

Shared Frailty Models for Recurrent Events and a Terminal Event

Lei Liu,^{1,*} Robert A. Wolfe,¹ and Xuelin Huang²

¹Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

²Department of Biostatistics, M. D. Anderson Cancer Center, University of Texas,
Houston, Texas 77030, U.S.A.

* *email*: liulei@umich.edu

SUMMARY. There has been an increasing interest in the analysis of recurrent event data (Cook and Lawless, 2002, *Statistical Methods in Medical Research* **11**, 141–166). In many situations, a terminating event such as death can happen during the follow-up period to preclude further occurrence of the recurrent events. Furthermore, the death time may be dependent on the recurrent event history. In this article we consider frailty proportional hazards models for the recurrent and terminal event processes. The dependence is modeled by conditioning on a shared frailty that is included in both hazard functions. Covariate effects can be taken into account in the model as well. Maximum likelihood estimation and inference are carried out through a Monte Carlo EM algorithm with Metropolis–Hastings sampler in the E-step. An analysis of hospitalization and death data for waitlisted dialysis patients is presented to illustrate the proposed methods. Methods to check the validity of the proposed model are also demonstrated. This model avoids the difficulties encountered in alternative approaches which attempt to specify a dependent joint distribution with marginal proportional hazards and yields an estimate of the degree of dependence.

KEY WORDS: Counting process; Dependent censoring; Hospitalization; Proportional hazards model; Survival analysis.

1. Introduction

Recurrent event data arise frequently in longitudinal medical studies. In many situations, the follow-up of recurrent events could be stopped by a terminal event, such as death. For example, patients may experience recurrent hospitalizations which are terminated by death. In the study of recurrent rejection episodes of kidney transplant patients, some patients may suffer total graft rejection which precludes further occurrence of rejection episodes.

In practice, recurrent and terminal events are often not independent. For example, the recurrence of serious events (like heart attacks) often increases the risk of death, which in turn makes any subsequent recurrent event impossible. This dependence should be accounted for in the joint modeling of recurrent and terminal events.

It is of interest to many researchers to study the dependent dropout mechanism in the repeated measure setting (De Gruttola and Tu, 1994; Little, 1995; Hogan and Laird, 1997; Sun and Song, 2001; Xu and Zeger, 2001). The analysis of recurrent event process in the presence of a terminal event bears much similarity to their studies. Two approaches are generally adopted: marginal models and frailty models.

A natural marginal approach was taken by Li and Lagakos (1997). They adapted Wei, Lin, and Weissfeld's marginal model (1989) and regarded the terminating event as a censoring event for each recurrent event, or treated the failure time for each recurrence as the minimum of the recurrent event time and survival time. Cook and Lawless (1997) intro-

duced the mean frequency/rate function of recurrent events among survivors at a given time. Ghosh and Lin (2000) presented asymptotic properties for the nonparametric estimate of the marginal mean of the cumulative number of recurrent events over time. Ghosh and Lin (2002) also developed semiparametric methods using IPCW (inverse probability of censoring weighting; Robins and Rotnitzky, 1992) and a similar technique, inverse probability of survival weighting (IPSW). These models are extensions to the proportional means/rates models (Lin et al., 2000) based on empirical process theory in the absence of a dependent terminating event.

These marginal models do not specify the dependence between recurrent events and death. The application of marginal semiparametric models such as Ghosh and Lin (2002) is limited due to strict conditions required for both IPCW and IPSW methods, which may not be satisfied in reality.

Shared random effects models or frailty models have been applied in this scenario. Lancaster and Intrator (1998) modeled the joint distribution of hospitalization and survival parametrically through a common unmeasured "frailty." They treated hospitalization as a Poisson process whose rate function shares the same frailty with the hazard function of the survival time. These two event processes were assumed to be independent given frailty. Wang, Qin, and Chiang (2001) modeled the occurrence rate function for recurrent events with informative censoring in semiparametric and nonparametric ways. They assumed a nonstationary Poisson process via a frailty for recurrent events likewise. Conditioning on frailty,

recurrent and terminal events are independent. They treated both the distribution of informative censoring and frailty as nuisance parameters. Thus, their models cannot be applied to situations where modeling both the recurrent and terminal events is of interest. Also the proposed semiparametric models cannot handle time-dependent covariates.

In this article, we propose a joint semiparametric model for the intensity functions of both recurrent events and death by a shared gamma frailty. Nielsen et al. (1992; see also Andersen et al., 1993, Section IX) presented a general frailty model which may be specialized to recurrent event data with a terminal event. In that model, the frailty effect on recurrent and terminal event rates is the same. This is a special case ($\gamma = 1$) of our model, in which the frailty can have different effects on the two hazards.

The remainder of this article is organized as follows. In Section 2, we introduce the underlying multiplicative intensity model which takes into account the covariate information. Statistical methods are developed in Section 3 using the EM algorithm with a Metropolis–Hastings E-step. Results from simulation and a real-data analysis are presented in Sections 4 and 5, respectively. Section 6 concludes the article with a summary of our results and a discussion of generalizations of the proposed models.

2. Models

Let C_i and D_i be the censoring and death times for subject i ($i = 1, 2, \dots, n$). Write $X_i = \min(C_i, D_i)$ as the follow-up time and $\Delta_i = I(D_i \leq C_i)$, where $I(\cdot)$ is the indicator function. Let $Y_i(t) = I(X_i \geq t)$ be the at-risk indicator. Denote by $N_i^{D^*}(t) = I(D_i \leq t)$ and $N_i^D(t) = I(X_i \leq t, \Delta_i = 1)$ the actual and observed death indicator by time t , respectively. Similarly, define by $N_i^{R^*}(t)$ and $N_i^R(t)$ the actual and observed number of recurrent events with $N_i^R(t) = N_i^{R^*}(\min(X_i, t))$, respectively. Write $dN_i^{R^*}(t) = N_i^{R^*}\{(t + dt) -\} - N_i^{R^*}(t-)$ as $dt \rightarrow 0$ and $dN_i^R(t) = I(X_i \geq t)dN_i^{R^*}(t)$. We introduce heterogeneity with observed covariate Z_i and unobserved frailty ν_i , which may measure the latent “health status” of the patient related to both recurrent events and the terminal event. The observation for subject i is $\mathbf{O}_i(t) \equiv \{Y_i(u), N_i^R(u), N_i^D(u), 0 \leq u \leq t\}$, an i.i.d. copy up to time t of the whole observed data $\mathbf{O} = \{\mathbf{O}(t), 0 \leq t \leq X\}$. Define by \mathcal{F}_0 the σ -field generated by (ν, Z) and $\mathcal{F}_t = \sigma\{\mathcal{F}_0, \mathbf{O}(u), 0 \leq u \leq t\}$.

The following assumptions are made on the underlying processes:

(1) The recurrent, terminating, and censoring processes all have continuous distribution so that recurrent events and death cannot happen at the same time. We adopt the convention that death happens first in the interval $[t, t + dt)$.

We assume that the terminal event stops the further occurrence of recurrent events in that $N_i^{R^*}(t)$ is constant after D_i . This is different from the conventional censoring event which only prevents us from observing further recurrent events but does not prevent their occurrence. Thus, the terminal and recurrent event processes are not independent even conditional upon frailty.

$$(2) P(dN^D(t) = 1 | \mathcal{F}_{t-}) = Y_i(t) d\Lambda_i(t) \equiv Y_i(t)\lambda_i(t) dt,$$

where $d\Lambda_i(t) = P(dN^{D^*}(t) = 1 | Z, \nu, D \geq t)$.

$$(3) P(dN^R(t) = 1 | \mathcal{F}_{t-}, D \geq t) = Y_i(t) dR_i(t) \equiv Y_i(t)r_i(t) dt,$$

where $dR_i(t) = P(dN_i^{R^*}(t) = 1 | Z_i, \nu_i, D_i \geq t)$.

Note that $P(dN^{R^*}(t) = 1 | \mathcal{F}_{t-}, D \leq t)$ is not estimable generally and is 0 in this setting where D is a terminal event. Events $(D < t)$ and $(D \leq t)$ are almost surely identical since D has a continuous distribution.

(4) Censoring is noninformative. In particular, as in type I or administrative censoring situations, censoring does not depend on ν .

(5) $P(N^R(X) > 1) > 0$, which assures that ν and γ can be identified.

As in Kalbfleisch and Prentice (2002), the full likelihood can be written as a product integral

$$\mathcal{L} = L(\mathcal{F}_0)L(\mathbf{O} | \mathcal{F}_0). \tag{1}$$

Ignoring the contribution from independent and noninformative censoring,

$$\begin{aligned} L(\mathbf{O}_i | \mathcal{F}_0) &= \mathcal{P}_0^\infty L(\mathcal{F}_{t-+dt} | \mathcal{F}_{t-}) \\ &\propto \mathcal{P}_0^\infty L(dN_i^R(t), dN_i^D(t) | \mathcal{F}_{t-}). \end{aligned} \tag{2}$$

Along the lines of the development of the competing risk likelihood, (2) can be written as the product of components given by (7) and (8). The detailed steps are

$$\begin{aligned} &L(dN_i^R(t), dN_i^D(t) | \mathcal{F}_{t-}) \\ &= [Y_i(t) d\Lambda_i(t)]^{dN_i^D(t)} [1 - Y_i(t) d\Lambda_i(t)]^{1-dN_i^D(t)} \\ &\quad \times \left\{ [Y_i(t) dR_i(t)]^{dN_i^R(t)} [1 - Y_i(t) dR_i(t)]^{1-dN_i^R(t)} \right\}^{1-dN_i^D(t)}. \end{aligned} \tag{3}$$

Note we adopt the convention $0^0 = 1$. It can be shown that for continuous event times, (3) is equivalent to

$$\begin{aligned} &[Y_i(t) d\Lambda_i(t)]^{dN_i^D(t)} [1 - Y_i(t) d\Lambda_i(t)]^{1-dN_i^D(t)} \\ &\quad \times [Y_i(t) dR_i(t)]^{dN_i^R(t)} [1 - Y_i(t) dR_i(t)]^{1-dN_i^R(t)}, \end{aligned}$$

or

$$\begin{aligned} &\mathcal{P}_0^\infty L(dN_i^R(t), dN_i^D(t) | \mathcal{F}_{t-}) \\ &= \mathcal{P}_0^\infty L(dN_i^R(t) | \mathcal{F}_{t-}, D > t) \mathcal{P}_0^\infty L(dN_i^D(t) | \mathcal{F}_{t-}). \end{aligned} \tag{4}$$

We extend the model in Huang and Wolfe (2002) for clustered survival with dependent censoring to the new setting as follows:

$$r_i(t) = \nu_i \exp(\beta^T Z_i) r_0(t), \tag{5}$$

$$\lambda_i(t) = \nu_i^\gamma \exp(\alpha^T Z_i) \lambda_0(t). \tag{6}$$

The presence of the common frailty parameter ν weakens the usual assumption of noninformative censoring of the recurrent event process by death. We adopt the gamma frailty function $f_\theta(\cdot)$ with unit mean and variance θ . The mean is unity to avoid the nonidentifiability problem, which otherwise might arise if we multiply and divide the frailty and the baseline hazard by the same constant. When $\gamma = 0$, $\lambda_i(t)$ does not depend on ν_i and is noninformative for the recurrent event

rate $r_i(t)$. $\theta = 0$ implies that the frailty terms ν_i 's are identically 1, i.e., the heterogeneity in both recurrent and terminal event rates is solely explained by Z_i .

Denote t_{ij} to be the j th recurrent event time for i th subject, δ_{ij} the indicator of the recurrent event at time t_{ij} . Let x_i be the observed follow-up time. The first factor of likelihood (4) is

$$\begin{aligned} & \exp \left\{ - \int_0^\infty Y_i(t) \nu_i \exp(\beta^T Z_i) dR_0(t) \right\} \\ & \times \prod_j [\nu_i \exp(\beta^T Z_i) dR_0(t_{ij})]^{\delta_{ij}}. \end{aligned} \quad (7)$$

Similarly, (6) suggests the second part of likelihood (4) is proportional to

$$\begin{aligned} & \exp \left\{ - \int_0^\infty Y_i(t) \nu_i^\gamma \exp(\alpha^T Z_i) d\Lambda_0(t) \right\} \\ & \times [\nu_i^\gamma \exp(\alpha^T Z_i) d\Lambda_0(x_i)]^{\Delta_i}. \end{aligned} \quad (8)$$

$L(\mathbf{O}_i | \mathcal{F}_0)$ can be written as the multiplication of (7) and (8). The full log likelihood for $\{(\mathbf{O}_i, \nu_i), i = 1, \dots, n\}$ is

$$\begin{aligned} l &= \log \prod_{i=1}^n L(\mathbf{O}_i, \nu_i | Z_i) = \log \left\{ \prod_{i=1}^n L(\mathbf{O}_i | \nu_i, Z_i) f_\theta(\nu_i) \right\} \\ &= \sum_{i=1}^n \left\{ \sum_j [\log \nu_i + \beta^T Z_i + \log dR_0(t_{ij})] \right. \\ & \quad \left. - \int_0^\infty Y_i(t) \nu_i \exp(\beta^T Z_i) dR_0(t) \right\} \\ &+ \sum_{i=1}^n \left\{ \Delta_i [\gamma \log \nu_i + \alpha^T Z_i + \log d\Lambda_0(x_i)] \right. \\ & \quad \left. - \int_0^\infty \nu_i^\gamma \exp(\alpha^T Z_i) d\Lambda_0(t) \right\} \\ &+ \sum_{i=1}^n \log f_\theta(\nu_i). \end{aligned} \quad (9)$$

Time-dependent covariates could be incorporated in our model. Analysis of internal time-dependent covariates yields the same likelihood form as in (6.11) of Kalbfleisch and Prentice (2002). In what follows we write $Z_i(t)$ to denote the time-dependent covariates.

3. Methods

Equation (9) gives the likelihood for the ‘‘complete data’’ with ‘‘known’’ frailties, which is more readily maximized than the ‘‘observed data likelihood.’’ This makes the EM algorithm a natural choice for parameter estimation. In the E-step, since there is no closed form for the density of $f(\nu_i | \mathbf{O}_i)$, Metropolis–Hastings algorithm can be used to generate M random numbers $\nu_i^{(m)} (m = 1, \dots, M)$ for the estimation of the expectation of the sufficient statistics involving frailties. Examples are $\hat{E}(\nu_i | \mathbf{O}_i) = \frac{1}{M} \sum_{m=1}^M \nu_i^{(m)}$ and $\hat{E}(\log \nu_i | \mathbf{O}_i) = \frac{1}{M} \sum_{m=1}^M \log \nu_i^{(m)}$. A brief introduction to the Metropolis–Hastings algorithm is given in Appendix A.

In the M-step, parameter estimates are obtained by maximizing of the likelihood (9) as if the frailty statistics are known. The components of the partial derivative for β and $r_0(\cdot)$ are

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \sum_{i=1}^n \left[\sum_j \delta_{ij} Z_i(t_{ij}) \right. \\ & \quad \left. - \int_0^\infty Y_i(t) Z_i(t) \hat{E}(\nu_i | \mathbf{O}_i) \exp(\beta^T Z_i(t)) dR_0(t) \right] \\ \frac{\partial l}{\partial r_0(t_{ij})} &= \frac{\delta_{ij}}{r_0(t_{ij})} - \sum_{k=1}^n Y_k(t_{ij}) \hat{E}(\nu_k | \mathbf{O}_k) \exp(\beta^T Z_k(t_{ij})). \end{aligned}$$

Parametric models for $r_0(t)$ can be estimated with standard asymptotic properties for fixed dimension MLE. The nonparametric estimate of baseline intensity for recurrent events is of the form of Breslow’s estimate,

$$\hat{r}_0(t_{ij}) = \frac{\delta_{ij}}{\sum_k Y_k(t_{ij}) \hat{E}(\nu_k | \mathbf{O}_k) \exp(\beta^T Z_k(t_{ij}))}. \quad (10)$$

$\hat{\beta}$ can be solved by substituting (10) into the component of partial derivatives $\partial l / \partial \beta$. The second partial derivative for covariate coefficient β is

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta^2} &= - \sum_{i=1}^n \int_0^\infty Y_i(t) Z_i(t) Z_i(t)^T \hat{E}(\nu_i | \mathbf{O}_i) \\ & \quad \times \exp(\beta^T Z_i(t)) dR_0(t). \end{aligned}$$

All other score components and second partial derivatives are given in Appendix B.

Since the EM algorithm does not provide the information matrix for the observed data likelihood directly, Louis’s formula (Louis, 1982) is used to obtain it. Let $\eta = (\beta, \alpha, \gamma, \theta, r_0, \lambda_0)$. The observed information matrix $I(\hat{\eta})$ is given by

$$\begin{aligned} I(\hat{\eta}) &= -\hat{E} \left\{ \frac{\partial^2 l}{\partial \eta \partial \eta'} \middle| \mathbf{O}, \hat{\eta} \right\} - \hat{E} \left\{ \frac{\partial l}{\partial \eta} \frac{\partial l}{\partial \eta'} \middle| \mathbf{O}, \hat{\eta} \right\} \\ & \quad + \hat{E} \left\{ \frac{\partial l}{\partial \eta} \middle| \mathbf{O}, \hat{\eta} \right\} \hat{E} \left\{ \frac{\partial l}{\partial \eta'} \middle| \mathbf{O}, \hat{\eta} \right\}. \end{aligned}$$

All of these terms are evaluated at the last iteration of the EM algorithm, when the last term becomes zero for the MLE $\hat{\eta}$. The first two expectations can be calculated by averaging over the corresponding terms involving Metropolis–Hastings values.

4. Simulation

In this section we report results from a simulation study of three settings to evaluate the performance of the proposed estimation procedures. In each setting, we considered a single binary covariate Z taking value 0 or 1, each with probability 1/2, sample size $n = 100$, and regression coefficients $\alpha = 1$ and $\beta = 1$. Frailties were generated from a gamma distribution with mean 1 and variance $\theta = 1$. The baseline intensity functions for both recurrent events and death were taken to be exponential with constant 2 and 1/2, respectively. We assumed a fixed censoring time $C = 0.8$ for all subjects. A total

Table 1
Simulation results: Parameter estimates for setting I

Parameter	Shared frailty model				Reduced model			
	Bias	SE	SEM	CP	Bias	SE	SEM	CP
$\beta = 1.0$	-0.009	0.272	0.274	95.1%	-0.061	0.270	0.273	94.9%
$\alpha = 1.0$	0.006	0.360	0.362	96.4%	-0.081	0.323	0.331	94.6%
$\gamma = 0.5$	0.022	0.285	0.276	94.3%				
$\theta = 1.0$	-0.001	0.244	0.253	93.6%	0.042	0.263	0.276	94.8%

Bias is the mean of the parameter estimates (based on 800 replicates) minus the true value; SE is the sampling standard error of the parameter estimate; SEM is the sampling mean of the standard error estimate; CP is the coverage probability of the corresponding 95% confidence interval.

Table 2
Simulation results: Parameter estimates for setting II

Parameter	Shared frailty model				Reduced model			
	Bias	SE	SEM	CP	Bias	SE	SEM	CP
$\beta = 1.0$	0.015	0.271	0.264	93.3%	0.063	0.260	0.247	93.0%
$\alpha = 1.0$	0.009	0.353	0.341	94.9%	-0.134	0.291	0.278	91.1%
$\gamma = -0.5$	-0.001	0.233	0.265	95.1%				
$\theta = 1.0$	-0.047	0.292	0.304	92.6%	-0.244	0.227	0.224	72.3%

of 800 replicates were generated. The only difference in these settings lies in γ , which takes 1/2, -1/2, and 0 in settings I, II, and III, respectively.

For setting I, subjects have 1.8 recurrent events on average. Of the subjects, 57% are censored and the rest experience death as the terminal event; 40% of subjects do not have any recurrent event. With $\gamma = -1/2$ in setting II, the incidence of recurrent events is preventive against the terminal event. For this setting, each subject has 2.1 recurrent events on average. Of the subjects, 42% are censored and 44% have no recurrent event. The results are shown in Tables 1 and 2, respectively.

It can be seen that in both settings I and II the magnitudes of the empirical biases of the estimates from the joint frailty model are very small. Only $\hat{\gamma}$ in setting I and $\hat{\theta}$ in setting II have significant biases of 0.022 and -0.047, respectively. The coverage probabilities are close to the nominal level 0.95, too. We observe only minor, if not negligible, biases for variance estimates.

We also used a reduced model estimator assuming that the hazards of recurrent and terminal events share no parameter, i.e., $\gamma = 0$. We fitted recurrent event data (by an algorithm similar to Klein, 1992) and survival data (by standard pro-

portional hazards model) separately. The results are shown in the right side of each table. The resulting parameter estimates are biased, especially for θ in setting II. It is clear that ignoring the dependence between the terminal and recurrent events can result in significant biases.

In setting III, we set $\gamma = 0$ to compare the reduced and full model estimates when both models are valid. The estimates are very close with the same accuracy and precision in this special case (Table 3).

Figure 1 shows histograms for $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\theta}$. In all settings, the distribution of $\hat{\alpha}$ and $\hat{\beta}$ are approximately symmetric and normal, while $\hat{\theta}$ is skewed to the right. The distribution of $\hat{\gamma}$ is skewed to the direction of its sign (positive or negative).

Figure 2 gives the estimates of cumulative baseline hazard functions for the recurrent and terminal events. For simplicity, we only include the points at times 0.1, 0.2, ..., 0.8. In each setting we draw the true cumulative baseline hazard functions $R_0(t) = 2t$ and $\Lambda_0(t) = 0.5t$ for comparison. It can be seen that $\hat{R}_0(t)$ for all settings are virtually unbiased. $\hat{\Lambda}_0(t)$ also have little bias in settings I and III, but it is biased upward and skewed in setting II.

We also plot the estimates of cumulative baseline hazards in the reduced model assuming $\gamma = 0$ (results not shown).

Table 3
Simulation results: Parameter estimates for setting III

Parameter	Shared frailty model				Reduced model			
	Bias	SE	SEM	CP	Bias	SE	SEM	CP
$\beta = 1.0$	-0.005	0.277	0.268	94.5%	-0.006	0.277	0.265	94.5%
$\alpha = 1.0$	0.011	0.319	0.314	94.9%	-0.006	0.312	0.308	93.0%
$\gamma = 0$	0.006	0.200	0.200	97.4%				
$\theta = 1.0$	-0.003	0.263	0.272	93.5%	-0.014	0.256	0.263	94.6%

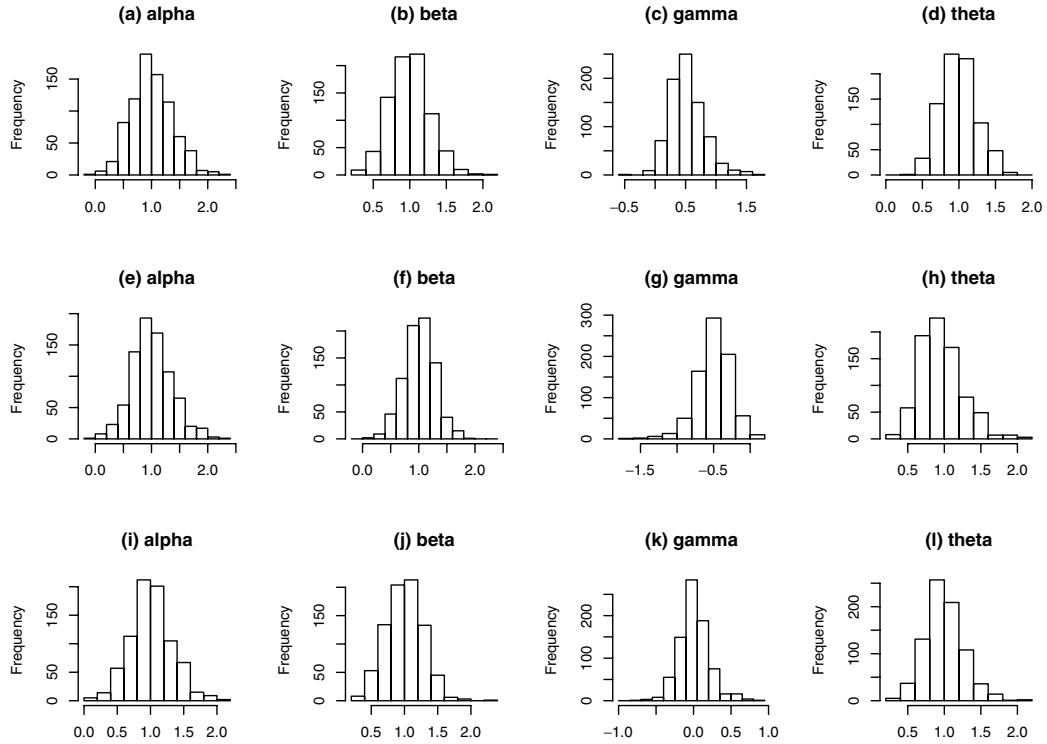


Figure 1. The histograms for the parameter estimates of α , β , γ , and θ . (a)–(d) Setting I; (e)–(h) setting II; (i)–(l) setting III.

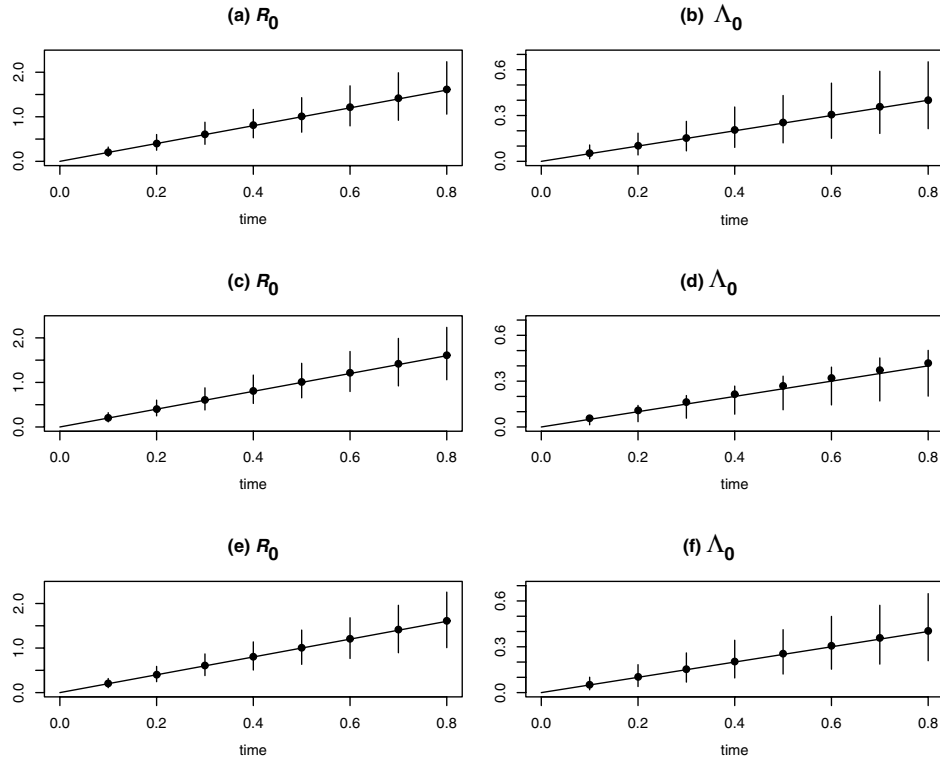


Figure 2. The estimates and 95% confidence interval of $R_0(t)$ and $\Lambda_0(t)$. (a and b) Setting I; (c and d) setting II; (e and f) setting III. Means of the hazard functions at each time point are denoted by a dot; pointwise 95% empirical confidence interval for the estimated cumulative baseline hazards are obtained from 800 replicates.

Both $\hat{R}_0(t)$ and $\hat{\Lambda}_0(t)$ are biased downward in setting I and upward in setting II, while unbiased in setting III. The bias in $\hat{\Lambda}_0(t)$ arises from the fact that the reduced model incorrectly estimates the death hazard in model (6) by the marginal proportional hazards method. The bias in $\hat{R}_0(t)$ results from the dependent censoring of the death event.

5. Application

We apply the proposed methods to hospitalization data for transplant candidates with kidney disease. Data are obtained from the Scientific Registry of Transplant Recipients (SRTR) database (Merion, 2003). This analysis includes all patients registered between January 1, 1999 and January 31, 1999 for kidney transplant. The follow-up for hospitalization and survival ended on December 31, 2002. We are interested in the joint modeling of hospitalization and survival event processes (in days), taking account of covariate information.

A total of 1121 patients were enrolled in this study. Among them there are 646 males (58%) and 743 whites (66%). The average age at registration is 47. The number of hospitalizations ranges from 0 to 33, averaging 2.6 per patient, whereas 460 (41%) patients were not hospitalized. The wide range of the number of hospitalizations suggests a large subject to subject variation. Kidney transplant was performed on 609, or 54% of patients during this period. At the end of the study 233 died and 375 were censored, while the remainder were censored before that date. In what follows we assumed censoring is independent of hospitalization and death. To avoid ties, we added a random number from Uniform(0, 1) distribution (in days) to each patient's event times.

We excluded two situations from the classification of the "recurrent event." First, the hospitalization for transplant is not counted as a hospitalization. Second, if the patient dies during hospitalization, it only counts as a terminal event, not as a hospitalization. The hospital discharge time is used as the recurrent event time.

Baseline covariates included in the analysis are age in years, race (1 = white, 0 = nonwhite), and gender (1 = male, 0 = female). Transplantation status Tx (1 = posttransplant, 0 = pretransplant) is a time-dependent covariate for survival. For hospitalization, we used two indicator variables to study the impact of transplant: Tx1 (1 = within 180 days after transplant, 0 = otherwise) and Tx2 (1 = more than 180 days after transplant, 0 = otherwise).

Preliminary studies showed that race, gender, and age effects were not significant for the hazard of hospitalization with this sample. Our final model is

$$\begin{aligned}
 r(t) &= \nu \exp(\beta_1 \text{Tx1} + \beta_2 \text{Tx2}) r_0(t), \\
 \lambda(t) &= \nu^\gamma \exp(\alpha_1 \text{Race} + \alpha_2 \text{Gender} \\
 &\quad + \alpha_3 \text{Age} + \alpha_4 \text{Tx}) \lambda_0(t).
 \end{aligned}
 \tag{11}$$

We summarize the results in Table 4. During the first 180 days after kidney transplant, the rate of hospitalization almost doubles ($HR = 2.03, p < 0.0001$). After 180 days the hospitalization rate decreases by 20% ($p = 0.005$) compared to that before transplant. Transplant also has a positive impact on patient's survival time. It decreases the death rate by 44% ($p = 0.0005$). Race and gender do not have any signif-

Table 4
Analysis of hospitalization data for kidney patients

Covariate	Estimate	SE	p-value	HR
<i>For hospitalization</i>				
Tx1: within 180 days after transplant	0.71	0.069	<0.0001	2.03
Tx2: more than 180 days after transplant	-0.22	0.079	0.006	0.80
<i>For survival</i>				
Race	-0.19	0.18	0.20	0.83
Gender	0.17	0.20	0.21	1.19
Age	0.049	0.0059	<0.0001	1.05
Tx: posttransplant	-0.58	0.17	0.0005	0.56
γ	0.56	0.078	<0.0001	1.76
θ	2.2	0.12		

icant effect on survival, whereas each 1-year increase in age elevates the death rate by about 5% ($p < 0.0001$).

The hospitalization rate varies greatly among patients ($\hat{\theta} = 2.2, p < 0.0001$), which results from the wide range of the number of hospitalization among patients. We also observe $\hat{\gamma} = 0.56$, which is significantly greater than 0 ($p < 0.0001$). This implies that the hospitalization and death rates are positively associated.

We checked the adequacy of the adopted model by evaluating the estimated cumulative hazard functions for various stratifications. First, we divided the subjects into two age groups: young (<47 years old) and old (≥ 47 years old). We fitted the age-stratified models with the same adjusting variables as in (11). Plots of $\log \hat{R}_0(t)$ and $\log \hat{\Lambda}_0(t)$ versus $\log t$ are displayed in Figure 3a and 3b, respectively. The parallelism of the curves in both figures suggests that the proportional hazards model for age is a very good approximation, after adjustment for other covariates. We observe no noticeable difference between young and old patients on $\hat{R}_0(t)$, which justifies the exclusion of age.

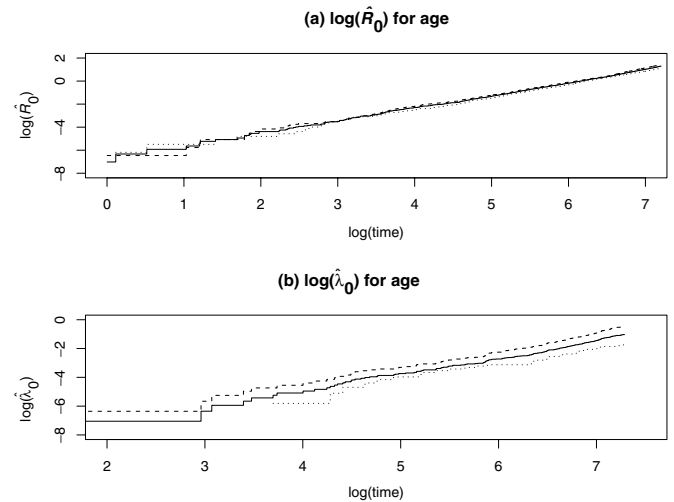


Figure 3. Model checking for age. (a) $\log \hat{R}_0(t)$. (b) $\log \hat{\Lambda}_0(t)$. (—) All; (···) young; and (- - -) old.

For the time-dependent transplant covariate Tx, model checking examines the appropriate choice of time scale, as well as the proportional hazards assumptions. The data were divided into pretransplant and posttransplant follow-up periods. Separating the pretransplant data corresponds to censoring the data at transplant. Since kidney transplantation is determined largely by external factors, such as waiting time and genetic matching of the donor organ rather than upon patient condition, this censoring is plausibly treated as independent. As shown below, the pretransplant rate is nearly constant, so a simple proportional hazards effect for transplant would correspond to a constant posttransplant rate that does not depend upon the time of transplant. For the posttransplant group, the time origin was redefined to be the time of transplant, in order to determine whether the ratio of pre- to posttransplant rates varied by time since transplant.

We show the resulting $\hat{R}_0(t)$ and $\hat{\Lambda}_0(t)$ versus t in Figure 4. For the pretransplant group, $\hat{R}_0(t)$ is nearly linear in t , suggesting that the event rate is approximately constant. The posttransplant hospitalization rate is not constant over time and has a change in slope around 180 days, which agrees with the inclusion of Tx1 and Tx2 in the hospitalization hazard model. Figure 4b shows that $\hat{\Lambda}_0(t)$ has a modest change in slope around 600 days in the pretransplant data, while the posttransplant death rate is nearly constant. A more detailed model could account for this nonproportionality if necessary.

We also considered the model with time (in months) from waitlist registration to transplant as a baseline covariate in the posttransplant group. Time to transplant is not a significant predictor for hospitalization rate ($HR = 1.00, p = 0.73$), indicating that the effect of transplant is constant with respect to time to transplant. But it is a significant predictor for survival ($HR = 1.03, p = 0.03$), suggesting that the data would be better fit by a model with an interaction term between time of transplant and the effect of transplant (Meier-Kriesche et al., 2001). The current results, shown without the

interaction term, represent the average effect of transplant over the range of times of transplant in this data set.

The diagnostic results suggest that our model could be simplified by using Weibull models for the baseline hazards as indicated by the linear relationship in Figure 3.

6. Discussion

In this article, we proposed a shared frailty model for recurrent events and a terminal event. Our model can be easily generalized. Other functional forms of the frailties can be incorporated in the joint models (5) and (6), such as exponential functions of ν in the mortality intensity. The corresponding likelihood and estimation equations are readily adapted with only a minor modification for the (expected) functions of frailty terms. Stratum-specific models or interaction between frailties and other covariates can be utilized. Other parametric families can be used for the frailty distributions, e.g., log-normal (Huang and Wolfe, 2002). These models are more flexible than those in Lancaster and Intrator (1998), which assumed a common gamma frailty effect for both the recurrent events and mortality. The proposed joint model can also incorporate time-dependent covariates. Parameters for terminal event can be estimated in our model, unlike that presented in Wang et al. (2001).

We did not use the profile likelihood method to estimate the variances because the resulting estimate for the variance of frailty distribution θ is biased downward. Andersen et al. (1997) studied this issue under the independent censoring situation and reached the same conclusion.

The use of fixed dimensional models for $\lambda_0(t)$ and $r_0(t)$ would lead to standard maximum likelihood asymptotic properties. We assume nonparametric forms for these functions, so the large sample properties of the estimators cannot be verified by the usual central limit theory. However, the simulation results suggest the asymptotic validity of this approach. Murphy (1995) studied asymptotic properties for the shared gamma frailty model in a simple setting with no covariate and no terminating events. Parner (1998) extended Murphy's result to a more general correlated frailty model with covariates. These results provide an approach to establishing the asymptotic properties of the proposed estimators.

Equations (3) and (4) are equivalent under the condition that recurrent events and death cannot happen simultaneously. In our application, if a recurrent event happened just before death (such as death while hospitalized), we treat it only as a terminal event and disregard the recurrent event. That is, our definition of recurrent event is actually "recurrent without death."

As we rewrite (5) as $r_i(t) = \exp(\beta^T Z_i + \log \nu_i) r_0(t)$, $\log \nu_i$ is a proportional hazards covariate for recurrent events which is scaled to have coefficient 1. Similarly it is also a covariate for the survival with coefficient γ . The true relationship between recurrent and terminal events might be a complicated function involving time, which is approximated in the proposed model to capture the association that is due to an unmeasured covariate. An alternative is to assume a robust model by replacing ν^γ in (6) by a function which can be estimated by nonparametric techniques, such as splines. The distributional assumption about the frailty may be checked similarly as Glidden (1999). Oakes (1989) developed diagnostic methods

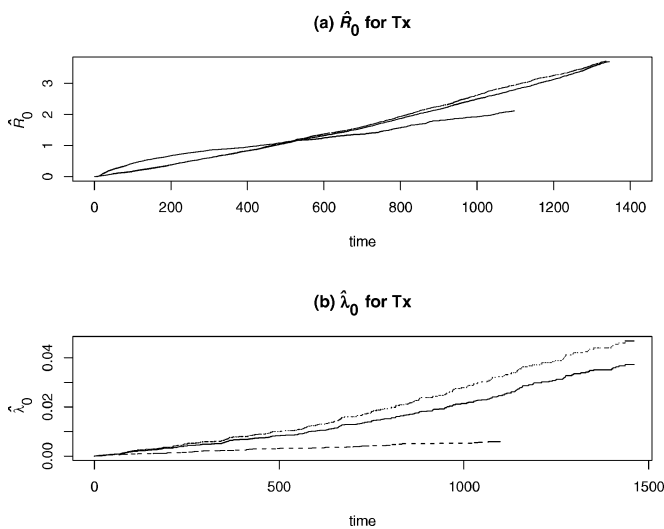


Figure 4. Model checking for transplant status Tx. (a) $\hat{R}_0(t)$ and (b) $\hat{\Lambda}_0(t)$. (—) All; (···) pretransplant; and (---) posttransplant.

for related models. His methods are not directly applicable to our model, but may provide valuable insight for future research. Other approaches to checking the validity of the proposed model are worthy of further investigation.

ACKNOWLEDGEMENTS

The authors thank the referees and an associate editor for their careful reading, and Drs J. Kalbfleisch, D. Ghosh, and D. Schaebel for helpful comments. We are also grateful to the Scientific Registry of Transplant Recipients for preparing the kidney hospitalization data.

RÉSUMÉ

L'intérêt porté à l'analyse des événements récurrents est croissant (Cook and Lawless, 2002). Dans de nombreuses situations, la survie d'un événement final, comme le décès, peut se produire pendant la période de suivi, empêchant ainsi que se produisent d'autres événements récurrents. Par ailleurs, le moment auquel survient le décès peut lui-même dépendre du passé concernant les événements récurrents. Dans cet article, nous considérons un modèle de fragilité avec taux proportionnels pour décrire le processus des événements récurrents et de l'événement final. La dépendance est modélisée en conditionnant par une fragilité partagée, incluse dans chacune des deux fonctions de risque. On peut aussi, dans ce modèle, prendre en compte l'effet de covariables. Les estimations et les tests associés au maximum de vraisemblance sont réalisés à partir de l'algorithme EM de la méthode de Monte Carlo, avec un échantillonnage de Métropolis-Hasting à l'étape E. L'analyse des données d'hospitalisation et du décès, chez des sujets en attente de dialyse, est présentée pour illustrer les méthodes proposées. Des méthodes, pour étudier la validité du modèle proposé, sont aussi présentées. Ce modèle évite les difficultés rencontrées dans les approches alternatives qui tentent de spécifier une distribution jointe, avec dépendance et des taux marginaux proportionnels, nécessitant alors d'estimer le degré de dépendance.

REFERENCES

- Andersen, P. K., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- Andersen, P. K., Klein, J., Knudsen, K., and Palacios, R. (1997). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics* **53**, 1475–1484.
- Cook, R. J. and Lawless, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* **16**, 911–924.
- Cook, R. J. and Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods in Medical Research* **11**, 141–166.
- De Gruttola, V. and Tu, X. M. (1994). Modeling progression of CD4+ lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.
- Glidden, D. V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika* **86**, 381–393.
- Ghosh, D. and Lin, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554–562.
- Ghosh, D. and Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12**, 663–688.
- Hogan, J. and Laird, N. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239–257.
- Huang, X. and Wolfe, R. A., (2002). A frailty model for informative censoring. *Biometrics* **58**, 510–520.
- Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. New York: Wiley.
- Klein, J. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795–806.
- Lancaster, A. and Intrator, O. (1998). Panel data with survival: Hospitalization of HIV-positive patients. *Journal of the American Statistical Association* **93**, 46–53.
- Li, Q. and Lagakos, S. (1997). Use of the Wei–Lin–Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine* **16**, 925–940.
- Lin, D. Y., Wei, L. J., Yang, L., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B* **62**, 711–730.
- Little, R. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.
- Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226–233.
- Meier-Kriesche, H.-U., Port, F. K., Ojo, A. O., Leichtman, A. B., Rudich, S. M., Arndorfer, J. A., PUNCH, J. D., and Kaplan, B. (2001). Deleterious effect of waiting time on renal transplant outcome. *Transplantation Proceedings* **33**, 1204–1206.
- Merion, R. (2003). Preface: SRTR report on the state of transplantation. *American Journal of Transplantation* **3**(Suppl. 4), 5–6.
- Murphy, S. (1995). Asymptotic theory for the frailty model. *Annals of Statistics* **23**, 182–198.
- Nielsen, G., Gill, R., Andersen, P., and Sorensen, T. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25–43.
- Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**, 487–493.
- Parner, E. (1998). Asymptotic theory for the correlated Gamma-frailty model. *Annals of Statistics* **28**, 183–214.
- Robins, J. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology-Methodological Issues*, N. Jewell, K. Dietz, and V. Farewell (eds), 297–331. Boston: Birkhauser.
- Sun, J. and Song, P. (2001). Statistical analysis of repeated measurements with informative censoring times. *Statistics in Medicine* **20**, 63–73.
- Wang, M., Qin, J., and Chiang, C. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96**, 1057–1065.
- Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.

Xu, J. and Zeger, S. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387.

Received June 2003. Revised March 2004.
Accepted March 2004.

APPENDIX A

It is difficult to sample directly from

$$f(\nu_i | \mathbf{O}_i) = \frac{f(\mathbf{O}_i | \nu_i)f(\nu_i)}{f(\mathbf{O}_i)} = \frac{f(\mathbf{O}_i | \nu_i)f(\nu_i)}{\int f(\mathbf{O}_i | \nu_i)f(\nu_i)d\nu_i},$$

where $f(\mathbf{O}_i | \nu_i)$ is given in Section 2. Metropolis–Hastings (M–H) algorithm is used to generate the random number chain $\nu_i^{(m)}$ ($m = 1, \dots, M$). Suppose we are at the k th E-step with current parameter estimates subscripted with (k) . The M–H chain starts with an initial value $\nu_i^{(1)}$. After we obtain $\nu_i^{(m)}$, a new value $\tilde{\nu}$ is sampled from gamma frailty with variance $\theta_{(k)}$. We also draw an independent random number u from Uniform(0, 1). $\nu_i^{(m+1)}$ is obtained as

$$\nu_i^{(m+1)} = \begin{cases} \tilde{\nu} & \text{if } u \leq \min\left(1, \frac{f(\mathbf{O}_i | \tilde{\nu})}{f(\mathbf{O}_i | \nu_i^{(m)})}\right), \\ \nu_i^{(m)} & \text{otherwise.} \end{cases}$$

Notice $f(\mathbf{O}_i)$ is cancelled in the ratio.

APPENDIX B

In the M-step, the score equations for $\alpha, \gamma, \lambda_0(\cdot)$, and θ are

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= \sum_{i=1}^n \left[Z_i(x_i)\Delta_i - \int_0^\infty Y_i(t)Z_i(t) \right. \\ &\quad \left. \times \exp(\alpha^T Z_i(t))\hat{E}(\nu_i^\gamma | \mathbf{O}_i)d\Lambda_0(t) \right], \\ \frac{\partial l}{\partial \gamma} &= \sum_{i=1}^n \left[\Delta_i\hat{E}(\log \nu_i | \mathbf{O}_i) - \int_0^\infty Y_i(t)\hat{E}(\nu_i^\gamma \log \nu_i | \mathbf{O}_i) \right. \\ &\quad \left. \times \exp(\alpha^T Z_i(t))d\Lambda_0(t) \right], \\ \frac{\partial l}{\partial \lambda_0(x_i)} &= \frac{\Delta_i}{\lambda_0(x_i)} - \sum_{k=1}^n Y_k(x_i) \exp(\alpha^T Z_k(x_i))\hat{E}(\nu_k^\gamma | \mathbf{O}_k), \\ \frac{\partial l}{\partial \theta} &= - \sum_{i=1}^n \hat{E} \left[\frac{1}{f_\theta(\nu_i)} \frac{\partial f_\theta(\nu_i)}{\partial \theta} \middle| \mathbf{O}_i \right]. \end{aligned}$$

We obtain the Breslow-type baseline hazard estimate as

$$\hat{\lambda}_0(x_i) = \frac{\Delta_i}{\sum_k Y_k(x_i)\hat{E}(\nu_k^\gamma | \mathbf{O}_k) \exp(\alpha^T Z_k(x_i))}.$$

The second derivatives for α, γ are

$$\begin{aligned} \frac{\partial^2 l}{\partial \alpha^2} &= - \sum_{i=1}^n \int_0^\infty Y_i(t)Z_i(t)^{\otimes 2} \\ &\quad \times \exp(\alpha^T Z_i(t))\hat{E}(\nu_i^\gamma | \mathbf{O}_i)d\Lambda_0(t), \\ \frac{\partial^2 l}{\partial \gamma^2} &= - \sum_{i=1}^n \int_0^\infty Y_i(t)\hat{E}(\log \nu_i | \mathbf{O}_i)^{\otimes 2}\hat{E}(\nu_i^\gamma | \mathbf{O}_i) \\ &\quad \times \exp(\alpha^T Z_i(t))d\Lambda_0(t), \end{aligned}$$

and

$$\frac{\partial^2 l}{\partial \alpha \partial \gamma} = - \sum_{i=1}^n \int_0^\infty Y_i(t)\hat{E}(\nu_i^\gamma \log \nu_i | \mathbf{O}_i) \exp(\alpha^T Z_i(t))d\Lambda_0(t),$$

where $a^{\otimes 0} = 1, a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$.

More components of the Information matrix are given below:

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial r_0(t_{ij})} &= -S^{(1)}(\beta, t_{ij}), \\ \frac{\partial^2 l}{\partial r_0(t_{ij})^2} &= -\frac{\delta_{ij}}{r_0(t_{ij})^2}, \\ \frac{\partial^2 l}{\partial \lambda_0(x_i)^2} &= -\frac{\Delta_i}{\lambda_0(x_i)^2}, \\ \frac{\partial^2 l}{\partial \alpha \partial \lambda_0(x_i)} &= -S^{(1)}(\alpha, x_i), \\ \frac{\partial^2 l}{\partial \gamma \partial \lambda_0(x_i)} &= -S^{(1)}(\gamma, x_i), \\ \frac{\partial^2 l}{\partial \theta^2} &= - \sum_{i=1}^n \hat{E} \left\{ \frac{1}{f_\theta(\nu_i)^2} \left[\frac{\partial^2 f_\theta(\nu_i)}{\partial \theta^2} f_\theta(\nu_i) \right. \right. \\ &\quad \left. \left. - \left(\frac{\partial f_\theta(\nu_i)}{\partial \theta} \right)^2 \right] \middle| \mathbf{O}_i \right\}, \end{aligned}$$

with

$$\begin{aligned} S^{(1)}(\beta, t) &= \sum_{k=1}^n Y_k(t)\hat{E}(\nu_k | \mathbf{O}_k)Z_k(t) \exp(\beta^T Z_k(t)), \\ S^{(1)}(\alpha, t) &= \sum_{k=1}^n Y_k(t)\hat{E}(\nu_k^\gamma | \mathbf{O}_k)Z_k(t) \exp(\alpha^T Z_k(t)), \end{aligned}$$

and

$$S^{(1)}(\gamma, t) = \sum_{k=1}^n Y_k(t)\hat{E}(\nu_k^\gamma \log \nu_k | \mathbf{O}_k) \exp(\alpha^T Z_k(t)).$$

All other off-diagonal terms are zero.

In the special case of gamma frailty with mean 1 and variance θ , we have

$$\frac{\partial l}{\partial \theta} = \frac{n}{\theta^2} \left[\log \theta - 1 + \Psi_1 \left(\frac{1}{\theta} \right) - \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{E}}(\log \nu_i | \mathbf{O}_i) + \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{E}}(\nu_i | \mathbf{O}_i) \right]$$

and

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{2n}{\theta^3} \left[\log \theta - 1 + \Psi_1 \left(\frac{1}{\theta} \right) - \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{E}}(\log \nu_i | \mathbf{O}_i) + \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{E}}(\nu_i | \mathbf{O}_i) \right] + \frac{n}{\theta^3} \left[1 - \frac{1}{\theta} \Psi_2 \left(\frac{1}{\theta} \right) \right],$$

where $\Psi_1(\cdot)$ and $\Psi_2(\cdot)$ are digamma and trigamma functions, i.e., the first and second derivative of $\log(\Gamma(\cdot))$.