

# Joint HGLM approach for repeated measures and survival data<sup>†</sup>

Il Do Ha<sup>1</sup>

<sup>1</sup>Department of Statistics, Pukyong National University

Received 17 June 2016, revised 11 July 2016, accepted 18 July 2016

## Abstract

In clinical studies, different types of outcomes (e.g. repeated measures data and time-to-event data) for the same subject tend to be observed, and these data can be correlated. For example, a response variable of interest can be measured repeatedly over time on the same subject and at the same time, an event time representing a terminating event is also obtained. Joint modelling using a shared random effect is useful for analyzing these data. Inferences based on marginal likelihood may involve the evaluation of analytically intractable integrations over the random-effect distributions. In this paper we propose a joint HGLM approach for analyzing such outcomes using the HGLM (hierarchical generalized linear model) method based on h-likelihood (i.e. hierarchical likelihood), which avoids these integration itself. The proposed method has been demonstrated using various numerical studies.

*Keywords:* Frailty model, H-likelihood, hierarchical generalized linear model, joint model, random effects.

## 1. Introduction

Consider the case that a subject has different types (e.g.  $y_1$  and  $y_2$ ) of outcomes. Then  $y_1$  and  $y_2$  may be correlated due to a shared but unobserved subject effect. Modelling jointly these outcomes has been widely studied (Henderson *et al.*, 2000; Ha *et al.*, 2003; Lee *et al.*, 2006; Rizopoulos, 2012) because a separated analysis to ignore the association can lead to a biased result (Guo and Carlin, 2004). For this purpose, it is useful to introduce an unobserved random effect into a joint model to describe the association between both outcomes (Ha *et al.*, 2003; Rizopoulos, 2012).

Let  $v$  be a common random effect for the same subject and  $\theta$  be an unknown parameter. For the analysis of these data, we may use marginal likelihood (i.e. observed-data likelihood), defined by

$$L(\theta) = f(y_1, y_2; \theta) = \int f(y_1, y_2 | v; \theta) g(v; \theta) dv, \quad (1.1)$$

where  $f(\cdot)$  is the joint density function of  $(y_1, y_2)$ , and  $f(\cdot | v)$  is the conditional density function of  $(y_1, y_2)$  given random effect  $v$ , and  $g(\cdot)$  is the density of  $v$ . However, the marginal

---

<sup>†</sup> This research was supported by a Research Grant of Pukyong National University (2015 year).

<sup>1</sup> Professor, Department of Statistics, Pukyong National University, Busan 608-737, Korea.  
E-mail: idha1353@pknu.ac.kr

likelihood in (1.1) may involve the evaluation of intractable integral over the random-effect distribution,  $g(v; \theta)$ . We thus use the h-likelihood (Lee and Nelder, 1996), defined by

$$h = \log f(y_1, y_2 | v; \theta) + \log g(v; \theta). \quad (1.2)$$

In particular, the h-likelihood obviates the need for intractable integrals over the random-effect terms, thereby providing a statistically efficient procedure for various random-effect models including mixed linear models and HGLMs (Lee and Nelder, 1996; Lee *et al.*, 2006; Paik *et al.*, 2015).

In this paper we consider repeated-measure response  $y (= y_1)$  and survival time  $T (= y_2)$ . It is thus interested to investigate the effects of covariates over these two types of responses. For this, joint modelling via random effects has been often used. Ha *et al.* (2003) presented a parametric joint model with mixed linear model for  $y$  and Weibull frailty model for  $T$ . Here, frailty is an unobserved random effect in the hazard model. In this paper we propose an improved modelling approach, which allows for semi-parametric frailty model with a nonparametric baseline hazard in the joint model of Ha *et al.* (2003). This joint model can be viewed as a joint HGLM to combine the mixed linear model (i.e. a normal HGLM) and the frailty model (i.e. a Poisson HGLM) (Lee *et al.*, 2006). For the inference we use the HGLM method based on h-likelihood.

The paper is organized as follows. In Section 2 we describe a joint model with both mixed linear model and semi-parametric frailty model, and derive the corresponding h-likelihood. In Section 3 we propose a new estimation procedure using the h-likelihood which leads to a simple fitting procedure. The proposed method is demonstrated using various numerical studies in Section 4. Finally, some discussions are given in Section 5.

## 2. Joint modelling approach

### 2.1. Joint model

Let  $y_{ij}$  be the  $j$ th repeated response of the  $i$ th patient ( $i = 1, \dots, q; j = 1, \dots, n_i$ ), and let  $T_i$  be a single event time of the  $i$ th patient and let  $C_i$  be the corresponding censoring time. Denote by  $v_i$  a shared random effect of the  $i$ th patient. Following Ha *et al.* (2003), we assume that  $y_i = (y_{i1}, \dots, y_{in_i})^T$  and  $T_i$  given  $v_i$  are conditionally independent, and that  $T_i$  and  $C_i$  given  $v_i$  are also conditionally independent. Then  $y_i$  and  $T_i$  are assumed to have the following joint model, which is specified in terms of the two linked components. That is, they are, respectively, (i) mixed linear model and (ii) semi-parametric frailty model, given by

$$(i) \ y_{ij} = x_{ij1}^T \beta_1 + v_i + \epsilon_{ij},$$

where  $v_i \sim N(0, \alpha)$  and  $\epsilon_{ij} \sim N(0, \phi)$  are independent, and

$$(ii) \ \lambda_i(t|v_i) = \lambda_0(t) \exp(x_{i2}^T \beta_2 + \gamma v_i),$$

where  $\lambda_0(\cdot)$  is a completely unspecified baseline hazard function. Here  $\beta_1$  and  $\beta_2$  are  $p_1 \times 1$  and  $p_2 \times 1$  regression-parameters vectors corresponding to the vectors of covariates  $x_{ij1}$  and  $x_{i2}$ , respectively. In particular, the frailty model of (ii) is an extension of Cox's (1972) proportional hazards model. Note that  $\gamma$  is a real-valued association parameter; if  $\gamma > 0$  ( $\gamma < 0$ ),  $y_{ij}$  and hazard rate tend to be positively (negatively) correlated, and if  $\gamma = 0$  they are not associated.

### 2.2. Construction of h-likelihood

All observable random variables are repeated-measure responses  $y_{ij}$  and survival-time responses  $(t_i^*, \delta_i)$ , given by

$$t_i^* = \min(T_i, C_i) \text{ and } \delta_i = I(T_i \leq C_i).$$

The construction of h-likelihood for the joint model above is immediate. From (1.2) we take  $y_1 = y$  and  $y_2 = (t^*, \delta)$ . Then by the two assumptions (i.e. conditional independency between  $y_i$  &  $T_i$  and non-informative censoring), the h-likelihood is constructed as follows:

$$\begin{aligned} h &= \log f(y, (t^*, \delta)|v; \theta) + \log g(v; \theta) \\ &= \log f(y|v; \theta) + f(t^*, \delta|v; \theta) + \log g(v; \theta). \end{aligned}$$

Along the lines of Ha *et al.* (2001), the h-likelihood for the joint model with all data is defined by

$$h = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i} + \sum_i \ell_{3i}, \tag{2.1}$$

with

$$\begin{aligned} \ell_{1ij} &= \ell_{1ij}(\beta_1, \phi; y_{ij}|v_i) = -\log(2\pi\phi)/2 - (y_{ij} - \eta_{1ij})^2/(2\phi), \\ \ell_{2i} &= \ell_{2i}(\beta_2, \lambda_0; (t_i^*, \delta_i)|v_i) = \delta_i(\log \lambda_0(t_i^*) + \eta_{2i}) - \Lambda_0(t_i^*) \exp(\eta_{2i}), \\ \ell_{3i} &= \ell_{3i}(\alpha; v_i) = -\log(2\pi\alpha)/2 - v_i^2/(2\alpha). \end{aligned}$$

where  $\ell_{1ij}$  is the conditional log-likelihoods for  $y_{ij}$  given  $v_i$ ,  $\ell_{2i}$  is that for  $(t_i^*, \delta_i)$  given  $v_i$ , and  $\ell_{3i}$  is the log-likelihood for  $v_i$ . Here

$$\eta_{1ij} = x_{ij1}^T \beta_1 + v_i \text{ and } \eta_{2i} = x_{i2}^T \beta_2 + \gamma v_i.$$

### 3. Estimation procedure

Notice that the functional form of  $\lambda_0(\cdot)$  in  $\ell_{2i}$  of (2.1) is unknown. Following Breslow (1972) and Ha *et al.* (2001), we define the baseline cumulative hazard function  $\Lambda_0(t) = \int_0^t \lambda_0(u) du$  to be a step function with jumps  $\lambda_{0r}$  at the observed event times  $t_{(r)}$ , defined by

$$\Lambda_0(t) = \sum_{r:t_{(r)} \leq t} \lambda_{0r}, \tag{3.1}$$

where  $t_{(r)}$  is the  $r$ th ( $r = 1, \dots, D$ ) smallest distinct event time among the  $t_i^*$ 's and  $\lambda_{0r} = \lambda_0(t_{(r)})$ . By substituting (3.1) into (2.1), the second term  $\sum_i \ell_{2i}$  in (2.1) becomes

$$\sum_i \ell_{2i} = \sum_r d_{(r)} \log \lambda_{0r} + \sum_i \delta_i \eta_{2i} - \sum_r \lambda_{0r} \left\{ \sum_{i \in R_{(r)}} \exp(\eta_{2i}) \right\},$$

where  $d_{(r)}$  is the number of events at  $t_{(r)}$  and

$$R_{(r)} = R(t_{(r)}) = \{i : t_i^* \geq t_{(r)}\}$$

is the risk set at  $t_{(r)}$ . As the number of  $\lambda_{0r}$ 's in  $\sum_i \ell_{2i}$  above increases with the number of events, the function  $\lambda_0(t)$  is potentially of high dimension. Following Ha *et al.* (2001), we use a profile h-likelihood after eliminating nuisances  $\lambda_{0r}$ , given by

$$h^* = h|_{\lambda_0 = \hat{\lambda}_0} = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}^* + \sum_i \ell_{3i}, \quad (3.2)$$

where

$$\sum_i \ell_{2i}^* = \sum_i \ell_{2i}|_{\lambda_0 = \hat{\lambda}_0} = \sum_r d_{(r)} \log \hat{\lambda}_{0r} + \sum_i \delta_i \eta_{2i} - \sum_r d_{(r)}.$$

Here

$$\hat{\lambda}_{0r} = \hat{\lambda}_{0r}(\beta_2, v) = \frac{d_{(r)}}{\sum_{i \in R_{(k)}} \exp(\eta_{2i})}$$

are the solutions of the estimating equations,  $\partial h / \partial \lambda_{0r} = 0$ , for  $r = 1, \dots, D$ . We thus see that  $h^*$  does not depend on  $\lambda_0$ , and that it becomes the penalized partial likelihood (Ripatti and Palmgren, 2000). Accordingly, below we show that the h-likelihood procedure can be derived via  $h^*$ .

### 3.1. Joint estimating equations

Let  $X_1$ ,  $X_2$  and  $Z$  be model matrices for vectors  $\beta_1$ ,  $\beta_2$  and  $v = (v_1, \dots, v_q)^T$ , respectively. The score equations for fixed and random effects  $(\beta_1, \beta_2, v)$  given dispersion parameters  $\psi = (\phi, \alpha, \gamma)^T$  are given by

$$\begin{aligned} \partial h^* / \partial \beta_1 &= X_1^T (y - \mu_1) / \phi, \\ \partial h^* / \partial \beta_2 &= X_2^T (\delta - \hat{\mu}_2), \\ \partial h^* / \partial v &= Z^T (y - \mu_1) / \phi + \gamma Z^T (\delta - \hat{\mu}_2) - v / \alpha. \end{aligned}$$

Here  $\mu_1 = X_1 \beta_1 + Zv = \eta_1$ ,

$$\hat{\mu}_2 = \exp(\log \hat{\Lambda}_0(t^*) + \eta_2) \text{ with } \eta_2 = X_2 \beta_2 + \gamma Zv,$$

and

$$\hat{\Lambda}_0(t) = \sum_{r: t_{(r)} \leq t} \hat{\lambda}_{0r}$$

is the Breslow-type estimator of cumulative baseline hazard.

Following Lee and Nelder (1996) and Ha and Lee (2003), we can show that the iterative weight least squares (IWLS) joint equations for  $\tau = (\beta_1^T, \beta_2^T, v^T)^T$  are given by

$$\begin{pmatrix} X_1^T W_1 X_1 & 0 & X_1^T W_1 Z \\ 0 & X_2^T W_2 X_2 & X_2^T (\gamma W_2) Z \\ Z^T W_1 X_1 & Z^T (\gamma W_2) X_2 & Z^T (W_1 + \gamma^2 W_2) Z + U \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{v} \end{pmatrix} = \begin{pmatrix} X_1^T W_1 w_1 \\ X_2^T w_2 \\ Z^T W_1 w_1 + \gamma Z^T w_2 \end{pmatrix},$$

where

$$W_1 = -\partial^2 h^* / \partial \eta_1 \partial \eta_1^T = I_{p_1} / \phi \text{ and } W_2 = -\partial^2 h^* / \partial \eta_2 \partial \eta_2^T.$$

Here  $I_{p_1}$  is a  $p_1 \times p_1$  identity matrix, and the detailed matrix form of  $W_2$  is given in Appendix B of Ha and Lee (2003), and  $U = -\partial^2 \ell_3 / \partial v \partial v^T = (1/\alpha) I_q$ ,  $w_1 = y$ , and

$$w_2 = W_2 \eta_2 + (\delta - \hat{\mu}_2).$$

### 3.2. Fitting procedure

Below we present that the IWLS equations above lead to a simple fitting procedure for the joint model in Section 2.1 via partition matrices.

$$\text{Let } \mathbf{P} = \begin{pmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{0} & I \end{pmatrix} \text{ and } \mathbf{\Sigma} = \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & U \end{pmatrix},$$

where

$$\mathbf{X} = \begin{pmatrix} X_1 & \mathbf{0} \\ \mathbf{0} & X_2 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} Z \\ \gamma Z \end{pmatrix} \quad \text{and} \quad \mathbf{W} = \begin{pmatrix} W_1 & \mathbf{0} \\ \mathbf{0} & W_2 \end{pmatrix}.$$

Then our fitting procedure consist of the following two steps:

(i) Estimation of fixed and random effects  $\tau = (\beta_1^T, \beta_2^T, v^T)^T$ :

Following Lee *et al.* (2006), the IWLS equations above reduces to a simple explicit form,

$$(\mathbf{P}^T \mathbf{\Sigma} \mathbf{P}) \hat{\tau} = \mathbf{P}^T \mathbf{\Sigma}^* \mathbf{y}^*, \tag{3.3}$$

where  $\mathbf{\Sigma}^* = \text{BD}(W_1, I_{p_2}, U)$  is a block diagonal matrix and  $\mathbf{y}^* = (w_1^T, w_2^T, 0^T)^T$ .

(ii) Estimation of dispersion parameters  $\psi = (\phi, \alpha, \gamma)^T$ :

For the estimation of  $\psi$  we use the adjusted profile h-likelihood (Ha and Lee, 2003; Lee *et al.*, 2006), given by

$$p_\tau(h^*) = [h^* - \frac{1}{2} \log \det\{H(h^*, \tau)/(2\pi)\}]|_{\tau=\hat{\tau}}, \tag{3.4}$$

where  $\hat{\tau} = \hat{\tau}(\psi)$  are solutions of  $\partial h^*/\partial \tau = 0$  for given  $\psi$ , and

$$H(h^*, \tau) = -\partial^2 h^*/\partial \tau \tau^T = \mathbf{P}^T \mathbf{\Sigma} \mathbf{P}$$

is observed information matrix for  $\tau$ . Thus the estimating equations of  $\psi$  are given by

$$\partial p_\tau(h^*)/\partial \psi = 0,$$

which leads to an extended restricted maximum likelihood (REML) estimator for  $\psi$  (Lee *et al.*, 2006).

## 4. Numerical study

A numerical study, based on 500 replications of simulated data, is presented to evaluate the performance of the proposed joint modelling approach. For this purpose, we consider a joint model for repeated measures and a single event time in Section 2. The simulation scheme is as follows. First we generate random effects  $v_i \sim N(0, \alpha = 0.5)$  for  $i = 1, \dots, 50$ . Then we generate repeated-measure responses  $y_{ij}$  for  $j = 1, 2, 3, 4$  given  $v_i$  from the mixed linear model with two covariates:

$$y_{ij}|v_i \sim N(\beta_{10} + \beta_{11} * \text{time}_{ij} + \beta_{12} \text{Trt}_i + v_i, \phi = 1),$$

where we set  $\beta_{10} = -0.5, \beta_{11} = 0.5, \beta_{12} = 1, \text{time}_{ij} = 0, 2, 4, 8$  (weeks) and  $\text{Trt}_i$  are generated from Bernoulli distribution with probability 0.5. Here ‘‘Trt’’ denotes a treatment group with

a new drug (coded as 1) or placebo (coded as 0). Next, we generate survival time  $T_i$  for  $i = 1, \dots, 50$  from the frailty model with one covariate:

$$\lambda_i(t|v_i) = \lambda_0(t) \exp(\text{Trt}_i\beta_{21} + \gamma v_i),$$

where we set  $\lambda_0(t) = 1$ ,  $\beta_{21} = -1$ , and  $\gamma = -0.5$  or  $0.5$ . The negative [positive] sign of association parameter  $\gamma$  means that the patient with the larger repeated-measure responses can show tendency with lower [higher] hazard rate as shown in Ha *et al.* (2003). Finally, the corresponding censoring time  $C_i$  are generated from an exponential distribution with about 30% and 50% censoring rates, respectively. We also considered a continuous covariate,  $x_i$ , in the frailty hazard model; the  $x_i$  was generated from a standard normal distribution.

**Table 4.1** Simulation results on the estimation of parameters with 500 replications in the joint model: Different covariates in the hazard model under two censoring rates.

Parameter	True	30% Cen.			50% Cen.		
		Mean	SD	SE (95% CP)	Mean	SD	SE (95% CP)
Mixed linear model for $y$							
$\beta_{10}$	-0.5	-0.494	0.205	0.193 (0.948)	-0.501	0.194	0.194 (0.950)
$\beta_{11}$	0.5	0.501	0.024	0.024 (0.942)	0.499	0.023	0.024 (0.954)
$\beta_{12}$	1	0.985	0.244	0.246 (0.958)	0.986	0.257	0.247 (0.932)
$\phi$	1	0.989	0.116	—	0.999	0.119	—
$\alpha$	0.5	0.504	0.160	—	0.505	0.155	—
Frailty model for $T$ (Binary covariate)							
$\beta_{21}$	-1	-1.033	0.406	0.395 (0.956)	-1.088	0.554	0.532 (0.936)
$\gamma$	-0.5	-0.522	0.429	—	-0.557	0.464	—
Mixed linear model for $y$							
$\beta_{10}$	-0.5	-0.488	0.191	0.190 (0.944)	-0.508	0.191	0.191 (0.942)
$\beta_{11}$	0.5	0.500	0.025	0.024 (0.932)	0.500	0.025	0.024 (0.938)
$\beta_{12}$	1	0.982	0.249	0.237 (0.934)	0.990	0.247	0.238 (0.940)
$\phi$	1	1.008	0.121	—	0.997	0.116	—
$\alpha$	0.5	0.494	0.150	—	0.497	0.156	—
Frailty model for $T$ (Normal covariate)							
$\beta_{21}$	-1	-1.069	0.257	0.244 (0.942)	-1.067	0.296	0.277 (0.952)
$\gamma$	-0.5	-0.553	0.450	—	-0.575	0.482	—

With 500 replications, we computed the mean, standard deviation (SD) and mean of the estimated standard error (SE) for  $\hat{\beta} = (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21})^T$ , and the mean and SD of  $\hat{\psi} = (\hat{\phi}, \hat{\alpha}, \hat{\gamma})^T$ . Here the SE is obtained from the inverse of the observed information matrix  $H(h^*, \tau)$  in (2.5) (Ha and Lee, 2003). In addition, we calculated the empirical coverage probability (CP) for a nominal 95% confidence interval for  $\beta$  based on the SE.

The simulation results are summarized in Table 4.1. Overall,  $\hat{\beta}$  work well. The standard-error estimators of  $\hat{\beta}$  also work well as judged by the good agreement between SE and SD. The coverage probabilities for  $\beta$  are also reasonable with the 93%-96% range. For estimators of dispersion parameters,  $\hat{\phi}$  and  $\hat{\alpha}$  perform well, but  $\hat{\gamma}$  seems to show a slight bias. A reason may be the relative lack of outcomes in the frailty model; i.e. the repeated-measure data have four outcomes for each subject, whereas the corresponding survival data have one single outcome for the same subject. For the increasing censoring rate from 30% to 50%, we observe that the estimates of parameters in the mixed linear model are not almost changed, but that they (i.e.  $\hat{\beta}_{21}$  and  $\hat{\gamma}$ ) in the frailty model are, as expected, sensitive. That is, the biases and

variations (SD and SE) of  $\hat{\beta}_{21}$  and  $\hat{\gamma}$  are overall increased with censoring rate. However, we find that the biases are relatively less sensitive to the types of covariate (i.e. binary and normal covariates) on the frailty models. Though not reported here, we found the similar results for  $\gamma = 0.5$ .

Furthermore, we conducted the simulation study above under various values for  $\gamma$ ;  $\gamma = -0.1, 0.1$  and  $\gamma = -0.7, 0.7$ . The simulation results are presented in Table 4.2. The behaviors of the estimates for  $\gamma$ , not  $\beta_{21}$ , are similar to those of censoring rates in Table 4.1, That is,  $\hat{\gamma}$  in the frailty model has larger biases and variations as the absolute value of  $\gamma$  increases.

In summary, our estimates for the joint model in Section 2.1 perform well for all parameters in the mixed linear model, but have slight biases for all parameters in the frailty model, particularly for  $\gamma$ . However, we have found that coverage probabilities of  $\beta_{21}$  in the frailty model work overall well.

**Table 4.2** Simulation results on the estimation of parameters with 500 replications in the joint model: Different association parameter  $\gamma$  under 30% censoring rate.

Parameter	True	Mean	SD	SE (95% CP)	Mean	SD	SE (95% CP)
		$\gamma = -0.1$			$\gamma = 0.1$		
		Mixed linear model for $y$					
$\beta_{10}$	-0.5	-0.494	0.188	0.194 (0.958)	-0.509	0.189	0.194 (0.936)
$\beta_{11}$	0.5	0.499	0.025	0.024 (0.940)	0.501	0.024	0.024 (0.948)
$\beta_{12}$	1	1.015	0.235	0.247 (0.950)	1.014	0.248	0.248 (0.940)
$\phi$	1	1.002	0.114	—	0.995	0.118	—
$\alpha$	0.5	0.504	0.154	—	0.512	0.158	—
		Frailty model for $T$ (Binary covariate)					
$\beta_{21}$	-1	-1.090	0.390	0.384 (0.938)	-1.046	0.404	0.385 (0.954)
$\gamma$		-0.095	0.352	—	0.102	0.395	—
		$\gamma = -0.7$			$\gamma = 0.7$		
		Mixed linear model for $y$					
$\beta_{10}$	-0.5	-0.496	0.191	0.193 (0.956)	-0.508	0.190	0.192 (0.958)
$\beta_{11}$	0.5	0.500	0.023	0.024 (0.960)	0.502	0.024	0.024 (0.940)
$\beta_{12}$	1	0.984	0.238	0.247 (0.952)	1.012	0.251	0.250 (0.950)
$\phi$	1	0.999	0.117	—	0.997	0.118	—
$\alpha$	0.5	0.502	0.156	—	0.509	0.159	—
		Frailty model for $T$ (Binary covariate)					
$\beta_{21}$	-1	-1.054	0.427	0.417 (0.950)	-1.032	0.464	0.445 (0.936)
$\gamma$		-0.773	0.467	—	-0.781	0.497	—

### 5. Discussion

An advantage of our joint HGLM method is ease of extensibility to models with random-effect terms, in which the integration to obtain the marginal likelihood is often intractable (Ha and Lee, 2003; Ha and Cho, 2012, 2015; Ha and Noh, 2013). The separate modelling does not consider the dependency between two types of responses, whereas the joint modelling does. Thus the joint modelling enables the information from the other responses to be exploited to improve the regression parameter estimators, which can be impossible from separate models (Ha *et al.*, 2003). Information gain from the other responses can be important for the analysis of scarce data.

We have investigated the behaviors of the proposed estimates via numerical studies under somewhat limited schemes. For example, it would be interested to consider various sample sizes with the number of clusters ( $q$ ) and/or cluster size ( $n_i$ ).

We assumed a shared frailty  $v_i$  in the joint models. Extension to correlated frailties may be more useful; for example, it would be interesting to allow for a bivariate normal distribution with correlation between  $v_{i1}$  in mixed linear model and  $v_{i2}$  in frailty model (Elashoff *et al.*, 2008). Furthermore, the development of an extended joint model allowing for time-dependent covariates (Rizopoulos, 2012) or competing risks (Christian *et al.*, 2016) would be also an interesting future work.

## References

- Breslow, N. E. (1972). Discussion of professor Cox's paper. *Journal of the Royal Statistical Society B*, **34**, 216-217.
- Christian, N. J., Ha, I. D. and Jeong, J. (2016). Hierarchical likelihood inference on clustered competing risks data. *Statistics in Medicine*, **35**, 251-267.
- Cox, D. R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society B*, **74**, 187-220.
- Elashoff, R. M., Li, G. and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, **64**, 762-771.
- Guo, X. and Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *American Statistician*, **58**, 16-24.
- Ha, I. D. and Cho, G.-H. (2012). H-likelihood approach for variable selection in gamma frailty models. *Journal of the Korean Data & Information Science Society*, **23**, 199-207.
- Ha, I. D. and Cho, G.-H. (2015). Variable selection in Poisson HGLMs using h-likelihood. *Journal of the Korean Data & Information Science Society*, **26**, 1513-1521.
- Ha, I. D., Lee, Y. and Song, J. K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, **88**, 233-243.
- Ha, I. D. and Lee, Y. (2003). Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, **12**, 663-681.
- Ha, I. D. and Noh, M. (2013). A visualizing method for investigating individual frailties using frailtyHL R-package. *Journal of the Korean Data & Information Science Society*, **24**, 931-940.
- Ha, I. D., Park, T. and Lee, Y. (2003). Joint modelling of repeated measures and survival time data. *Biometrical Journal*, **45**, 647-658.
- Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465-480.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society B*, **58**, 619-678.
- Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalised Linear Models with Random Effects: Unified Analysis via h-Likelihood*, Chapman and Hall, London.
- Paik, M. C., Lee, Y. and Ha, I. D. (2015). Frequentist inference on random effects based on summarizability. *Statistica Sinica*, **25**, 1107-1132.
- Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**, 1016-1022.
- Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data with applications in R*, Chapman and Hall, London.