Journal of the Korean Data & Information Science Society
2011, **22**(3), 605–612

한국데이터정보과학회지

Inference for heterogeneity of treatment effect in multi-center clinical trial[†]

Il Do Ha[1]

[1]Department of Asset Management, Daegu Haany University

Received 2 April 2011, revised 4 May 2011, accepted 12 May 2011

Abstract

In multi-center randomized clinical trial the treatment effect may be changed over centers. It is thus important to investigate the heterogeneity in treatment effect between centers. For this, uncorrelated random-effect models assuming independence between random-effect terms have been often used, which may be a strong assumption. In this paper we propose a correlated frailty modelling approach of investigating such heterogeneity using the hierarchical-likelihood method when the outcome is time-to-event. In particular, we show how to construct a proper prediction interval for frailty, which explores graphically the potential heterogeneity for a treatment-by-center interaction term. The proposed method is illustrated via numerical studies based on data from the design of a multi-center clinical trial.

Keywords: Frailty Models, hierarchical likelihood, multi-center clinical trial, prediction interval, random effects, randomization, treatment-by-center interaction.

1. Introduction

A multi-center clinical trial with a standardized protocol involves collaborative research efforts between participating centers in tasks of enrolling and following patients (Friedman *et al.*, 1998, p.345). The two main reasons for conducting multi-center trials are to recruit an adequate number of participants within a reasonable time frame and to provide a sound basis for generalizing any observed treatment effect. The latter reason owes much to Bradford Hill's (1962) ideas on consistency of effects. However, participating centers may vary from specialized treatment centers to community hospitals and accordingly there may be differences between the patient populations sampled which may lead to unanticipated treatment-by-center interaction. Alternatively, even when the trial protocol renders the center samples homogeneous, clinical experience and practice may vary between centers, leading, again, to a treatment-by-center interaction (Gray, 1994; Vaida and Xu, 2000). For investigation of this heterogeneity the use of random-effect models, rather than fixed-effect models, is useful (Gray, 1994; Andersen *et al.*, 1999; Legrand *et al.*, 2005; Ha, 2008a).

[†] This research was supported by a grant from Daegu Haany University Ky·lin Foundation in 2009.
[1] Professor, Department of Asset Management, Daegu Haany University, Gyeongsan 712-715, Korea. E-mail: idha@dhu.ac.kr

In this paper we propose a general frailty modelling approach based on hierarchical likelihood (h-likelihood) method (Ha *et al.*, 2007; Ha, 2009; Kim *et al.*, 2011) when the interesting outcome is time-to-event (i.e. survival time). Vaida and Xu (2000) and Legrand *et al.* (2005) have assumed the independence between frailty terms, but in this paper we allow their correlation. For the inference, marginal likelihood approach (Vaida and Xu, 2000) or Bayesian approach (Gray, 1994; Legrand *et al.*, 2005) has been used. However, the marginal likelihood approach often requires intractable computations because of integrating out random effects. The h-likelihood avoids the integration itself and provides a statistically efficient estimation procedure (Lee *et al.*, 2006; Ha *et al*, 2001, 2010). In this paper, we also show that the h-likelihood method gives a proper prediction interval of frailties (random effects), leading to a graphical investigation for heterogeneity of main treatment effect over centers. That is, plots based on these intervals are useful when investigating the heterogeneity of random center and treatment effects. The Statistical Principles for Clinical Trials contained in the international conference of harmonization (ICH, 1998) guidelines also suggest the need to explore (e.g. graphically) potential heterogeneity via a treatment-by-center interaction term.

The paper is organized as follows. In Section 2 we present a general frailty model allowing correlation between two-component random effects. In Section 3 we derive the h-likelihood estimation procedure for fixed parameters and prediction of random effects, and then show how to construct the prediction intervals. Simulation study for the proposed method is given in Section 4. Finally, we discuss our approach in Section 5.

2. Correlated frailty models

Suppose that data consist of survival times of patients collected from q centers. Let T_{ij} ($i = 1, \cdots, q,\ j = 1, \cdots, n_i,\ n = \sum_i n_i$) be the survival time for the jth observation of the ith center (or cluster). For each patient, we observe several covariates, denoted by $x_{ij1}, \cdots, x_{ijp}$. In particular, let x_{ij1} be the binary covariate representing the main treatment arm to which the patient has been randomized with $x_{ij1} = 1$ if the patient is in the standard group and $x_{ij1} = 0$ if the patient is in the experiment group. Denote by v_{i0} and v_{i1} the unobserved random intercepts and slopes (or random effects) for the ith center, respectively.

Following Vaida and Xu (2000) and Legrand *et al.* (2005), frailty models analyzing survival data from for multi-center trials are described as follows. Given log-frailties v_{i0} and v_{i1}, the conditional hazard function of T_{ij} takes the form

$$\lambda_{ij}(t|v_{i0}, v_{i1}) = \lambda_0(t) \exp(\eta_{ij}) \tag{2.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function,

$$\eta_{ij} = v_{i0} + (\beta_1 + v_{i1})x_{ij1} + \sum_{l=2}^{p} \beta_l x_{ijl} \tag{2.2}$$

is the linear predictor for the hazards, and $x_{ij} = (x_{ij1}, \cdots, x_{ijp})^T$ is covariates vectors corresponding to unknown regression parameters $\beta = (\beta_1, \cdots, \beta_p)^T$. The distribution of frailties $v_i = (v_{i0}, v_{i1})^T$ is usually assumed to follow a normal distribution (Vaida and Xu, 2000; Legrand *et al.*, 2005) with mean $E(v_i) = 0$ and var $(v_i) = \Sigma$, which is useful for modelling multi-component frailties (Ha *et al.*, 2007). Here the variance-covariance matrix of v_{i0} and v_{i1} is given by

$$\Sigma = \Sigma(\sigma) = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix},$$

which depends on $\sigma = (\sigma_0^2, \sigma_1^2, \sigma_{01})^T$, a vector of unknown parameters (i.e. frailty parameters). However, Vaida and Xu (2000) and Legrand *et al.* (2005) assumed independence between both frailties (v_{i0} and v_{i1}), leading that $\Sigma = \text{diag}(\sigma_0^2, \sigma_1^2)$ is diagonal with $\rho = 0$; this can be a strong assumption. Our model (2.1) is a general one allowing the correlation, $\rho = \sigma_{01}/(\sigma_0\sigma_1)$, in Σ.

In model (2.1), v_{i0} means the random deviation of the $i$the center from the overall underlying baseline risk. Similarly, v_{i1} represents the random deviation of the i the center from the overall treatment effect. Note here that v_{i0} is called random baseline risk or random center effect and that v_{i1} is called random treatment effect or random treatment-by-center interaction (Legrand *et al.*, 2005).

3. Estimation and prediction

In this section we derive the estimation procedure for correlated models (2.1), and then show how to construct the prediction interval of random effects.

3.1. Derivation of estimation procedure

Let C_{ij} be the censoring time corresponding to survival time T_{ij}. Then we have the following observable random variables,

$$y_{ij} = \min(T_{ij}, C_{ij}) \text{ and } \delta_{ij} = I(T_{ij} \leq C_{ij}),$$

where $I(\cdot)$ is the indicator function. Since the functional form of $\lambda_0(t)$ is unknown, following Breslow (1972) and Ha *et al.*, (2001) we consider the baseline cumulative hazard function $\Lambda_0(t)$ to be a step function with jumps at the r distinct observed death times, $\Lambda_0(t) = \sum_{k:y_{(k)} \leq t} \lambda_{0k}$, where $y_{(k)}$ is the kth ($k = 1, \cdots, r$) smallest distinct death time among the y'_{ij}s, and $\lambda_{0k} = \lambda_0(y_{(k)})$. Following Ha *et al.* (2001), the h-likelihood for the frailty models (2.1) is defined by

$$h = h(\beta, \lambda_0, \sigma) = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i},$$

where

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

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

$$\ell_{2i} = \ell_{2i}(\sigma; v_i) = -\frac{1}{2}\log\,\det(2\pi\Sigma(\sigma)) - \frac{1}{2}\,v_i^T\Sigma(\sigma)^{-1}v_i$$

is the logarithm of the density function for v_i with parameters $\sigma = (\sigma_0^2, \sigma_1^2, \sigma_{01})^T$, $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0r})^T$, $d_{(k)}$ is the number of deaths at $y_{(k)}$ and $R_{(k)} = R(y_{(k)}) = \left\{(i,j) : y_{ij} \geq y_{(k)}\right\}$ is the risk set at $y_{(k)}$.

Notice here that the dimension of λ_0 increases with sample size n. For the estimation of (β, v) with $v = (v_1^T, \ldots, v_q^T)^T$ we use the profile h-likelihood h^* (Ha *et al.*, 2001; Ha and Lee, 2003) with λ_0 eliminated:

$$h^* = h|_{\lambda_0=\widehat{\lambda_0}} = \sum_{ij}\ell_{1ij}^* + \sum_i \ell_{2i},$$

where $\widehat{\lambda}_{0k}$ are solutions of the estimating equations, $\partial h/\partial\lambda_{0k} = 0$, for $k = 1, \cdots, r$. Here, $\Sigma_{ij}\ell_{1ij}^* = \Sigma_{ij}\ell_{1ij}|_{\lambda_0=\widehat{\lambda_0}} = \sum_{ij}\delta_{ij}\eta_{ij} - \sum_k d_k \log\left\{\sum_{(i,j)\in R_{(k)}}\exp(\eta_{ij})\right\}$ does not depend on λ_0. Thus, the estimation of $\tau = (\beta^T, v^T)^T$ given σ is obtained by solving

$$\partial h^*/\partial\tau = (\partial h/\partial\tau)|_{\lambda_0=\widehat{\lambda_0}} = 0. \tag{3.1}$$

Next, the estimation of the frailty parameters σ can be carried out by using the adjusted profile h-likelihood (Lee and Nelder, 2001), defined by

$$p_\tau(h^*) = \left[h^* - \frac{1}{2}\times\log\det\left\{H(h^*,\tau)/(2\pi)\right\}\right]\Bigg|_{\tau=\widehat{\tau}}, \tag{3.2}$$

where $H(h^*,\tau) = -\partial^2 h^*/\partial\tau^2$ is given in (3.3) and $\widehat{\tau}$ solves $\partial h^*/\partial\tau = 0$ in (3.1). The resulting estimators by maximizing (3.2) give extended restricted maximum likelihood (REML) ones (Ha *et al.*, 2001; Lee *et al.*, 2006).

3.2. Prediction of random effects and their intervals

Lee and Ha (2010) proposed how to construct a proper prediction interval based on h-likelihood for Poisson random-effect models. Thus we show that such interval can be extended to correlated frailty models (2.1). Following Ha *et al.* (2001) and Ha and Lee (2003), the asymptotic covariance matrix of $\widehat{\beta}$ and $\widehat{v} - v$ is the inverse of Hessian matrix, $H = H(h^*,\tau) = -\partial^2 h^*/\partial\tau^2$, without nuisance parameters λ_0, given by

$$H = -\begin{pmatrix}\partial^2 h^*/\partial\beta^2 & \partial^2 h^*/\partial\beta\partial v\\ \partial^2 h^*/\partial v\partial\beta & \partial^2 h^*/\partial v^2\end{pmatrix} = \begin{pmatrix}X^TW^*X & X^TW^*Z\\ Z^TW^*X & Z^TW^*Z + R\end{pmatrix} \tag{3.3}$$

where X is the $n\times p$ matrix whose ith row vector is x_{ij}^T, Z is the $n\times q^*$ group indicator matrix whose ith row vector is z_{ij}^T, W^* is the $n\times n$ symmetric matrix given in Appendix 2

of Ha and Lee (2003) and $R = \text{BD}\left\{-\partial^2 \ell_{2i}/\partial v_i^2\right\}$ is the $q^* \times q^*$ block diagonal matrix. Here $q^* = 2q$.

Let v be random effects. The bottom right-hand corner of H^{-1} in (3.3) gives a variance of $\widehat{v} - v$, given by

$$\text{var}(\widehat{v} - v) = \left\{(Z^T W^* Z + R) - (Z^T W^* X)(X^T W^* X)^{-1}(X^T W^* Z)\right\}^{-1}. \tag{3.4}$$

Following Lee and Ha (2010), we can show that the variance (3.4) also becomes an approximation of conditional mean-square error of prediction (CMSEP) by Booth and Hobert (1998). Thus we can construct the prediction interval for random effects as follows. The 95% h-likelihood prediction intervals for random effects $v_{ik}(i = 1, \cdots, q; k = 0, 1)$ under asymptotic normality of the estimators, are given by

$$\widehat{v}_{ik} \pm 1.96 \times \text{SE}(\widehat{v}_{ik} - v_{ik}), \tag{3.5}$$

where $\widehat{v}_{ik}$ is obtained from (3.1) and $\text{SE}(\widehat{v}_{ik} - v_{ik}) = \sqrt{\text{var}(\widehat{v}_{ik} - v_{ik})}$ is the estimated standard error obtained from (3.4). Note that $\widehat{v} \approx E(v|y, \delta)$ (Lee and Nelder, 1996; Vaida and Xu, 2000). For the prediction of random effects, Vaida and Xu (2000) used empirical Bayes (EB) method based on conditional posterior distribution of random effects v given (y, δ), leading to

$$\text{var}(\widehat{v} - v) \approx (-\partial h^2/\partial v^2)^{-1} = (Z^T W^* Z + U)^{-1}. \tag{3.6}$$

Thus, the corresponding 95% EB prediction intervals are constructed from (3.5) with (3.6). However, the EB method can underestimate the variance of $\widehat{v} - v$ because it ignores covariance term between $\widehat{v} - v$ and $\widehat{\beta}$ in (3.3), leading to a misleading prediction interval (Ha, 2008b; Lee and Ha, 2010).

4. Simulation study

We conducted numerical studies, based upon 500 replications of simulated data, in order to evaluate the performance of the proposed method. For comparison of prediction of random effect, we include the EB method (Vaida and Xu, 2000).

Following the setups of Vaida and Xu's (2000) data analysis for a multi-center clinical trial, we generate data from the correlated frailty model (2.1) with two covariates in (2.2):

$$\lambda_{ij}(t|v_{i0}, v_{i1}) = \lambda_0(t)\exp(\eta_{ij}) \text{ with } \eta_{ij} = v_{i0} + (\beta_1 + v_{i1})x_{ij1} + \beta_2 x_{ij2}. \tag{4.1}$$

Here we assume $\lambda_0(t) = 1$, $\beta_1 = -0.5$, $\beta_2 = 0.5$, and $\sigma_0^2 = \sigma_1^2 = 1$ with $\rho = -0.5$ (i.e. $\sigma_{01} = -0.5$). Though not reported here, we found the similar results for $\rho = 0.5$. The binary covariates x_{ij1} and x_{ij2} are each generated from a Bernoulli distribution with success probability 0.5. In Vaida and Xu's (2000) data the number of centers is 31 and the average number of patients per center is 19. Thus, we consider the following sample sizes: $n = \sum_{i=1}^{q} n_i = 150, 600, 1200$ with $(q, n_i) = (30,\ 5), (30,\ 20), (60,\ 20)$. The censoring times are generated from an exponential distribution with parameter empirically determined to achieve approximately the right censoring rate, around 15%. For the model fitting and computation we used SAS/IML.

The simulation results for h-likelihood estimates of frailty parameters $(\sigma_0^2, \sigma_1^2, \rho)$ are summarized in Table 4.1. Overall, the estimates perform well. The bias and variations (i.e. SD and MSE) are decreased as sample size n increases. Even though unreported, the estimates for fixed effects (β_1, β_2) also work well in all cases considered.

Table 4.1 Simulations results for h-likelihood estimators of frailty parameters

n	$\widehat{\sigma}_0^2$			$\widehat{\sigma}_1^2$			$\widehat{\sigma}_{01}$		
	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
150	0.027	0.525	0.276	0.079	0.766	0.593	-0.009	0.461	0.213
600	0.011	0.302	0.091	0.027	0.342	0.118	-0.008	0.247	0.061
1200	0.010	0.201	0.041	0.006	0.209	0.044	0.003	0.167	0.028

Note: The simulation is conducted with 500 replications for the correlated frailty models parassuming the true frailty parameters $(\sigma_0^2, \sigma_1^2, \sigma_{01})$=$(1, 1, -0.5)$, with three sample sizes n=$\sum_{i=1}^{q} n_i$=$150, 600, 1200$ with (q, n_i)=$(30,\ 5), (30,\ 20), (60,\ 20)$. SD and MSE, standard deviation and mean squared error of estimates over 500 simulations.

Figure 4.1 shows the coverage probabilities of the nominal 95% prediction intervals for all random effects (v'_{i1}s) in model (4.1). The h-likelihood (HL) intervals are overall better than the EB intervals in terms of maintaining the nominal level. In particular, the HL intervals preserve well the nominal level, except for a sample size, $n = 150$ with $(q, n_i) = (30, 5)$, with a small center size compared to the number of centers, wheras the EB intervals are not. The results about v'_{i0}s are similar to those of v'_{i1}s (not shown).

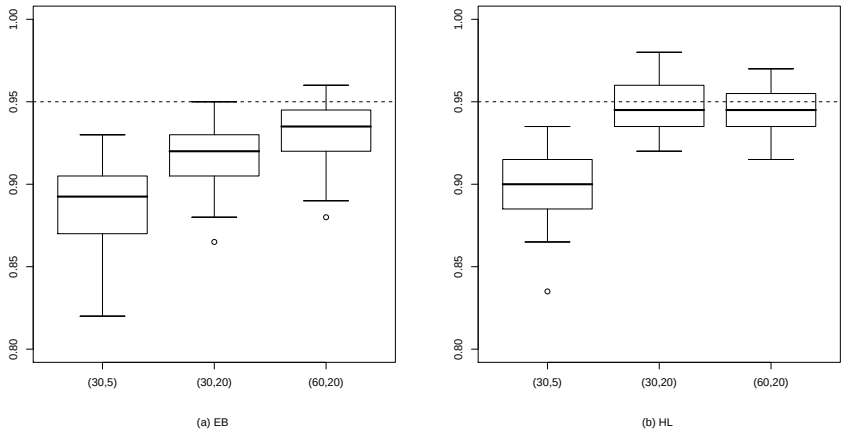

Figure 4.1 Simulation results for coverage probabilities of the nominal 95% (dotted line) (a) EB and (b) HL (h-likelihood) prediction intervals of all random effects (v'_{i1}'s) in correlated frailty models; (30, 5), (30, 20) and (60, 20) in x-label indicate the corresponding sample size (q, n_i); q is the number of centers and n_i is the center size.

Furthermore, we also conducted a numerical study in order to demonstrate how to investigate graphically the heterogeneity of random effects (i.e. random baseline risk and random treatment effects) over centers. For this, we use a simulated data set which is generated from model (4.1) under a sample size n=600 with center size n_i=20 for all center i and a smaller variance $(\sigma_0^2 = \sigma_1^2 = 0.1)$ or a larger variance $(\sigma_0^2 = \sigma_1^2 = 1)$. Notice that the size of

variance may be proportional to that of heterogeneity across centers. We thus considered the four combinations, (σ_0^2, σ_1^2) = (0.1, 0.1), (0.1, 1), (1, 0.1) and (1, 1). The remaining simulation schemes are the same as before. To save space, we report the results of only a case, (σ_0^2, σ_1^2) = (0.1, 1). As expected, Figure 4.2 shows that the random baseline risk (v'_{i0}s) are homogeneous over centers, whereas the random treatment effect (v'_{i1}s) are substantially heterogeneous across centers. In particular, all intervals of v_{i0} in Figure 2 (a) include zero, indicating to homogeneity over centers. However, many intervals (e.g. center number 1, 3 and 4) of v_{i1} in Figure 2 (b) do not include zero, indicating that the random treatment effect is changed over centers, that is, there is the random treatment-by-center interaction in this data set. The results suggest that care is necessary in assessing the main treatment effect (β_1).

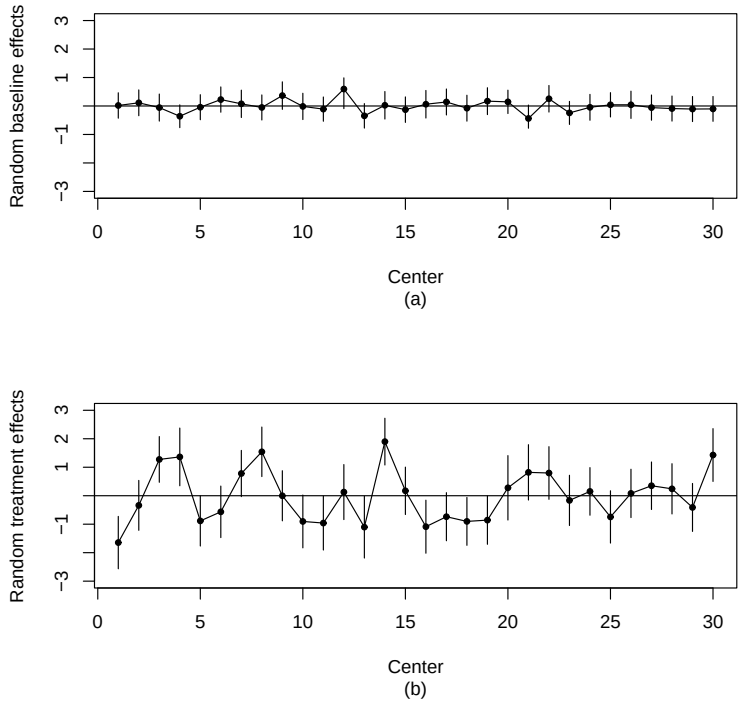

Figure 4.2 Random effects and their 95% prediction intervals using h-likelihood, based on a simulated data set generated from a correlated frailty model assuming (σ_0^2, σ_1^2, ρ) = (0.1, 1, -0.5) and (q, n_i)= (30, 20). (a) random baseline risk (v_0); (b) random treatment effect (v_1).

5. Discussion

We have shown that h-likelihood gives a unified framework for both estimation of parameters and prediction of random effects for correlated frailty models. In particular, we demonstrated via a numerical study that the proposed method is useful in investigating graphically the heterogeneity of treatment effect in multi-center trial. However, some prac-

tical study (e.g. Vaida and Xu, 2000; Legrand *et al.*, 2005) using a real dataset from a multi-center clinical trialis is required as a further study.

The simulation results in Section 4 show that our procedure overall performs well. However, another further study is necessary in developing a proper prediction interval in multi-center trial with a smaller center size compared to the number of centers.

References

Andersen, P. K., Klein, J. P. and Zhang M.-J. (1999). Testing for centre effects in multi-centre survival studies: A monte carlo comparison of fixed and random effects tests. *Statistics in Medicine*, **18**, 1489-1500.

Booth, J. G. and Hobert, J. P. (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*, **93**, 262-272.

Breslow, N. E. (1972). Discussion of Professor Cox's paper. *Journal of the Royal Statistical Society B,* **34**, 216-217.

Friedman, L. M., Furberg, C. D. and DeMets, D. L. (1998). *Fundamentals of clinical trials,* Springer, New York.

Gray, R. J. (1994). A Bayesian analysis of institutional effects in multicenter cancer clinical trial. *Biometrics*, **5**, 244-253.

Ha, I. D. (2008a). A HGLM framework for meta-analysis of clinical trials with binary outcomes. *Journal of the Korean Data & Information Science Society*, **19**, 1429-1440.

Ha, I. D. (2008b). On estimation of random effect in Poisson HGLMs. *Journal of the Korean Data & Information Science Society*, **19**, 375-383.

Ha, I. D. (2009). A correction of SE from penalized partial likelihood in frailty models. *Journal of the Korean Data & Information Science Society*, **20**, 895-903.

Ha, I. D. and Lee, Y. (2003). Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational Graphical Statistics,* **12**, 663-681.

Ha, I. D., Lee, Y and MacKenzie, G. (2007) Model selection for multi-component frailty models. *Statistics in Medicine*, **26**, 4790-4807.

Ha, I. D., Lee, Y. and Song, J.-K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika,* **88**, 233-243.

Ha, I. D., Noh, M and Lee, Y. (2010). Bias reduction of likelihood estimators in semiparametric frailty models. *Scandinavian Journal of Statistics*, **37**, 307-320.

Hill, A. B. (1962). *Statistical methods in clinical and preventive medicine,* Edinburgh, Livingstone.

Kim, H. K., Noh, M. and Ha, I. D. (2011). A study using HGLM on regional difference of the dead due to injuries. *Journal of the Korean Data & Information Science Society*, **22**, 137-148.

Lee, Y and Ha, I. D. (2010). Orthodox BLUP versus h-likelihood methods for inferences about random effects in Tweedie mixed models. *Statistics and Computing*, **20**, 295-303.

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 and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions, *Biometrika,* **88**, 987-1006.

Lee, Y. and Nelder, J. A. and Pawitan, Y. (2006). *Generalised linear models with random effects: Unified analysis via h-likelihood,* London, Chapman and Hall.

Legrand, C., Ducrocq, V., Janssen, P., Sylvester, R. and Duchateau, L. (2005). A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. *Statistics in Medicine*, **24**, 3789-3804.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine*, **19**, 3309-3324.