Lifetime Data Analysis (2020) 26:109–133
https://doi.org/10.1007/s10985-019-09464-2

Frailty modelling approaches for semi-competing risks data

Il Do Ha[1] · Liming Xiang[2] · Mengjiao Peng[2] · Jong-Hyeon Jeong[3] · Youngjo Lee[4]

Received: 14 October 2017 / Accepted: 29 January 2019 / Published online: 7 February 2019

Abstract
In the semi-competing risks situation where only a terminal event censors a non-terminal event, observed event times can be correlated. Recently, frailty models with an arbitrary baseline hazard have been studied for the analysis of such semi-competing risks data. However, their maximum likelihood estimator can be substantially biased in the finite samples. In this paper, we propose effective modifications to reduce such bias using the hierarchical likelihood. We also investigate the relationship between marginal and hierarchical likelihood approaches. Simulation results are provided to validate performance of the proposed method. The proposed method is illustrated through analysis of semi-competing risks data from a breast cancer study.

Keywords Frailty models · Hierarchical likelihood · Marginal likelihood · Modified likelihood · Semi-competing risks

1 Introduction

In clinical studies, we often observe semi-competing risks data (Fine et al. 2001), which involve two-types of events; a non-terminal event (e.g. disease recurrence) and a terminal event (e.g. death). Here, a subject may experience both events that may be dependent (Chen 2012). Thus, several authors have recently studied semi-parametric

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10985-019-09464-2) contains supplementary material, which is available to authorized users.

✉ Il Do Ha
idha1353@pknu.ac.kr

[1] Department of Statistics, Pukyong National University, Busan 608-737, South Korea

[2] School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore

[3] Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, USA

[4] Department of Statistics, Seoul National University, Seoul 151-742, South Korea

frailty models for semi-competing risks data (Xu et al. 2010; Zhang et al. 2014; Varadhan et al. 2014; Lee et al. 2015). In particular, Xu et al. (2010) proposed a marginal-likelihood approach under the gamma frailty model. Zhang et al. (2014) and Lee et al. (2015) have proposed Bayesian approaches. However, the marginal likelihood or Bayesian approaches may often involve evaluation of intractable integrals over the frailty distributions, especially for the non-gamma frailties.

Unlike the classical likelihood for fixed parameters only, the hierarchical likelihood (h-likelihood; Lee and Nelder 1996; Lee et al. 2017) is constructed for both fixed parameters and unobserved frailties at the same time. Ha et al. (2001) proposed the h-likelihood for the frailty models where the maximum likelihood (ML) estimator can be substantially biased in the finite samples, particularly for the frailty parameter (Barker and Henderson 2005; Ha et al. 2010). This is because the number of nuisance parameters associated with the baseline hazard function increases with the number of events. In addition, the bias can also occur when the cluster size n_i for subject i is very small such as $n_i = 1$ or 2 (Ha et al. 2010). Simulation results from Ha et al. (2010) showed that the bias of ML estimator reduces slowly when $n_i = 1$ or 2 as the sample size (i.e. the number of clusters or subjects) grows.

In this paper, we extend the h-likelihood (Ha et al. 2001) for one shared frailty model to semi-competing risks data. Specifically three shared frailty models are considered to incorporate three states in the semi-competing risks setting; state 0 for on study, state 1 for non-terminal event, and state 2 for terminal event. We treat each subject as a cluster, which would generate realizations of two semi-competing event times. This formulates the semi-competing risks problem into a multivariate survival data setting with cluster size of $n_i = 2$. With this small cluster size, therefore, the aforementioned bias problem would still exist when the h-likelihood method is directly applied to this case. To overcome this, we propose a modified estimation procedure of Ha et al. (2001) where the h-likelihoods are constructed to consider the dependency and left-truncation as well as a Markov specification for the terminal event following non-terminal event (multi-state modelling). We also investigate the relationship between marginal likelihood and proposed h-likelihood approaches. In geneal, the h-likelihood method provides efficient statistical inference for various univariate and multivariate survival models (Ha et al. 2017), as well as other statistical models such as generalized linear models with random effects, joint models for different types of responses, and missing or incomplete-data problems, etc. (Lee et al. 2017).

This paper is organized as follows. In Sect. 2 we derive the h-likelihood procedure. We also propose the modifications of likelihoods via the h-likelihood and study their relationships. In Sect. 3, we present simulation studies to assess performance of the proposed method. Section 4 illustrates the proposed method using a real data set from a breast cancer study conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP) (Fisher et al. 1989, 1996). Section 5 concludes with a discussion. Technical details for the estimation procedures are provided in "Appendix".

2 Semi-competing risks frailty models and estimation

2.1 The model

Suppose that a subject may experience a terminal event (e.g. death) with or without a non-terminal event (e.g. disease recurrence). Let T_{i1} and T_{i2} be the non-terminal and terminal event times for the ith subject, respectively ($i = 1, \ldots, n$).

A schematic diagram for semi-competing risks data as shown in Xu et al. (2010) has three states (state 0, on study; state 1, recurrence; state 2, death). To model these three states under semi-competing risks, we consider the following specification of the hazard functions:

$$\lambda_1(t_1) = \lim_{\Delta t \to 0} Pr\{t_1 \le T_1 \le t_1 + \Delta t | T_1 \ge t_1, T_2 \ge t_1\}/\Delta t, \ t_1 > 0, \tag{1}$$

$$\lambda_2(t_2) = \lim_{\Delta t \to 0} Pr\{t_2 \le T_2 \le t_2 + \Delta t | T_1 \ge t_2, T_2 \ge t_2\}/\Delta t, \ t_2 > 0, \tag{2}$$

$$\lambda_{12}(t_2|t_1) = \lim_{\Delta t \to 0} Pr\{t_2 \le T_2 \le t_2 + \Delta t | T_1 = t_1, T_2 \ge t_2\}/\Delta t, \ 0 < t_1 < t_2. \tag{3}$$

For $\lambda_{12}(t_2|t_1)$ we assume a Markov process where the transition probability from state 1 to state 2 does not depend on the duration in state 1 (Aalen et al. 2008). That is, we assume

$$\lambda_{12}(t_2|t_1) = \lambda_{12}(t_2), 0 < t_1 < t_2. \tag{4}$$

Note that for transition from state 1 to state 2, the left-truncation time is t_1, the time at which the recurrence occurred.

Let x_i be a p-dimensional vector of covariates of the ith subject. The classical semi-competing risks model (i.e. illness-death model; Lawless 2003) can be extended to the frailty models which induce the dependency between non-terminal and terminal event times. For simplicity, we consider the semi-competing risks frailty models (Xu et al. 2010) with the common frailty (random effect), denote by u_i for the ith subject. That is, given u_i, the conditional hazards functions extended from Eqs. (1)–(3) can be expressed as three shared frailty models

$$\lambda_{1i}(t_1|u_i; x_i) = \lambda_{01}(t_1) \exp(x_i^T \beta_1) u_i, \ t_1 > 0, \tag{5}$$

$$\lambda_{2i}(t_2|u_i; x_i) = \lambda_{02}(t_2) \exp(x_i^T \beta_2) u_i, \ t_2 > 0, \tag{6}$$

$$\lambda_{12i}(t_2|t_1, u_i; x_i) = \lambda_{03}(t_2) \exp(x_i^T \beta_3) u_i, \ 0 < t_1 < t_2, \tag{7}$$

where $\lambda_{01}(\cdot)$, $\lambda_{02}(\cdot)$ and $\lambda_{03}(\cdot)$ are the unspecified baseline hazard functions and the frailties u_i are assumed to be independent and identically distributed with a density function having a frailty parameter α. Recall that from (4) and (7) $\lambda_{12i}(t_2|t_1, u_i; x_i) = \lambda_{12i}(t_2|u_i; x_i)$ depends only on t_2 and u_i, not t_1. The common distributions assumed for u_i are gamma and log-normal (Therneau and Grambsch 2000; Ha et al. 2001), where $E(u_i) = 1$ and $\text{var}(u_i) = \alpha$ for the gamma frailty model, and $v_i = \log u_i \sim N(0, \alpha)$ for the lognormal one.

2.2 Estimation procedures

Let C_i denote the right censoring time for the ith subject ($i = 1, \ldots, n$). If the subject experiences a terminal event before the non-terminal event occurs, we define $T_{i1} = \infty$. Then, for the ith subject we have the following observable data:

$$y_{i1} = T_{i1} \wedge y_{i2},\ y_{i2} = T_{i2} \wedge C_i,\ \delta_{i1} = I(T_{i1} \le y_{i2}) \text{ and } \delta_{i2} = I(T_{i2} \le C_i),$$

where $0 \le y_{i1} \le y_{i2}$.

For the likelihood-based inference, the h-loglikelihood for the semi-competing risks frailty models (5)–(7) is defined by (Ha et al. 2001)

$$h = h(\beta, v, \lambda_0, \alpha) = \sum_i \ell_{1i} + \sum_i \ell_{2i}, \tag{8}$$

where $\ell_{1i} = \ell_{1i}(\beta, \lambda_0; y_i^o|u_i)$ is the logarithm of the conditional density function for the observed data $y_i^o = (y_{i1}, y_{i2}, \delta_{i1}, \delta_{i2})$ given u_i. Following Xu et al. (2010), we have

$$\begin{aligned}
\ell_{1i} &= \log\Bigg[\lambda_{1i}(y_{i1}|u_i)^{\delta_{i1}}\lambda_{2i}(y_{i2}|u_i)^{\delta_{i2}(1-\delta_{i1})}\lambda_{12i}(y_{i2}|u_i)^{\delta_{i1}\delta_{i2}} \\
&\quad \times \exp\Bigg[-\int_0^{y_{i1}} \{\lambda_{1i}(t|u_i) + \lambda_{2i}(t|u_i)\}\, dt\Bigg] \times \exp\Bigg\{-\int_{y_{i1}}^{y_{i2}} \lambda_{12i}(t|u_i)\, dt\Bigg\}\Bigg] \\
&= \delta_{i1}\{\log\lambda_{01}(y_{i1}) + \eta_{i1}\} + \delta_{i2}(1-\delta_{i1})\{\log\lambda_{02}(y_{i2}) + \eta_{i2}\} \\
&\quad + \delta_{i1}\delta_{i2}\{\log\lambda_{03}(y_{i2}) + \eta_{i3}\} \\
&\quad - \{\Lambda_{01}(y_{i1})\exp(\eta_{i1}) + \Lambda_{02}(y_{i1})\exp(\eta_{i2}) + \Lambda_{03}(y_{i1}, y_{i2})\exp(\eta_{i3})\},
\end{aligned}$$

and

$$\ell_{2i} = \ell_{2i}(\alpha; v_i)$$

is the logarithm of the density function of $v_i = \log u_i$ with a parameter α. Here

$$\eta_{i1} = x_i^T\beta_1 + v_i,\ \eta_{i2} = x_i^T\beta_2 + v_i \text{ and } \eta_{i3} = x_i^T\beta_3 + v_i,$$

$\beta = (\beta_1^T, \beta_2^T, \beta_3^T)^T$ with $\beta_j = (\beta_{j1}, \ldots, \beta_{jp})^T$, $v = (v_1, \ldots, v_n)^T$, and $\lambda_0 = (\lambda_{01}, \lambda_{02}, \lambda_{03})^T$, and $\Lambda_{03}(s,t) = \Lambda_{03}(t) - \Lambda_{03}(s)$ reflecting the left truncation. We have $\ell_{2i} = \alpha^{-1}(v_i - u_i) - \log\Gamma(\alpha^{-1}) - \alpha^{-1}\log\alpha$ for the gamma frailty, and $\ell_{2i} = -\log(2\pi\alpha)/2 - v_i^2/(2\alpha)$ for the log-normal frailty. By adding the frailty in the likelihood, the likelihood in (8) can account for the dependency between the non-terminal and terminal events. Unlike a single frailty model studied in Ha et al. (2001), the proposed method focuses on three shared frailty models (5)–(7) for semi-competing risks data.

Note that the functional forms of λ_{0j} ($j = 1, 2, 3$) from $\sum_i \ell_{1i}$ in (8) are unknown. Let $y_{1(1)}, y_{1(2)}, \ldots, y_{1(D_1)}$ be ordered distinct recurrence times with $(\delta_{i1}, \delta_{i2})$=(1, 0) or

(1, 1), and $y_{2(1)}, y_{2(2)}, \ldots, y_{2(D_2)}$ be ordered distinct death times without recurrence with $(\delta_{i1}, \delta_{i2}) = (0, 1)$, and $y_{3(1)}, y_{3(2)}, \ldots, y_{3(D_3)}$ be ordered distinct death times following recurrence with $(\delta_{i1}, \delta_{i2}) = (1, 1)$. Thus, following Fan and Li (2002) and Ha et al. (2011), we approximate the baseline cumulative hazard function $\Lambda_{0j}(t)$ $(j = 1, 2, 3)$ by a step function with jumps λ_{0jk_j} at the observed distinct event times;

$$\Lambda_{0j}(t) = \sum_{k_j : y_{j(k_j)} \le t} \lambda_{0jk_j} \ (j = 1, 2, 3), \tag{9}$$

where $y_{j(k_j)}$ is the k_jth $(k_j = 1, \ldots, D_j)$ smallest distinct event time for each j, and $\lambda_{0jk_j} = \lambda_{0j}(y_{j(k_j)})$. Then $\sum_i \ell_{1i}$ in (8) can be rewritten as

$$\begin{aligned}\sum_i \ell_{1i} = &\sum_{k_1} d_{1(k_1)} \log \lambda_{01k_1} + \sum_i \delta_{i1}\eta_{i1} - \sum_{k_1} \lambda_{01k_1} \left\{ \sum_{i \in R_{(k_1)}} \exp(\eta_{i1}) \right\} \\ &+ \sum_{k_2} d_{2(k_2)} \log \lambda_{02k_2} + \sum_i \delta_{i2}(1 - \delta_{i1})\eta_{i2} - \sum_{k_2} \lambda_{02k_2} \left\{ \sum_{i \in R_{(k_2)}} \exp(\eta_{i2}) \right\} \\ &+ \sum_{k_3} d_{3(k_3)} \log \lambda_{03k_3} + \sum_i \delta_{i1}\delta_{i2}\eta_{i3} - \sum_{k_3} \lambda_{03k_3} \left\{ \sum_{i \in R_{(k_3)}} \exp(\eta_{i3}) \right\}, \end{aligned} \tag{10}$$

where $d_{j(k_j)}$ $(j = 1, 2, 3)$ is the number of the events at $y_{j(k_j)}$, and

$$\begin{aligned} R_{(k_1)} &= R(y_{1(k_1)}) = \{i : y_{i1} \ge y_{1(k_1)}\}, \\ R_{(k_2)} &= R(y_{2(k_2)}) = \{i : y_{i1} \ge y_{2(k_2)}\}, \text{ and} \\ R_{(k_3)} &= R(y_{3(k_3)}) = \{i : y_{i1} < y_{3(k_3)} \le y_{i2}\}, \end{aligned}$$

are the risk sets at $y_{1(k_1)}$, $y_{2(k_2)}$ and $y_{3(k_3)}$, respectively. Define $\lambda_{01} = (\lambda_{011}, \ldots, \lambda_{01D_1})^T$, $\lambda_{02} = (\lambda_{021}, \ldots, \lambda_{02D_2})^T$, and $\lambda_{03} = (\lambda_{031}, \ldots, \lambda_{03D_3})^T$. Note that the number of nuisance parameters λ_{0j}'s can increase with the number of events, which can be viewed as the Neyman-Scott problem (Neyman and Scott 1948; Lee and Nelder 2009). Accordingly, for estimation of (β, v), Ha et al. (2001) proposed the use of the profile h-likelihood $h^* = h^*(\beta, v, \alpha)$, where λ_{0j} $(j = 1, 2, 3)$ are replaced by their non-parametric estimates as follows:

$$h^* = h|_{\lambda_{0j} = \widehat{\lambda}_{0j}} = \sum_i \ell^*_{1i} + \sum_i \ell_{2i},$$

where $\ell^*_{1i} = \ell^*_{1i}(\beta, v) = \ell_{1i}|_{\lambda_{0j} = \widehat{\lambda}_{0j}}$ and

$$\widehat{\lambda}_{0jk_j}(\beta, v) = \frac{d_{j(k_j)}}{\sum_{i \in R_{(k_j)}} \exp(\eta_{ij})}, \ (j = 1, 2, 3)$$

are the solutions of the estimating equations, $\partial h/\partial \lambda_{0jk_j} = 0$, for $k_j = 1, \ldots, D_j$. Now, from (10) we have

$$\begin{aligned}\sum_i \ell_{1i}^* = &\sum_{k_1} d_{1(k_1)} \log \widehat{\lambda}_{01k_1} + \sum_i \delta_{i1}\eta_{i1} - \sum_{k_1} d_{1(k_1)} \\ &+ \sum_{k_2} d_{2(k_2)} \log \widehat{\lambda}_{02k_2} + \sum_i \delta_{i2}(1-\delta_{i1})\eta_{i2} - \sum_{k_2} d_{2(k_2)} \\ &+ \sum_{k_3} d_{3(k_3)} \log \widehat{\lambda}_{03k_3} + \sum_i \delta_{i1}\delta_{i2}\eta_{i3} - \sum_{k_3} d_{3(k_3)},\end{aligned}$$

which is proportional to the conditional log-partial likelihood $\ell_p = \ell_p(\beta, v)$, given by

$$\begin{aligned}\ell_p = &\sum_i \delta_{i1}\eta_{i1} - \sum_{k_1} d_{1(k_1)} \log\Big\{ \sum_{i \in R_{(k_1)}} \exp(\eta_{i1}) \Big\} \\ &+ \sum_i \delta_{i2}(1-\delta_{i1})\eta_{i2} - \sum_{k_2} d_{2(k_2)} \log\Big\{ \sum_{i \in R_{(k_2)}} \exp(\eta_{i2}) \Big\} \\ &+ \sum_i \delta_{i1}\delta_{i2}\eta_{i3} - \sum_{k_3} d_{3(k_3)} \log\Big\{ \sum_{i \in R_{(k_3)}} \exp(\eta_{i3}) \Big\}\end{aligned}$$

with the constant terms omitted. This leads to the partial h-likelihood (Ha et al. 2010)

$$h_p = h_p(\beta, v, \alpha) = \ell_p + \sum_i \ell_{2i}, \tag{11}$$

which is proportional to the profile h-likelihood h^*. Thus, once we have h_p, the h-likelihood method (Ha and Lee 2003; Ha et al. 2010) can be directly extended to the semi-competing risks frailty model. Note that Therneau and Grambsch (2000) and Ripatti and Palmgren (2000) defined a penalized partial likelihood (PPL) by substituting the log-partial likelihood ℓ_p for $\sum_i \ell_{1i}$ in the h-likelihood (8), where $\sum_i \ell_{2i}$ was regarded as a penalty term. Our h-likelihood method and their PPL procedure are essentially the same in estimation of (β, v) given α, but are different in estimation of α (Ha et al. 2010, 2011).

Specifically, the h-likelihood procedure based on h_p is as follows. Following Ha et al. (2001), given α we estimate $\tau = (\beta^T, v^T)^T$ by maximizing h_p with respect to τ. The estimating equations for $\tau = (\beta^T, v^T)^T$ are then in form of

$$\partial h_p/\partial \tau = (\partial h/\partial \tau)|_{\lambda_0 = \widehat{\lambda}_0} = 0, \tag{12}$$

where $\lambda_0 = (\lambda_{01}^T, \lambda_{02}^T, \lambda_{03}^T)^T$. Note that the asymptotic covariance matrix for $\widehat{\tau} - \tau$ is given by the inverse of $H(h_p, \tau) = -\partial^2 h_p/\partial \tau^2$. This asymptotic covariance is constructed along the same lines of work by Lee and Nelder (1996, Section 3.3), which

proved that with the h-likelihood h, the inverse of $H(h, \tau)$ is an asymptotic covariance matrix of $\widehat{\tau}-\tau$ under the generalized linear models with random effects. By substituting h with h_p given in (11), their asymptotic result can be applied straightforwardly to a class of semiparametric frailty models because the partial h-likelihood h_p does not depend on the nuisance quantities λ_0 (Ha and Lee 2003; Ha et al. 2001, 2017).

Let $\ell = \ell(\theta, \psi)$ be a likelihood function, either an h-likelihood h or a marginal likelihood m (defined in (15) in Section 2.3), with nuisance parameters θ and the parameters of interest ψ. Here, if ψ is α, the nuisance parameters θ can be either fixed effects (β, λ_0) or random effects v or both. Consider a function $p_\theta^\ell(\psi)$, defined by

$$p_\theta^\ell(\psi) = \left[\ell - \frac{1}{2}\log\det\{H(\ell, \theta)/(2\pi)\}\right]\Big|_{\theta=\widehat{\theta}}, \tag{13}$$

where $H(\ell, \theta) = -\partial^2\ell/\partial\theta^2$ is an adjustment term to eliminate θ, and $\widehat{\theta}$ solves $\partial\ell/\partial\theta = 0$. The function $p_\theta^\ell(\psi)$ produces an adjusted profile likelihood for ψ evaluated at $\theta = \hat{\theta}$. Following Lee and Nelder (2001) and Ha et al. (2010), we abbreviate $p_\theta^\ell(\psi)$ in (13) by $p_\theta(\ell)$. Next, to estimate the frailty parameter α, we use the adjusted profile h-likelihood $p_\tau(h_p)$ (Ha and Lee 2003; Ha et al. 2017), given by

$$p_\tau(h_p) = \left[h_p - \frac{1}{2}\log\det\left\{H(h_p, \tau)/(2\pi)\right\}\right]\Big|_{\tau=\widehat{\tau}},$$

where $\widehat{\tau} = \widehat{\tau}(\alpha) = (\widehat{\beta}^T(\alpha), \widehat{v}^T(\alpha))^T$ and $\widehat{\tau}$ solves $\partial h_p/\partial\tau = 0$. Thus the estimating equation for α is given by

$$\partial p_\tau(h_p)/\partial\alpha = 0.$$

This procedure may work well for the log-normal frailty (Ha et al. 2011), but not for the gamma frailty, so we use the second-order approximation (Lee and Nelder 2001; Lee et al. 2017), defined by

$$s_{\beta,v}(h_p) = p_\tau(h_p) - \{F(h_p)/24\},$$

where $F(h_p) = \mathrm{tr}[-\{3(\partial^4 h_p/\partial v^4) + 5(\partial^3 h_p/\partial v^3)H(h_p, v)^{-1}(\partial^3 h_p/\partial v^3)\}$ $H(h_p; v)^{-2}]|_{v=\widehat{v}}$. To reduce the computational burden, Ha et al. (2010) used $F(h)$ instead of $F(h_p)$, and Noh and Lee (2007) showed the resulting dispersion-parameter estimators from two restricted likelihoods, $p_\tau(h)$ and $p_v(h)$, are asymptotically equivalent (Ha et al. 2007). Therefore, we can use

$$s_v(h_p) = p_v(h_p) - \{F(h)/24\}, \tag{14}$$

to replace $s_{\beta,v}(h_p)$ and define

$$p_v(h_p) = \left[h_p - \frac{1}{2}\log\det\{H(h_p, v)/(2\pi)\}\right]\Big|_{v=\widehat{v}},$$

where $H(h_p, v) = -\partial^2 h_p/\partial v^2$ is an adjustment term to eliminate v and $\widehat{v} = \widehat{v}(\alpha)$ solves $\partial h_p/\partial v = 0$. In particular, under the gamma frailty $F(h)$ has a simple form,

$$F(h) = \sum_i \{-2(\alpha^{-1} + \delta_{i+})^{-1}\},$$

where $\delta_{i+} = \delta_{i1} + \delta_{i2}$.

The classical clustered survival data have a cluster size of $n_i \geq 2$, and the bias becomes smaller as n_i increases. In the semi-competing risks setting, however, we consider each cluster contains only two observations for the non-terminal and terminal event times, that is $n_i = 2$ for all i. Thus, further modifications on the h-likelihood given in the following subsection are necessary for a bias correction.

2.3 Modification of the likelihood

In the standard semi-parametric frailty models, Ha et al. (2010) showed that the ML estimators can be substantially biased in the finite samples. In this section we propose likelihood-based modifications to reduce such biases in the semi-parametric frailty models (5)–(7) under semi-competing risks. We consider a case where parameters of interest are (β, α) and nuisance quantities are λ_0 (fixed parameters) and v (random effects). For inference on (β, α), we need to eliminate the nuisance quantities (λ_0, v), the dimension of which increase with sample size and number of events. There are typically two ways of eliminating the nuisance parameters using the h-likelihood: one is to integrate v out of the h-likelihood and the other is to profile out λ_0. In this paper, we propose the following two methods to eliminate such nuisance quantities efficiently from the h-likelihood:

Method 1 Eliminate v first and then λ_0,

Method 2 Eliminate λ_0 first and then v.

We now show how to construct the likelihoods depending on the order of the quantities being eliminated. Firstly, in Method 1 we consider the marginal log-likelihood, denoted by m, which can be obtained by integrating out the frailties v from the h-likelihood, i.e.

$$m = m(\beta, \lambda_0, \alpha) = \sum_i \log\left\{\int \exp(h_i)\, dv_i\right\}, \tag{15}$$

where $h_i = \ell_{1i} + \ell_{2i}$ is the contribution of the ith individual to h in (8). Then we construct a profile marginal log-likelihood m^* by plugging in the estimates of λ_0, defined by

$$m^* = m^*(\beta, \alpha) = m|_{\lambda_{0j} = \widetilde{\lambda}_{0j}}, \tag{16}$$

where $\widetilde{\lambda}_{0jk_j}(\beta, v)$, $(j = 1, 2, 3)$ are the solutions of the estimating equations, $\partial m/\partial \lambda_{0jk_j} = 0$, for $k_j = 1, \ldots, D_j$. Secondly, in Method 2 we consider the partial

h-likelihood $h_p = h_p(v, \beta, \alpha)$ in (11), where λ_0 has been already eliminated by profiling. Thus we can construct a partial marginal log-likelihood m_p (Ha et al. 2010) by integrating out the frailty v from h_p,

$$m_p = m_p(\beta, \alpha) = \log\left\{\int \exp(h_p)\, dv\right\}. \tag{17}$$

The marginal log-likelihood m often requires a numerical integration (e.g. for the log-normal frailty) except for the gamma frailty. Note that the resulting ML estimators by maximizing m are equivalent to those by maximizing m^*. The partial marginal log-likelihood m_p gives a partial ML estimator without finite sample biases (Therneau et al. 2003; Gu et al. 2004) due to an efficient elimination of the nuisance quantities. However, m_p is is not easy to use due to intractable integration that does not allow a closed form even with the gamma frailty. Moreover, m_p involves high dimensional integration where the dimension increases with the number of frailties (Gu et al. 2004).

As an adequate approximation to m_p, Ha et al. (2010) proposed to use an adjusted profile marginal likelihood $p_\omega(m)$, which was defined as a function of (β, α),

$$p_\omega(m) = \left[m - \frac{1}{2}\log\det\{H(m, \omega)/(2\pi)\}\right]\Big|_{\omega=\widetilde{\omega}}, \tag{18}$$

where $\omega = (\omega_1, \ldots, \omega_r)^T$ with $\omega_k = \log\lambda_{0k}$, $H(m, \omega) = -\partial^2 m/\partial\omega^2$ is an adjustment term to eliminate λ_0, and $\widetilde{\omega}$ solves $\partial m/\partial\omega = 0$. Under the standard gamma frailty models, Ha et al. (2010) showed that as $n^* = \min_{1\le i\le q} n_i \to \infty$ for cluster size n_i for subject i,

$$p_\omega(m) \approx s_v(h_p).$$

Remark 1 For the semi-competing risks models (5)–(7) with gamma frailty, the marginal likelihood has an explicit form. In "Appendix A", we show that the marginal likelihood procedure is equivalent to that of Xu et al. (2010). In "Appendix B", we show that given α, the h-likelihood and marginal likelihood procedures give the same estimators as in the standard gamma frailty models (Ha et al. 2001; Ha and Lee 2003) and they are compared in terms of the EM, which provides the ML estimators. □

For simplicity, we consider the gamma frailty models for semi-competing risks data where the marginal likelihood m has an explicit form. We have found that $s_v(h_p)$ sometimes gives a convergence problem in fitting the semi-competing risks gamma frailty models, particularly for a large α (e.g. $\alpha \ge 1$). To overcome this problem, we further consider a higher-order approximation based on the h-likelihood. Following Tierney and Kadane (1986) and Lee et al. (2017), we can show that with gamma frailty, the fourth-order Laplace approximation (denoted by $m_v(h)$) to the marginal likelihood m is given by

$$m_v(h) = s_v(h) - F^*(h),$$

where $s_v(h) = p_v(h) - F(h)$ is the second-order Laplace approximation to m in (15) and $F^*(h) = (1/360)\sum_i(\alpha^{-1} + \delta_{i+})^{-3}$, which is equivalent to approximating m by the fourth-order Stirling approximation

$$\log\Gamma(x) \doteq (x - 1/2)\log(x) + \log(2\pi)/2 - x + 1/(12x) - 1/(360x^3).$$

Accordingly we define a modified h-likelihood based on h_p,

$$m_v(h_p) = s_v(h_p) - F^*(h), \tag{19}$$

where $s_v(h_p)$ is a function of (β, α) as is given in (14). The modified h-likelihood $m_v(h_p)$ is also a function of (β, α) and is a higher-order approximation to m_p in (17). Note that $s_v(h)$ and $s_v(h_p)$ are the second-order Laplace approximations to m in (15) and m_p in (17), respectively and that $m_v(h)$ and $m_v(h_p)$ are the fourth-order Laplace approximations to m and m_p, respectively.

The first-order Laplace approximation becomes exact as $n^* = \min_{1\le i\le q} n_i \to \infty$. However, it may lead to a serious bias when cluster size n_i is small. The second-order approximation generally reduces such bias to some extent (Lee et al. 2017). In the semi-competing risks setting, we suggest an even higher order approximation $m_v(h_p)$ (fourth order) for more effective bias correction.

In summary, we consider four likelihoods for $\psi = (\beta, \alpha)^T$ constructed by **Methods 1** and **2** as follows:

m^* in (16): a profile marginal likelihood by **Method 1**,
$p_\omega(m)$ in (18) : an adjusted profile marginal likelihood by **Method 1**,
$s_v(h_p)$ in (14): the second-order Laplace approximation to m_p by **Method 2**,
$m_v(h_p)$ in (19): a modified h-likelihood based on h_p by **Method 2**,

where the h-likelihood methods have been mainly used in **Method 2**. Notice that simultaneous eliminations (e.g. $p_{\lambda_0,v}(h)$) of the nuisance quantities (λ_0, v) may require a heavy computation because the dimension of the corresponding Hessian matrix $H(h, (\lambda_0, v))$ increases with sample size and the number of events.

We call the estimators maximizing the marginal likelihoods m^* and $p_\omega(m)$ the maximum marginal likelihood 1 (MML1) and 2 (MML2) estimators, respectively. It can be shown that with gamma frailty, the MML1 estimator is equivalent to the ML estimator provided by Xu et al. (2010): see "Appendix A" for more details. We also call the estimators maximizing the partial likelihoods $s_v(h_p)$ and $m_v(h_p)$ the maximum partial likelihood 1 (MPL1) and 2 (MPL2) estimators, respectively. Accordingly, the four estimators of $\psi = (\beta, \alpha)^T$ are summarized as follows:

$$\textbf{Method 1}\quad \hat{\psi}^{\mathrm{MML1}} = \arg\max_{\psi} m^* \ \text{ and } \ \hat{\psi}^{\mathrm{MML2}} = \arg\max_{\psi} p_\omega(m),$$

$$\textbf{Method 2}\quad \hat{\psi}^{\mathrm{MPL1}} = \arg\max_{\psi} s_v(h_p) \ \text{ and } \ \hat{\psi}^{\mathrm{MPL2}} = \arg\max_{\psi} m_v(h_p).$$

The fitting algorithm for the four estimation methods is summarized as follows:

Step 0 Find the initial estimates of (β, v); i.e. take (0,…,0, 0,…,0) as all initial estimates of components of (β, v).

Step 1 In the inner loop, given α, we maximize h_p for β and v.
Step 2 In the outer loop, given β and v the four likelihoods, m^*, $p_\omega(m)$, $s_v(h_p)$ and $m_v(h_p)$ are maximized for α. A simple grid search method is used to estimate α (Therneau and Grambsch 2000; Ha et al. 2010).

After convergence, we compute the estimated standard errors of $\hat{\beta}$ from the inverse of the observed information, $-\partial^2 h_p/\partial\tau^2$, in (12).

3 Simulation study

We have performed simulation studies to assess the finite sample performance of the proposed estimation methods, MML2, MPL1 and MPL2, in comparison with the MML1 method developed by Xu et al. (2010).

We generated data under the semi-competing risks frailty models (5)–(7) as follows: (i) generate the common frailty u_i from a gamma distribution with mean 1 and variance $\alpha = 0.5$ or 1.0, and (ii) given the frailty, generate two event times t_{i1} and t_{i2} independently from the proportional hazards models (5) and (6), i.e.

$$\lambda_{1i}(t_1|u_i; x_i) = \lambda_{01}(t_1)\exp(x_i^T\beta_1)u_i \quad \text{and} \quad \lambda_{2i}(t_2|u_i; x_i) = \lambda_{02}(t_2)\exp(x_i^T\beta_2)u_i,$$

respectively, and (iii) the transition time t_{i2}^* is generated based on the transition model from state 1 to state 2 in the form of

$$\lambda_{12i}(t_2|u_i; x_i) = \lambda_{03}(t_2)\exp(x_i^T\beta_3)u_i.$$

First, we consider a single covariate case where $x_i = x_{i1}$, which follows the standard normal distribution, for $i = 1, \ldots, n$. The baseline hazard rates are set to be $\lambda_{01}(t_1) = \lambda_{02}(t_2) = 1$, $\lambda_{03}(t_2) = 2$, and the regression coefficients to be $\beta_1 = \beta_2 = \beta_3 = 0.5$.

We evaluate the proposed methods under two scenarios with fixed and random censoring times, denoted by c_i, as follows:

(i) **Scenario 1** c_i is fixed as the duration of the follow-up, $c_i = 3$, yielding a censoring rate ranging from 40% to 60% for the non-terminal event time y_{i1} and 10% to 30% for terminal event time y_{i2}, and
(ii) **Scenario 2** c_i is randomly generated from a 50-50 mixture of a uniform distribution on interval (1.5, 3) and a degenerate distribution concentrated at 3, according to the censoring scheme used in Xu et al. (2010), resulting in the similar ranges of censoring rates for y_{i1} and y_{i2} as in Scenario 1.

We considered various sample sizes of $n = 100, 250$ and 500, and implemented 200 replications for each simulation setting. Simulation results from different methods are reported in terms of percentage of relative bias (%rbias) and mean squared error (MSE) for $\hat{\alpha}$ and $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^T$. Based on 200 replications, we also computed the standard deviation (SD) and the mean of the estimated standard errors (SE) for $\hat{\beta}$. The SEs were obtained from the inverse of the observed information, $-\partial^2 h_p/\partial\tau^2$, in (12). The results are presented in Tables 1, 2 for Scenario 1 and Tables 3, 4 for Scenario 2.

Table 1 Simulation results under Scenario 1 with fixed censoring time

α	n	Parameter	MML1		MML2		MPL1		MPL2	
			%rbias	MSE	%rbias	MSE	%rbias	MSE	%rbias	MSE
0.5	100	α	− 26.4	0.132	− 17.8	0.202	18.6	0.299	13.4	0.254
		β_1	5.4	0.048	0.4	0.053	6.8	0.055	6.0	0.054
		β_2	4.2	0.059	− 3.0	0.044	3.6	0.047	2.8	0.046
		β_3	− 4.8	0.070	− 6.8	0.083	3.8	0.088	2.6	0.085
0.5	250	α	− 25.4	0.087	2.6	0.091	15.6	0.120	10.0	0.102
		β_1	− 3.8	0.020	− 0.2	0.018	0.8	0.019	− 0.2	0.019
		β_2	− 4.6	0.016	− 0.2	0.015	1.8	0.016	0.8	0.015
		β_3	− 7.4	0.027	1.0	0.033	3.4	0.032	1.8	0.031
0.5	500	α	− 24.2	0.059	− 6.0	0.047	2.4	0.052	− 1.4	0.048
		β_1	− 6.2	0.011	− 2.0	0.010	− 0.4	0.010	− 1.0	0.010
		β_2	− 3.2	0.009	0.6	0.008	2.4	0.009	1.6	0.008
		β_3	− 6.4	0.016	0.2	0.015	2.6	0.015	1.4	0.015
1	100	α	− 48.2	0.424	− 19.7	0.331	4.0	0.445	− 8.3	0.342
		β_1	0.4	0.063	− 0.5	0.066	3.0	0.070	1.2	0.065
		β_2	0.4	0.062	1.5	0.060	3.3	0.065	1.5	0.060
		β_3	− 7.9	0.089	− 6.2	0.100	− 1.7	0.107	− 4.1	0.100
1	250	α	− 37.7	0.319	− 13.9	0.220	6.8	0.293	− 7.8	0.197
		β_1	− 3.2	0.022	− 1.4	0.022	1.5	0.023	− 0.5	0.021
		β_2	− 3.4	0.026	− 1.2	0.025	1.6	0.027	− 0.4	0.025
		β_3	− 5.7	0.043	− 1.2	0.038	3.1	0.043	0.2	0.037
1	500	α	− 21.1	0.187	− 8.7	0.157	6.8	0.208	− 7.3	0.138
		β_1	− 2.4	0.015	− 0.6	0.015	1.7	0.016	− 0.4	0.014
		β_2	− 2.6	0.014	− 0.4	0.013	1.9	0.014	− 0.2	0.013
		β_3	− 4.6	0.023	− 2.0	0.020	1.3	0.022	− 1.6	0.019

MML maximum marginal likelihood, *MPL* maximum partial likelihood
%rbias, percentage of relative bias
MML1, m^* in (16); MML2, $p_\omega(m)$ in (18)
MPL1, $s_v(h_p)$ in (14); MPL2, $m_v(h_p)$ in (19)

From Table 1, we find that the relative biases of the proposed MML2, MPL1 and MPL2 estimates are smaller than those of the MML1 estimates in most of the settings. The MML1 method exhibits non-negligible biases in estimating the parameters, especially for the frailty parameter α, which confirms the simulation results in Ha et al. (2010) for the standard gamma frailty models. Table 2 shows that the proposed SEs are generally slightly underestimated as compared to the SDs. The similar findings are observed in Scenario 2 under the random censoring scheme, as shown in Tables 3, 4.

In Tables 1 and 3, the existing MML1 method gives somewhat smaller MSEs compared to other three methods (MML2, MPL1 and MPL2), especially when the frailty parameter α is as small as $\alpha = 0.5$, possibly due to underestimation of α by the MML1 method. It is worth noting that the proposed MPL2 estimation outperforms all the other proposed methods in terms of the MSE when the frailty parameter α is as

Table 2 Simulation results for comparison between estimated standard error (SE) and sample standard deviation (SD) for β under Scenario 1 with fixed censoring

α	n	Parameter	MML1		MML2		MPL1		MPL2	
			SD	SE	SD	SE	SD	SE	SD	SE
0.5	100	β_1	0.217	0.187	0.230	0.191	0.233	0.203	0.231	0.201
		β_2	0.241	0.187	0.210	0.191	0.217	0.203	0.213	0.201
		β_3	0.263	0.232	0.286	0.235	0.296	0.253	0.292	0.250
0.5	250	β_1	0.140	0.115	0.134	0.120	0.139	0.123	0.138	0.121
		β_2	0.126	0.115	0.123	0.120	0.126	0.122	0.124	0.121
		β_3	0.159	0.138	0.181	0.147	0.178	0.151	0.175	0.149
0.5	500	β_1	0.099	0.081	0.097	0.084	0.099	0.085	0.098	0.084
		β_2	0.095	0.081	0.092	0.084	0.092	0.085	0.091	0.084
		β_3	0.121	0.097	0.123	0.101	0.123	0.103	0.122	0.102
1	100	β_1	0.251	0.196	0.257	0.212	0.263	0.225	0.254	0.218
		β_2	0.248	0.197	0.244	0.213	0.253	0.227	0.244	0.220
		β_3	0.288	0.256	0.310	0.273	0.326	0.289	0.314	0.280
1	250	β_1	0.144	0.126	0.146	0.135	0.150	0.142	0.144	0.137
		β_2	0.157	0.126	0.157	0.135	0.164	0.142	0.158	0.137
		β_3	0.199	0.158	0.195	0.171	0.205	0.181	0.193	0.174
1	500	β_1	0.122	0.093	0.121	0.097	0.125	0.101	0.119	0.097
		β_2	0.115	0.093	0.114	0.096	0.118	0.100	0.112	0.097
		β_3	0.144	0.117	0.139	0.121	0.148	0.127	0.137	0.122

large as $\alpha = 1$. This advantage might be from a reduced number of tied observations under Scenario 2 and improvememt in bias correction of the MPL2 estimation using the higher-order approximation (19) to a modified marginal likelihood m_p as described in Sect. 2.

We have conducted further simulations to assess performance of the proposed estimation methods when there are multiple covariates in each transition model. The simulation scheme is the same as before, except that three additional covariates were considered in each model, that is, $x_i = (x_{i1}, x_{i2}, x_{i3})^T$, where x_{i1} and x_{i2} follow the standard normal distribution, x_{i3} follows a Bernoulli(0.5) distribution, for $i = 1, \ldots, n$. The corresponding coefficients were set to be $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13}) = (0.5, 0.5, -0.5)$, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23}) = (0.5, 0.5, -0.5)$ and $\beta_3 = (\beta_{31}, \beta_{32}, \beta_{33}) = (1, 1, -1)$.

Again, we consider the following two scenarios with fixed and random censoring times c_i:

(i) **Scenario 3** c_i is fixed as the duration of the follow-up, $c_i = 3$, which yields a censoring rate ranging from 50% to 60% for the non-terminal event time y_{i1} and 20% to 30% for terminal event time y_{i2}, and

(ii) **Scenario 4** c_i is randomly generated from a 50-50 mixture of a uniform distribution on interval (1.5, 3) and a degenerate distribution concentrated at 3, resulting in the similar ranges of censoring rates for y_{i1} and y_{i2} as in Scenario 3.

Table 3 Simulation results under Scenario 2 with random censoring

α	n	Parameter	MML1		MML2		MPL1		MPL2	
			%rbias	MSE	%rbias	MSE	%rbias	MSE	%rbias	MSE
0.5	100	α	− 24.0	0.107	− 11.4	0.225	19.2	0.307	14.8	0.282
		β_1	10.4	0.040	0.2	0.051	6.2	0.055	5.0	0.054
		β_2	11.8	0.040	2.0	0.047	8.0	0.052	6.8	0.050
		β_3	− 7.2	0.071	− 6.8	0.089	2.4	0.098	0.6	0.094
0.5	250	α	− 30.4	0.101	− 3.0	0.101	13.4	0.160	6.8	0.122
		β_1	− 4.0	0.017	0.8	0.019	4.0	0.021	3.0	0.020
		β_2	− 3.2	0.015	2.2	0.017	5.4	0.019	4.4	0.018
		β_3	− 10.8	0.028	− 3.4	0.026	0.8	0.028	− 0.6	0.027
0.5	500	α	− 24.0	0.073	− 4.6	0.064	3.6	0.080	0.0	0.067
		β_1	− 4.4	0.010	− 0.2	0.009	1.4	0.010	0.8	0.009
		β_2	− 5.0	0.009	− 0.6	0.008	0.8	0.009	0.2	0.008
		β_3	− 9.4	0.022	− 3.4	0.021	− 1.2	0.022	− 2.6	0.021
1	100	α	− 48.5	0.453	− 21.4	0.367	4.9	0.477	− 9.1	0.352
		β_1	− 0.9	0.056	0.6	0.052	4.3	0.058	2.4	0.053
		β_2	− 0.4	0.059	− 2.5	0.062	0.8	0.066	− 1.1	0.063
		β_3	− 10.3	0.100	− 7.2	0.094	− 1.4	0.103	− 4.0	0.097
1	250	α	− 32.9	0.351	− 8.1	0.298	12.1	0.370	− 4.5	0.243
		β_1	− 2.7	0.029	0.0	0.027	2.8	0.029	0.5	0.025
		β_2	− 1.7	0.028	1.0	0.026	3.8	0.028	1.6	0.025
		β_3	− 7.5	0.049	− 2.7	0.049	1.6	0.052	− 1.6	0.046
1	500	α	− 24.6	0.198	− 11.9	0.170	6.2	0.246	− 9.3	0.145
		β_1	− 3.6	0.015	− 1.8	0.014	0.8	0.016	− 1.0	0.014
		β_2	− 2.9	0.015	− 1.2	0.013	1.3	0.015	− 0.6	0.013
		β_3	− 4.3	0.028	− 1.1	0.026	2.6	0.030	− 0.7	0.025

Simulation results based on 200 replications over different values of α (frailty variance) are presented in the Supplementary Material, where Tables S1-S4 are under Scenario 3 and Tables S5-S8 under Scenario 4. From Tables S1-S2, we can see that the relative biases of the proposed MML2, MPL1 and MPL2 estimates are smaller than those of the MML1 estimates under all settings. The MML1 method exhibits non-negligible biases in the parameter estimates, especially for the frailty parameter α. These results confirm the simulation results from Ha et al. (2010) for the standard gamma frailty models. Tables S3-S4 show that the proposed estimated standard errors are closer to their corresponding sample SDs compared to the SEs for the MML1 estimator in most cases, implying that the estimated SEs for the proposed estimators work more effectively than that for MML1 estimator.

The similar observations can be made from Tables S5-S8 under Scenario 4 with a random censoring scheme. From Tables S1, S2, S5 and S6, we find that all of those four methods produce comparable estimation for β in terms of the MSEs, while the

Table 4 Simulation results for comparison between estimated standard error (SE) and sample standard deviation (SD) for β under Scenario 2 with random censoring

α	n	Parameter	MML1		MML2		MPL1		MPL2	
			SD	SE	SD	SE	SD	SE	SD	SE
0.5	100	β_1	0.193	0.188	0.226	0.191	0.232	0.201	0.230	0.200
		β_2	0.190	0.189	0.216	0.192	0.224	0.202	0.220	0.200
		β_3	0.264	0.238	0.297	0.243	0.313	0.258	0.307	0.255
0.5	250	β_1	0.127	0.114	0.139	0.120	0.144	0.124	0.142	0.122
		β_2	0.123	0.115	0.129	0.121	0.135	0.124	0.132	0.123
		β_3	0.157	0.139	0.159	0.148	0.168	0.153	0.164	0.151
0.5	500	β_1	0.095	0.081	0.096	0.084	0.099	0.085	0.096	0.085
		β_2	0.093	0.082	0.090	0.084	0.093	0.086	0.090	0.085
		β_3	0.142	0.098	0.144	0.103	0.148	0.105	0.144	0.104
1	100	β_1	0.236	0.197	0.227	0.214	0.238	0.228	0.229	0.221
		β_2	0.243	0.199	0.248	0.213	0.256	0.228	0.250	0.220
		β_3	0.299	0.255	0.298	0.277	0.321	0.297	0.309	0.287
1	250	β_1	0.168	0.128	0.164	0.137	0.167	0.144	0.159	0.139
		β_2	0.167	0.129	0.162	0.138	0.163	0.145	0.156	0.139
		β_3	0.208	0.163	0.219	0.175	0.227	0.185	0.215	0.177
1	500	β_1	0.118	0.093	0.118	0.096	0.126	0.101	0.116	0.097
		β_2	0.117	0.093	0.114	0.096	0.123	0.101	0.112	0.097
		β_3	0.162	0.117	0.160	0.122	0.172	0.128	0.158	0.123

MPL2 method slightly outperforms the other methods in estimating α. Overall, the above findings indicate that the performance of the proposed estimators is sustainable for the semi-competing risks models with multiple covariates.

Remark 2 Our simulation experience indicates all of these four methods may encounter a convergence problem, caused by a monotone likelihood (Heinze and Schemper, 2001) as shown in the Cox's PH model. For example, the plot of the profile likelihood m^* against α shows a monotone function as in Ha et al. (2017, pp. 79). We have observed that the MML1 method (m^*) has a serious convergence problem for a small sample case (e.g. $n = 100$ with $(q, n_i) = (50, 2)$), while the other three methods, i.e. MML2, MPL1 and MPL2, generally overcome such problems. In particular, the MPL2 method converges most of the time except for a few cases with a small sample size (e.g. $n = 100$ with $(q, n_i) = (50, 2)$), leading to a bias reduction.

4 A practical example

For an illustration, we consider a data set from the B-14 phase III breast cancer clinical trial conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP, Fisher et al. 1989, 1996). Total 2572 eligible patients were followed up for five years since randomization. Patients were randomized to one of two treatment

Table 5 Observed event types by two treatment arms ($n = 2{,}572$ patients)

Types of event	Placebo	Tamoxifen	Total
Type 1 (State 0 → State 1): recurrence from study	108	72	180 (7.00%)
Type 2 (State 0 → State 2): death without recurrence	242	293	535 (20.80%)
Type 3 (State 1 → State 2): death after recurrence	331	209	540 (21.00%)
No event (Censoring)	613	704	1317 (51.21%)

arms, tamoxifen (1278 patients) or placebo (1294 patients). The patients' median age was 56 (range 25–75) and their average tumor size was about 2.2 cm.

The aim of this analysis is to investigate the effect of a hormonal treatment on a cancer recurrence and/or death, considering three event types. Type 1 is breast cancer recurrence, Type 2 is death without recurrence, and Type 3 is death after recurrence. Table 5 gives the observed numbers of event types in this data set. Here 180 patients (7.00%) experienced Type 1, 535 patients (20.80%) did Type 2, 540 patients (21.00%) did Type 3, and the remaining 1317 patients (51.21%) had no events. Table 5 also shows the numbers of observed event types by treatment arm.

In this analysis, we consider three covariates of interest: treatment ($x_{i1} = 1$ for tamoxifen and 0 for placebo), tumor size (x_{i2}) and age (x_{i3}). We first apply the naive transition model (5)–(7) without a frailty and then one with the gamma frailty. For estimation under the gamma frailty model, we use four likelihood methods (MML1, MML2, MPL1 and MPL2) as in the simulation study. The fitted results are listed in Table 6. The results from the naive model and frailty models using the four likelihood methods are very similar because the frailty-parameter estimates are all very small ($\hat{\alpha} = 0.087$ for MPL1, $\hat{\alpha} = 0.090$ for MPL2, $\hat{\alpha} = 0.059$ for MML1 and $\hat{\alpha} = 0.085$ for MML2). Moreover, to test the frailty effect $H_0 : \alpha \equiv \text{var}(u_i) = 0$ which is on the boundary of the parameter space, an asymptotic null distribution of the likelihood ratio test follows a 50:50 chi-square mixture, denoted by $\chi^2_{0:1}$ (Self and Liang 1987; Stram and Lee 1994; Ha et al. 2011) with its critical value equal to $\chi^2_{1,0.1} = 2.71$ at a 5% significance level. Let ℓ_B be the Breslow's log-likelihood (1974) for the naive model above, i.e. $u_i = 1$ for all i in the gamma frailty model with (5)–(7), defined by (Lee and Nelder 1996)

$$\ell_B = \lim_{\alpha \to 0} s_v(h_p).$$

The difference between the likelihood functions from the naive model and frailty model (MPL1) is $2\{s_v(h_p) - \ell_B\} = 0.7 (< 2.71)$, indicating that the frailty effect is not significant. The marginal likelihood method (MML1) also gives $2\{m^* - \ell_B\} = 0.4$.

In Table 6, the treatment effect (x_1) is significant for time to recurrence and time to death after recurrence, but not for time to death without recurrence. For time to death without recurrence, the sign of treatment effect is positive, which may be explained from the fact that more patients died without cancer recurrence in the tamoxifen group (293/535) than the placebo group (242/535). We also see that the use of tamoxifen significantly reduces breast cancer recurrence (Type 1), but it is not beneficial in

Table 6 Fitted results from the semi-competing risks models for NSABP B-14 data

Model	Time to recurrence Est. (SE)	Time to death without recurrence Est. (SE)	Time to death after recurrence Est. (SE)
Naive model			
Treatment (x_1)	−0.543 (0.077)	0.058 (0.087)	0.329 (0.089)
Age (x_2)	−0.015 (0.004)	0.090 (0.006)	0.007 (0.004)
Tumor size (x_3)	0.018 (0.002)	0.007 (0.003)	0.010 (0.003)
$-2\ell_B = 23886.1$			
Frailty model (MPL1)			
Treatment (x_1)	−0.552 (0.078)	0.046 (0.089)	0.332 (0.094)
Age (x_2)	−0.015 (0.004)	0.090 (0.006)	0.008 (0.004)
Tumor size (x_3)	0.018 (0.003)	0.008 (0.004)	0.011 (0.003)
Frailty α	0.087		
$-2s_v(h_p) = 23885.4$			
Frailty model (MPL2)			
Treatment (x_1)	−0.552 (0.078)	0.046 (0.089)	0.332 (0.094)
Age (x_2)	−0.015 (0.004)	0.090 (0.006)	0.008 (0.004)
Tumor size (x_3)	0.018 (0.003)	0.008 (0.004)	0.011 (0.003)
Frailty $\hat{\alpha}$	0.090		
$-2m_v(h_p) = 23885.4$			
Frailty model (MML1)			
Treatment (x_1)	−0.549 (0.077)	0.050 (0.088)	0.332 (0.093)
Age (x_2)	−0.015 (0.004)	0.090 (0.006)	0.007 (0.004)
Tumor size (x_3)	0.018 (0.002)	0.008 (0.004)	0.011 (0.003)
$\hat{\alpha}$	0.059		
$-2m^* = 23885.7$			
Frailty model (MML2)			
Treatment (x_1)	−0.551 (0.078)	0.046 (0.089)	0.332 (0.094)
Age (x_2)	−0.015 (0.004)	0.090 (0.006)	0.008 (0.004)
Tumor size (x_3)	0.018 (0.003)	0.008 (0.004)	0.011 (0.003)
Frailty $\hat{\alpha}$	0.085		
$-2p_\omega(m) = 23885.4$			

Naive model, semi-competing risks model without frailty
Frailty model, semi-competing risks model with gamma frailty
α, variance of gamma frailty

death after recurrence (Type 3). In terms of the other covariates, the age effect (x_2) is very significant for event types 1 and 2. The effect of tumor size (x_3) is positively significant for all three event types, implying that the event rate is significantly higher among patients whose tumor sizes were larger at surgery.

Next we restricted the data analysis only to older patients (1,776 patients with age $\geq$ 50). The results are summarized in Table 7. Here we present the results only

from three methods (MPL2, MML1 and MML2) because the MPL1 method did not converge. We find that the frailty parameter estimates are all larger compared to those in Table 5. The likelihood difference from the naive model is $2\{m_v(h_p) - \ell_B\} = 9.8 > 2.71$, indicating that the frailty effect is significantly large, i.e. $\alpha > 0$. The difference between the naive and profile marginal likelihoods also gives $2\{m^* - \ell_B\} = 9.3$, selecting the frailty model again. Thus, the results from the naive and frailty models are expected to be somewhat different. The treatment effects are overall similar to those in Table 6 even though their signs in the frailty model have changed for time to death without recurrence in Table 7. However, for time to death without recurrence and time to death after recurrence, the tumor size effect is not significant in the naive model, whereas it is in the frailty model.

In addition, we also considered older patients only (484 patients with age ≥ 65). The results are also summarized in Table 7. In particular, we find that the MML1 estimate for α ($\hat{\alpha} = 0.482$) is somewhat smaller compared to other two methods ($\hat{\alpha} = 1.021$ from MPL2 and $\hat{\alpha} = 1.238$ from MML2), which was also demonstrated in the simulation study. This underestimation is also reflected in the likelihood ratio test for $H_0 : \alpha = 0$. That is, the likelihood difference is $2\{m^* - \ell_B\} = 1.5 < 2.71$, whereas $2\{m_v(h_p) - \ell_B\} = 4.3 > 2.71$ and $2\{p_\omega(m) - \ell_B\} = 4.4 > 2.71$. This implies that the MPL2 and MML2 methods are sensitive enough to detect the significance of the frailty effect, but the MML1 method does not. We also observe that the sign of tumor-size effect is negative from the MML1 method, but becomes positive in the MPL2 and MML2 methods.

5 Discussion

We have shown how to eliminate nuisance quantities from the h-likelihood and thus how to find effective modifications (MML2, MPL1 and MPL2) to reduce the bias of the maximum likelihood estimators (MML1). In general, the adjusted profile marginal likelihood ($p_\omega(m)$) is hard to use because an explicit form of the marginal likelihood (m) is not available. For the models such as the lognormal frailty or with correlated frailties, we recommend using the modified likelihoods, $s_v(h_p)$ or $m_v(h_p)$, based on the partial h-likelihood h_p. This implies that elimination of the nuisance quantities by **Method 2** proposed in Sect. 2.3 is practically effective. Based on our experience in numerical studies in the current and previous work for the frailty models, the proposed h-likelihood based methods (i.e. the modified likelihood approaches using $s_v(h_p)$ or $m_v(h_p)$) often provide estimators with smaller biases than that from the marginal likelihood method. Theoretical justification of this property as inquired by a referee would merit future research.

Section 3 presents simulation results for the parameters of interest (α, β) only, while the estimates of the nuisance parameters, i.e. three baseline cumulative hazards $\Lambda_{01}(t)$, $\Lambda_{02}(t)$ and $\Lambda_{03}(t)$, are excluded. In fact, even though our simulations also included estimation of those three functions, their estimates from all of the four methods tends to be biased as the time t increases, yet had only minimal impact on the estimation of (α, β). It would be worth a further investigation to improve the accuracy of the baseline cumulative hazard estimates.

Table 7 Fitted results from the semi-competing risks models for older patients in the NSABP B-14 data

Model Age $\geq$ 50	Time to recurrence Est. (SE)	Time to death without recurrence Est. (SE)	Time to death after recurrence Est. (SE)
Naive model			
Treatment (x_1)	-0.631 (0.097)	0.081 (0.092)	0.471 (0.112)
Tumor size (x_3)	0.023 (0.003)	0.005 (0.004)	0.006 (0.004)
$-2\ell_B = 16417.1$			
Frailty model (MPL2)			
Treatment (x_1)	-0.775 (0.116)	-0.101 (0.119)	0.472 (0.150)
Tumor size (x_3)	0.032 (0.004)	0.016 (0.005)	0.015 (0.006)
$\hat{\alpha}$	1.315		
$-2m_v(h_p) = 16407.3$			
Frailty model (MML1)			
Treatment (x_1)	-0.787 (0.120)	-0.116 (0.126)	0.466 (0.160)
Tumor size (x_3)	0.033 (0.005)	0.017 (0.006)	0.016 (0.006)
$\hat{\alpha}$	1.444		
$-2m^* = 16407.8$			
Frailty model (MML2)			
Treatment (x_1)	-0.814 (0.124)	-0.152 (0.134)	0.452 (0.165)
Tumor size (x_3)	0.035 (0.005)	0.019 (0.006)	0.018 (0.007)
$\hat{\alpha}$	1.751		
$-2p_\omega(m) = 16404.3$			
Age $\geq$ 65	Est. (SE)	Est. (SE)	Est. (SE)
Naive model			
Treatment (x_1)	-0.621 (0.200)	0.070 (0.139)	0.318 (0.232)
Tumor size (x_3)	0.026 (0.007)	0.003 (0.006)	-0.005 (0.009)
$-2\ell_B = 4112.5$			
Frailty model (MPL2)			
Treatment (x_1)	-0.793 (0.230)	-0.116 (0.194)	0.425 (0.320)
Tumor size (x_3)	0.034 (0.009)	0.009 (0.009)	0.003 (0.014)
$\hat{\alpha}$	1.021		
$-2m_v(h_p) = 4108.2$			
Frailty model (MML1)			
Treatment (x_1)	-0.709 (0.214)	-0.016 (0.164)	0.412 (0.287)
Tumor size (x_3)	0.030 (0.008)	0.006 (0.007)	-0.001 (0.012)
$\hat{\alpha}$	0.482		
$-2m^* = 4111.0$			
Frailty model (MML2)			
Treatment (x_1)	-0.823 (0.237)	-0.156 (0.206)	0.424 (0.330)
Tumor size (x_3)	0.036 (0.010)	0.010 (0.010)	0.004 (0.015)
$\hat{\alpha}$	1.238		
$-2p_\omega(m) = 4108.1$			

For modelling the semi-competing risks data, we used a shared frailty in three transition models. Extension to a model, where the transitions are affected by differnt frailties that are correlated, would be an interesting further work. Furthermore, we have assumed a Markov process for such transitions, but comparison with a semi-Markov assumption, i.e. $\lambda_{12}(t_2|t_1) = \lambda_{12}(t_2 - t_1)$, may be also interesting. The marginal likelihood may involve evaluation of analytically intractable integrals over the frailty distribution (e.g. log-normal distribution), whereas the h-likelihood obviates such integration. Extension of the proposed h-likelihood method to general frailty distributions including the log-normal distribution would be also an interesting future topic.

Acknowledgements Dr. Ha's research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2015R1D1A3A01015663). Dr. Xiang's research was supported in part by the Singapore MOE AcRF (MOE2013-T2-2-118). Dr. Jeong's research was supported in part by National Institute of Health (NIH) grants 5-U10-CA69651-11. Dr. Lee's research was supported by an NRF grant funded by Korea government (MEST) (No. 2011-0030810).

Appendix

Appendix A: Marginal-likelihood estimation procedure

For gamma frailty models with $E(u_i) = 1$ and $\text{var}(u_i) = \alpha$, we have an explicit marginal likelihood as follows. Since the second term of h-likelihood in (8) under the gamma frailty is given by

$$\ell_{2i} = \ell_{2i}(\alpha; v_i) = \alpha^{-1}(v_i - u_i) + c(\alpha),$$

with $c(\alpha) = -\log\Gamma(\alpha^{-1}) - \alpha^{-1}\log\alpha$, from (8) and (15) we have that

$$\begin{aligned} m = & \sum_i [\delta_{i1}\{\log\lambda_{01}(y_{i1}) + x_i^T\beta_1\} + \delta_{i2}(1-\delta_{i1})\{\log\lambda_{02}(y_{i2}) + x_i^T\beta_2\} \\ & + \delta_{i1}\delta_{i2}\{\log\lambda_{03}(y_{i2}) + x_i^T\beta_3\}] \\ & - \sum_i [(\alpha^{-1} + \delta_{i+})\log(1+\alpha\mu_{i+}) - \log\{\alpha^{\delta_{i+}}\Gamma(\alpha^{-1}+\delta_{i+})/\Gamma(\alpha^{-1})\}] \\ = & \sum_{k_1} d_{1(k_1)}\log\lambda_{01k_1} + \sum_i \delta_{i1}(x_i^T\beta_1) + \sum_{k_2} d_{2(k_2)}\log\lambda_{02k_2} \\ & + \sum_i \delta_{i2}(1-\delta_{i1})(x_i^T\beta_2) + \sum_{k_3} d_{3(k_3)}\log\lambda_{03k_3} + \sum_i \delta_{i1}\delta_{i2}(x_i^T\beta_3) \\ & - \sum_i [(\alpha^{-1} + \delta_{i+})\log(1+\alpha\mu_{i+}) - \delta_{i1}\delta_{i2}\log(1+\alpha)], \end{aligned} \tag{A.1}$$

where $\delta_{i+} = \delta_{i1} + \delta_{i2}$ and $\mu_{i+} = \sum_{j=1}^{3} \mu_{ij}$ with

$$\mu_{i1} = \Lambda_{01}(y_{i1}) \exp(x_i^T \beta_1) = \sum_{k_1} \lambda_{01k_1} I(y_{1(k_1)} \le y_{i1}) \exp(x_i^T \beta_1),$$
$$\mu_{i2} = \Lambda_{02}(y_{i2}) \exp(x_i^T \beta_2) = \sum_{k_2} \lambda_{02k_2} I(y_{2(k_2)} \le y_{i1}) \exp(x_i^T \beta_2),$$
$$\mu_{i3} = \Lambda_{03}(y_{i1}, y_{i2}) \exp(x_i^T \beta_3) = \sum_{k_3} \lambda_{03k_3} I(y_{i1} < y_{3(k_3)} \le y_{i2}) \exp(x_i^T \beta_3).$$

In fact, the marginal likelihood (A.1) is the same as that of Xu et al. (2010).

Under the gamma frailty, the score equations for β are given by

$$\frac{\partial m}{\partial \beta_1} = \sum_i \left\{ \delta_{i1} - \left(\frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}} \right) \mu_{i1} \right\} x_i, \tag{A.2}$$

$$\frac{\partial m}{\partial \beta_2} = \sum_i \left\{ \delta_{i2}(1 - \delta_{i1}) - \left(\frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}} \right) \mu_{i2} \right\} x_i, \tag{A.3}$$

$$\frac{\partial m}{\partial \beta_3} = \sum_i \left\{ \delta_{i1}\delta_{i2} - \left(\frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}} \right) \mu_{i3} \right\} x_i. \tag{A.4}$$

In particular, the solutions of $\partial m / \partial \lambda_{0jk_j} = 0$ $(j = 1, 2, 3)$ lead to closed forms:

$$\widetilde{\lambda}_{0jk_j}(\beta, \alpha) = \frac{d_{j(k_j)}}{\sum_{i \in R_{(k_j)}} \exp(x_i^T \beta_j) \tilde{u}_i}, \tag{A.5}$$

where $\tilde{u}_i = (\alpha^{-1} + \delta_{i+})/(\alpha^{-1} + \mu_{i+})$. We see that the score equations of (β, λ_{0j}) in (A.2)–(A.4) and (A.5) are extensions of those in the shared gamma frailty models (Andersen et al. 1997). Finally, the score equation for the frailty parameter α is given by

$$\frac{\partial m}{\partial \alpha} = \sum_i \left\{ \delta_{i1}\delta_{i2}(1 + \alpha)^{-1} + \alpha^{-2} \log(1 + \alpha\mu_{i+}) - (\alpha^{-1} + \delta_{i+})\mu_{i+}(1 + \alpha\mu_{i+})^{-1} \right\}.$$

Then the estimates of fixed parameters $(\beta, \alpha, \lambda_0)$ can be obtained using a numerical iterative method such as the Newton-Raphson method. Note that the maximum likelihood estimating equations, $\partial m / \partial(\beta, \alpha, \lambda_0) = 0$, by Xu et al. (2010) are equivalent to $\partial m^* / \partial(\beta, \alpha) = 0$, where m^* is the profile marginal likelihood in (16).

Appendix B: Comparison of h-likelihood with marginal likelihood

We assume that α is known. Recall that given (β, v), the score equations $\partial h / \partial \lambda_{0jk_j} = 0$ $(j = 1, 2, 3)$ provide the non-parametric maximum h-likelihood estimators in Sect. 2.2, i.e.

$$\widehat{\lambda}_{0jk_j}(\beta, v) = \frac{d_{j(k_j)}}{\sum_{i \in R_{(k_j)}} \exp(x_i^T \beta_j) u_i}.$$

The maximum h-likelihood estimating equations for β, under the gamma frailty, become

$$\frac{\partial h}{\partial \beta_1}\Big|_{\lambda_{01}=\widehat{\lambda}_{01}} = \sum_i \left\{\delta_{i1} - \mu_{i1} u_i\right\} x_i \Big|_{\lambda_{01}=\widehat{\lambda}_{01}} = 0 \ , \tag{B.1}$$

$$\frac{\partial h}{\partial \beta_2}\Big|_{\lambda_{02}=\widehat{\lambda}_{02}} = \sum_i \left\{\delta_{i2}(1-\delta_{i1}) - \mu_{i2} u_i\right\} x_i \Big|_{\lambda_{02}=\widehat{\lambda}_{02}} = 0 \ , \tag{B.2}$$

$$\frac{\partial h}{\partial \beta_3}\Big|_{\lambda_{03}=\widehat{\lambda}_{03}} = \sum_i \left\{\delta_{i1}\delta_{i2} - \mu_{i3} u_i\right\} x_i \Big|_{\lambda_{03}=\widehat{\lambda}_{03}} = 0 \ . \tag{B.3}$$

From

$$\frac{\partial h}{\partial v_i} = (\delta_{i+} - \mu_{i+} u_i) + \alpha^{-1} - \alpha^{-1} u_i = 0,$$

we have that

$$\hat{u}_i = \frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}}, \tag{B.4}$$

which also becomes $E(u_i | y_i^o)$ because the conditional distribution of u_i given the observed data $y_i^o = (y_{i1}, y_{i2}, \delta_{i1}, \delta_{i2})$ is gamma. Here $\delta_{i+} = \delta_{i1} + \delta_{i2}$ and $\mu_{i+} = \mu_{i1} + \mu_{i2} + \mu_{i3}$. From (12) we see that the estimating Eqs. (B.1)–(B.3) with (B.4) are equivalent to the estimating Eqs. (A.2)–(A.4) with (A.5), given by

$$\frac{\partial m}{\partial \beta_1}\Big|_{\lambda_{01}=\widetilde{\lambda}_{01}} = \sum_i \left\{\delta_{i1} - \mu_{i1}\left(\frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}}\right)\right\} x_i, \Big|_{\lambda_{01}=\widetilde{\lambda}_{01}} = 0 \tag{B.5}$$

$$\frac{\partial m}{\partial \beta_2}\Big|_{\lambda_{02}=\widetilde{\lambda}_{02}} = \sum_i \left\{\delta_{i2}(1-\delta_{i1}) - \mu_{i2}\left(\frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}}\right)\right\} x_i, \Big|_{\lambda_{02}=\widetilde{\lambda}_{02}} = 0 \tag{B.6}$$

$$\frac{\partial m}{\partial \beta_3}\Big|_{\lambda_{03}=\widetilde{\lambda}_{03}} = \sum_i \left\{\delta_{i1}\delta_{i2} - \mu_{i3}\left(\frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}}\right)\right\} x_i, \Big|_{\lambda_{03}=\widetilde{\lambda}_{03}} = 0 \ . \tag{B.7}$$

Accordingly, the maximum h-likelihood (MHL) estimator for β given α is the same as the marginal maximum likelihood (ML) estimator as shown in the standard gamma frailty models (Ha et al. 2001; Ha and Lee 2003). Note, however, that both methods give different estimators for α.

The marginal likelihood does not often have an analytic form (e.g. log-normal frailty model), so that the natural approach to the maximum likelihood estimator (MLE) is to use the EM treating the random effects as missing data. Below we present the comparison of the proposed h-likelihood method with the EM method for obtaining the

MLE. The EM equations for fixed parameters θ can be expressed via the h-likelihood as follows:

$$E(\partial h/\partial\theta|\text{data}) = 0,$$

which is equivalent to the ML equations, i.e. $\partial m/\partial\theta = 0$ (Lee and Nelder 1996; Engel and Keen 1996; Ha et al. 2001).

In the semi-competing risks frailty models (5)–(7), the EM equations for (β, α) are given by

$$E\left(\partial h/\partial(\beta,\alpha)|\ y_i^o, \widetilde{\lambda}^*_{0jk_j}\right) = 0.$$

Here, the EM equations of the baseline hazards λ_{0jk_j} are given by

$$E(\partial h/\partial\lambda_{0jk_j}|\ y_i^o) = 0,$$

which lead to

$$\widetilde{\lambda}^*_{0jk_j}(\beta, v) = \frac{d_{j(k_j)}}{\sum_{i\in R_{(k_j)}} \exp(x_i^T\beta_j)E(u_i|y_i^o)}.$$

Following Ha et al. (2001), in the gamma frailty models, given α the MHL equations for β

$$\partial h_p/\partial\beta = (\partial h/\partial\beta)|_{\lambda_0=\widehat{\lambda}_0} = 0,$$

with (B.4) are equivalent to the EM equations

$$E(\partial h/\partial\beta|\ y_i^o, \widetilde{\lambda}^*_{0jk_j}) = 0,$$

since $E(u_i|y_i^o)$ becomes $\tilde{u}_i$ in (A.5) and thus $\widetilde{\lambda}^*_{0jk_j}$ is identical to $\tilde{\lambda}_{0jk_j}$ in (A.5) as well as to $\widehat{\lambda}_{0jk_j}$. However, in general the EM may be difficult to apply because the conditional distribution of u_i given y_i^o is not trivial to be evaluated. For example, in the log-normal frailty with $v_i = \log u_i \sim N(0, \alpha)$, the EM equation for β_1 is given by

$$E\left(\partial h/\partial\beta_1|\ y_i^o, \widetilde{\lambda}^*_{0jk_j}\right) = \sum_i \left\{\delta_{i1} - \widetilde{\mu}^*_{i1}E(u_i|y_i^o)\right\}x_i = 0,$$

where

$$\widetilde{\mu}^*_{i1} = \widetilde{\Lambda}^*_{01}(y_{i1})\exp(x_i^T\beta_1) = \sum_{k_1}\widetilde{\lambda}^*_{01k_1}I(y_{1(k_1)} \le y_{i1})\exp(x_i^T\beta_1).$$

Note here that the computation of $E(u_i|y_i^o)$ requires a numerical integration.

References

Aalen O, Borgan O, Gjessing HK (2008) Survival and event history analysis. Springer, New York

Andersen PK, Klein JP, Knudsen K, Palacios RT (1997) Estimation of variance in Cox's regression model with shared gamma frailties. Biometrics 53:1475–1484

Barker P, Henderson R (2005) Small sample bias in the gamma frailty model for univariate survival. Lifetime Data Anal 11:265–284

Breslow NE (1974) Covariance analysis of censored survival data. Biometrics 30:89–99

Chen YH (2012) Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. Lifetime Data Anal 18:36–57

Engel E, Keen A (1996) Discussion of Lee and Nelder's paper. J R Stat Soc Ser B 58:656–657

Fan J, Li R (2002) Variable selection for Cox's proportional hazards model and frailty model. Ann Stat 30:74–99

Fine JP, Jiang H, Chappell R (2001) On semi-competing risks data. Biometrika 88:907–919

Fisher B, Costantino J, Redmond C et al (1989) A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen receptor-positive tumors. N Engl J Med 320:479–484

Fisher B, Dignam J, Bryant J et al (1996) Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor- positive tumors. J Natl Cancer Inst 88:1529–1542

Gu MG, Sun L, Huang C (2004) A universal procedure for parametric frailty models. J Stat Comput Simul 74:1–13

Ha ID, Lee Y (2003) Estimating frailty models via Poisson hierarchical generalized linear models. J Comput Graph Stat 12:663–681

Ha ID, Lee Y, Song JK (2001) Hierarchical likelihood approach for frailty models. Biometrika 88:233–243

Ha ID, Lee Y, Pawitan Y (2007) Genetic mixed linear models for twin survival data. Behav Genet 37:621–630

Ha ID, Noh M, Lee Y (2010) Bias reduction of likelihood estimators in semiparametric frailty models. Scand J of Stat 37:307–320

Ha ID, Sylvester R, Legrand C, MacKenzie G (2011) Frailty modelling for survival data from multi-centre clinical trials. Stat Med 30:2144–2159

Ha ID, Jeong J-H, Lee Y (2017) Statistical modelling of survival data with random effects: h-likelihood approach. Springer, Singapore

Heize G, Schemper M (2001) A solution to the problem of monotone likelihood in Cox regression. Biometrics 57:114–119

Lawless JF (2003) Statistical models and methods for lifetime data, 2nd edn. Wiley, New York

Lee Y, Nelder JA (1996) Hierarchical generalized linear models (with discussion). J R Stat Soc Ser B 58:619–678

Lee Y, Nelder JA (2001) Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika 88:987–1006

Lee Y, Nelder JA (2009) Likelihood inference for models with unobservables: another view (with discussion). Stat Sci 24:255–293

Lee KH, Haneuse S, Schrag D, Dominici F (2015) Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. J R Stat Soc Ser C 64:253–273

Lee Y, Nelder JA, Pawitan Y (2017) Generalised linear models with random effects: unified analysis via h-likelihood, 2nd edn. Chapman and Hall, Baca Raton

Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. Econometrica 16:1–32

Noh M, Lee Y (2007) REML estimation for binary data in GLMMs. J Multivar Anal 98:896–915

Ripatti S, Palmgren J (2000) Estimation of multivariate frailty models using penalized partial likelihood. Biometrics 56:1016–1022

Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Stram DO, Lee JW (1994) Variance components testing in the longitudinal mixed effects model. Biometrics 50:1171–1177

Therneau TM, Grambsch PM (2000) Modelling survival data: extending the Cox model. Springer, New York
Therneau TM, Grambsch PM, Pankratz VS (2003) Penalized survival models and frailty. J Comput Graph Stat 12:156–175
Tierney L, Kadane JB (1986) Accurate approximations for posterior moments and marginal densities. J Am Stat Assoc 81:82–86
Varadhan R, Xue QL, Bandeen-Roche K (2014) Semicompeting risks in aging research: methods, issues and needs. Lifetime Data Anal 20:538–562
Xu J, Kalbfleisch JD, Tai B (2010) Statistical analysis of illness-death processes and semicompeting risks data. Biometrics 66:716–725
Zhang Y, Chen MH, Ibrahim JG, Zeng D, Chen Q, Pan Z, Xue X (2014) Bayesian gamma frailty models for survival data with semi-competing risks and treatment switching. Lifetime Data Anal 20:76–105