathbf{z}_i), \qquad \lambda_i(t|\nu_i, \mathbf{z}_i) = \nu_i^{\gamma} \lambda_i(t|\mathbf{z}_i),$$

which is the analogue of the frailty model of Liu, Wolfe and Huang (2004). An additional parameter $\gamma \in \mathcal{R}$ controls the degree of dependence between two event processes and alleviates potential misspecification issues, including the reduced model (3.2) with $\gamma = 0$. The proposed EM algorithm can be modified to accommodate such extra parametrization.

We applied our joint analysis approach to data from a sarcoma cancer study in which patients may experience local and distant tumor recurrences. Although we do not distinguish local versus distant recurrences, they may indicate different levels of risk and associations with death. If this is the case, it is desirable to incorporate a series of frailty variables into the model to adjust the individual recurrence intensities and account for the dependence among different types of recurrent events and that between the recurrent events and the terminal event. Recently, Ning et al. (2015) proposed a time-dependent measure to assess the local dependence between two types of recurrent event processes. They modeled the rate ratio as a parametric function of time, leaving unspecified all other aspects of the distribution, and applied their methods to the same sarcoma data. This may provide additional insight and facilitate a better understanding of the interactive associations between different types of cancer recurrences.

Acknowledgement

This research was supported in part by grants from Korea University Future Research Grant (KUFRG) and the National Institutes of Health (NIH).

References

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

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

Chen, H. Y. and Little, R. J. A. (1999). Proportional hazards regression with missing covariates. *J. Amer. Statist. Assoc.* **94**, 896-908.

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

Cormier, J. N., Huang, X., Xing, Y., Thall, P. F., Wang, X., Benjamin, R. S., Pollock, R. E., Antonescu, C. R., Maki, R. G., Brennan, M. F. and Pisters, P. W. T. (2004). Cohort analysis of 674 patients with localized high-risk extremity soft tissue sarcoma treated at two cancer centers: chemotherapy-associated outcomes. *J. Clin. Oncol.* **22**, 4567-4574.

Huang, C. Y. and Wang, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc.* **99**, 1153-1165.

Huang, X., Cormier, J. N. and Pisters, P. W. T. (2006). Estimation of the causal effects on survival of two-stage nonrandomized treatment sequences for recurrent diseases. *Biometrics* **62**, 901-909.

Huang, X.. and Wolfe, R. A. (2002). A frailty model for informative censoring. *Biometrics* **58**, 510-520.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341-353.

Jin, Z., Lin, D. Y. and Ying, Z. (2006). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scan. J. Stat.* **33**, 1-13.

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statist. Probab. Let.* **9**, 129-132.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd edition, Wiley.

Kalbfleisch, J. D., Schaubel, D. E., Ye, Y. and Gong, Q. (2013). An estimating function approach to the analysis of recurrent and terminal events. *Biometrics* **69**, 366-374.

Lin, D. Y., Wei, L. J. and Ying, Z. L. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605-618.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate function of recurrent events. *J. Roy. Statist. Soc. Ser. B.* **69**, 711-730.

Liu, B., Lu, W. and Zhang, J. (2013). Kernel smoothed profile likelihood estimation in AFT frailty model for clustered survival data, *Biometrika* **100**, 741-755.

Liu, B., Lu, W. and Zhang, J. (2014). Accelerated intensity frailty model for recurrent events data. *Biometrics* **70**, 579-587.

Liu, L. and Huang, X. (2007). The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statist. Medicine* **27**, 2665-2683.

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

Liu, L. and Ying, Z. (2007). Joint analysis of longitudinal data with informative right censoring. *Biometrics* **63**, 363-371.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sorensen, T. I. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25-43.

Ning, J., Chen, Y., Cai, C., Huang, X. and Wang, M. C. (2015). On the dependence structure of bivariate recurrent event processes: inference and estimation. *Biometrika* **102**, 345-358.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Assoc.* **88**, 811-820.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statist. Sinica* **14**, 809-834.

Wang, M. C., Qin, J. and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *J. Amer. Statist. Assoc.* **96**, 1057-1065.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.

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

Ying, Z. (1993). A large samples study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.

Zeng, D. and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *J. Amer. Statist. Assoc.* **102**, 1387-1396.

Zeng, D. and Lin, D. Y. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* **65**, 746-752.

Zhang, J. and Peng, Y. (2007). An alternative estimation method for the accelerated failure time frailty model. *Comput. Statist. Data Anal.* **51**, 4413-4423.

Zhu, L., Sun, J., Srivastava, D. K., Tong, X. Leisenring, W., Zhang, H. and Robison, L. L. (2011). Semiparametric transformation models for joint analysis of multivariate recurrent and terminal events. *Statist. Medicine* **30**, 3010-3023.

Department of Statistics, Korea University, 145 Anam-dong, Seongbuk-gu, Seoul, South Korea.

E-mail: choisang@korea.ac.kr

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030, U.S.A.

E-mail: xlhuang@mdanderson.org

Department of Preventive Medicine, The University of Texas Medical Branch, 700 Harborside Drive, Galveston, TX 77550, U.S.A.

E-mail: hyju@utmb.edu

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030, U.S.A.

E-mail: jning@mdanderson.org

(Received April 2015; accepted February 2016)