



Published in final edited form as:

Stat Med. 2011 July 30; 30(17): 2144–2159. doi:10.1002/sim.4250.

Frailty Modelling for Survival Data from Multi-Centre Clinical Trial

Il Do Ha,

Department of Asset Management, Daegu Haany University, Gyeongsan, 712-715, South Korea

Richard Sylvester,

European Organisation for Research and Treatment of Cancer, Brussels, Belgium

Catherine Legrand, and

Institut de statistique, biostatistique et sciences actuarielles (ISBA), Université catholique de Louvain, Louvain-la-neuve, Belgium

Gilbert MacKenzie

CREST, ENSAI, Rennes, France and Centre for Biostatistics, University of Limerick, Ireland

Il Do Ha: idha@dhu.ac.kr; Richard Sylvester: richard.sylvester@eortc.be; Catherine Legrand: catherine.legrand@uclouvain.be; Gilbert MacKenzie: gibert.mackenzie@ul.ie

Summary

Despite the use of standardized protocols in, multicentre, randomised clinical trials (RCTs), outcome may vary between centres. Such heterogeneity may alter the interpretation and reporting of the treatment effect. Below, we propose a general frailty modelling approach for investigating, *inter alia*, putative treatment-by-centre interactions in time-to-event data in multi-centre clinical trials. A correlated random effects model is used to model the baseline risk and the treatment effect across centres. It may be based on shared, individual or correlated random-effects. For inference we develop the hierarchical-likelihood (or h-likelihood) approach which facilitates computation of prediction intervals for the random effects with proper precision. We illustrate our methods using disease-free time-to-event data on bladder cancer patients participating in an European Organization for Research and Treatment of Cancer (EORTC) trial, and a simulation study. We also demonstrate model selection using h-likelihood criteria.

Keywords

Correlated random effects; Focussed model selection; Frailty models; Hierarchical likelihood; Prediction interval; Random treatment-by-centre interaction

1. Introduction

In this paper we focus on multi-centre trials with time to event endpoints. We are interested in investigating potential heterogeneity in outcomes between centres. In this context, the use of proportional hazards (PH) frailty models with random effects, rather than PH models with fixed (centre) effects, are useful [1–4].

Our approach is to model: (a) the between-centre variation in the baseline risk and (b) the treatment effect across centres [2, 5], using random effects. Thus, our model incorporates a random centre effect and a random treatment-by-centre interaction. These two random components (or frailty terms) have usually been assumed to be independent [2, 5]. However, independence may be un-necessarily restrictive [6–8]. In particular, Legrand *et al.* [4] has recommended using correlated random effects. Furthermore, our approach also models

individual-specific frailty terms [9–11], because covariates specified in the protocol, or involved in the minimization procedure [12, p.71], may not account for all prior differences between patients. Thus, by deploying classical frailty concepts, we hope to improve conventional strategies for analyzing RCTs trials.

Usually, inference in frailty models requires a marginal-likelihood approach, whereby the random effects are integrated out of the joint density consisting of response variables and random effects. This may involve the evaluation of analytically intractable integrals over the random effect distributions. To avoid these difficulties, several methods (e.g. Monte Carlo EM and Markov chain Monte Carlo) have been suggested [13, 14], but these remain computationally intensive, particularly when the number of random components is large or when their correlation structure is modelled [15–19].

Another important issue is that of estimating the standard errors for the *prediction* of random effects, which is required in order to construct $100(1 - \alpha)\%$ prediction intervals. Plots based on these intervals are useful, especially when investigating the heterogeneity of random centre and treatment effects. However, estimating the standard errors of random effects using “plug-in” methods, such as empirical Bayes (EB, [13, 20]), may underestimate the true variability of the estimated random effects. Thus, the development of an *integral* method of inference for frailty models is required.

Accordingly, we propose a unified method of inference within the h-likelihood framework [21–23]. The h-likelihood consists of data, parameters and unobserved random effects, and obviates integration over the random-effect distributions. Thus, the h-likelihood can be used directly for inference on random effects, while the marginal likelihood cannot because it eliminates them by integration. The h-likelihood approach also gives a statistically efficient estimation procedure for various random-effect models [11, 19, 24–26]. We derive, via the h-likelihood approach, improved methods for estimating the standard errors of the predictor of the random effects and the frailty parameters. In particular, we emphasize inference on the random effects rather than on just estimating the frailty parameters. Predictions and their intervals are useful in investigating heterogeneity over centers. We illustrate the methodology by analyzing time to first recurrence in patients with bladder cancer from an EORTC trial [27] and by a simulation study. We also employ the data to illustrate model selection using criterion [11] based on h-likelihood.

The paper is organized as follows. In Section 2 we review a formulation of frailty models, present an extension, and show how to interpret the random-effect terms. The h-likelihood estimation procedure for fitting the model is derived and an improved method for estimating the standard-error of the frailty parameters is proposed. Next, a new prediction method for random effects is proposed in Section 3. The new method is illustrated using the bladder cancer data set in Section 4. A simulation study is conducted to evaluate the performance of the proposed method in Section 5. And, finally, we discuss the approach in Section 6. The technical details are given in Appendices.

2. The model and estimation

2.1. Model formulation and interpretation

In general, suppose that data consist of right censored time-to-event observations collected from q centres. Let T_{ij} ($i = 1, \dots, q, j = 1, \dots, n_i, n = \sum_i n_i$) be the survival time for the j th observation in the i th centre (or cluster) and let C_{ij} be the corresponding censoring time. Then observable data become $y_{ij} = \min\{T_{ij}, C_{ij}\}$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where $I(\cdot)$ is the indicator function.

Denote by v_i a s -dimensional vector of unobserved log-frailties (random effects) associated with the i th cluster. Given v_i , the conditional hazard function of T_{ij} is of the form

$$\lambda_{ij}(t|v_i) = \lambda_0(t) \exp(\eta_{ij}), \quad (1)$$

where $\lambda_0(\cdot)$ is a unknown baseline hazard function,

$$\eta_{ij} = x_{ij}^T \beta + z_{ij}^T v_i$$

is the linear predictor for the hazards, and $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ and $z_{ij} = (z_{ij1}, \dots, z_{ijs})^T$ are $p \times 1$ and $s \times 1$ covariate vectors corresponding to fixed effects $\beta = (\beta_1, \dots, \beta_p)^T$ and log-frailties v_i , respectively. Here z_{ij} is often a subset of x_{ij} . Alternatively, it may be the constant (unity) representing a cluster effect on the baseline hazard [13]. In this paper, we assume the normal distribution for v_i :

$$v_i \sim N_s(0, \Sigma_i),$$

which is useful for modelling multi-component [11, 28] or correlated frailties [13, 29]. Here the covariance matrix $\Sigma_i = \Sigma_i(\theta)$ depends on θ , a vector of unknown parameters. We note that the formulation of model (1) is actually the same as that of Vaida and Xu [13] but that their covariance matrix for Σ_i is diagonal [8].

Equation (1) includes some well-known models. Let v_{i0} be a random baseline intercept (representing the random baseline risk) and let v_{i1} be a random slope (i.e. random treatment effect or random treatment-by-center interaction). If in model (1) $z_{ij} = 1$ and $v_i = v_{i0}$ for all i, j , it becomes a random intercept or shared model [30, 31] with

$$\eta_{ij} = x_{ij}^T \beta + v_{i0}, \quad (2)$$

where $v_{i0} \sim N(0, \Sigma_i)$ with $\Sigma_i \equiv \sigma_0^2$ for all i . Let β_1 be the effect of primary covariate x_{ij1} such as the main treatment effect and let β_m ($m = 2, \dots, p$) be the fixed effects corresponding to the covariates x_{ijm} . Our two random components lead to a bivariate model [8, 32] with

$$\eta_{ij} = v_{i0} + (\beta_1 + v_{i1})x_{ij1} + \sum_{m=2}^p \beta_m x_{ijm}, \quad (3)$$

which is easily derived by taking $z_{ij} = (1, x_{ij1})^T$ and $v_i = (v_{i0}, v_{i1})^T$ in (1). Here

$$\begin{pmatrix} v_{i0} \\ v_{i1} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_i \equiv \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \right\}.$$

Model (3) allows a correlation term, $\rho = \sigma_{01}/(\sigma_0\sigma_1)$, between two random effects (v_{i0} and v_{i1}) within a centre and potentially extends the independent frailty model in which $\rho = 0$ [2, 5].

Furthermore, we note the model (1) may be asymmetric (or unbalanced) as it does not contain a generic individual-specific frailty term, v_{ij} to match the individual-level fixed effects, x_{ij} . Following Ha *et al.* [11], the one-component model (1) with $\eta_{ij} = x_{ij}^T \beta + z_{ij}^T v_i$ can be extended easily to a two-component model with

$$\eta_{ij} = x_{ij}^T \beta + z_{ij}^{(1)T} v_i^{(1)} + z_{ij}^{(2)T} v_i^{(2)}, \quad (4)$$

where $v_i^{(1)}$ and $v_i^{(2)}$ are independent, and $z_{ij}^{(1)}$ and $z_{ij}^{(2)}$ are random-covariate vectors corresponding to $v_i^{(1)}$ and $v_i^{(2)}$, respectively. In fact, the model (4) can be written as in (1) by taking $z_{ij} = (z_{ij}^{(1)T}, z_{ij}^{(2)T})^T$ and $v_i = (v_i^{(1)}, v_i^{(2)})^T$. Thus, the extension of results from one-component model (1) to two-component model (4) is straightforward [11, 21].

In order to interpret the fixed and random effects, we consider model (3) with a single binary-treatment indicator, x_{ij} . Then,

$$\lambda_{ij}(t|v_{i0}, v_{i1}; x_{ij}) = \lambda_0(t) \exp\{v_{i0} + (\beta_1 + v_{i1})x_{ij}\}.$$

Now, the time-dependent relative risk for treatment becomes

$$\psi_{ij}(t|x=1, x=0) = \frac{\lambda_0(t) \exp\{v_{i0} + (\beta_1 + v_{i1}) \cdot 1\}}{\lambda_0(t) \exp\{v_{i0} + (\beta_1 + v_{i1}) \cdot 0\}} = \exp(\beta_1 + v_{i1}),$$

which is free of time t and holds for all patients in centre i . Here $\exp(\beta_1)$ is the usual expression for the relative risk in a standard PH model. Thus, $\psi_{ij}(t|x=1, x=0)$, represents a random multiplicative divergence from the standard relative risk in a PH model which is homogeneous with respect to centres. Note that $\exp(\beta_1 + v_{i1})$ is often called the treatment hazard ratio in the i th centre [2, 5]. We also have that

$$\frac{\exp(\beta_1 + v_{i1})}{\exp(\beta_1)} = \exp(v_{i1}).$$

Thus v_{i1} means the random deviation of the i th centre from the overall treatment effect. Similarly, in order to interpret v_{i0} we consider the model without the covariate x_{ij}

$$\lambda_{ij}(t|v_{i0}) = \lambda_0(t) \exp(v_{i0})$$

whence,

$$\varphi_{ij}(t) = \frac{\lambda_0(t)\exp(v_{i0})}{\lambda_0(t)\exp(0)} = \exp(v_{i0})$$

which is free of time t and holds for all patients in centre i , and v_{i0} represents the random deviation of the i th centre from the overall underlying baseline risk.

2.2. H-likelihood estimation

We now show how to derive the h-likelihood estimation procedure for fitting a correlated semiparametric model (1) and also propose how to obtain valid standard-error estimates for frailty parameters (i.e. dispersion parameters).

Since the functional form of $\lambda_0(t)$ is unknown, following Breslow [33], we approximate the baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ by a step function with jumps at the observed death times [23, 34];

$$\Lambda_0(t) = \sum_{k: y_{(k)} \leq t} \lambda_{0k},$$

where $y_{(k)}$ is the k th ($k = 1, \dots, D$) smallest distinct death time among the y_{ij} 's, and $\lambda_{0k} = \lambda_0(y_{(k)})$.

Following Ha *et al.* [23], the hierarchical log likelihood (h-likelihood) for frailty models (1) is defined by

$$h = h(\beta, v, \lambda_0, \theta) = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}, \quad (5)$$

where

$$\begin{aligned} \sum_{ij} \ell_{1ij} &= \sum_{ij} \delta_{ij} \{ \log \lambda_0(y_{ij}) + \eta_{ij} \} - \sum_{ij} \{ \Lambda_0(y_{ij}) \exp(\eta_{ij}) \} \\ &= \sum_k d_{(k)} \log \lambda_{0k} + \sum_{ij} \delta_{ij} \eta_{ij} - \sum_k \lambda_{0k} \left\{ \sum_{(i,j) \in R_{(k)}} \exp(\eta_{ij}) \right\}, \end{aligned}$$

$\ell_{1ij} = \ell_{1ij}(\beta, \lambda_0; y_{ij}, \delta_{ij} | v_i)$ is the logarithm of the conditional density function for y_{ij} and δ_{ij} given v_i ,

$$\ell_{2i} = \ell_{2i}(\theta; v_i) = -\frac{1}{2} [\log \det \{ 2\pi \sum_i(\theta) \}] - \frac{1}{2} v_i^T \sum_i(\theta)^{-1} v_i$$

is the logarithm of the density function for v_i with parameters θ , $\Lambda_0(\cdot)$ is the baseline cumulative hazard function, and $\eta_{ij} = x_{ij}^T \beta + z_{ij}^T v_i$. Here $\beta = (\beta_1, \dots, \beta_p)^T$, $v = (v_1^T, \dots, v_q^T)^T$, $\lambda_0 = (\lambda_{01}, \dots, \lambda_{0D})^T$, $d_{(k)}$ is the number of deaths at $y_{(k)}$ and $R_{(k)} = R(y_{(k)}) = \{(i, j) : y_{ij} \geq y_{(k)}\}$ is the

risk set at $y_{(k)}$. In (5) log likelihood of the i th cluster is the logarithm of the joint density of (y_i, δ_i) and v_i , where $y_i = (y_{i1}, \dots, y_{im_i})^T$ and $\delta_i = (\delta_{i1}, \dots, \delta_{im_i})^T$. As the number of λ_{0k} s can increase with the number of events, the function $\lambda_0(t)$ is potentially of high dimension. Accordingly, for estimation of (β, v) Ha *et al.* [23] proposed the use of the profiled h-likelihood h^* from which λ_0 is eliminated:

$$h^* = h|_{\lambda_0 = \hat{\lambda}_0} = \sum_{ij} \ell_{1ij}^* + \sum_i \ell_{2i}, \quad (6)$$

where

$$\hat{\lambda}_{0k}(\beta, v) = \frac{d_{(k)}}{\sum_{(i,j) \in R_{(k)}} \exp(\eta_{ij})},$$

are solutions of the estimating equations, $\partial h / \partial \lambda_{0k} = 0$, for $k = 1, \dots, D$. Note here that

$\sum_{ij} \ell_{1ij}^* = \sum_{ij} \ell_{1ij}|_{\lambda_0 = \hat{\lambda}_0} = \sum_k d_{(k)} \log \hat{\lambda}_{0k} + \sum_{ij} \delta_{ij} \eta_{ij} - \sum_k d_{(k)}$ does not depend on λ_0 . Thus Lee and Nelder's [21–22] h-likelihood procedure for hierarchical generalized linear models (HGLMs) can be extended to the frailty models based on h^* [11, 19]. Here, for the estimation of frailty parameters θ we use an adjusted profile h-likelihood [18, 35], $p_{\beta, v}(h^*)$ defined in (A2); the details for estimation procedure are given in Appendix A.

We have shown that the approximated standard-error estimates for β are obtained from the inverse of $-\partial^2 h^* / \partial \beta \partial \beta^T$, given in (7) [18, 23]. In this paper we propose that the approximated standard-error estimates for θ are directly obtained using the inverse of $-\partial^2 p_{\beta, v}(h^*) / \partial \theta^2$; the technical details are given in Appendix B. We also show the performance of these estimates by simulation below. Some conceptual differences between h-likelihood and other estimation methods for frailty models (1) are described in Appendix C.

3. Prediction of random effects

In HGLMs location parameters and dispersion parameter are asymptotically orthogonal [21]. Thus, very recently Lee and Ha [36] showed that a proper standard-error (SE) estimate for the prediction interval of random effects v can be computed from the inverse of the information matrix for (β, v) based on the h-likelihood. Here, the SE becomes the squared root of the approximation of the conditional mean-square error of prediction (CMSEP) of Booth and Horbert [37]. This is a general measure of predictive uncertainty and, following Lee and Ha [36], its extension to frailty models is straightforward as shown below.

In frailty models (1), as in HGLMs, location parameters (β, λ_0, v) and frailty parameters θ are asymptotically orthogonal. For a moment, assume that θ is known. Accordingly, we need only focus on (β, v) after eliminating λ_0 i.e. by using h^* . Following Ha *et al.* [23] and Ha and Lee [18], the asymptotic covariances for β and $\hat{v} - v$ are obtained from the inverse of information matrix, $J(h^*; \beta, v) = -\partial^2 h^* / \partial (\beta, v) \partial (\beta, v)^T$, of β and v based on h^* :

$$J(h^*; \beta, v) = - \begin{pmatrix} \partial^2 h^* / \partial \beta \partial \beta^T & \partial^2 h^* / \partial \beta \partial v^T \\ \partial^2 h^* / \partial v \partial \beta^T & \partial^2 h^* / \partial v \partial v^T \end{pmatrix} = \begin{pmatrix} X^T W^* X & X^T W^* Z \\ Z^T W^* X & Z^T W^* Z + U \end{pmatrix}, \quad (7)$$

where X and Z are the $n \times p$ and $n \times q$ model matrices for β and v whose ij th row vectors are x_{ij}^T and z_{ij}^T , respectively, $U = -\partial^2 \ell_2 / \partial v^2 = \sum^{-1} = \text{BD}(\sum_1^{-1}, \dots, \sum_q^{-1})$, and the weight matrix $W^* = W^*(\beta, v)$ is given in (B4). Here, $\text{BD}(\cdot)$ denotes a block diagonal matrix. This means that the SEs of $\hat{v} - v$ can be computed from the information matrix, J in (7), of the profile h-likelihood h^* .

Let $y_o = (y^T, \delta^T)^T$ and $\psi = (\beta^T, \lambda_0^T)^T$. Here y and δ are the vectors of y_{ij} 's and δ_{ij} 's, respectively. Following Booth and Horbert [37], the CMSEP based on y_o and ψ is defined by

$$\text{CMSEP} = E_{\psi}[\{\hat{v}(\hat{\psi}) - v\} \{\hat{v}(\hat{\psi}) - v\}^T | y_o],$$

where $\hat{v}(\hat{\psi}) \equiv \hat{v}(\psi)|_{\psi=\hat{\psi}}$ and $\hat{v}(\psi)$ is the solution to $\partial h / \partial v = 0$ for a given ψ . Note here that $\hat{v}(\psi) = E(v|y_o)$ asymptotically. Along the lines of Lee and Ha [36], $J(h^*; \hat{\beta}, \hat{v})^{-1} \equiv J(h^*; \beta, v)^{-1}|_{\beta=\hat{\beta}, v=\hat{v}}$ gives the first-order approximation to the CMSEP, leading to a SE for $\hat{v} - v$:

$$\begin{aligned} \text{CMSEP} &= \text{var}_{\psi}(v|y_o) + D(\psi) \\ &\approx \{(-\partial^2 h / \partial v \partial v^T)^{-1} + (\partial \hat{v} / \partial \psi) \text{var}(\hat{\psi}) (\partial \hat{v} / \partial \psi)^T\}|_{\psi=\hat{\psi}, v=\hat{v}} \end{aligned} \quad (8)$$

$$\begin{aligned} &= \text{bottom right hand corner of } J(h^*; \hat{\beta}, \hat{v})^{-1} \\ &= \{(Z^T W^* Z + U) - (Z^T W^* X)(X^T W^* X)^{-1}(X^T W^* Z)\}^{-1}|_{\beta=\hat{\beta}, v=\hat{v}} \end{aligned} \quad (9)$$

where $\text{var}_{\psi}(v|y_o) = E[\{\hat{v}(\psi) - v\} \{\hat{v}(\psi) - v\}^T | y_o]$ and $D(\psi) = E[\{\hat{v}(\hat{\psi}) - \hat{v}(\psi)\} \{\hat{v}(\hat{\psi}) - \hat{v}(\psi)\}^T | y_o]$ is a nonnegative correction that accounts for the variability of parameter estimates $\hat{\psi}$. Here $\partial \hat{v} / \partial \psi = -(-\partial^2 h / \partial v \partial v^T)^{-1} (-\partial^2 h / \partial v \partial \psi^T)|_{v=\hat{v}}$. For the SE of prediction of random effects Vaida and Xu [13] and Othus and Li [20] used the EB method based on conditional posterior distribution of v given y_o , leading to

$$\text{var}_{\psi}(v|y_o) \approx (-\partial^2 h / \partial v \partial v^T)^{-1}|_{\psi=\hat{\psi}, v=\hat{v}} \quad (10)$$

which corresponds to the first term on the right hand side of (8) [21, 37]. The EB method can underestimate the SE of $\hat{v} - v$ because it ignores the term above, $D(\psi)$, which accounts for the inflation of the CMSEP caused by estimating $\psi = (\beta^T, \lambda_0^T)^T$ [36, 37]. Following Lee and Ha [36], for the 95% h-likelihood and EB prediction intervals for v we use

$$\{\hat{v} - 1.96\text{SE}(\hat{v} - v), \hat{v} + 1.96\text{SE}(\hat{v} - v)\},$$

where \hat{v} are obtained from (A1). Here the estimated h-likelihood and EB standard errors, $\text{SE}(\hat{v} - v)$, are also obtained from the square roots of (9) and (10), respectively.

4. Practical example

4.1. The data and correlated model

The duration of the Disease Free Interval (DFI) in non muscle invasive bladder cancer patients, treated in various centres in Europe, is analysed. The DFI is defined as the time from randomization to the date of the first recurrence. Patients without recurrence at the end of the follow-up period were censored at their last date of follow-up. Patients were enrolled in 7 studies conducted by the EORTC [27]. For simplicity of analysis, we consider only 410 patients from 21 centres included in EORTC trial 30791 (Table 1). The two covariates of interest are: CHEMO x_{ij1} (0=No, 1=Yes) and TUSTAT x_{ij2} (0=Primary, 1=Recurrent). Notice that x_{ij1} is the main treatment covariate. Patients with missing values for x_{ij2} were excluded. The numbers of patients per centre varied from 3 to 78, with mean 19.5 and median 15. Of the 410 patients, 204 patients (49.8 per cent) without recurrence were censored at the date of last follow up.

For the purpose of analysis, we consider the three submodels of (3):

M1 (Cox): Cox model without frailties (basic hazard),

M2 (Indep): Cox models, with two independent frailty terms ($\rho = 0$),

M3 (Corr): Cox models, with two correlated frailty terms ($\rho \neq 0$).

Models M2 and M3 contain the random baseline risk v_{i0} and the random treatment-by-centre interaction term, $v_{i1}x_{ij1}$. The models were fitted using SAS/IML. The results are summarized in Table 2. In all three models the two fixed effects ($\beta_j, j = 1, 2$) are significant. In particular, the use of chemotherapy (CHEMO = 1) significantly prolongs the time to first recurrence as compared to patients who do not receive chemotherapy (CHEMO = 0): see also [4]. The two nested models (M1 and M2) ignoring random components or their correlation show similar results for β_j ($j = 1, 2$). However, the absolute magnitude and SE of the estimate for the main treatment effect β_1 in M1 and M2 are smaller than those for the correlated model (M3). In M2 and M3, the variances (σ_0^2 and σ_1^2) indicate the amount of variation between centres in the baseline risk and in the treatment effect, respectively. Here, the estimate of σ_0^2 is relatively larger than that of σ_1^2 . This does not seem surprising since differences in outcome according to treatment effect are typically smaller than differences due to patient characteristics which often vary across centres. However, care may be necessary in comparing the two variances because these two values should not be interpreted on the same scale.

Moreover, the correlated model M3 explains the degree of dependency between the two random components (i.e. the random centre effect v_0 and the random treatment-by-centre interaction v_1). The estimate of ρ ($\hat{\rho} = -0.893$) gives a large negative value, indicating that the two predicted random components (\hat{v}_0 and \hat{v}_1) have a strong negative correlation. It is clear from the plot (not shown) of \hat{v}_1 against \hat{v}_0 that as v_{i0} increases (i.e. the baseline risk increases), v_{i1} decreases. Note here that $\exp(v_{i1})$ represents the ratio of treatment hazard rate in the i th centre (i.e. $\exp(\beta_1 + v_{i1})$) to overall hazard rate (i.e. $\exp(\beta_1)$). In particular, the estimate of β_1 in M3 is negative; we see that a decreasing value of v_{i1} corresponds to an increased treatment effect. Thus, the negative correlation leads to the conclusion that treatment confers more benefit in centres with a higher baseline risk. This is consistent with the findings by Turner *et al.* [6] and Rondeau *et al.* [8] in the context of meta-analysis.

Figure 1 compares SE estimates of h-likelihood (HL) versus EB under M3. As expected, the EB estimates are smaller than HL estimates in both v_{i0} and v_{i1} , leading to a lower coverage probability of prediction interval than the nominal level. Accordingly, below, we conduct

detailed analyses of the random effects using HL. Figure 2 shows the estimates and 95% prediction intervals for the random effects in the 21 centres using M3. It shows the variations of the two random components (v_{i0} , v_{i1}) over centres, ordered by the number of patients entered. In particular, Figure 2(a) shows that centres 12 and 19 provide the highest and the lowest baseline risk, respectively. From Figure 2(b) we see that the corresponding centres give lowest and highest treatment hazards, respectively, which leads in this case to a negative correlation ($\hat{\rho} = -0.893$), as shown in Table 2.

Figures 2(a) and 2(b) also give the prediction intervals for the random baseline risk (v_{i0}) and the random treatment-by-centre interaction (v_{i1}), respectively. Overall, the lengths of the intervals are seen to decrease as the number of patients per centre increases, particularly for Figure 2(a): see also [13]. Figure 2(a) indicates substantial variation in the baseline risk across centres. However, Figure 2(b) shows overall homogeneity in the effect of treatment across centres, that is, there is little treatment-by-centre interaction in this data set. Thus, in this multicentre trial there is little difference in the treatment effects across centres and the treatment is shown to be effective, while there appears to be substantial variation in the baseline risk of DFI across centres. These results suggest that the treatment effect may be generalized to a broader patient population as in the findings by Yamaguchi and Ohashi [2].

In addition, the prediction intervals for the log treatment hazard rates (i.e. $b_{i1} = \beta_1 + v_{i1}$) in the different centres are also useful to check the variations over centres. Similarly, the 95% prediction interval of b_{i1} is given by

$$\{\widehat{b}_{i1} - 1.96SE(\widehat{b}_{i1} - b_{i1}), \widehat{b}_{i1} + 1.96SE(\widehat{b}_{i1} - b_{i1})\},$$

where $\widehat{b}_{i1} = \widehat{\beta}_1 + \widehat{v}_{i1}$ and $SE(\widehat{b}_{i1} - b_{i1}) = \sqrt{\text{var}(\widehat{b}_{i1} - b_{i1})}$. Here $\text{var}(\widehat{b}_{i1} - b_{i1}) = \text{var}(\widehat{\beta}_1) + \text{var}(\widehat{v}_{i1} - v_{i1}) + 2\text{cov}(\widehat{\beta}_1, \widehat{v}_{i1} - v_{i1})$ is obtained from J^{-1} in (7). Figure 2(c) shows wider interval lengths than in Figure 2(b) due to the additional variance and covariance terms, but again confirms there is little difference in the treatment effects over centres.

4.2. Model selection

A thorough analysis will involve us in enlarging the potential model space beyond M1, M2 or M3. We consider a number of extensions below and show how to select an appropriate model using an Akaike information criterion (AIC) [11] based on the focussed extended restricted likelihood (ERL, [35]);

$$AIC(T_d^*) = T_d^* + 2p_T.$$

Notice that $T_d^* = -2p_{\beta,v}(h^*)$ is a deviance based on the ERL $p_{\beta,v}(h^*)$ in (A2) which eliminates (β, v) from h^* , the profile h-likelihood from which the nuisance function $\lambda_0(t)$ has already been eliminated. Thus, T_d^* is a function only of the frailty parameters θ and $AIC(T_d^*)$ is used to select the frailty structure best supported by the data. Here p_T is the number of frailty parameters (i.e. the parameters governing the frailty distribution), not the number of all fitted parameters or frailties. Notice that the focussed $AIC(T_d^*)$ is a sharper model selection tool than the more usual unfocussed AIC [11].

Recall that v_{i0} and v_{i1} are the random baseline risk and random treatment effect of the i th centre, respectively. For the purpose of analysis, we consider the following five models

including M1–M3, $\lambda_{ij}(t|v) = \lambda_0(t) \exp(\eta_{ij})$ with η_{ij} allowing several frailty structures in models M2–M5: Here $(v_{i0}, v_{i1}) \sim BN$ means that $v_{i0} \sim N(0, \sigma_0^2)$, $v_{i1} \sim N(0, \sigma_1^2)$ and $\rho = \text{Corr}(v_{i0}, v_{i1})$; $(v_{i0}, v_{i1}) \sim IN$ also means BN with $\rho = 0$.

$$\begin{aligned} \text{M1 (Cox): } \eta_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2}, \\ \text{M2 (Indep): } \eta_{ij} &= (\beta_1 + v_{i1}) x_{ij1} + \beta_2 x_{ij2} + v_{i0}, \text{ with } (v_{i0}, v_{i1}) \sim IN, \\ \text{M3 (Corr): } \eta_{ij} &= (\beta_1 + v_{i1}) x_{ij1} + \beta_2 x_{ij2} + v_{i0}, \text{ with } (v_{i0}, v_{i1}) \sim BN, \\ \text{M4 (B): } \eta_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + v_{i0}, \text{ with } v_{i0} \sim N(0, \sigma_0^2), \\ \text{M5 (T): } \eta_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + v_{i1} x_{ij1}, \text{ with } v_{i1} \sim N(0, \sigma_1^2), \end{aligned}$$

where B and T denote random baseline risk and random treatment effect, respectively. Here M3 is our full model and the others are various simplifications of it by assuming null components, i.e. M1 ($v_{i0} = 0, v_{i1} = 0$), M2 ($\rho = 0$), M4 ($v_{i1} = 0$) and M5 ($v_{i0} = 0$). For ease of comparison and ranking of candidate models, we have set the smallest AIC to be zero and the other AIC values are shifted accordingly. In Table 3 we report the AIC differences, not the AIC values themselves. The deviance from model M2 is very similar to that obtained in M4 because in M2 the variance of the v_{i1} is very small, i.e. $\widehat{\sigma}_1^2 \approx 0$ in Table 2. If the AIC difference is larger than 1 the choice can be made [38, p.84]. Under this empirical criterion, we note that $\text{AIC}(T_d^*)$ selects M4 as an appropriate model; its estimation results are also presented in Table 2. In particular, it clearly rejects more complex models M2 and M3 than M4, indicating that it reflects model complexity properly [11].

However, the T_{ij} may also depend on the individual-specific random effects as in Ha *et al.* [11]. If this is the case, some of the observed variation between centres is attributable to the heterogeneity between patients. We account for this properly, by introducing an appropriate patient-specific frailty component. Let v_{ij} be the random effects of the j th patient in the i th centre, satisfying $v_{ij} \sim N(0, \sigma^2)$. The extra random term v_{ij} , which is matched with individual-level event time T_{ij} and fixed effect x_{ij} , can be viewed as modelling heterogeneity between patients at the individual patient level [9]. Accordingly, we consider the following additional models:

$$\begin{aligned} \text{M6 (I): } \eta_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + v_{ij}, \text{ with } v_{ij} \sim N(0, \sigma^2), \\ \text{M7 (B+I): } \eta_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + v_{i0} + v_{ij}, \text{ with } v_{i0} \sim N(0, \sigma_0^2) \text{ and } v_{ij} \sim N(0, \sigma^2), \\ \text{M8 (T+I): } \eta_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + v_{i1} x_{ij1} + v_{ij}, \text{ with } v_{i1} \sim N(0, \sigma_1^2) \text{ and } v_{ij} \sim N(0, \sigma^2), \\ \text{M9 (Indep+I): } \eta_{ij} &= (\beta_1 + v_{i1}) x_{ij1} + \beta_2 x_{ij2} + v_{i0} + v_{ij}, \text{ with } (v_{i0}, v_{i1}) \sim IN \text{ and } v_{ij} \sim N(0, \sigma^2), \\ \text{M10 (Corr+I): } \eta_{ij} &= (\beta_1 + v_{i1}) x_{ij1} + \beta_2 x_{ij2} + v_{i0} + v_{ij}, \text{ with } (v_{i0}, v_{i1}) \sim BN \text{ and } v_{ij} \sim N(0, \sigma^2), \end{aligned}$$

where I denotes individual random effect. Now, M10 is the full model which combines models M3 and M6 and the others are various simplifications of it as before, i.e. M9 ($\rho = 0$), M8 ($v_{i0} = 0$), M7 ($v_{i1} = 0$) and M6 ($v_{i0} = 0, v_{i1} = 0$). Note that M6 has independence between the survival times within centres. However, comparing model M6 with M4, M6 is rejected. We also see that additional random effects v_{ij} for B, T, Indep and Corr do not lead to any improvement in deviances. Here, the $\text{AIC}(T_d^*)$ again rejects the additional complexity implied by models M7–M10. Thus, for the bladder-cancer data set the focussed AIC chooses M4 as the best model among those considered. Under M4 the predicted random effects (i.e. random baseline risks) and 95% prediction intervals for each centre are plotted in Figure 3. It shows substantial variations in the baseline risk over centre as evident in Figure 2(a). In particular, the three centres (12, 16) and 19 stand out as having the highest and lowest baseline risks, respectively. Note that although we report the SEs of the σ^2 s, one should not

use them for testing $\sigma^2 = 0$ [13]. Now we are also interested in testing the hypothesis $H_0: \sigma_0^2 = 0$, no centre effect (i.e. no variation in random-baseline risk). Such a null hypothesis is on the boundary of the parameter space, so that the critical value of an asymptotic $(\chi_0^2 + \chi_1^2)/2$ distribution is 2.71 at 5% significant level [25, 39, 40]. The difference in deviance $(-2p_\tau(h^*))$ in Table 3) between M1 and M4 is 3.2 (> 2.71), indicating that the centre effect is significant, i.e. $\sigma_0^2 > 0$.

5. Simulation study

Numerical studies, using 200 replications of simulated data, were conducted to evaluate the performance of the proposed method. Here we consider the two interesting models (2) and (3), which correspond to M4 and M3, respectively. The structure of bladder-cancer data in Table 1 is assumed in order to generate the data from each model. That is, the simulated data structures consist of the total patients $n = 410$ coming from 21 centres, with the number (n_i) of different patients.

Firstly, data are generated from the model (2) with $\lambda_0(t) = 1$ and the two different binary covariates, the main treatment x_{ij1} and x_{ij2} ;

$$\lambda_{ij}(t|v_{i0}) = \exp\{\beta_1 x_{ij1} + \beta_2 x_{ij2} + v_{i0}\}.$$

Here x_{ij1} and x_{ij2} are generated from a Bernoulli distribution with success probability 0.5, respectively. The corresponding true parameters are $\beta_1 = -0.5$ and $\beta_2 = 0.5$. The random effects v_{i0} are also generated from $N(0, \sigma_0^2)$ with $\sigma_0^2 = 0.2$ and 1.0. The corresponding censoring times were, respectively, generated from exponential distribution with parameter values empirically determined to achieve approximately the right censoring rate in each centre of Table 1.

For the 200 replications we computed the mean, standard deviation (SD), the mean of the estimated SE for $\hat{\beta}_j$ ($j = 1, 2$) and $\hat{\sigma}_0^2$, respectively. The corresponding SEs are, respectively, obtained from J^{-1} in (7) and $\{-\partial^2 p_\tau(h^*)/\partial\theta^2\}^{-1}$ in (B2). The results of fitting the model (2) are summarized in Table 4. Here, to save space we report only the results about $\sigma_0^2 = 1$ which give similar results to $\sigma_0^2 = 0.2$. Overall, the h-likelihood estimates of β_j and σ_0^2 perform well even though the simulated data consist of somewhat high censoring. In Table 4 SD is the estimates of the true $\{\text{var}(\hat{\xi})\}^{1/2}$, and SEM is the average of SE estimates for $\hat{\xi}$, where $\hat{\xi} = (\beta_0, \beta_1, \sigma_0^2)^T$. Our SE estimates work well as judged by the very good agreement between SEM and SD.

Next, data were generated from the model (3) with $\lambda_0(t) = 1$:

$$\lambda_{ij}(t|v_{i0}, v_{i1}) = \exp\{\beta_1 x_{ij1} + \beta_2 x_{ij2} + v_{i0} + v_{i1} x_{ij1}\}. \quad (11)$$

The random effects v_{i0} and v_{i1} are generated from the bivariate normal distribution with four combinations of frailty parameters; $(\sigma_0^2, \sigma_1^2, \rho) = (0.2, 0.2, -0.5), (1.0, 1.0, -0.5), (0.2, 0.2, 0.5)$ and $(1.0, 1.0, 0.5)$, leading to $\sigma_{01} = -0.1, -0.5, 0.1, 0.5$, respectively. The remaining simulation schemes including (x_{ij1}, x_{ij2}) and (β_1, β_2) are the same as before. The results of fitting the model (3) with $\rho = -0.5$ are also given in Table 4. Though not reported here, we found the similar results for $\rho = 0.5$. Overall, our approach again works well. However, the

estimates of the frailty parameters $(\sigma_0^2, \sigma_1^2, \sigma_{01})$ are slightly biased when the variances are large as in $(\sigma_0^2, \sigma_1^2) = (1.0, 1.0)$.

In addition, we investigated the performance of our h-likelihood procedure when the normal assumption of log-frailties v_{i0} and v_{i1} in (11) is violated. For linear mixed models Ha *et al.* [41] and Verbeke and Lesaffre [42] have shown that misspecifying the normal random-effect distribution has little effect on the fixed effects estimates. Following Verbeke and Lesaffre [42], for simplicity we consider a mixture (Johnson and Kotz, [43, p.73]) of two bivariate normal distributions. That is, $v_i = (v_{i0}, v_{i1})^T$ are generated from one of the following two cases:

$$\begin{aligned} \text{Case 1: } & \frac{1}{2}N \left\{ \left(\begin{array}{c} -0.3 \\ 0.1 \end{array} \right), \left(\begin{array}{cc} 0.11 & -0.07 \\ -0.07 & 0.19 \end{array} \right) \right\} + \frac{1}{2}N \left\{ \left(\begin{array}{c} 0.3 \\ -0.1 \end{array} \right), \left(\begin{array}{cc} 0.11 & -0.07 \\ -0.07 & 0.19 \end{array} \right) \right\}, \\ \text{Case 2: } & \frac{1}{2}N \left\{ \left(\begin{array}{c} -0.8 \\ 0.4 \end{array} \right), \left(\begin{array}{cc} 0.36 & -0.18 \\ -0.18 & 0.84 \end{array} \right) \right\} + \frac{1}{2}N \left\{ \left(\begin{array}{c} 0.8 \\ -0.4 \end{array} \right), \left(\begin{array}{cc} 0.36 & -0.18 \\ -0.18 & 0.84 \end{array} \right) \right\}. \end{aligned}$$

Two non-normal distributions with Cases 1 and 2 have been chosen such that $E(v_i) = 0$ and such that $\text{var}(v_i)$ equals the random-effect variance-covariance parameters in the first and second settings given in M3 of Table 4, respectively. Note that Cases 1 and 2 produce unimodal and bimodal distributions, respectively (not shown). The results in the third and fourth settings given in M3 of Table 4 again confirm that the h-likelihood method gives robust results for the estimation of parameters, particularly for β , when the distribution of frailty is misspecified.

The SAS/IML program for a correlated model (11) with a simulated data set is available from the website: <http://stat.snu.ac.kr/hglmlab>.

6. Discussion

We have shown that the proposed method provides a unified framework for the inference. The data-directed simulation results have demonstrated that our procedure performs well for the estimation of parameters, including the estimated SEs. Using h-likelihood, we have also shown how to investigate potential sources of the heterogeneity related to treatment effect over centres in multi-centre clinical trial. The proposed method can be also employed when studying such heterogeneity in a meta-analysis [8] which combines survival data from different clinical trials.

The heterogeneity of treatment effect could also arise in other situations besides treatment-by-centre interaction. For example, it could arise in the case that the treatment effect affects the variances of the frailty terms [44]; a simple dispersion model is a model (2) with $v_{i0} \sim N(0, \sigma_{i0}^2)$ allowing a regression model for σ_{i0}^2 , given by $\log \sigma_{i0}^2 = \gamma_0 + \gamma_1 x_{ij1}$ where x_{ij1} is a main treatment covariate. Pan and MacKenzie [45, 46] have developed appropriate structural dispersion methods for testing this hypothesis in the repeated measures setting with Gaussian response variables and with and without random effects. Thus, we are currently working on an extension of our method to models with structural dispersion.

In the data set in Section 4 we coded the main treatment as $x_{ij1} = 0, 1$ to indicate control or treatment group. However, the coding of $x_{ij1} = \pm \frac{1}{2}$ may give a flexible covariance structure for the random effects [6, 8]. Though not reported here, both codings give similar estimation results for all random-effect models (M2–M10) considered and select M4 as the best model. Furthermore, we investigated how the small size (e.g. $n_i = 3$) of some centres in Table 1

influences the inference results. Here, the centres, centre numbers 1 and 2, with less than 4 patients were combined into one new centre. We have also observed (not shown) that the results obtained from fitting a correlated model (M3) under the combined data set are very similar to those of M3 in Table 2.

The focussed AIC(T_d^*) in Section 4.2 is a criterion for the frailty parameters only and it cannot be used for model selection involving the β parameter because the restricted likelihood T_d^* eliminates the β . However, if β is the subject of the model selection process we may use the AIC based upon an adjusted profile h-likelihood $p_v(h_p)$ in (C1) [11]. Thus, there is clearly scope for further research on the development of a criterion for selecting the best model globally. We have ignored missing covariates in the data set analysed because their frequency is too small (i.e. 4/414=1%), but the original data set with 7 studies includes more missing covariates. The development of h-likelihood methods for frailty models allowing for missing covariates would be an interesting topic for future work.

Acknowledgments

The authors thank the European Organization for Research and Treatment of Cancer Genito-Urinary Tract Cancer Group for permission to use the data from EORTC trial 30791 for this research. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-521-C00057). Professor MacKenzie was supported by the BIO-SI project (SFI 07/MI/012) and was funded wholly by ENSAI, Rennes, France when this paper was completed. This publication was also supported by grants number 5U10 CA011488-38 through 5U10 CA011488-39 from the National Cancer Institute (Bethesda, Maryland, USA) and by the EORTC Charitable Trust. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute. Professor Legrand was supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy).

References

1. Andersen PK, Klein JP, Zhang M-J. Testing for centre effects in multi-centre survival studies: a monte carlo comparison of fixed and random effects tests. *Statistics in Medicine*. 1999; 18:1489–1500. [PubMed: 10398287]
2. Yamaguchi T, Ohashi Y. Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer. *Statistics in Medicine*. 1999; 18:1961–1971. [PubMed: 10440879]
3. Glidden DV, Vittinghoff E. Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine*. 2004; 23:369–388. [PubMed: 14748034]
4. Legrand C, Ducrocq V, Janssen P, Sylvester R, Duchateau L. A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. *Statistics in Medicine*. 2005; 24:3789–3804. [PubMed: 16320274]
5. Gray RJ. A Bayesian analysis of institutional effects in multicenter cancer clinical trial. *Biometrics*. 1994; 50:244–253. [PubMed: 8086608]
6. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*. 2000; 19:3417–3432. [PubMed: 11122505]
7. Komarek A, Lesaffre E, Legrand C. Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Statistics in Medicine*. 2007; 26:5457–5472. [PubMed: 17910009]
8. Rondeau V, Michiels S, Liqueur B, Pignon JP. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine*. 2008; 27:1894–1910. [PubMed: 18069745]
9. Yau KKW, Kuk AYC. Robust estimation in generalized linear mixed models. *Journal of the Royal Statistical Society, Series B*. 2002; 64:101–117.
10. Song X-Y, Lee S-Y. Model comparison of generalized linear mixed models. *Statistics in Medicine*. 2006; 25:1685–1698. [PubMed: 16220521]

11. Ha ID, Lee Y, MacKenzie G. Model selection for multi-component frailty models. *Statistics in Medicine*. 2007; 26:4790–4807. [PubMed: 17476647]
12. Friedman, LM.; Furberg, CD.; DeMets, DL. *Fundamentals of clinical trials*. Springer; New York: 1998.
13. Vaida F, Xu R. Proportional hazards model with random effects. *Statistics in Medicine*. 2000; 19:3309–3324. [PubMed: 11122497]
14. Gamst A, Donohue M, Xu R. Asymptotic properties and empirical evaluation of the NPMLE in the proportional hazards mixed-effects model. *Statistica Sinica*. 2009; 19:997–1011.
15. Abrahantes JC, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L. Comparison of different estimation procedures for proportional hazards model with random effects. *Computational Statistics and Data Analysis*. 2007; 51:3913–3930.
16. Duchateau, L.; Janssen, P. *The Frailty Models*. Springer; New York: 2008.
17. Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*. 2000; 56:1016–1022. [PubMed: 11129456]
18. Ha ID, Lee Y. Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*. 2003; 12:663–681.
19. Ha ID, Noh M, Lee Y. Bias reduction of likelihood estimators in semi-parametric frailty models. *Scandinavian Journal of Statistics*. 2010; 37:307–320.
20. Othus, M.; Li, Y. Marginalized frailty models for multivariate survival data. Harvard University Biostatistics Working Paper Series, paper. <http://www.bepress.com/harvardbiostat/paper104>
21. Lee Y, Nelder JA. Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*. 1996; 58:619–678.
22. Lee Y, Nelder JA. Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*. 2001; 88:987–1006.
23. Ha ID, Lee Y, Song J-K. Hierarchical likelihood approach for frailty models. *Biometrika*. 2001; 88:233–243.
24. Ha ID, Lee Y. Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models. *Biometrika*. 2005; 92:717–723.
25. Ha ID, Lee Y. Multilevel mixed linear models for survival data. *Lifetime Data Analysis*. 2005; 11:131–142. [PubMed: 15747594]
26. Lee, Y.; Nelder, JA.; Pawitan, Y. *Generalised Linear Models with Random Effects: unified analysis via h-likelihood*. Chapman and Hall; London: 2006.
27. Sylvester R, van der Meijden APM, Oosterlinck W, Witjes J, Bouffieux C, Denis L, Newling DWW, Kurth K. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *European Urology*. 2006; 49:466–477. [PubMed: 16442208]
28. Yau KKW. Multilevel models for survival analysis with random effects. *Biometrics*. 2001; 57:96–102. [PubMed: 11252624]
29. Yau KKW, McGilchrist CA. ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine*. 1998; 17:1201–1213. [PubMed: 9670410]
30. McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. *Biometrics*. 1991; 47:461–466. [PubMed: 1912255]
31. Hougaard, P. *Analysis of multivariate survival data*. Springer; New York: 2000.
32. Longford, NT. *Random coefficient models*. Oxford University Press; New York: 1993.
33. Breslow NE. Discussion of Professor Cox's paper. *Journal of the Royal Statistical Society, Series B*. 1972; 34:216–217.
34. Nielsen GG, Gill RD, Andersen PK, Sørensen TIA. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*. 1992; 19:25–44.
35. Lee Y, Nelder JA. Extended-REML estimators. *Journal of Applied Statistics*. 2003; 30:845–856.
36. Lee Y, Ha ID. Orthodox BLUP versus h-likelihood methods for inferences about random effects in Tweedie mixed models. *Statistics and Computing*. 2010; 20:295–303.
37. Booth JG, Hobert JP. Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*. 1998; 93:262–272.

38. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. Akaike information criterion statistics. KTK Scientific Publisher; Tokyo, Japan: 1986.
39. Self SG, Liang KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*. 1987; 82:605–610.
40. Stram DO, Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics*. 1994; 50:1171–1177. [PubMed: 7786999]
41. Ha ID, Lee Y, Song J-K. Hierarchical-likelihood approach for mixed linear models with censored data. *Lifetime Data Analysis*. 2002; 8:163–176. [PubMed: 12048865]
42. Verbeke G, Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis*. 1997; 23:541–556.
43. Johnson, NL.; Kotz, S. Continuous multivariate distributions. John Wiley & Sons; New York: 1972.
44. Noh M, Ha ID, Lee Y. Dispersion frailty models and HGLMs. *Statistics in Medicine*. 2006; 25:341–1354.
45. Pan JX, MacKenzie G. Regression models for covariance structures in longitudinal studies. *Statistical Modelling*. 2006; 6:43–57.
46. Pan JX, MacKenzie G. Modelling conditional covariance in the linear mixed model. *Statistical Modelling*. 2007; 7:49–71.
47. Cox DR. Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B*. 1972; 74:187–220.
48. Breslow NE. Covariance analysis of censored survival data. *Biometrics*. 1974; 30:89–99. [PubMed: 4813387]

Appendix A

H-likelihood estimation procedure

With h^* in (6) we estimate the fixed parameters (β, θ) and random effects v as follows. Ha *et al.* [23] further showed that given θ the estimation of $\tau = (\beta^T, v^T)^T$ is obtained by solving

$$\frac{\partial h^*}{\partial \tau} = \frac{\partial h}{\partial \tau} \Big|_{\lambda_0 = \hat{\lambda}_0} = 0. \quad (\text{A1})$$

Here the first partial derivatives, $\partial h / \partial \tau$, are given by the simple forms:

$$\frac{\partial h}{\partial \beta} = \sum_{ij} (\delta_{ij} - \mu_{ij}) x_{ij} \quad \text{and} \quad \frac{\partial h}{\partial v_i} = \sum_j (\delta_{ij} - \mu_{ij}) z_{ij} - \sum_i^{-1} v_i \quad (i=1, \dots, q),$$

where $\mu_{ij} = \Lambda_0(y_{ij}) \exp(\eta_{ij})$. Next, for the estimation of the frailty parameters θ , we use Lee and Nelder's [21] adjusted profile h-likelihood [18] which eliminates (β, v) from h^* , defined by

$$p_{\tau}(h^*) = \left[h^* - \frac{1}{2} \log \det \left\{ \frac{J(h^*; \tau)}{(2\pi)} \right\} \right] \Big|_{\tau = \hat{\tau}}, \quad (\text{A2})$$

where $\hat{\tau} = \hat{\tau}(\theta) = (\beta^T(\theta), v^T(\theta))^T$ and $J(h^*; \tau) = -\partial^2 h^* / \partial \tau^2$ is an information matrix for τ with a detailed form in (7). The restricted maximum likelihood (REML) estimator for θ are obtained by solving iteratively

$$\frac{\partial p_{\tau}(h^*)}{\partial \theta} = 0. \quad (\text{A3})$$

Note here that

$$\frac{\partial p_{\tau}(h^*)}{\partial \theta} = -\frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta} \right) - \frac{1}{2} \widehat{v}^T \left(\frac{\partial \Sigma^{-1}}{\partial \theta} \right) \widehat{v} - \frac{1}{2} \text{tr} \left(\widehat{J}^{-1} \frac{\partial \widehat{J}}{\partial \theta} \right),$$

where $\Sigma = \text{BD}(\Sigma_1, \dots, \Sigma_q)$ is the $q \times q$ block diagonal matrix and $\widehat{J} = \widehat{J}(\theta) = J(h^*; \tau)|_{\tau=\widehat{\tau}(\theta)}$. Note also that in implementing (A3) we allow the $\partial \widehat{J} / \partial \theta$ term [18, 24, 26]; the computations of the $\partial \widehat{J} / \partial \theta$ term including the $\partial \widehat{v} / \partial \theta$ term are given in Appendix B.

In summary, the estimates of τ and θ are obtained by alternating between the two estimating equations (A1) and (A3) until convergence is achieved [18, 23]. The two equations are, respectively, solved using the Newton-Raphson method with the corresponding Hessian matrices, $-\mathcal{J}^2 h^* / \partial \tau^2$ and $-\mathcal{J}^2 p_{\tau}(h^*) / \partial \theta^2$. After convergence, we directly compute the estimates of $\text{var}(\widehat{\tau} - \tau)$ and $\text{var}(\widehat{\theta})$ using the inverses of $-\mathcal{J}^2 h^* / \partial \tau^2$ and $-\mathcal{J}^2 p_{\tau}(h^*) / \partial \theta^2$, respectively.

Appendix B

The computation of $-\partial^2 p_{\tau}(h^*) / \partial \theta^2$

The adjusted profile h-likelihood in (A2) can be expressed as

$$p_{\tau}(h^*) = \widehat{h} - \frac{1}{2} \log \det(\widehat{J}) + \frac{(p+q)}{2} \log(2\pi),$$

where $\tau = (\beta^T, v^T)^T$, $\widehat{h} = h^*|_{\tau=\widehat{\tau}(\theta)} = h^*(\widehat{\tau}(\theta), \theta)$ and $\widehat{J} = J(h^*; \tau)|_{\tau=\widehat{\tau}(\theta)} = J(\widehat{\tau}(\theta), \theta)$. Since

$$\frac{\partial p_{\tau}(h^*)}{\partial \theta_r} = \frac{\partial \widehat{h}}{\partial \theta_r} - \frac{1}{2} \text{tr} \left(\widehat{J}^{-1} \frac{\partial \widehat{J}}{\partial \theta_r} \right), \quad (\text{B1})$$

we have

$$-\frac{\partial^2 p_{\tau}(h^*)}{\partial \theta_r \partial \theta_s} = -\frac{\partial^2 \widehat{h}}{\partial \theta_r \partial \theta_s} + \frac{1}{2} \text{tr} \left(-\widehat{J}^{-1} \frac{\partial \widehat{J}}{\partial \theta_r} \widehat{J}^{-1} \frac{\partial \widehat{J}}{\partial \theta_s} + \frac{\partial^2 \widehat{J}}{\partial \theta_r \partial \theta_s} \right). \quad (\text{B2})$$

We now show how to compute equation (B2). Following Lee and Nelder [22] and Ha and Lee [18], we allow for $\partial \widehat{J} / \partial \theta_r$ in computing the two equations, (B1) and (B2), but not for $\partial \widehat{\beta} / \partial \theta_r$. Then we have

$$\begin{aligned} \frac{\widehat{\partial h}}{\partial \theta_r} &= \left\{ \left(\frac{\partial h^*}{\partial \theta_r} \right) + \left(\frac{\partial h^*}{\partial v} \right) \left(\frac{\partial v}{\partial \theta_r} \right) \right\} \Big|_{v=\widehat{v}} \\ &= \frac{\partial h^*}{\partial \theta_r} \Big|_{v=\widehat{v}} \end{aligned}$$

since $(\partial h^* / \partial v)|_{v=\widehat{v}} = 0$: see also Appendix 2 of Ha *et al.* [23]. Along the lines of Appendix C of Lee and Nelder [21], we can show that

$$\begin{aligned} \frac{\partial v}{\partial \theta_r} &= - \left(- \frac{\partial^2 h^*}{\partial v^2} \right)^{-1} \left(- \frac{\partial^2 h^*}{\partial v \partial \theta_r} \right) \Big|_{v=\widehat{v}} \\ &= - (Z^T \widehat{W} Z + U)^{-1} U_r' \widehat{v}, \end{aligned}$$

where \widehat{W} is given in (B3), $U = \Sigma^{-1}$ and $U_r' = \partial \Sigma^{-1} / \partial \theta_r = - \Sigma^{-1} (\partial \Sigma / \partial \theta_r) \Sigma^{-1}$. From these results the first term on the right hand side (RHS) of (B2) becomes

$$\begin{aligned} - \frac{\partial^2 \widehat{h}}{\partial \theta_r \partial \theta_s} &= \left\{ \left(- \frac{\partial^2 h^*}{\partial \theta_r \partial \theta_s} \right) - \left(\frac{\partial^2 h^*}{\partial \theta_r \partial v} \right) \left(- \frac{\partial v}{\partial \theta_r} \right) \right\} \Big|_{v=\widehat{v}} \\ &= - \frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_r} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_s} - \Sigma^{-1} \frac{\partial^2 \Sigma}{\partial \theta_r \partial \theta_s} \right) + \frac{1}{2} \widehat{v}^T U_{rs}'' \widehat{v} + \widehat{v}^T U_s' \left(\frac{\partial v}{\partial \theta_r} \right), \end{aligned}$$

where $U_{rs}'' = \partial^2 U / \partial \theta_r \partial \theta_s = - U_s' (\partial \Sigma / \partial \theta_r) \Sigma^{-1} - \Sigma^{-1} (\partial \Sigma / \partial \theta_r) U_s' - \Sigma^{-1} (\partial^2 \Sigma / \partial \theta_r \partial \theta_s) \Sigma^{-1}$. From (7) we have

$$\widehat{J} = \begin{pmatrix} X^T \widehat{W} X & X^T \widehat{W} Z \\ Z^T \widehat{W} X & Z^T \widehat{W} Z + U \end{pmatrix}, \tag{B3}$$

where $\widehat{W} = W^*|_{\tau=\widehat{\tau}(\theta)} = W^*(\widehat{\tau}(\theta), \theta)$. Note here that following Appendix B of Ha and Lee [18], $W^* = W^*(\beta, v)$ is given by

$$W^* = W_1 - W_2, \tag{B4}$$

where $W_1 = \text{diag}\{\widehat{\Lambda}_{0ij} \exp(\eta_{ij})\}$ is the $n \times n$ diagonal matrix with $\widehat{\Lambda}_{0ij} = \widehat{\Lambda}_0(y_{ij})$ and $W_2 = (W_3 M) C^{-1} (W_3 M)^T$ is the $n \times n$ symmetric matrix. Here $W_3 = \text{diag}\{\exp(\eta_{ij})\}$,

$C = \text{diag}\{d_{(k)} / \widehat{\lambda}_{0k}^2\}$ is the $D \times D$ diagonal matrix, and $M = (M_1, \dots, M_D)^T$ is the $n \times D$ indicator matrix whose (ij, k) th element is 1 if $y_{ij} \geq y_{(k)}$ and 0 otherwise. Notice that $\widehat{\Lambda}_{0ij}$ and $\widehat{\lambda}_{0k}$ also depend on (β, v) only and that the corresponding matrix forms are available in Ha and Lee [18]. Thus, the two derivatives in the second term on the RHS of (B2) are computed as follows:

$$\frac{\partial \widehat{J}}{\partial \theta_r} = \begin{pmatrix} X^T W_r' X & X^T W_r' Z \\ Z^T W_r' X & Z^T W_r' Z + U_r' \end{pmatrix} \text{ and } \frac{\partial^2 \widehat{J}}{\partial \theta_r \partial \theta_s} = \begin{pmatrix} X^T W_{rs}'' X & X^T W_{rs}'' Z \\ Z^T W_{rs}'' X & Z^T W_{rs}'' Z + U_{rs}'' \end{pmatrix}.$$

Here $W_r' = \partial \widehat{W} / \partial \theta_r$ and $W_{rs}'' = \partial^2 \widehat{W} / \partial \theta_r \partial \theta_s$ are calculated by the following procedures.

$$\frac{\partial \widehat{W}}{\partial \theta_r} = \left\{ \left(\frac{\partial W^*}{\partial \theta_r} \right) + \left(\frac{\partial W^*}{\partial v} \right) \left(\frac{\partial \widehat{v}}{\partial \theta_r} \right) \right\} \Big|_{v=\widehat{v}} = \left\{ \left(\frac{\partial W^*}{\partial v} \right) \left(\frac{\partial \widehat{v}}{\partial \theta_r} \right) \right\} \Big|_{v=\widehat{v}}$$

since $\partial W^*/\partial \theta_r = 0$, and

$$\frac{\partial^2 \widehat{W}}{\partial \theta_r \partial \theta_s} = \left\{ \left(\frac{\partial \widehat{v}}{\partial \theta_r} \right) \left(\frac{\partial^2 W^*}{\partial v^2} \right) \left(\frac{\partial \widehat{v}}{\partial \theta_s} \right) + \left(\frac{\partial W^*}{\partial v} \right) \left(\frac{\partial^2 \widehat{v}}{\partial \theta_r \partial \theta_s} \right) \right\} \Big|_{v=\widehat{v}},$$

where

$$\frac{\partial^2 \widehat{v}}{\partial \theta_r \partial \theta_s} = - (Z^T \widehat{W} Z + U)^{-1} \left\{ (Z^T W'_r Z + U'_r) \frac{\partial \widehat{v}}{\partial \theta_s} + U'_s \frac{\partial \widehat{v}}{\partial \theta_r} + U''_{rs} \widehat{v} \right\},$$

and $\partial W^*/\partial v$ and $\partial^2 W^*/\partial v^2$ can be calculated by repeatedly differentiating (B4) with respect to v .

Appendix C

Comparison of different estimation methods

Ha *et al.* [19, 23] have showed that the profile h-likelihood h^* in (6) is proportional to the penalized partial likelihood h_p [PPL, 17], which uses the partial likelihood [47–48] for ℓ_{1ij} in h ; $h^* = h_p + \text{constant}$. The h-likelihood and PPL procedures are the same for the estimation of β and v , given frailty parameters θ , but are different for that of θ . For the estimation of θ , the h-likelihood method uses the restricted likelihood $p_{\beta,v}(h^*)$, whereas the PPL method uses an adjusted profile h-likelihood

$$p_v(h_p) = \left[h_p - \frac{1}{2} \log \det \left\{ \frac{J(h_p; v)}{(2\pi)} \right\} \right] \Big|_{v=\widehat{v}}, \quad (C1)$$

where $J(h_p; v) = -\partial^2 h_p / \partial v^2$, which is a Laplace approximation to the marginal likelihood [19]; notice that $p_v(h_p) - p_v(h^*) = \text{constant}$. However, the PPL ignores the $\partial \widehat{v} / \partial \theta$ term in solving the score equations $\partial p_v(h_p) / \partial \theta = 0$; this leads to an underestimation of the parameters and/or SEs, particularly when the cluster size n_i is small [14, 18, 19].

Recently, the penalized maximum likelihood approach [8], which penalizes the baseline hazard $\lambda_0(t)$ in marginal likelihood, has been proposed for the inference of parameters, but it can not directly use for inference of frailties because it eliminates them by integration as in standard marginal-likelihood approach [13, 34]. Furthermore, Bayesian approaches [4, 7] have been also suggested. Legrand *et al.* [4] proposed a Bayesian approach using a Laplace integration technique to approximate the marginal posterior density, $\pi(\theta|y, \delta)$; it can be shown that under uniform priors (i.e. flat priors) for β and θ , $\log\{\pi(\theta|y, \delta)\} \simeq p_{\beta,v}(h^*)$. Thus, we see that the h-likelihood method is equivalent to Legrand *et al.*'s method under uniform priors - a choice, however, which is unlikely to be adopted in practice by Bayesians.

Komarek *et al.* [7] also proposed to use a Markov chain Monte Carlo algorithm but in an accelerated failure time model with Gaussian random effects.

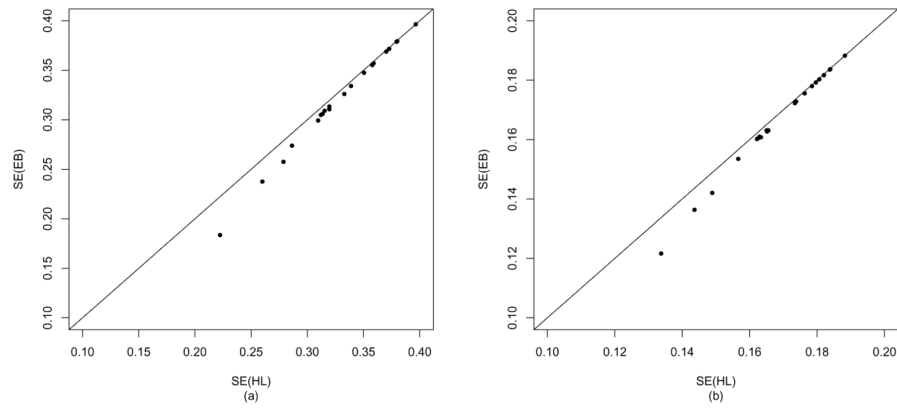


Figure 1. Standard error (SE) estimates of empirical Bayes (EB) versus h-likelihood (HL) analyses for random effects in the bladder cancer data; (a) v_0 and (b) v_1 in correlated model (M3).

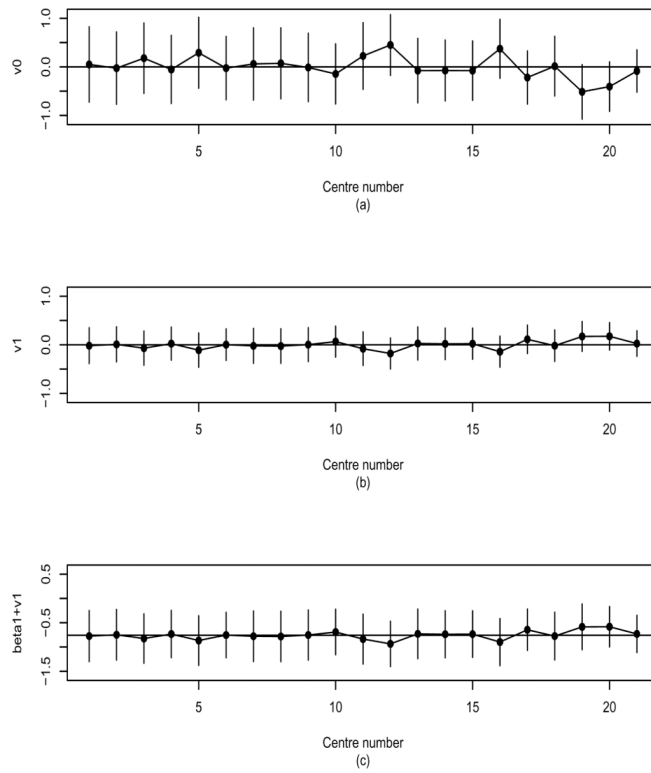


Figure 2. Random effects of 21 centres in the bladder cancer data and their 95% prediction intervals, under correlated model (M3); (a) baseline risk (v_{i0}); (b) treatment-by-centre interaction (v_{i1}); (c) log treatment hazard ratio ($b_{i1} = \beta_1 + v_{i1}$). Centres are sorted in increasing order of number of patients.

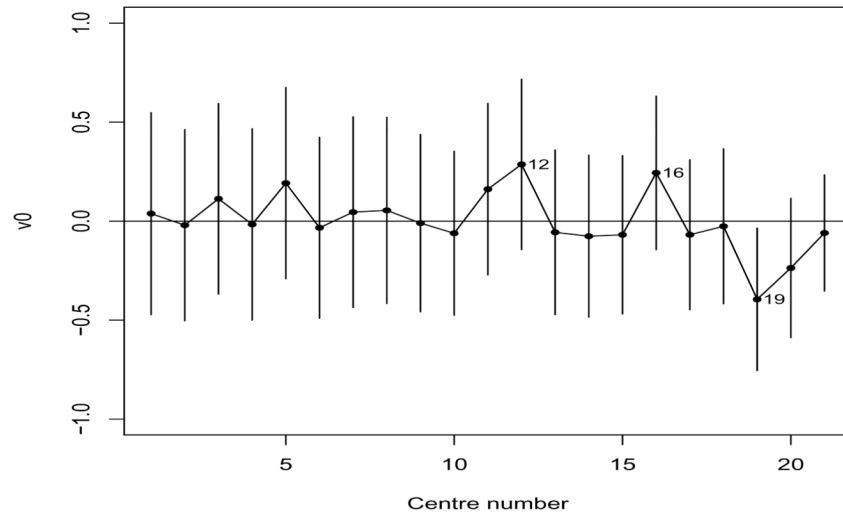


Figure 3. Random baseline risks (v_{i0}) of 21 centres in the bladder cancer data and their 95% prediction intervals, under the final model (M4). Centres are sorted in increasing order of number of patients.

Table 1

Numbers of patients and recurrences per centre in the bladder cancer data

Centre	n_p	n_e	r_c	Centre	n_p	n_e	r_c
1	3	1	0.67	12	15	11	0.27
2	3	2	0.33	13	18	8	0.56
3	4	4	0	14	18	9	0.50
4	5	2	0.60	15	21	10	0.52
5	5	5	0	16	27	17	0.37
6	6	4	0.33	17	28	15	0.46
7	7	3	0.57	18	30	12	0.60
8	8	4	0.50	19	42	13	0.69
9	11	5	0.55	20	52	18	0.65
10	14	8	0.43	21	78	46	0.41
11	15	9	0.40				

n_p = No. of patients in centre; n_e = No. of events; $r_c = 1 - (n_e/n_p)$; Centres ordered by increasing n_p .

Table 2

Results for fitting the four models to the bladder cancer data

Model	β_1 (SE)	β_2 (SE)	$\tilde{\sigma}_0^2$ (SE)	$\tilde{\sigma}_1^2$ (SE)	$\hat{\sigma}_{01}$ (SE)	$[\hat{\rho}]$
M1 (Cox)	-0.667 (0.170)	0.509 (0.144)	-	-	-	-
M2 (Indep)	-0.695 (0.175)	0.544 (0.149)	0.070 (0.058)	3×10^{-12} (1×10^{-4})	-	-
M3 (Corr)	-0.757 (0.191)	0.532 (0.150)	0.161 (0.178)	0.036 (0.170)	-0.068 (0.149)	[-0.893]
M4 (B)	-0.695 (0.175)	0.544 (0.149)	0.070 (0.058)	-	-	-

M1: Cox model without frailties; M2: independent frailty model with $\rho = 0$;M3: correlated frailty model with $\rho \neq 0$; M4: shared frailty model with random baseline risk (B) only; β_1 and β_2 , effects of treatment and tumor status, respectively; σ_0^2 and σ_1^2 , the variances of random baseline risk and random treatment effect, respectively; σ_{01} and ρ , the corresponding covariance and correlation with $\rho = \sigma_{01}/(\sigma_0\sigma_1)$;

SE, the estimated standard error for parameters.

Table 3

Deviance results: AICs for selecting frailty structures in the bladder cancer data

Model	T_d^*	p_T	AIC(T_d^*)
M1 (Cox)	2196.2	0	1.2
M2 (Indep)	2193.0	2	2.0
M3 (Corr)	2192.7	3	3.7
M4 (B)	2193.0	1	0
M5 (T)	2194.2	1	1.2
M6 (I)	2195.6	1	2.6
M7 (B+I)	2192.3	2	1.3
M8 (T+I)	2193.5	2	2.5
M9 (Indep+I)	2192.3	3	3.3
M10 (Corr+I)	2192.1	4	5.1

AIC, differences where the smallest AIC is adjusted to be zero; T, random treatment effect (v_{i1}); I, individual random effect (v_{ij}); Indep, B & T are independent; Corr, B & T are correlated; $T_d^* = -2\{p_\tau(h^*)\}$; p_T , the number of frailty parameters.

Simulation results for the estimation of fixed parameters over 200 replications under random baseline (M4) and correlated frailty models (M3) when the true distribution of log-frailty is correctly specified or misspecified (i.e. a mixture of two bivariate normals).

Table 4

Fitted model	Setting	Parameter	True	Mean	SD	SEM		
M4	Correct	$\sigma_0^2=1$						
		β_1	-0.5	-0.505	0.150	0.150		
		β_2	0.5	0.504	0.156	0.149		
		$\widehat{\sigma}_0^2$	1	1.005	0.426	0.418		
M3	Correct	$\sigma_0^2=\sigma_1^2=0.2, \sigma_{01} = -0.1$						
		β_1	-0.5	-0.506	0.191	0.186		
		β_2	0.5	0.494	0.153	0.148		
		$\widehat{\sigma}_0^2$	0.2	0.211	0.138	0.138		
		$\widehat{\sigma}_1^2$	0.2	0.212	0.221	0.209		
		$\hat{\sigma}_{01}$	-0.1	-0.104	0.135	0.133		
		$\hat{\rho}$	-0.5	-0.493	-	-		
		Correct	$\sigma_0^2=\sigma_1^2=1, \sigma_{01} = -0.5$	β_1	-0.5	-0.480	0.261	0.283
				β_2	0.5	0.502	0.155	0.153
				$\widehat{\sigma}_0^2$	1	1.021	0.472	0.457
$\widehat{\sigma}_1^2$	1			1.029	0.569	0.559		
$\hat{\sigma}_{01}$	-0.5			-0.519	0.398	0.397		
$\hat{\rho}$	-0.5			-0.506	-	-		

Fitted model	Setting	Parameter	True	Mean	SD	SEM
Misspecified						
		$\sigma_0^2 = \sigma_1^2 = 0.2, \sigma_{01} = -0.1$				
		β_1	-0.5	0.502	0.198	0.185
		β_2	0.5	0.488	0.155	0.148
		$\tilde{\sigma}_0^2$	0.2	0.209	0.140	0.144
		$\tilde{\sigma}_1^2$	0.2	0.208	0.211	0.205
		$\hat{\sigma}_{01}$	-0.1	-0.102	0.139	0.138
		$\hat{\rho}$	-0.5	-0.490	-	-
Misspecified						
		$\sigma_0^2 = \sigma_1^2 = 1, \sigma_{01} = -0.5$				
		β_1	-0.5	-0.476	0.309	0.293
		β_2	0.5	0.512	0.156	0.153
		$\tilde{\sigma}_0^2$	1	1.048	0.483	0.461
		$\tilde{\sigma}_1^2$	1	1.089	0.600	0.588
		$\hat{\sigma}_{01}$	-0.5	-0.525	0.415	0.404
		$\hat{\rho}$	-0.5	-0.491	-	-

SD, standard deviation of estimates over 200 simulations, is defined by $\{\sum_i (\hat{\kappa}^{(i)} - \bar{\kappa})^2 / 199\}^{1/2}$, where $\hat{\kappa}^{(i)}$ is the estimate of κ in the i th replication and $\bar{\kappa} = \sum_i \hat{\kappa}^{(i)} / 200$ is the mean of $\hat{\kappa}^{(i)}$'s, and $\kappa = \beta_1, \beta_2, \sigma_0^2, \sigma_1^2$, or σ_{01} .

SEM, the mean of estimated standard errors over 200 simulations.