Statistics and Its Interface Volume 10 (2017) 699–710

Joint modeling of recurrent and terminal events using additive models

Lianqiang Qu, Liuquan Sun*, and Lei Liu

In this article, we propose a joint modeling approach for the analysis of recurrent event data with a terminal event. We specify an additive rates model with a multiplicative frailty for the conditional recurrent event rate, and an additive hazards frailty model for the terminal event. A shared frailty is used to account for the association between recurrent and terminal events. An estimating equation approach is developed for the marginal and association parameters in the joint model, and the asymptotic properties of the proposed estimators are established. The finite sample performance of the proposed estimators is examined through simulation studies, and an application to a medical cost study of chronic heart failure patients is illustrated.

AMS 2000 subject classifications: Primary 62N01, 62N02; secondary 62G05.
Keywords and phrases: Additive models, Estimating equations, Frailty, Partial marginal models, Recurrent events, Terminal event.

1. INTRODUCTION

Recurrent event data are frequently encountered in a wide variety of fields such as biomedicine, public health, engineering and economics, where each subject may experience a particular event repeatedly over time. Examples include cancer tumor recurrences, repeated hospitalization, multiple infection episodes, repeated drug use, recurrent economic recessions and many others (Cook and Lawless, 2007). Interest of recurrent event analysis often focuses on assessing the effects of covariates on certain features of the recurrent event times. Various methods have been considered for the analysis of recurrent event data based on the intensity or rate functions of recurrent events (e.g. Anderson and Gill, 1982; Lin et al., 2000; Lin et al., 2001; Zeng and Lin, 2007; Cook et al., 2009; Zeng and Cai, 2010; Sun et al., 2011). A comprehensive review of the existing statistical methods can be found in Kalbfleisch and Prentice (2002) and Cook and Lawless (2007).

In many applications, there may exist a terminal event such as death that stops the follow-up. For example, patients may experience recurrent hospitalizations that are terminated by death. Furthermore, it is often the case that the terminal event is likely to be strongly correlated with recurrent events of interest. Some efforts have been made recently on the analysis of recurrent events with a terminal event, and the existing methods generally fall into three approaches: intensity models, marginal models and partial marginal models (Pan and Schaubel, 2009; Liu et al., 2012). Intensity models use frailties to account for the dependence between the recurrent and terminal events, and the intensity functions of the recurrent and terminal events are fully specified by the observed covariates and the unobserved frailties (Wang et al., 2001; Liu et al., 2004; Huang and Wang, 2004; Zeng and Lin, 2009). For this approach, it is assumed that given the frailties, the recurrent event process is a Poisson process, and the estimation procedures would generally be sensitive to deviations from the Poisson assumption. Marginal models consider the marginal rates of the recurrent and terminal events, and the association between the recurrent and terminal events is left unspecified (Ghosh and Lin, 2002; Cook et al., 2009; Schaubel and Zhang, 2010; Zhao et al., 2011). In this case, the rate of the recurrent events averages over surviving and deceased subjects, and the model parameters are somewhat hard to interpret (Kalbfleisch et al., 2013).

Partial marginal models focus on the recurrent event rates among survivors, and a variation of this approach uses a frailty to specify the dependence between the recurrent and terminal events. More detailed discussion on this approach can be found in Cook and Lawless (1997), Liu et al. (2004), Ye et al. (2007), Pan and Schaubel (2009), Zeng and Cai (2010), Liu et al. (2012) and Sun and Kang (2013). For example, Cook and Lawless (1997) proposed the mean and rate of the recurrent events among survivors at certain time points. Ye et al. (2007) and Kalbfleisch et al. (2013) studied a joint semiparametric model through a shared gamma frailty which is used to account for the dependence between the conditional recurrent event rate and terminal event hazard. Partial marginal models offer great flexibilities in formulating the effects of covariates on the conditional recurrent event rate among survivors, and the model parameters can be interpreted as the marginal effects on the conditional recurrent event rate given survival.

For partial marginal models, most existing methods assume multiplicative covariate effects on the conditional recurrent event rate and terminal event hazard. In practice, a useful and important alternative to the multiplicative model

*Corresponding author.

is the additive model, in which the covariate effects are add to the conditional recurrent event rate and terminal event hazard. When the additive and multiplicative models fit the data equally well, the additive model may be preferred due to the interpretation of the regression parameters (Lin and Ying, 1994; Zeng and Cai, 2010). Recently, Pan and Schaubel (2009) proposed a treatment effect measure by combining an additive model for the conditional recurrent event rate and a proportional hazards model for the terminal event. Zeng and Cai (2010) and Sun and Kang (2013) studied an additive rates model and an additive-multiplicative rates model, respectively, wherein a proportional hazards model is used to model the terminal event, and the association between the recurrent and terminal events is treated as a nuisance because the baseline rate function depends on a frailty nonparametrically. Chen et al. (2016) considered a partly Aalen's additive model with a multiplicative frailty for the recurrent event rate and assumed a proportional hazards frailty model for the terminal event. But they did not establish the asymptotic distributions of the proposed estimators, because additional work is needed in developing a full asymptotic treatment of this approach. To our knowledge, there is no existing work that simultaneously uses the additive models with frailty to analyze the recurrent and terminal events. In this article, we propose the additive models with a shared frailty for the recurrent and terminal events simultaneously. To be specific, an additive rates model with a multiplicative frailty for the conditional recurrent event rate, and an additive hazards frailty model for the terminal event. A shared frailty is used to account for the association between recurrent and terminal events. An estimating equation approach is developed for the marginal and association parameters in the joint model. A rigorous justification is provided for the consistency and asymptotic normality of the resulting estimators.

The remainder of the paper is organized as follows. Section 2 describes the joint models. Section 3 presents an estimating procedure for the marginal and association parameters, and the asymptotic properties of the proposed estimators are established. Some simulation results for evaluating the proposed methods are reported in Section 4. An application to a medical cost study of chronic heart failure patients from the University of Virginia Health System is provided in Section 5, and some concluding remarks are given in Section 6. All proofs are relegated to the Appendix.

2. JOINT MODELING

Let $\tilde{N}^R(t)$ denote the number of recurrent events over the time interval $(0, t]$, and let $Z(t)$ be the $p\times 1$ vector of external time-dependent covariates (Kalbfleisch and Prentice, 2002). Let D be the terminal event time (e.g., death) and C be the censoring time, where the terminal event stops further recurrent events in that $d\tilde{N}^R(t)$ takes the value 0 for $t > D$. Write $T = C \wedge D$ and $\Delta(t) = I(T \geq t)$, where $a \wedge b = \min(a, b)$, and $I(\cdot)$ is the indictor function. Due to censoring, $\tilde{N}^R(t)$ is not fully observed, and the observed number of recurrent events is denoted by $N^R(t) = \tilde{N}^R(t \wedge T)$. Also, let $N^D(t) = \tilde{N}^D(t \wedge T)$ denote the observed number of the terminal event, where $\tilde{N}^D(t) = I(D \leq t)$. For a random sample of n subjects, the observed data consist of $\{N_i^R(t), N_i^D(t), T_i, \Delta_i(t), Z_i(t), 0 \leq t \leq T_i, i = 1, ..., n\}$.

Let υ be a nonnegative unobserved frailty that is assumed to be independent of $Z(t)$. Following Ye et al. (2007) and Kalbfleisch et al. (2013), we consider a (partial) marginal rate of the recurrent events given $Z(t)$, $D = s$ and υ, which is defined as

$$d\Lambda_R(t|\upsilon) = P\{d\tilde{N}^R(t) = 1|Z(t), D = s, \upsilon\}, \quad s \geq t.$$

Note that $d\Lambda_R(t|\upsilon)$ may depend on $Z(t)$ and the frailty υ, but does not depend on the terminal event time $D = s \geq t$. This implies that given covariates, υ accounts for the correlation between the recurrent and terminal events. Also, it follows that $d\Lambda_R(t|\upsilon) = P\{d\tilde{N}^R(t) = 1|Z(t), D \geq t, \upsilon\}$ (e.g. Kalbfleisch et al., 2013), which indicates that given $Z(t)$ and υ, $d\Lambda_R(t|\upsilon)$ specifies the marginal rate of the recurrent events among those subjects surviving to time t. For the analysis, we consider the following marginal additive rate model for the recurrent events:

$$d\Lambda_R(t|\upsilon) = \upsilon\{d\Lambda_0^R(t) + \beta^T Z(t)dt\}, \tag{1}$$

where β is a $p\times 1$ vector of regression parameters, and $d\Lambda_0^R(t)$ is an unspecified baseline rate function.

Let $d\Lambda_D(t|\upsilon) = P\{d\tilde{N}^D(t) = 1|Z(t), D \geq t, \upsilon\}$ be the hazard function for the terminal event given $Z(t)$ and υ. We specify the following additive hazards model for the terminal event:

$$d\Lambda_D(t|\upsilon) = \upsilon\{d\Lambda_0^D(t) + \alpha^T Z(t)dt\}, \tag{2}$$

where α is a $p \times 1$ vector of regression parameters, and $d\Lambda_0^D(t)$ is an unspecified baseline hazard function. For notational convenience, models (1) and (2) assume the same set of covariates $Z(t)$. The proposed estimation procedure can be extended in a straightforward manner to deal with different set of covariates for these two models. In addition, as in Ye et al. (2007), we assume that the frailty υ has a gamma distribution with mean 1 and variance θ, where $E(\upsilon) = 1$ is fixed for identifiability reasons. In what follows, we assume that given $Z(\cdot)$, the censoring time C is independent of $\{\tilde{N}^R(\cdot), \tilde{N}^D(\cdot), D, \upsilon\}$.

3. ESTIMATION PROCEDURES

Note that if the frailty υ is known, then by using the approaches of Schaubel et al. (2006) and Lin and Ying (1994) to models (1) and (2), respectively, we can get the estimating equations for β and α. However, in reality, υ is not observed, and thus we cannot directly use these approaches. For this, we consider an induced marginal model for β and α by taking the conditional expectation of (1) and (2) given $D \geq t$ and $Z(t)$. Under the assumed gamma distribution for υ, we have

$$
\begin{aligned}
(3) \qquad d\Lambda_R(t) &= P\{d\tilde{N}^R(t) = 1|Z(t), D \geq t\} \\
&= \psi(t)\{d\Lambda_0^R(t) + \beta^T Z(t)dt\},
\end{aligned}
$$

and

$$
\begin{aligned}
(4) \qquad d\Lambda_D(t) &= P\{d\tilde{N}^D(t) = 1|Z(t), D \geq t\} \\
&= \psi(t)\{d\Lambda_0^D(t) + \alpha^T Z(t)dt\},
\end{aligned}
$$

where

$$
\begin{aligned}
\psi(t) &= E[\upsilon|Z(t), D \geq t] \\
&= \Big[1 + \theta \int_0^t \{d\Lambda_0^D(u) + \alpha^T Z(u)du\}\Big]^{-1}.
\end{aligned}
$$

Define

$$dM_i^R(t) = dN_i^R(t) - \Delta_i(t)\psi_i(t)\{d\Lambda_0^R(t) + \beta^T Z_i(t)dt\},$$

and

$$dM_i^D(t) = dN_i^D(t) - \Delta_i(t)\psi_i(t)\{d\Lambda_0^D(t) + \alpha^T Z_i(t)dt\},$$

where

$$\psi_i(t) \equiv \psi_i(t; \theta, \alpha, \Lambda_0^D) = \frac{1}{1 + \theta \int_0^t \{d\Lambda_0^D(u) + \alpha^T Z_i(u)du\}}.$$

Under the assumed models, it follows from (3) and (4) that $M_i^R(t)$ and $M_i^D(t)$ are zero-mean stochastic processes. Thus, for given $\psi(t)$, similar to Schaubel et al. (2006) and Lin and Ying (1994), we can use the following estimating equations to estimate $\Lambda_0^R(t)$, $\Lambda_0^D(t)$, β and α:

$$
\begin{aligned}
&\sum_{i=1}^n \Big[dN_i^R(t) - \Delta_i(t)\psi_i(t)\{d\Lambda_0^R(t) + \beta^T Z_i(t)dt\}\Big] \\
&= 0, \quad 0 \leq t \leq \tau, \\
&\sum_{i=1}^n \Big[dN_i^D(t) - \Delta_i(t)\psi_i(t)\{d\Lambda_0^D(t) + \alpha^T Z_i(t)dt\}\Big] \\
&= 0, \quad 0 \leq t \leq \tau, \\
&\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\}\{dN_i^R(t) - \Delta_i(t)\psi_i(t)\beta^T Z_i(t)dt\} \\
&= 0, \\
&\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\}\{dN_i^D(t) - \Delta_i(t)\psi_i(t)\alpha^T Z_i(t)dt\} \\
&= 0,
\end{aligned}
$$

where τ is a prespecified constant such that $P(T_i \geq \tau) > 0$, and

$$\bar{Z}(t) = \frac{\sum_{i=1}^n \Delta_i(t)\psi_i(t)Z_i(t)}{\sum_{i=1}^n \Delta_i(t)\psi_i(t)}.$$

However, the weight function $\psi_i(t)$ also includes unknown parameters θ, α and $\Lambda_0^D(t)$, which must be estimated. In order to estimate θ, let $\tilde{N}^R(t)$ and D_i be defined as earlier, but with subject $i, i = 1, ..., n$. Define $\omega_{1i}(t) = E[\tilde{N}_i^R(t)|Z_i(t), D_i = t]$ and $\omega_{2i}(t) = E[\tilde{N}_i^R(t)|Z_i(t), D_i > t]$. Under the assumed models, we obtain

$$\omega_{1i}(t) = (\theta + 1)\psi_i(t) \int_0^t \{d\Lambda_0^R(u) + \beta^T Z_i(u)du\},$$

and

$$\omega_{2i}(t) = \psi_i(t) \int_0^t \{d\Lambda_0^R(u) + \beta^T Z_i(u)du\}.$$

Thus,

$$(5) \qquad \frac{\omega_{1i}(t)}{\omega_{2i}(t)} = \theta + 1,$$

which is a local dependence measure between recurrent and terminal events (Kalbfleisch et al., 2013). This also means that the subject with a termination event at time t is expected to have more recurrent events than the one with the termination event after t. In view of (5), as discussed in Kalbfleisch et al. (2013), we specify the following estimating equation for θ:

$$\sum_{i=1}^n \int_0^\tau \{N_i^R(t) - (\theta + 1)Q(t)\omega_{2i}(t)\}dN_i^D(t) = 0,$$

where

$$Q(t) = \frac{\sum_{i=1}^n \omega_{2i}(t)^{-1}\Delta_i^*(t)N_i^R(t)}{\sum_{i=1}^n \Delta_i^*(t)},$$

and $\Delta_i^*(t) = \Delta_i(t)\{1 - N_i^D(t)\}$ is an indictor that subject i is at risk at t and dies after t.

Let $\gamma = (\beta^T, \alpha^T, \theta, \Lambda_0^R, \Lambda_0^D)^T$. We propose to estimate γ using the solutions to the equations $U(\gamma) = (U_1^T, U_2^T, U_3, U_4, U_5)^T = 0$, where

$$
\begin{aligned}
U_1 &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \\
&\quad \times \{dN_i^R(t) - \Delta_i(t)\psi_i(t)\beta^T Z_i(t)dt\}, \\
U_2 &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \\
&\quad \times \{dN_i^D(t) - \Delta_i(t)\psi_i(t)\alpha^T Z_i(t)dt\}, \\
U_3 &= \sum_{i=1}^n \int_0^\tau \{N_i^R(t) - (\theta + 1)Q(t)\omega_{2i}(t)\}dN_i^D(t), \\
U_4 &= \sum_{i=1}^n \Big[dN_i^R(t) - \Delta_i(t)\psi_i(t) \\
&\quad \times \{d\Lambda_0^R(t) + \beta^T Z_i(t)dt\}\Big], \quad 0 \leq t \leq \tau,
\end{aligned}
$$

$$U_5 = \sum_{i=1}^{n} \Big[dN_i^D(t) - \Delta_i(t)\psi_i(t) \times \big\{ d\Lambda_0^D(t) + \alpha^T Z_i(t)dt \big\} \Big], \quad 0 \le t \le \tau.$$

Let $\hat{\beta}, \hat{\alpha}, \hat{\theta}, \hat{\Lambda}_0^R(t)$ and $\hat{\Lambda}_0^D(t)$ denote the solutions to $U(\gamma) = 0$, where the estimates $\hat{\Lambda}_0^R(t)$ and $\hat{\Lambda}_0^D(t)$ will be a piecewise constant function with jumps only at the observed recurrent event times (across all subjects) and the observed death times, respectively. Since estimation of each parameter depends on a subset of the other parameters, the solutions to the estimating equations can be obtained through a recursive procedure.

Here, we propose the following iterative algorithm to solve $U(\gamma) = 0$, which is robust and effective in the simulation studies in Section 4.

Step 0. Choose initial estimates $\theta^{(0)}$, $\alpha^{(0)}$ and $\Lambda_0^{D(0)}(t)$.

Step 1. Let $\psi_i^{(0)}(t) = \psi_i(t; \theta^{(0)}, \alpha^{(0)}, \Lambda_0^{D(0)})$. Put $\psi_i^{(0)}(t)$ into $U_1 = 0$, $U_2 = 0$, $U_4 = 0$ and $U_5 = 0$, and solve the resulting equations for updated estimates $\beta^{(1)}$, $\alpha^{(1)}$, $\Lambda_0^{R(1)}(t)$ and $\Lambda_0^{D(1)}(t)$.

Step 2. For given $\beta^{(1)}$ and $\Lambda_0^{R(1)}(t)$ with $\psi_i^{(1)}(t) = \psi_i(t; \theta^{(0)}, \alpha^{(1)}, \Lambda_0^{D(1)})$, obtain $\theta^{(1)}$ by solving $U_3 = 0$.

Step 3. Return to Step 1 with updated estimates until convergence.

Note that many choices can be used for the initial estimates $\theta^{(0)}$, $\alpha^{(0)}$ and $\Lambda_0^{D(0)}(t)$. Usually we can take $\theta^{(0)} = 1$, $\alpha^{(0)} = 0$, and set $\Lambda_0^{D(0)}(t)$ to be the Nelson-Aalen type estimate of the cumulative baseline hazard function. For the convergence, also several criteria can be applied, and in the simulation studies below, we used the absolute differences $\le 10^{-3}$ between the iterative estimates of the parameters. The algorithm converges most times in general, but non-convergence could occur occasionally depending on the set-ups. In the simulation studies below, the percentage of non-convergence is about 0.6% under different set-ups with sample size $n = 200$, and the algorithm always converges for the case of sample size $n = 400$.

Let $\eta = (\beta^T, \alpha^T, \theta)^T$, $\hat{\eta} = (\hat{\beta}^T, \hat{\alpha}^T, \hat{\theta})^T$, and $\eta_0 = (\beta_0^T, \alpha_0^T, \theta_0)^T$ be the true value of η. We now describe the asymptotic properties of the proposed estimators. First we consider the existence, uniqueness, and strong consistency of $\hat{\eta}$, $\hat{\Lambda}_0^R(t)$ and $\hat{\Lambda}_0^D(t)$. The results are summarized in the following theorem with the proof given in the Appendix.

Theorem 1. *Under the regularity conditions (C1)–(C4) stated in the Appendix,* $\hat{\eta}$, $\hat{\Lambda}_0^R(t)$ *and* $\hat{\Lambda}_0^D(t)$ *exist and are unique. Moreover,* $\hat{\eta}$ *is strongly consistent to* η_0, *and* $\hat{\Lambda}_0^R(t) \to \Lambda_0^R(t)$ *and* $\hat{\Lambda}_0^D(t) \to \Lambda_0^D(t)$ *almost surely uniformly in* $t \in [0, \tau]$.

The asymptotic distributions of $\hat{\eta}$, $\hat{\Lambda}_0^R(t)$ and $\hat{\Lambda}_0^D(t)$ are given in the next theorem.

Theorem 2. *Under the regularity conditions (C1)–(C4) stated in the Appendix,* $n^{1/2}(\hat{\eta} - \eta_0)$ *converges in distribution to a normal random vector with mean zero and covariance matrix* $\Gamma^{-1}\Sigma(\Gamma^T)^{-1}$, *where* Σ *and* Γ *are given in the Appendix. Furthermore,* $n^{1/2}\{\hat{\Lambda}_0^R(t) - \Lambda_0^R(t)\}$ *and* $n^{1/2}\{\hat{\Lambda}_0^D(t) - \Lambda_0^D(t)\}$ *jointly converge weakly to a zero-mean bivariate Gaussian process for* $t \in [0, \tau]$.

The asymptotic covariance matrix can be consistently estimated by the usual plug-in method. However, Σ involves the Hadamard derivatives of some functions with respect to $\Lambda_0^D(t)$ and $\Lambda_0^R(t)$, and has complicated analytic form. Therefore, it may be unstable to estimate Σ using the usual plug-in method with moderate sample size. Here, we adopt the bootstrap method to estimate the covariance matrix of $\hat{\eta}$. The accuracy of the bootstrap method depends on the sample size and the number of bootstrap samples. In the following simulation studies with sample sizes $n = 200$ and 400, we find that the covariance estimation is fairly accurate when 100 bootstrap samples are used.

4. SIMULATION STUDIES

We conducted simulation studies to examine the finite sample properties of the proposed estimators. In the study, the covariate Z was generated from a Bernoulli distribution with success probability 0.5, and the frailty υ was generated from a gamma distribution with unit mean and variance $\theta = 0$ or 0.5. For given the frailty υ and the covariates Z, the terminal event time D was generated from model (2) with $\Lambda_0^D(t) = 0.2t$ and $\alpha = 0.5$. The recurrent event times were generated from a Poisson process with the intensity $\upsilon\{\lambda_0(t) + \beta_0 Z\}$, where $\lambda_0(t) \equiv 1.8$, and $\beta_0 = 0.5$ or 1. The censoring time was generated from a uniform distribution on (1,6), and the censoring rate was about 40%. The average number of the observed events per subject was about 3 under the preceding settings. The results presented below are based on 500 replications with sample sizes $n = 200$ or 400, and final estimators were reached at convergence. The asymptotic variance was estimated using the bootstrap method with 100 bootstrap samples, which were found to be adequate. We also found that variance estimators from 100 and 1000 bootstrap samples are close to each other.

The simulation results for estimation of β, α and θ are summarized in the first half of Table 1 with $n = 400$, which includes the bias (Bias) given by the sample means of proposed estimates minus the true values, the sample standard errors (SE), the sample mean of the standard error estimate (SEE), and the 95% empirical coverage probabilities (CP) based on the normal approximation. The results indicate that our proposed method performed well for the situations considered here. Specifically, the proposed estimators were nearly unbiased, and the standard error estimators were very accurate based on the bootstrap method. The 95% empirical coverage probabilities were reasonable.

Table 1. Simulation results for the estimation of α, β and θ

β	Estimate	$\theta = 0$				$\theta = 0.5$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
		$Z \sim$ Bernoulli(0.5) with $n = 400$							
0.25	β	0.0023	0.1328	0.1277	0.936	0.0121	0.2292	0.2316	0.950
	α	0.0064	0.0656	0.0645	0.938	0.0075	0.0891	0.0913	0.957
	θ	0.0019	0.0434	0.0436	0.951	0.0010	0.1038	0.0962	0.937
0.5	β	0.0044	0.1355	0.1352	0.950	0.0073	0.2383	0.2496	0.954
	α	0.0020	0.0637	0.0646	0.954	0.0108	0.0878	0.0920	0.948
	θ	0.0021	0.0424	0.0418	0.951	0.0039	0.0961	0.0947	0.945
1	β	-0.0007	0.1450	0.1507	0.960	0.0181	0.2804	0.2832	0.959
	α	0.0034	0.0631	0.0648	0.954	0.0081	0.0854	0.0913	0.953
	θ	0.0001	0.0377	0.0392	0.966	0.0025	0.0914	0.0912	0.944
		$Z(t) = \tilde{Z}t$ with $\tilde{Z} \sim$ Uniform(0, 1) with $n = 200$							
0.25	β	0.0284	0.2449	0.2521	0.946	0.0334	0.4750	0.4893	0.952
	α	0.0321	0.1587	0.1677	0.954	0.0249	0.2479	0.2632	0.944
	θ	0.0155	0.0646	0.0661	0.956	0.0013	0.1245	0.1382	0.956
0.5	β	0.0348	0.2607	0.2678	0.938	0.0315	0.5188	0.5208	0.942
	α	0.0273	0.1619	0.1681	0.950	0.0285	0.2523	0.2707	0.942
	θ	0.0170	0.0593	0.0570	0.966	-0.0013	0.1305	0.1388	0.946
1	β	0.0367	0.2700	0.2937	0.956	0.0036	0.5855	0.5930	0.952
	α	0.0375	0.1502	0.1720	0.960	0.0148	0.2411	0.2594	0.940
	θ	0.0159	0.0548	0.0660	0.952	-0.0018	0.1176	0.1355	0.950

Table 2. Comparison results for our method and the naïve method with $n = 200$

θ	β	Estimate	Our method				Naïve method			
			Bias	SE	SEE	CP	Bias	SE	SEE	CP
0	0.25	β	0.0066	0.1769	0.1791	0.950	0.0066	0.1587	0.1606	0.950
		α	-0.0001	0.0838	0.0898	0.962	-0.0001	0.0770	0.0835	0.964
		θ	0.0020	0.0586	0.0587	0.932				
	0.5	β	0.0136	0.1866	0.1880	0.948	0.0107	0.1729	0.1660	0.942
		α	0.0002	0.0875	0.0904	0.952	-0.0006	0.0814	0.0838	0.946
		θ	0.0007	0.0576	0.0570	0.962				
	1	β	0.0019	0.1970	0.2121	0.956	0.0011	0.1760	0.1786	0.948
		α	0.0129	0.0911	0.0917	0.956	0.0119	0.0863	0.0844	0.950
		θ	0.0008	0.0537	0.0557	0.964				
0.5	0.25	β	0.0007	0.3239	0.3280	0.942	-0.3891	0.2168	0.2120	0.552
		α	0.0078	0.1218	0.1283	0.954	-0.1772	0.0675	0.0701	0.290
		θ	-0.0001	0.1288	0.1328	0.948				
	0.5	β	0.0167	0.3566	0.3516	0.950	-0.4521	0.2187	0.2246	0.474
		α	0.0111	0.1226	0.1282	0.942	-0.1760	0.0682	0.0700	0.298
		θ	0.0020	0.1335	0.1331	0.948				
	1	β	0.0183	0.4021	0.3961	0.936	-0.5900	0.2526	0.2500	0.354
		α	0.0087	0.1239	0.1260	0.948	-0.1752	0.0689	0.0709	0.326
		θ	-0.0074	0.1292	0.1259	0.930				

Next, we conducted simulation studies when the baseline rate and hazard functions and the covariate are all time-dependent. In the study, the setup was the same as in the first half of Table 1, except that $\Lambda_0^D(t) = 0.1t^2$, $\lambda_0(t) = t$, and the time-dependent covariate $Z(t)$ was taken as $\tilde{Z}t$, where $\tilde{Z}$ was generated from a uniform distribution on $(0, 1)$. The results are summarized in the second half of Table 1 with $n = 200$. Simulation results show that the proposed method still performed well for the situations considered here.

For comparison, we also considered the naïve method regarding the terminal event as an independent censoring time (Schaubel et al., 2006). We used the same setup as in the first half of Table 1, and the comparison results are given in Table 2 with $n = 200$. We can observe that the naïve method yielded consistent estimators when the terminal event was

Table 3. Sensitivity analysis for the misspecification of the frailty with $n = 200$

Frailty	Estimate	Bias	SE	SEE	CP
Uniform	β	0.0270	0.3635	0.3647	0.950
	α	0.0339	0.1328	0.1361	0.964
Log-normal	β	0.0151	0.3303	0.3358	0.947
	α	0.0103	0.1287	0.1312	0.955
Poisson	β	0.0013	0.2331	0.2340	0.948
	α	0.0104	0.1006	0.1030	0.952

Table 4. Analysis results for the medical cost data of heart failure patients

	β			α			θ
	Male	White	Age	Male	White	Age	Correlation
				Our method			
Est	-0.0108	-1.3587	1.9264	0.0298	-0.0239	0.0592	1.2792
SE	0.0507	0.0514	0.0492	0.0066	0.0054	0.0037	0.0612
p-value	0.4159	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
				Naïve method			
Est	-0.3846	-0.5820	0.4612	0.0161	-0.0303	0.0472	
SE	0.2099	0.2388	0.1208	0.0141	0.0157	0.0087	
p-value	0.0335	0.0074	0.0001	0.1264	0.0269	<0.0001	

Note: Est is the estimate of the parameter, and SE is the standard error estimate.

independent of the recurrent event (i.e., $\theta = 0$). Under such situation, both methods provided reasonable and comparable estimates for β and α, and the variances of our method were only slightly larger than those of the naïve method. This was because the latter utilized the independent assumption in estimation. However, when such independent assumption was violated (i.e., $\theta \neq 0$), the naïve method may have led to biases, and the proposed method worked well.

Finally, we conducted simulation studies to examine the performance of the proposed estimators when the gamma distribution was misspecified. We considered three scenarios for the frailty v: (i) v had a uniform distribution on $(0, 2)$; (ii) v followed a log-normal distribution with unit mean and variance 0.6; (iii) v was generated as one-tenth of a Poisson variable with mean 10. The other setups were the same as in the first half of Table 1 with $\beta = 0.5$ and $\alpha = 0.5$. The results are presented in Table 3 with $n = 200$. It can be seen that the proposed estimators still performed reasonably well for the three scenarios considered, and the proposed method was robust to misspecification of the frailty distribution.

5. AN APPLICATION

For illustration purposes, we applied the proposed methods to the medical cost data of chronic heart failure patients from the clinical data repository at the University of Virginia Health System. The data set included a total of 1475 patients aged 60–89 years who were first diagnosed with heart failure and treated in 2004. The follow-up ended with each patient's last hospital admission up to July 31, 2006, or death date, which was obtained from the Death Certificate Data at the Virginia Department of Vital Statistics. During follow-up, 297 patients (20%) died and others were censored. For each patient, three baseline covariates were recorded: gender, race and age. Preliminary studies indicated that patients visiting the hospital more often tended to have a higher mortality rate. That is, the death (terminal event) was likely to be strongly correlated with the hospital visits (recurrent events) of interest (Liu et al., 2008; Sun et al., 2012). Here, we applied the proposed methods to jointly analyze the death hazard and the rate of hospital visits, and focused on the effects of gender, race and age on the hospital visits and death.

For the analysis, let Z_{i1} be a binary indicator of gender (male$\,=1$, female$\,=0$), and Z_{i2} be a binary indicator of race (white$\,=1$, nonwhite$\,=0$). Following Sun et al. (2012), we defined Z_{i3} as the age group, taking values 0, 1 and 2 for 60–69, 70–79 and 80–89 years, respectively. Let τ be the longest follow-up time. The asymptotic variance was estimated by the bootstrap method with 100 bootstrap samples. The analysis results are summarized in Table 4. These results show that both race and age had significant effects on the rate of hospital visits, but gender did not seem to be significantly related to the hospital visits. In particular, white patients tended to visit hospital at less risk, and older patients were more likely to visit the hospital. In addition, gender, race and age had significant effects on the mortality rate. Specifically, male and older patients tended to have higher mortality rates, while white patients had a lower mortality rate. The estimate $\hat{\theta} = 1.2792$ (p-value < 0.0001) indicates that there was a significantly positive association between the hospital visits and the death. That is, patients who tended to visit hospital more frequently had a higher mortality rate. Moreover, in view of (5), a patient who is known to die at time t is expected to have more than twice (2.279) many hospital visits as a patient with identical covariates who has not died by the time t.

For comparison, we also analyzed the data with the naïve method regarding the terminal event as an independent censoring time (Schaubel et al., 2006). The comparison results are also given in Table 4. It can be seen that although the covariate effects in the recurrent and terminal event models had the same directions between the naïve model and the joint model across the various covariates, the covariate effects in the two models had different sizes. Specifically, for the hospital visits, the naïve method overestimated significantly the effect of gender, and underestimated the effects of race and age. For the death, the naïve estimators for all the effects were significantly different from ours. This is because, as shown in the simulation, the naïve method ignores the correlation between the hospital visits and the death, and leads to biases. In contrast, the proposed method uses a shared frailty to account for the association between the hospital visits and the death, and yields unbiased estimates.

6. DISCUSSION

In this article, we proposed a joint model for analyzing recurrent event data with a terminal event, and a shared frailty was used to account for the association between recurrent and terminal events. An estimating equation approach was developed for the model parameters, which yielded consistent and asymptotically normal estimators. The simulation results showed that the proposed estimation approach performs well, and the method was robust to misspecification of the frailty distribution, at least for the situations considered. An application to a medical cost study of chronic heart failure patients was provided to illustrate our method.

Note that models (1) and (2) allow a positive association between recurrent and terminal events. Although these models fit the example discussed in Section 5 well, the negative association may exist in some situations. For this case, we could apply the model

$$d\Lambda_R(t|\upsilon) = \upsilon^{-1}\{d\Lambda_0^R(t) + \beta^T Z(t)dt\},$$

and model (2), where υ has a gamma distribution with mean 1 and variance $\theta < 1$. It can be checked that $\psi^*(t) = E[\upsilon^{-1}|Z(t), D \geq t] = 1/(\psi(t)(1-\theta))$, and $\omega_{1i}(t)/\omega_{2i}(t) = 1-\theta$. This implies that the association between recurrent and terminal events is negative, since the subject with a terminal event at time t is expected to have less recurrent events than the one with the terminal event after t. In addition, by replacing $\psi(t)$ and $1+\theta$ with $\psi^*(t)$ and $1-\theta$, respectively, in U_1 and U_3 and U_4, the same estimating equations can be constructed as in previous sections.

In the joint model, the multiplicative frailty does not have an interpretation as an unobserved covariate. To obtain this interpretation, similarly to Pipper and Martinussen (2004) and Liu and Wu (2011), we could consider an additive frailty for the recurrent event rate as follows:

$$d\Lambda_R(t|\upsilon) = d\Lambda_0^R(t) + \beta^T Z(t)dt + \upsilon dt.$$

The proposed estimation procedure can be easily extended to the additive frailty model with a slight modification for the equations U_1 and U_4 as

$$U_1^* = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}^*(t)\} \times\{dN_i^R(t) - \Delta_i(t)[\beta^T Z_i(t) + \psi_i(t)]dt\},$$

and

$$U_4^* = \sum_{i=1}^n \Big[dN_i^R(t) - \Delta_i(t) \times \{d\Lambda_0^R(t) + \beta^T Z_i(t)dt + \psi_i(t)dt\}\Big], \quad 0 \leq t \leq \tau,$$

where $\bar{Z}^*(t) = \sum_{i=1}^n \Delta_i(t)Z_i(t)/\sum_{i=1}^n \Delta_i(t)$. Furthermore, we also could consider the additive frailty for both the recurrent event rate and terminal event hazard functions as follows:

$$d\Lambda_R(t|\upsilon) = d\Lambda_0^R(t) + \beta^T Z(t)dt + \upsilon dt,$$

and

$$d\Lambda_D(t|\upsilon) = d\Lambda_0^D(t) + \alpha^T Z(t)dt + \upsilon dt.$$

In a similar manner, we can obtain the following estimating equations for γ:

$$\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}^*(t)\} \times\{dN_i^R(t) - \Delta_i(t)\big[\beta^T Z_i(t) + \psi(t)\big]dt\} = 0,$$

$$\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}^*(t)\} \times\{dN_i^D(t) - \Delta_i(t)\big[\alpha^T Z_i(t) + \psi(t)\big]dt\} = 0,$$

$$\sum_{i=1}^n \int_0^\tau \Big\{N_i^R(t) - \big[\frac{\theta\psi(t)^2 t}{\lambda_0^D(t) + \alpha^T Z_i(t) + \psi(t)} + Q(t)\omega_{2i}(t)\big]\Big\} dN_i^D(t) = 0,$$

$$\sum_{i=1}^n \Big[dN_i^R(t) - \Delta_i(t) \times\{d\Lambda_0^R(t) + [\beta^T Z_i(t) + \psi(t)]dt\}\Big] = 0, \ 0 \leq t \leq \tau,$$

$$\sum_{i=1}^n \Big[dN_i^D(t) - \Delta_i(t) \times\{d\Lambda_0^D(t) + [\alpha^T Z_i(t) + \psi(t)]dt\}\Big] = 0, \ 0 \leq t \leq \tau,$$

where $\psi(t) = E[\upsilon|D \geq t, Z(t)] = (1+\theta t)^{-1}$, $\lambda_0^D(t) = d\Lambda_0^D(t)/dt$, and

$$\omega_{2i}(t) = E[\tilde{N}_i^R(t)|Z_i(t), D_i > t] = \psi(t)t + \int_0^t \{d\Lambda_0^R(u) + \beta^T Z_i(u)du\}.$$

The above estimating equations involve estimating $\lambda_0^D(t)$, which could be obtained by using some kernel-smoothing methods. The asymptotic properties of the resulting estimators would be worthy of investigation in future studies.

Since the proposed estimation procedure was given by using the generalized estimating equation approach, it may not be the most efficient method in the framework of semiparametric models. Further research is needed to develop a simple and more efficient inference procedure. In addition, in order to assess the adequacy of models (1) and (2) for the data, we can consider some simple graphical and numerical procedures based on the residuals $d\hat{M}_i^R(t)$ and $d\hat{M}_i^D(t)$ as in Lin et al. (1993) and Zeng and Cai (2010), where $d\hat{M}_i^R(t) = dN_i^R(t) - \Delta_i(t)\psi_i(t;\hat{\theta},\hat{\alpha},\hat{\Lambda}_0^D)\{d\hat{\Lambda}_0^R(t) + \hat{\beta}^T Z_i(t)dt\}$, and $d\hat{M}_i^D(t) = dN_i^D(t) - \Delta_i(t)\psi_i(t;\hat{\theta},\hat{\alpha},\hat{\Lambda}_0^D)\{d\hat{\Lambda}_0^D(t) + \hat{\alpha}^T Z_i(t)dt\}$. It would be worthwhile to further address this issue both theoretically and numerically.

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, Professor Heping Zhang, and two referees for their constructive and insightful comments and suggestions that greatly improved the article. This research was partly supported by the National Natural Science Foundation of China (Grant Nos. 11231010 and 11171330), Key Laboratory of RCSDS, CAS (No. 2008DP173182), BCMIIS and R01 HS 020263 from Agency of Healthcare Research and Quality of USA.

APPENDIX: PROOF OF ASYMPTOTIC RESULTS

In order to study the asymptotic properties of the proposed estimators, we need the following regularity conditions:

(C1) $\{N_i^R(\cdot), N_i^D(\cdot), T_i, Z_i(\cdot)\}$, $i = 1, ..., n$, are independent and identically distributed.
(C2) $N_i^R(\tau)$ is bounded almost surely, and $\Pr(T \geq \tau) > 0$.
(C3) $Z_i(t)$ is almost surely of bounded variation on $[0, \tau]$.
(C4) There exist a compact set $\mathcal{B}$ of η_0 such that for $\eta \in \mathcal{B}$, $\Gamma(\eta)$ is nonsingular, where $\Gamma(\eta)$ is the limit of $-\partial\tilde{U}(\eta)/\partial\eta^T$ with $\tilde{U}(\eta)$ defined in (A4).

Proof of Theorem 1. Define

$$S_k(t;\eta,\Lambda^D) = \frac{1}{n}\sum_{i=1}^n \Delta_i(t)\psi_i(t;\theta,\alpha,\Lambda^D)Z_i(t)^{\otimes k}, \quad k = 0, 1,$$

where $a^{\otimes 0} = 1$ and $a^{\otimes 1} = a$ for any vector a. Let $\bar{N}^D(t) = n^{-1}\sum_{i=1}^n N_i^D(t)$, $\bar{N}^R(t) = n^{-1}\sum_{i=1}^n N_i^R(t)$, and $s_k(t;\eta,\Lambda^D)$ be the limits of $S_k(t;\eta,\Lambda^D)$ $(k = 0, 1)$. Also for given η, let $\tilde{\Lambda}_0^R(t;\eta)$ and $\tilde{\Lambda}_0^D(t;\eta)$ denote the solutions to $U_4 = 0$ and $U_5 = 0$. Then

$$\tilde{\Lambda}_0^R(t;\eta) = \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dN_i^R(u) - \Delta_i(u)\psi_i(u;\theta,\alpha,\tilde{\Lambda}_0^D)\beta^T Z_i(u)du}{S_0(u;\eta,\tilde{\Lambda}^D)},$$

and $\tilde{\Lambda}_0^D(t;\eta)$ satisfies the following integral equation:

$$\text{(A1)} \quad \tilde{\Lambda}_0^D(t;\eta) = \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dN_i^D(u) - \Delta_i(u)\psi_i(u;\theta,\alpha,\tilde{\Lambda}_0^D)\alpha^T Z_i(u)du}{S_0(u;\eta,\tilde{\Lambda}_0^D)},$$

which is a nonlinear Volterra integral equation and has a unique solution (Polyanin and Manzhirov, 2008). For given η, let $\Lambda^D(t;\eta)$ denote the solution to the following nonlinear Volterra integral equation:

$$\text{(A2)} \quad \Lambda^D(t;\eta) = \int_0^t \frac{E\{d\bar{N}^D(u)\} - \alpha^T s_1(u;\eta,\Lambda^D)du}{s_0(u;\eta,\Lambda^D)},$$

which also is a nonlinear Volterra integral equation and has a unique solution with $\Lambda^D(t;\eta_0) \equiv \Lambda_0^D(t)$. Let

$$H_n(t) = \frac{1}{n}\sum_{i=1}^n \int_0^t \theta\Delta_i(u)\tilde{\psi}_i(u;\theta,\alpha,\tilde{\Lambda}^D)\psi_i(u;\theta,\alpha,\Lambda^D) \times \frac{d\bar{N}^D(u) - \alpha^T S_1(u;\eta,\Lambda^D)du}{S_0(u;\eta,\tilde{\Lambda}_0^D)S_0(u;\eta,\Lambda^D)}.$$

Then it follows from (A1) and (A2) that

$$\tilde{\Lambda}_0^D(t;\eta) - \Lambda^D(t;\eta) = \int_0^t \left[\tilde{\Lambda}_0^D(u;\eta) - \Lambda^D(u;\eta)\right] dH_n(u) + \mathcal{E}_n(t;\eta),$$

where

$$\mathcal{E}_n(t;\eta) = \int_0^t \frac{d\bar{N}^D(u) - \alpha^T S_1(u;\eta,\Lambda^D)du}{S_0(u;\eta,\Lambda^D)} - \Lambda^D(t;\eta).$$

By the uniform strong law of large numbers, we have that $\mathcal{E}_n(t;\eta) \to 0$ almost surely uniformly in $t \in [0, \tau]$ and a compact set $\mathcal{B}$ of η_0. which is a linear Volterra integral equation, and the solution is

$$\text{(A3)} \quad \tilde{\Lambda}_0^D(t;\eta) - \Lambda^D(t;\eta) = \frac{1}{P_n(t)}\int_0^t P_n(u-)d\mathcal{E}_n(u;\eta),$$

where $P_n(t) = \prod_{s \leq t}\{1 - dH_n(s)\}$ is the product-integral of $H_n(s)$ over $[0, t]$ (Andersen et al., 1993, Theorem II.6.3), and $P_n(u-)$ is the left-continuous version of $P_n(u)$. Using the asymptotic properties of the product-integral (Gill and Johansen, 1990), the uniform strong law of large numbers (Pollard, 1990), and Lemma A.1 of Lin and Ying (2001), we obtain that $\tilde{\Lambda}_0^D(t;\eta)$ converges to $\Lambda^D(t;\eta)$ almost surely uniformly in $t \in [0,\tau]$ and $\eta \in \mathcal{B}$. Let

$$\Lambda^R(t;\eta) = \int_0^t \frac{E\{d\bar{N}^R(u)\} - \beta^T s_1(u;\eta,\Lambda^D)du}{s_0(u;\eta,\Lambda^D)},$$

with $\Lambda^R(t;\eta_0) \equiv \Lambda_0^R(t)$. In a similar manner, we get that $\tilde{\Lambda}_0^R(t;\eta)$ converges to $\Lambda^R(t;\eta)$ almost surely uniformly in $t \in [0,\tau]$ and $\eta \in \mathcal{B}$. Thus, to prove the existence and uniqueness of $\hat{\eta}$, $\hat{\Lambda}_0^D(t)$ and $\hat{\Lambda}_0^R(t)$, it suffices to show that there exists a unique solution to

$$\text{(A4)} \qquad \tilde{U}(\eta) = (\tilde{U}_1(\eta)^T, \tilde{U}_2(\eta)^T, \tilde{U}_3(\eta))^T = 0,$$

where

$$\begin{aligned}
\tilde{U}_1(\eta) &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t;\eta)\} \\
&\quad \times \{dN_i^R(t) - \Delta_i(t)\psi_i(t;\theta,\alpha,\tilde{\Lambda}_0^D)\beta^T Z_i(t)dt\}, \\
\tilde{U}_2(\eta) &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t;\eta)\} \\
&\quad \times \{dN_i^D(t) - \Delta_i(t)\psi_i(t;\theta,\alpha,\tilde{\Lambda}_0^D)\alpha^T Z_i(t)dt\}, \\
\tilde{U}_3(\eta) &= \sum_{i=1}^n \int_0^\tau \{N_i^R(t) - (\theta+1)Q(t;\eta)\omega_{2i}(t;\eta)\} \\
&\quad \times dN_i^D(t),
\end{aligned}$$

and $\bar{Z}(t;\eta)$, $\omega_{2i}(t;\eta)$ and $Q(t;\eta)$ are defined as $\bar{Z}(t)$, $\omega_{2i}(t)$ and $Q(t)$ with $\psi_i(t)$ replaced by $\psi_i(t;\eta,\tilde{\Lambda}_0^D)$, respectively. Let $\hat{\Gamma}(\eta) = -n^{-1}\partial\tilde{U}(\eta)/\partial\eta^T$. Then by the uniform strong law of large numbers and the uniform convergence of $\tilde{\Lambda}_0^D(t;\hat{\eta})$ and $\tilde{\Lambda}_0^R(t;\hat{\eta})$, we get that $\hat{\Gamma}(\eta)$ converges to a nonrandom function $\Gamma(\eta)$ uniformly in $\eta \in \mathcal{B}$. Also it can be checked that $n^{-1}\tilde{U}(\eta_0) \to 0$ almost surely. Thus, the convergence of $\hat{\Gamma}(\eta)$ and condition (C4) imply that for $\eta \in \mathcal{B}$, $\hat{\Gamma}(\eta)$ is nonsingular when n is large enough. Hence it follows from the inverse function theorem (Rudin, 1976, page 221) that within $\mathcal{B}$, there exist a unique solution $\hat{\eta}$ to $\tilde{U}(\eta) = 0$ for every sufficiently large n. Therefore, there exist unique estimators $\hat{\eta}$, $\hat{\Lambda}_0^D(t) \equiv \tilde{\Lambda}_0^D(t;\hat{\eta})$ and $\hat{\Lambda}_0^R(t) \equiv \tilde{\Lambda}_0^R(t;\hat{\eta})$ $(0 \le t \le \tau)$. Note that by the Taylor expansion,

$$n^{-1}\tilde{U}(\hat{\eta}) - n^{-1}\tilde{U}(\eta_0) = -\hat{\Gamma}(\eta_0)(\hat{\eta} - \eta_0) + o(\|\hat{\eta} - \eta_0\|).$$

It then follows that almost surely,

$$\hat{\Gamma}(\eta_0)(\hat{\eta} - \eta_0) + o(\|\hat{\eta} - \eta_0\|) = o(1).$$

Due to the nonsingularity of $\hat{\Gamma}(\eta)$ in $\mathcal{B}$, the above equality implies that $\hat{\eta}$ is strongly consistent. In addition, the uniform convergence of $\tilde{\Lambda}_0^D(t;\eta)$ and $\tilde{\Lambda}_0^R(t;\eta)$ implies that $\hat{\Lambda}_0^D(t) \to \Lambda^D(t;\eta_0) \equiv \Lambda_0^D(t)$ and $\hat{\Lambda}_0^R(t) \to \Lambda^R(t;\eta_0) \equiv \Lambda_0^R(t)$ almost surely uniformly in $t \in [0,\tau]$.

Proof of Theorem 2. Let $\tilde{\psi}_i(u;\Lambda^D) = \psi_i(u;\theta_0,\alpha_0,\Lambda^D)$. Define

$$\begin{aligned}
\tilde{G}_1(t) &= \frac{1}{n}\sum_{i=1}^n \int_0^t \theta_0\Delta_i(u)\tilde{\psi}_i(u;\tilde{\Lambda}_0^D)\tilde{\psi}_i(u;\Lambda_0^D) \\
&\qquad \times \frac{\{\alpha_0^T Z_i(u)du + d\Lambda_0^D(u)\}}{S_0(u;\eta_0,\tilde{\Lambda}_0^D)}.
\end{aligned}$$

It can be checked that

$$\begin{aligned}
\tilde{\Lambda}_0^D(t;\eta_0) - \Lambda_0^D(t) &= \int_0^t \{\tilde{\Lambda}_0^D(u;\eta_0) - \Lambda_0^D(u)\}d\tilde{G}_1(u) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dM_i^D(u)}{S_0(u;\eta_0,\tilde{\Lambda}_0^D)}.
\end{aligned}$$

Similarly to (A3), we have

$$\tilde{\Lambda}_0^D(t;\eta_0) - \Lambda_0^D(t) = \frac{1}{\tilde{R}(t)}\int_0^t \tilde{R}(u-)\frac{\sum_{i=1}^n dM_i^D(u)}{nS_0(u;\eta_0,\tilde{\Lambda}_0^D)},$$

where $\tilde{R}(t) = \prod_{s \le t}\{1 - d\tilde{G}_1(s)\}$ is the product-integral of $\tilde{G}_1(s)$ over $[0,t]$. Using the uniform convergence of $\tilde{\Lambda}_0^D(t;\eta_0)$, the uniform strong law of large numbers and and Lemma A.1 of Lin and Ying (2001), we get that uniformly in $t \in [0,\tau]$,

$$\text{(A5)} \qquad \tilde{\Lambda}_0^D(t;\eta_0) - \Lambda_0^D(t) = \frac{1}{n}\sum_{i=1}^n \phi_{1i}(t) + o_p(n^{-1/2}),$$

where

$$\phi_{1i}(t) = \frac{1}{R(t)}\int_0^t R(u-)\frac{dM_i^D(u)}{s_1^{(0)}(u)},$$

$R(t)$ is the limit of $\tilde{R}(t)$, and $s_1^{(0)}(u) \equiv s_1^{(0)}(u;\eta_0,\Lambda_0^D)$. Let

$$\begin{aligned}
\tilde{G}_2(t) &= \frac{1}{n}\sum_{i=1}^n \int_0^t \theta_0\Delta_i(u)\tilde{\psi}_i(u;\tilde{\Lambda}_0^D)\tilde{\psi}_i(u;\Lambda_0^D) \\
&\qquad \times \frac{\{\beta_0^T Z_i(u)du + d\Lambda_0^R(u)\}}{S_0(u;\eta_0,\tilde{\Lambda}_0^D)},
\end{aligned}$$

and $G_2(t)$ be the limit of $\tilde{G}_2(t)$. It then follows from (A5) that

$$\begin{aligned}
\text{(A6)} \qquad & \tilde{\Lambda}_0^R(t;\eta_0) - \Lambda_0^R(t) \\
&= \int_0^t \{\tilde{\Lambda}_0^D(u;\eta_0) - \Lambda_0^D(u)\}d\tilde{G}_2(u) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dM_i^R(u)}{S_0(u;\eta_0,\tilde{\Lambda}_0^D)} \\
&= \frac{1}{n}\sum_{i=1}^n \phi_{2i}(t) + o_p(n^{-1/2}),
\end{aligned}$$

where

$$\phi_{2i}(t) = \int_0^t \phi_{1i}(u)dG_2(u) + \int_0^t \frac{dM_i^R(u)}{s_1^{(0)}(u)}.$$

Note that

$$\begin{aligned}
&\tilde{U}_1(\eta_0)\\
=\;& \sum_{i=1}^{n}\int_0^{\tau}\Big\{Z_i(t)-\bar{Z}(t;\eta_0)\Big\}\\
&\times\Big[dN_i^R(t)-\Delta_i(t)\tilde{\psi}_i(t;\tilde{\Lambda}_0^D)\{\beta_0^T Z_i(t)dt+d\Lambda_0^R(t)\}\Big]\\
=\;& \sum_{i=1}^{n}\int_0^{\tau}\Big\{Z_i(t)-\bar{Z}(t;\eta_0)\Big\}\\
&\times\Big[dN_i^R(t)-\Delta_i(t)\tilde{\psi}_i(t;\Lambda_0^D)\{\beta_0^T Z_i(t)dt+d\Lambda_0^R(t)\}\Big]\\
&-\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t)-\bar{Z}(t;\eta_0)\}\\
&\times\Delta_i(t)\big[\tilde{\psi}_i(t;\tilde{\Lambda}_0^D)-\tilde{\psi}_i(t;\Lambda_0^D)\big]\{\beta_0^T Z_i(t)dt+d\Lambda_0^R(t)\}.
\end{aligned}$$

Let

$$\begin{aligned}
\tilde{G}_3(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\theta_0\Delta_i(u)\tilde{\psi}_i(u;\tilde{\Lambda}_0^D)\tilde{\psi}_i(u;\Lambda_0^D)\\
&\quad\times\{Z_i(u)-\bar{Z}(u;\eta_0)\}\big[\beta_0^T Z_i(u)du+d\Lambda_0^R(u)\big],
\end{aligned}$$

and $\bar{z}(t)$ and $G_3(t)$ are the limits of $\bar{Z}(t;\eta_0)$ and $\tilde{G}_3(t)$. Similarly to (A6), we obtain

$$\text{(A7)}\qquad \tilde{U}_1(\eta_0)=\sum_{i=1}^{n}\xi_{1i}+o_p(n^{1/2}),$$

where

$$\xi_{1i}=\int_0^{\tau}\{Z_i(t)-\bar{z}(t)\}\,dM_i^R(t)+\int_0^{\tau}\phi_{1i}(t)dG_3(t).$$

Likewise,

$$\text{(A8)}\qquad \tilde{U}_2(\eta_0)=\sum_{i=1}^{n}\xi_{2i}+o_p(n^{1/2}),$$

where

$$\xi_{2i}=\int_0^{\tau}\{Z_i(t)-\bar{z}(t)\}\,dM_i^D(t)+\int_0^{\tau}\phi_{2i}(t)dG_3(t),$$

and $G_4(t)$ is the limit of $\tilde{G}_4(t)$ with

$$\begin{aligned}
\tilde{G}_4(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\theta_0\Delta_i(u)\tilde{\psi}_i(u;\tilde{\Lambda}_0^D)\tilde{\psi}_i(u;\Lambda_0^D)\\
&\quad\times\{Z_i(u)-\bar{Z}(u;\eta_0)\}\big[\alpha_0^T Z_i(u)du+d\Lambda_0^D(u)\big].
\end{aligned}$$

Let

$$\omega_{2i}^*(t)=\tilde{\psi}_i(t;\Lambda_0^D)\int_0^{t}\{d\Lambda_0^R(u)+\beta_0^T Z_i(u)du\},$$

$$\tilde{\omega}_{2i}(t)=\tilde{\psi}_i(t;\tilde{\Lambda}_0^D)\int_0^{t}\{d\tilde{\Lambda}_0^R(u)+\beta_0^T Z_i(u)du\},$$

$$\tilde{Q}(t)=\frac{\sum_{i=1}^{n}\tilde{\omega}_{2i}(t)^{-1}\Delta_i^*(t)N_i^R(t)}{\sum_{i=1}^{n}\Delta_i^*(t)},$$

and $q(t)$ be the limit of $\tilde{Q}(t)$. Note that

$$\begin{aligned}
\text{(A9)}\qquad &\tilde{U}_3(\eta_0)\\
=\;&\sum_{i=1}^{n}\int_0^{\tau}\{N_i^R(t)-(\theta_0+1)\omega_{2i}^*(t)q(t)\}dN_i^D(t)\\
&-(\theta_0+1)\sum_{i=1}^{n}\int_0^{\tau}\tilde{Q}(t)\{\tilde{\omega}_{2i}(t)-\omega_{2i}^*(t)\}dN_i^D(t)\\
&-(\theta_0+1)\sum_{i=1}^{n}\int_0^{\tau}\{\tilde{Q}(t)-q(t)\}\omega_{2i}^*(t)dN_i^D(t).
\end{aligned}$$

As in the proof of (A7), the second term on the right-hand side of (A9) equals

$$\begin{aligned}
\text{(A10)}\quad &-(\theta_0+1)\sum_{i=1}^{n}\int_0^{\tau}\Big[\tilde{Q}(t)\{\tilde{\psi}_i(t;\tilde{\Lambda}_0^D)-\tilde{\psi}_i(t;\Lambda_0^D)\}\\
&\qquad\times\int_0^{t}\{\beta_0^T Z_i(u)du+d\Lambda_0^R(u)\}\Big]dN_i^D(t)\\
&-(\theta_0+1)\sum_{i=1}^{n}\int_0^{\tau}\tilde{Q}(t)\tilde{\psi}_i(t;\Lambda_0^D)\}\big[\tilde{\Lambda}_0^R(t;\eta_0)-\Lambda_0^R(t)\big]\\
&\qquad\times dN_i^D(t)\\
=\;&-(\theta_0+1)\sum_{i=1}^{n}\int_0^{\tau}\big[\phi_{1i}(t)dG_6(t)+\phi_{2i}(t)dG_7(t)\big]\\
&+o_p(n^{1/2}),
\end{aligned}$$

where $G_6(t)$ and $G_7(t)$ are the limits of $\tilde{G}_6(t)$ and $\tilde{G}_7(t)$ with

$$\begin{aligned}
\tilde{G}_6(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\theta_0\Big[\tilde{Q}(s)\tilde{\psi}_i(s;\tilde{\Lambda}_0^D)\tilde{\psi}_i(s;\Lambda_0^D)\\
&\quad\times\int_0^{s}\{\beta_0^T Z_i(u)du+d\Lambda_0^R(u)\}\Big]dN_i^D(s),
\end{aligned}$$

and

$$\tilde{G}_7(t)=\frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\theta_0\tilde{Q}(s)\tilde{\psi}_i(s;\Lambda_0^D)dN_i^D(s).$$

Let $d\Phi(t)=E\{\omega_{2i}^*(t)dN_i^D(t)\}$,

$$\begin{aligned}
\tilde{G}_8(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\theta_0\tilde{\psi}_i(t;\tilde{\Lambda}_0^D)\tilde{\psi}_i(t;\Lambda_0^D)\\
&\quad\times\int_0^{t}\{\beta_0^T Z_i(u)du+d\Lambda_0^R(u)\}\frac{\Delta_i^*(t)N_i^R(t)d\Phi(t)}{\tilde{\omega}_{2i}(t)\omega_{2i}^*(t)},\\
\tilde{G}_9(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\theta_0\tilde{\psi}_i(t;\Lambda_0^D)\frac{\Delta_i^*(t)N_i^R(t)}{\tilde{\omega}_{2i}(t)\omega_{2i}^*(t)},
\end{aligned}$$

and $G_8(t)$ and $G_9(t)$ be the limits of $\tilde{G}_8(t)$ and $\tilde{G}_9(t)$. In a similar manner, the third term on the right-hand side of (A9) is

$$(\theta_0+1)\sum_{i=1}^{n}\int_0^{\tau}\left[\phi_{1i}(t)G_8(t)+\phi_{2i}(t)G_9(t)\right]d\Phi(t)$$
$$\text{(A11)}\quad -(\theta_0+1)\sum_{i=1}^{n}\int_0^{\tau}\Big[\frac{\omega_{2i}^*(t)^{-1}\Delta_i^*(t)N_i^R(t)}{E\{\Delta_i^*(t)\}} -\frac{q(t)}{E\{\Delta_i^*(t)\}}\Delta_i^*(t)\Big]d\Phi(t)+o_p(n^{1/2}).$$

Using (A9)–(A11), we have

$$\text{(A12)}\qquad \tilde{U}_3(\eta_0)=\sum_{i=1}^{n}\xi_{3i}+o_p(n^{1/2}),$$

where

$$\begin{aligned}\xi_{3i} &= \int_0^{\tau}\left\{N_i^R(t)-(\theta_0+1)\omega_{2i}^*(t)q(t)\right\}dN_i^D(t)\\ &- (\theta_0+1)\int_0^{\tau}\left[\phi_{1i}(t)dG_6(t)+\phi_{2i}(t)dG_7(t)\right]\\ &+ (\theta_0+1)\int_0^{\tau}\left[\phi_{1i}(t)G_8(t)+\phi_{2i}(t)G_9(t)\right]d\Phi(t)\\ &- (\theta_0+1)\int_0^{\tau}\Big[\frac{\omega_{2i}^*(t)^{-1}\Delta_i^*(t)N_i^R(t)}{E\{\Delta_i^*(t)\}}-\frac{q(t)}{E\{\Delta_i^*(t)\}}\Delta_i^*(t)\Big]\\ &\times d\Phi(t).\end{aligned}$$

Let $\xi_i=(\xi_{1i}^T,\xi_{2i}^T,\xi_{3i})^T$. Then it follows from (A7), (A8), (A12) and the Taylor expansion that

$$\text{(A13)}\qquad n^{1/2}(\hat{\eta}-\eta_0)=n^{-1/2}\Gamma^{-1}\sum_{i=1}^{n}\xi_i+o_p(1).$$

By the multivariate central limit theorem, $n^{1/2}n^{1/2}(\hat{\eta}-\eta_0)$ is asymptotically normal with mean zero and covariance matrix $\Gamma^{-1}\Sigma(\Gamma^T)^{-1}$, where $\Sigma=E\{\xi_i\xi_i^T\}$.

To show the weak convergence of $n^{1/2}\{\hat{\Lambda}_0^D(t)-\Lambda_0^D(t)\}$ and $n^{1/2}\{\hat{\Lambda}_0^R(t)-\Lambda_0^R(t)\}$, first note that

$$\text{(A14)}\quad \begin{aligned}\hat{\Lambda}_0^D(t)-\Lambda_0^D(t) &= \{\tilde{\Lambda}_0^D(t;\hat{\eta})-\tilde{\Lambda}_0^D(t;\eta_0)\}\\ &+ \{\tilde{\Lambda}_0^D(t;\eta_0)-\Lambda_0^R(t)\}.\end{aligned}$$

Let $\tilde{\Upsilon}_1(t;\eta)=\partial\tilde{\Lambda}_0^D(t;\eta)/\partial\eta$. By the uniform strong law of large numbers, it can be shown that $\tilde{\Upsilon}_1(t;\eta)$ converge to a nonrandom function $\Upsilon_1(t;\eta)$ almost surely uniformly in $t\in[0,\tau]$ and η. It follows from the Taylor expansion, (A5), (A13) and (A14) that

$$\hat{\Lambda}_0^D(t)-\Lambda_0^D(t)=n^{-1/2}\sum_{i=1}^{n}\Psi_{1i}(t)+o_p(1),$$

where

$$\Psi_{1i}(t)=\Upsilon_1(t;\eta_0)\Gamma^{-1}\xi_i+\phi_{1i}(t).$$

Similarly,

$$\hat{\Lambda}_0^R(t)-\Lambda_0^R(t)=n^{-1/2}\sum_{i=1}^{n}\Psi_{2i}(t)+o_p(1),$$

where

$$\Psi_{2i}(t)=\Upsilon_2(t;\eta_0)\Gamma^{-1}\xi_i+\phi_{2i}(t),$$

and $\Upsilon_2(t;\eta_0)$ is the limit of $\partial\tilde{\Lambda}_0^R(t;\eta_0)/\partial\eta$. Let $\Psi_i(t)=(\Psi_{1i}(t),\Psi_{1i}(t))^T$. Because $\Psi_i(t)(i=1,...,n)$ are independent zero-mean random variables for each t, the multivariate central limit theorem implies that $n^{1/2}\{\hat{\Lambda}_0^D(t)-\Lambda_0^D(t)\}$ and $n^{1/2}\{\hat{\Lambda}_0^R(t)-\Lambda_0^R(t)\}$ jointly converge in finite-dimensional distributions to a zero-mean Gaussian process. Since $\Psi_i(t)$ can be written as sums or products of monotone functions of t and are thus tight (van der Vaart and Wellner, 1996). Thus, $n^{1/2}\{\hat{\Lambda}_0^D(t)-\Lambda_0^D(t)\}$ and $n^{1/2}\{\hat{\Lambda}_0^R(t)-\Lambda_0^R(t)\}$ are tight and jointly converge weakly to a zero-mean bivariate Gaussian process whose covariance function at (s,t) is given by $E\{\Psi_i(s)\Psi_i(t)^T\}$.

Received 21 March 2016

REFERENCES

Andersen, P., Borgan, O., Gill, R. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York. MR1198884

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

Chen, C.-M., Shen, P.-S. and Chuang, Y.-W. (2016). The partly Aalen's model for recurrent event data with a dependent terminal event. *Stat. Med.* **35** 268–281. MR3438219

Cook, R. and Lawless, J. (1997). Marginal analysis of recurrent events and a terminating event. *Stat. Med.* **16** 911–924.

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

Cook, R., Lawless, J., Lakhal-Chaieb, L. and Lee, K.-A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: Application to skeletal complications in cancer metastatic to bone. *J. Amer. Statist. Assoc.* **104** 60–75. MR2502850

Ghosh, D. and Lin, D. (2002). Marginal regression models for recurrent and terminal events. *Statist. Sinica* **12** 663–688. MR1929958

Gill, R. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** 1501–1555. MR1074422

Huang, C. and Wang, M. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc.* **99** 1153–1165. MR2109503

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. Wiley, New York. MR1924807

Kalbfleisch J., Schaubel, D., Ye, Y. and Gong, Q. (2013). An estimating function approach to the analysis of recurrent and terminal events. *Biometrics* **69** 366–374. MR3071055

Lin, D., Wei, L., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate function of recurrent events. *J. Royal Stat. Soc. Ser. B* **62** 711–730. MR1796287

Lin, D., Wei, L. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80** 557–572. MR1248021

Lin, D., Wei, L. and Ying, Z. (2001). Semiparametric transformation models for point processes. *J. Amer. Statist. Assoc.* **96** 620–628. MR1946429

Lin D. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81** 61–71. MR1279656

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96** 103–126. MR1952726

Liu, L., Huang, X. and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64** 950–958. MR2526647

Liu, D., Schaubel, D. and Kalbfleisch, J. (2012). Computationally efficient marginal models for clustered recurrent event data. *Biometrics* **68** 637–647. MR2959631

Liu, L., Wolfe, R. and Huang, X. (2004). Shared frailty model for recurrent events and a terminal event. *Biometrics* **60** 747–756. MR2089451

Liu, Y. and Wu, Y. (2011). Semiparametric additive intensity model with frailty for recurrent events. *Acta Math. Sin. Engl. Ser.* **27** 1831–1842. MR2827077

Pan, Q. and Schaubel, D. (2009). Flexible estimation of differences in treatment-specific recurrent event means in the presence of a terminating event. *Biometrics* **65** 753–761. MR2649848

Pipper, C. and Martinussen, T. (2004). An estimating equation for parametric shared frailty models with marginal additive hazards. *J. Royal Stat. Soc. Ser. B* **66** 207–220. MR2035767

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, Institute of Mathematical Statistics, Hayward, CA. MR1089429

Polyanin, A. and Manzhirov, A. (2008). *Handbook of Integral Equations*, 2nd edition. Boca Chapman and Hall/CRC, Raton, FL. MR2404728

Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd edition. McGraw-Hill, New York. MR0385023

Schaubel, D. and Zhang, M. (2010). Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event. *Lifetime Data Anal.* **16** 451–477. MR2726219

Schaubel, D., Zeng, D. and Cai, J. (2006). A semiparametric additive rates model for recurrent event data. *Lifetime Data Anal.* **12** 389–406. MR2338958

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *J. Amer. Statist. Assoc.* **107** 688–700. MR2980077

Sun, L. and Kang, F. (2013). An additive-multiplicative rates model for recurrent event data with informative terminal event. *Lifetime Data Anal.* **19** 117–137. MR3016952

Sun, L., Zhao, X. and Zhou, J. (2011). A class of mixed models for recurrent event data. *Canad. J. Statist.* **39** 578–590. MR2860828

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

Wang, M., Qin, J. and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *J. Amer. Statist. Assoc.* **96** 1057–1065. MR1947253

Ye, Y., Kalbfleisch, J. and Schaubel, D. E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63** 78–87. MR2345577

Zeng, D. and Cai, J. (2010). A semiparametric additive rate model for recurrent events with informative terminal event. *Biometrika* **97** 699–712. MR2672492

Zeng, D. and Lin, D. (2007). Semiparametric transformation models with random effects for recurrent events. *J. Amer. Statist. Assoc.* **102** 167–180. MR2345550

Zeng, D. and Lin, D. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* **65** 746–752. MR2649847

Zhao, X., Zhou, J. and Sun, L. (2011). Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. *Biometrics* **67** 404–414. MR2829009

Lianqiang Qu
Institute of Applied Mathematics
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing, 100190
People's Republic of China
E-mail address: qulianq@amss.ac.cn

Liuquan Sun
Institute of Applied Mathematics
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing, 100190
People's Republic of China
E-mail address: slq@amt.ac.cn

Lei Liu
Department of Preventive Medicine
Northwestern University
Chicago, IL 60611
USA
E-mail address: lei.liu@northwestern.edu