Published in final edited form as:

Stat Med. 2016 January 30; 35(2): 251–267. doi:10.1002/sim.6628.

Hierarchical Likelihood Inference on Clustered Competing Risks Data

Nicholas J. Christian¹, Il Do Ha², and Jong-Hyeon Jeong^{3,*}

¹Departments of Epidemiology, University of Pittsburgh, Pittsburgh, PA 15261, U.S.A

²Department of Statistics, Pukyong National University, Busan, 609-737, South Korea

³Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, U.S.A

Abstract

The frailty model, an extension of the proportional hazards model, is often used to model clustered survival data. However, some extension of the ordinary frailty model is required when there exist competing risks within a cluster. Under competing risks, the underlying processes affecting the events of interest and competing events could be different but correlated. In this paper, the hierarchical likelihood method is proposed to infer the cause-specific hazard frailty model for clustered competing risks data. The hierarchical likelihood (h-likelihood) incorporates fixed effects as well as random effects into an extended likelihood function, so that the method does not require intensive numerical methods to find the marginal distribution. Simulation studies are performed to assess the behavior of the estimators for the regression coefficients and the correlation structure among the bivariate frailty distribution for competing events. The proposed method is illustrated with a breast cancer dataset.

Keywords

Cause-specific hazard; Clustered data; Competing risks; Frailty models; Hierarchical likelihood

1 Introduction

Modern medical practice guidelines emphasize the need for individualized or targeted therapy by evaluating both potential harm and benefit of a treatment at the population and individual levels. The frailty model [1], an extension of the proportional hazards model [2] for clustered survival data, is often used to account for dependent event times in a cluster by including a latent random effect. Efficient inference procedures on the random effects in survival data might be useful in guiding the medical practice at an individual level.

The shared frailty model [1] assumes that within a cluster, censoring times are independent of event times, but this assumption is no longer reasonable under competing risks [3] where factors that affect one type of events may also influence the probability of other types of events. Huang and Wolfe [4] proposed a model that extended the ordinary frailty model to

* jjeong@pitt.edu.

competing risks. The model was fitted using the EM algorithm, along with Markov Chain Monte Carlo (MCMC) simulation. Liu and Huang [5] proposed a parametric method using the Gaussian quadrature to fit the competing risks frailty model proposed by Huang and Wolfe [4]. Another approach to handling clustered competing risks data is given by Katsahian *et al.* [6] and Katsahian and Boudreau [7] who extends the subdistribution hazard regression model [8] to clustered data by including a frailty term. Zhou *et al.* [9] proposed an alternative procedure for modeling the subdistribution hazard by stratification. Gorfine and Hsu [10] considered a flexible cause-specific hazard frailty model, which is an extension of the bivariate frailty model [11] to the competing risks setting. They used the EM algorithm for inference on the parameters from the model under parametric and nonparametric settings. Zhou *et al.* [27] also proposed a marginal subdistribution hazards model where clustering is handled via a robust/sandwich variance estimator.

The hierarchical likelihood (h-likelihood) [12] incorporates fixed effects as well as random effects into an extended likelihood function. As a result, the method does not require intensive numerical methods to find the marginal distribution like the EM-algorithm. Instead parameters are estimated by using a method similar to the Newton-Raphson to maximize the adjusted profile h-likelihoods assuming the nonparametric baseline hazard. Since the frailties are not integrated out, this approach also allows for direct inference on the random effects. Ha *et al.* [13] and Ha and Lee [14] extended the h-likelihood approaches to the shared frailty models with gamma and log-normal frailties for parametric and non-parametric baseline hazard functions. Recently, the h-likelihood has been used to estimate the subdistribution hazard model with multivariate frailties [15]. In this paper, we propose a joint inference procedure based on the h-likelihood for the regression coefficients, the frailties for clusters, the frailties for event types, and their correlation structure under the cause-specific hazard frailty model. To the best of our knowledge, this extension to the competing risks data has not been considered previously in the literature.

This paper is organized as follows: Section 2 presents the cause-specific hazard frailty model. Section 3 formulates the h-likelihood for clustered competing risks data. Sections 4 through 6 propose an estimation procedure for the regression coefficients, frailties, and the variance components of the assumed frailty distribution. Sections 7 and 8 present the simulation results and application to a real dataset, respectively. Finally Section 9 discusses the results and future research areas.

2 Model and likelihood function

Suppose there are $k = 1, 2, \dots, m$ event types and assume $V_i = (V_{i1}, V_{i2}, \dots, V_{im})^T$, where V_{ik} is the random effect for event k in cluster i . A natural choice for a distribution of V_i would be the multivariate normal distribution with mean 0 and $m \times m$ covariance matrix Σ . Then the cause-specific hazard function conditional on the log-frailty $V_{ik} = v_{ik}$ for the j th observation in cluster i who failed from cause k (or type k) is

$$\lambda_{ijk}(t|v_i) = \lambda_{0k}(t) \exp(X_{ij}^T \beta_k + v_{ik}), \quad (1)$$

where $\lambda_{0k}(t)$ is the baseline hazard function for event type k , $\beta_k = (\beta_{k1}, \beta_{k2}, \dots, \beta_{kp})^T$ is a $p \times 1$ vector of regression coefficients for event type k , and $X_{ij} = (X_{ij1}, X_{ij2}, \dots, X_{ijp})^T$ is a $p \times 1$ vector of fixed covariates corresponding to β_k . Let $\beta = (\beta_1, \beta_2, \dots, \beta_m)$ be a $mp \times 1$ vector of all the regression coefficients for all event types. Similarly let $\lambda_0 = (\lambda_{01}, \lambda_{02}, \dots, \lambda_{0m})$ denote the collection of all cause-specific baseline hazard functions.

Model (1) provides a good deal of flexibility in that the frailty effects can vary over different types of events within a cluster. The model also allows for a negative association within a cluster as well as a positive one [11]. The case of negative association can arise in practice as reducing the risk of dying from cancer can increase the risk of dying from some other disease.

Model (1) has an analogy to one proposed by Huang and Wolfe [4] in that it can account for a correlation between failure and informative censoring, where failure can be an event type of main interest (type I) and informative censoring may be caused by competing events (type II). Specifically, event times from cause 1 would follow a cause-specific proportional hazards model

$$\lambda_{ij1}(t|v_i) = \lambda_{01}(t) \exp(X_{ij}^T \beta_1 + v_{i1}),$$

and event times from cause 2 would follow similarly a model

$$\lambda_{ij2}(t|v_i) = \lambda_{02}(t) \exp(X_{ij}^T \beta_2 + v_{i2}),$$

where v_{i1} and v_{i2} might be correlated. In the traditional cause-specific analysis, patients who failed from cause 2 are treated as censored for the analysis of type I events, which ignores a potential correlation between v_{i1} and v_{i2} ; see also Gorfine and Hsu [10].

Since there is an association between the random effects, it is not possible to model each event type separately as is normally done when modeling the cause-specific hazard rates without clustering. Instead, the covariate effects for all event types as well as all of the random effects need to be estimated jointly. For the j th observation in the i th cluster, let T_{ijk} , $k = 1, \dots, m$, denote the time to event for event type k and let C_{ij} denote the independent censoring time. Then the observed event time is $T_{ij} = \min(T_{ij1}, T_{ij2}, \dots, T_{ijm}, C_{ij})$ and define the event indicator $\delta_{ijk} = 1$ if $T_{ij} = T_{ijk}$ and 0 otherwise. The conditional likelihood for cluster i given the frailty v_{ik} is given by

$$L_i(\beta, \lambda_0 | v_{ik}) = \prod_{k=1}^m \prod_{j=1}^{n_i} \left[\lambda_{0k}(t_{ij}) e^{X_{ij}^T \beta_k + v_{ik}} \right]^{\delta_{ijk}} \exp \left(-\Lambda_{0k}(t_{ij}) e^{X_{ij}^T \beta_k + v_{ik}} \right), \quad (2)$$

where $\Lambda_{0k}(\cdot)$ is the baseline cumulative hazard function for cause k .

3 Hierarchical likelihood for clustered competing risks

To simplify the notation, we assume there are just two event types $k = 1, 2$. The results can be easily generalized to m event types. First, let us define the h-likelihood function. Suppose there are $i = 1, 2, \dots, n$ clusters where each cluster has $j = 1, 2, \dots, n_i$ observations, so that

the total sample size is $N = \sum_{i=1}^n n_i$. The two indices i and j denote a unique observation from the overall sample of size N . Under competing risks, the contribution of the j th observation in the i th cluster to the h-likelihood for event k is given by the logarithm of the joint density function of $(T_{ij}, \delta_{ijk}, v_{ik})$ written as a function of the parameters for event type k , $(\beta_k, \lambda_{0k}, \theta)$,

$$h_{ijk} \equiv h_{ijk}(\beta_k, \lambda_{0k}, \theta; t_{ij}, \delta_{ijk}, v_{ik}) = \log [f_{ijk}(t_{ij}, \delta_{ijk}; \beta_k, \lambda_{0k} | v_{ik}) f_i(v_i; \theta)], \quad (3)$$

where f_{ijk} is the conditional density function of T_{ij} and δ_{ijk} given $V_{ik} = v_{ik}$ with parameters (β_k, λ_{0k}) and f_i is the density function of $V_i = (V_{i1}, V_{i2})^T$ with a parameter vector θ . The conditional density function of (T_{ij}, δ_{ijk}) given $V_{ik} = v_{ik}$ is

$$f_{ijk}(t_{ij}, \delta_{ijk}; \beta_k, \lambda_{0k} | v_{ik}) = \left[\lambda_{0k}(t_{ij}) e^{X_{ij}^T \beta_k + v_{ik}} \right]^{\delta_{ijk}} \exp \left(-\Lambda_{0k}(t_{ij}) e^{X_{ij}^T \beta_k + v_{ik}} \right). \quad (4)$$

Assuming that $V_i = (V_{i1}, V_{i2})^T$ follows a bivariate normal distribution with mean 0 and covariance matrix Σ , the probability density function is given by

$$f_i(v_i) = \frac{1}{(2\pi)^{1/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} v_i^T \Sigma^{-1} v_i \right). \quad (5)$$

Let $\theta = (\sigma_{11}, \sigma_{22}, \sigma_{12})^T$ be the dispersion-parameter vector for $f_i(v_i)$ that contains the variance components of Σ .

Notationally it will be helpful to distinguish between random effects associated with a cluster and random effects for a particular event type. So, let $V_i = (V_{i1}, V_{i2})^T$ denote the random effects vector for both event types for cluster i and let $\tilde{V}_k = (V_{1k}, V_{2k}, \dots, V_{nk})^T$ denote an n -dimensional vector of the random effects from all clusters just for event k , V_{ik} being the random effect for event k in cluster i . Let $V = (V_{11}, V_{21}, \dots, V_{n1}, V_{12}, V_{22}, \dots, V_{n2})^T$ be a $2n$ -dimensional vector of all random effects, for all clusters and event types. Notice that the random effects are arranged by event type so that all of the random effects for the same event type are adjacent. The structure of V is important later in this section because it determines how the observed information matrices are structured. Similarly, let $\beta = (\beta_1, \beta_2)^T$ be a vector of regression coefficients for both event types and let $\lambda_0 = (\lambda_{01}, \lambda_{02})$ be a collection of all the baseline hazards.

Since event times within a cluster are conditionally independent given the frailty $V_i = v_i$ and the frailties V_i are independent and identically distributed random variables, the h-likelihood for the cause-specific hazard frailty model is,

$$h(\beta, \lambda_0, v, \theta) = \sum_{ijk} h_{ijk} = \sum_{ijk} l_{ijk}(\beta_k, \lambda_{0k}; t_{ij}, \delta_{ijk} | v_{ik}) + \sum_i l_i(\theta; v_i), \quad (6)$$

where

$$l_{ijk}(\beta_k, \lambda_{0k}; t_{ij}, \delta_{ijk} | v_{ik}) = \delta_{ijk} \left(\log \lambda_{0k}(t_{ij}) + x_{ij}^T \beta_k + v_{ik} \right) - \Lambda_{0k}(t_{ij}) \exp \left(x_{ij}^T \beta_k + v_{ik} \right)$$

is the log of the conditional density function for T_{ij} and δ_{ijk} given $V_{ik} = v_{ik}$ and

$$l_i(\theta; v_i) = -\log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} v_i^T \Sigma^{-1} v_i$$

is the log of the bivariate normal density function for V_i with parameter θ .

It will be nonparametrically assumed that the cumulative baseline hazard function for event type k is a step function with jumps at the observed event times,

$$\Lambda_{0k}(t) = \sum_{r: t_{(kr)} \leq t} \lambda_{0kr} \quad (7)$$

where $t_{(k1)} < t_{(k2)} < \dots < t_{(kD_k)}$ denote the D_k ordered unique event times for type k events among all of the t_{ij} 's that refer to the time of a type k event and $\lambda_{0kr} = \lambda_{0k}(t_{(kr)})$. Also let d_{kr} be the number of events that occur at time $t_{(kr)}$.

Let $Z_{ij} = (Z_{ij1}, Z_{ij2}, \dots, Z_{ijD_k})^T$ be a $n \times 1$ cluster indicator vector where $Z_{ijq} = 1$ if $i = q$ and 0 otherwise and let Z be a $N \times n$ matrix whose ij row is Z_{ij}^T , which leads to $v_{ik} = Z_{ij}^T v_k$ for any j . By replacing v_{ik} with $Z_{ij}^T v_k$ in (6) and summing over the unique event times for each event type, the h-likelihood (6) can be rewritten as

$$h = \sum_{k=1}^2 \left[\sum_{r=1}^{D_k} d_{kr} \log \lambda_{0kr} + S_{X_{kr}}^T \beta_k + S_{Z_{kr}}^T v_k - \lambda_{0kr} \sum_{ij \in \mathcal{D}_{kr}} \exp \left(X_{ij}^T \beta_k + Z_{ij}^T v_k \right) \right] + \sum_{i=1}^n l_i(\theta; v_i), \quad (8)$$

where $S_{X_{kr}}^T = \sum_{ij \in \mathcal{D}_{kr}} X_{ij}^T$ and $S_{Z_{kr}}^T = \sum_{ij \in \mathcal{D}_{kr}} Z_{ij}^T$ are the sums of the vectors X_{ij}^T and Z_{ij}^T over the set $\mathcal{D}_{kr} = \{ij : \delta_{ijk} = 1 \text{ and } t_{ij} = t_{(kr)}\}$ of all individuals who have a type k event at

time $t_{(kr)}$ and $\mathcal{R}_{kr} = \{ij : t_{ij} \leq t_{(kr)}\}$ is the risk set at time $t_{(kr)}$, i.e. the set of all individuals who are still at risk to experience an event.

Using an approach similar to Johansen [16], by fixing β_k , v_k and θ and maximizing (8) as a function of λ_{0kr} gives the nonparametric maximum hierarchical likelihood estimator of λ_{0kr} ,

$$\hat{\lambda}_{0kr} = \frac{d_{kr}}{\sum_{ij \in \mathcal{R}_{kr}} \exp(X_{ij}^T \beta_k + Z_{ij}^T v_k)}. \quad (9)$$

Thus, $\hat{\Lambda}_{0k}(t) = \sum_{r: t_{(kr)} \leq t} \hat{\lambda}_{0kr}$ is an extension of the Breslow estimator [17] of the baseline cumulative hazard function. Replacing λ_{0kr} with $\hat{\lambda}_{0kr}$ in (8) gives the profile h-likelihood as a function of β , v , and θ only,

$$h_p(\beta, v, \theta) = \sum_{kr} \left\{ S_{Xkr}^T \beta_k + S_{Zkr}^T v_k - d_{kr} \log \left(\sum_{ij \in \mathcal{R}_{kr}} \exp(X_{ij}^T \beta_k + Z_{ij}^T v_k) \right) \right\} + \sum_{i=1}^n l_i(\theta; v_i), \quad (10)$$

omitting a constant term of $\sum_{kr} d_{kr} \{\log d_{kr} - 1\}$.

4 Estimation of regression coefficients and frailties

Because the profile likelihood function (h_p) involves the variance components θ of the bivariate frailty distribution, first we will find the maximum hierarchical likelihood estimators (MHLE) of β and v by maximizing the profile h-likelihood for a *fixed* θ using a method similar to the Newton-Raphson. Starting at initial values $\hat{\beta}^{(0)}$ and $\hat{v}^{(0)}$, the approximate maximums are updated iteratively until convergence is achieved by,

$$\begin{pmatrix} \hat{\beta}^{(i+1)} \\ \hat{v}^{(i+1)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}^{(i)} \\ \hat{v}^{(i)} \end{pmatrix} + \left[\begin{pmatrix} \frac{-\partial^2 h_p}{\partial \beta^2} & \frac{-\partial^2 h_p}{\partial \beta \partial v} \\ \frac{-\partial^2 h_p}{\partial v \partial \beta} & \frac{-\partial^2 h_p}{\partial v^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial h_p}{\partial \beta} \\ \frac{\partial h_p}{\partial v} \end{pmatrix} \right] \bigg|_{(\beta, v) = (\hat{\beta}^{(i)}, \hat{v}^{(i)})} \quad (11)$$

where $\hat{\beta}^{(i)}$ and $\hat{v}^{(i)}$ represent the estimates of β and v at the i th iteration. Reasonable starting values for $\hat{\beta}^{(0)} = (\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)})^T$ are the estimates from the cause-specific hazard model for each event type with no random effects. A possible starting value for $\hat{v}^{(0)}$ is a random sample of size n from the bivariate normal distribution (5) with the initial values of the variance components, $\hat{\theta}^{(0)}$.

Now the elements of the gradient vector $(h_p/\beta, h_p/v)^T$ can be calculated as $h_p/\beta = (h_p/\beta_1, h_p/\beta_2)^T$, where

$$\frac{\partial h_p}{\partial \beta_k} = \sum_{ij} \left\{ X_{ij} \delta_{ijk} - X_{ij} \hat{\Lambda}_{0k}(t_{ij}) \exp \left(X_{ij}^T \beta_k + Z_{ij}^T v_k \right) \right\}, \quad (12)$$

and $h_p/v = (h_p/v_1, h_p/v_2)^T$ where,

$$\frac{\partial h_p}{\partial v_k} = \sum_{ij} \left\{ Z_{ij} \delta_{ijk} - Z_{ij} \hat{\Lambda}_{0k}(t_{ij}) \exp \left(X_{ij}^T \beta_k + Z_{ij}^T v_k \right) \right\} - \sum_{i=1}^n v_i \bullet (\sigma^{kk}, \sigma^{12}), \quad (13)$$

the $\bullet$ denoting the inner product of two vectors. Here σ^{kk} and σ^{12} are elements of the precision matrix Σ^{-1} .

The following matrix notations are used for the remainder of this section. Let $R_k = (R_1, R_2, \dots, R_{D_k})$ be a $N \times D_k$ at risk indicator matrix where the ij^{th} element in column r is one if $t_{ij} \leq t_{kr}$ and zero otherwise. Define E_k as a $N \times 1$ type k event indicator vector with ij^{th} element δ_{ijk} . Let M_k be a $N \times N$ diagonal matrix with elements $\hat{\Lambda}_{0k}(t_{ij}) \exp(X_{ij}^T \beta_k + Z_{ij}^T v_k)$, N_k be a $N \times N$ diagonal matrix with elements $\exp(X_{ij}^T \beta_k + Z_{ij}^T v_k)$, and C_k be a diagonal $D_k \times D_k$ matrix where the r th element is $\hat{\lambda}_{0kr}^2 / d_{kr}$. Finally, let I_n be a $n \times n$ identity matrix and let $\otimes$ denote the Kronecker product. Recall that X is a $N \times p$ matrix of p covariates and Z is a $N \times n$ cluster indicator matrix.

Using these notations, equation (12) can be expressed as

$$\frac{\partial h_p}{\partial \beta_k} = X^T (E_k - M_k), \quad (14)$$

and the derivative of h_p with respect to all random effects v is

$$\frac{\partial h_p}{\partial v} = \begin{pmatrix} Z^T (E_1 - M_1) \\ Z^T (E_2 - M_2) \end{pmatrix} - (\Sigma^{-1} \otimes I_n) v. \quad (15)$$

Now to obtain the observed information matrix H of β and v for fixed θ , let us define $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$ as block diagonal matrices such that,

$$\mathbf{X} = \begin{pmatrix} X & \mathbf{0} \\ \mathbf{0} & X \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} Z & \mathbf{0} \\ \mathbf{0} & Z \end{pmatrix} \quad \text{and} \quad \mathbf{W} = \begin{pmatrix} W_1 & \mathbf{0} \\ \mathbf{0} & W_2 \end{pmatrix}, \quad (16)$$

where $\mathbf{0}$ is a conformable matrix of zeros and $W_k = W_k(\beta_k, v_k) = M_k - N_k R_k C_k (R_k N_k)^T$ for $k = 1, 2$. Then the observed information matrix H is a $(mp + mn) \times (mp + mn)$ matrix

$$H = H(\beta, v, \theta) = \begin{pmatrix} \frac{-\partial^2 h_p}{\partial \beta^2} & \frac{-\partial^2 h_p}{\partial \beta \partial v} \\ \frac{-\partial^2 h_p}{\partial v \partial \beta} & \frac{-\partial^2 h_p}{\partial v^2} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W} \mathbf{X} & \mathbf{Z}^T \mathbf{W} \mathbf{Z} + Q \end{pmatrix}, \quad (17)$$

where Q is a $n \times n$ matrix that is the negative second derivative of the log of the joint density function for all random effects with respect to the vector v , i.e.

$$Q(v, \theta) = -\frac{\partial^2}{\partial v^2} \sum_{i=1}^n l_i(\theta; v_i) = \Sigma^{-1} \otimes I_n. \quad (18)$$

5 Estimation of variance components

In the previous section, we evaluated the gradient vector and the observed information matrix for β and v given the variance components θ for the profile h-likelihood (10). The next step is to find the MHLE of θ by maximizing the following adjusted profile h-likelihood [12–14]:

$$h_A(\theta) = \left[h_p - \frac{1}{2} \log(\det(H/2\pi)) \right] \Big|_{(\beta, v) = (\hat{\beta}(\theta), \hat{v}(\theta))} \quad (19)$$

given $\hat{\beta} = \hat{\beta}(\theta)$ and $\hat{v} = \hat{v}(\theta)$, the “current” estimates of β and v conditional on θ . The adjusted profile h-likelihood is used to approximate the restricted likelihood of θ that takes into account the estimation of β and v as well as ensures unbiased estimation of the variance components. A Newton-Raphson-type method is also used to find the MHLE of θ , denoted as $\hat{\theta}$. This requires finding the first and second derivatives of h_A with respect to every variance component of $\theta = (\sigma_{11}, \sigma_{22}, \sigma_{12})^T$.

Using the properties of determinants, equation (19) can be rewritten as

$$h_A = \hat{h}_p - \frac{1}{2} \log(\det(\hat{H})) + \frac{(p+n)}{2} \log(2\pi), \quad (20)$$

where $\hat{h}_p = h_p(\hat{\beta}(\theta), \hat{v}(\theta), \theta)$ and $\hat{H} = H(\hat{\beta}(\theta), \hat{v}(\theta), \theta)$ are the profile h-likelihood and observed information matrix evaluated at the current estimates of β and v , respectively. Then, from the matrix calculus, the q th component of the gradient vector h_A/θ is

$$\frac{\partial h_A}{\partial \theta_q} = \frac{\partial \hat{h}_p}{\partial \theta_q} - \frac{1}{2} \text{trace} \left(\hat{H}^{-1} \frac{\partial \hat{H}}{\partial \theta_q} \right). \quad (21)$$

Furthermore, the element in row q and column s of the 3×3 observed information matrix $^2 h_A/\theta^2$ for the frailty parameter θ is

$$-\frac{\partial^2 h_A}{\partial \theta_q \partial \theta_s} = -\frac{\partial^2 \hat{h}_p}{\partial \theta_q \partial \theta_s} + \frac{1}{2} \text{trace} \left(-\hat{H}^{-1} \frac{\partial \hat{H}}{\partial \theta_q} \hat{H}^{-1} \frac{\partial \hat{H}}{\partial \theta_s} + \hat{H}^{-1} \frac{\partial^2 \hat{H}}{\partial \theta_q \partial \theta_s} \right). \quad (22)$$

Since $\hat{\beta}(\theta)$ and $\hat{v}(\theta)$ are functions of θ , it is not appropriate to use only the partial derivatives in (21) and (22). Instead the total derivative should be used. The total derivative of h_A with respect to θ_q is,

$$\frac{\partial h_A}{\partial \theta_q} = \frac{\partial h_A}{\partial \theta_q} + \left(\frac{\partial h_A}{\partial \beta} \Big|_{\beta=\hat{\beta}} \right) \frac{\partial \hat{\beta}}{\partial \theta_q} + \left(\frac{\partial h_A}{\partial v} \Big|_{v=\hat{v}} \right) \frac{\partial \hat{v}}{\partial \theta_q}. \quad (23)$$

Originally, Lee and Nelder [12] and Ha *et al.* [13] ignored $\hat{\beta}/\theta_q$ and $\hat{v}/\theta_q$ when differentiating $\hat{h}_p$ and $\hat{H}$ with respect to θ because the parameters are asymptotically orthogonal. However, this approach does not work in some cases such as data with binary covariates and small cluster sizes. Following Ha and Lee [14], only $\hat{\beta}/\theta_q$ is ignored here because there is an indirect dependency between $\hat{\beta}$ and θ_q while $\hat{v}/\theta_q$ is included because there is a direct dependency between $\hat{v}$ and θ_q , which is clear from (14) and (15). With this adjustment, it is known [14] that the restricted maximum likelihood estimator for the variance of the frailty distribution based on the h-likelihood is less biased than the marginal maximum likelihood estimator such as one from the penalized likelihood [18–19]. Detailed steps of calculating the gradient vector in (21) and the elements of the information matrix in (22) are provided in the Appendix. Estimates of θ can be now updated using the Newton-Raphson method,

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + \left[\frac{-\partial^2 h_A}{\partial \theta^2} \right]^{-1} \frac{\partial h_A}{\partial \theta} \Big|_{\theta=\hat{\theta}^{(i)}} \quad (24)$$

where $\hat{\theta}^{(i)}$ is the estimate of θ at the i th iteration.

6 Iterative algorithm

The MHLE of β , v and θ are found by maximizing simultaneously the profile h-likelihood (h_p) and the adjusted profile h-likelihood (h_A). First, the estimates of β and v are updated with a single Newton-Raphson step with the profile h-likelihood conditional on the current estimate of θ . Then the estimate of θ is updated with a single step of the Newton-Raphson to maximize the adjusted profile h-likelihood, given the current estimates of β and v . Continue alternating between (11) and (24) until convergence is achieved. Convergence is defined as,

$$\max \left\{ \left| \hat{\beta}^{(i+1)} - \hat{\beta}^{(i)} \right|, \left| \hat{\theta}^{(i+1)} - \hat{\theta}^{(i)} \right| \right\} < \Delta,$$

where Δ is a predetermined tolerance limit.

After convergence has been achieved, the estimated covariance matrix of the parameter estimates is evaluated by taking the inverse of the observed information matrix (17) for the profile h-likelihood, which can be used for an asymptotic Wald-type test and to construct confidence intervals for β and v .

The univariate and multivariate cause-specific hazard frailty models are non-nested models. To select which model is more appropriate Akaike's information criterion (AIC) can be used [22]. Define the AIC criteria as, $AIC = -2h_A(\hat{\theta}) + 2s$, where s is the number of dispersion parameters, i.e. the number of variance components in Σ . Note that s here is not the total number of parameters. The model with the smaller AIC indicates a better fitting model. The AIC criteria can also be used to select which correlation structure for Σ gives the best fit when using the multivariate model, as well as choose between a model with random effects and without random effects. This AIC criteria cannot be used to select fixed effects β since they have been eliminated from h_A , so that the adjusted profile h-likelihood is just a function of the frailty parameters θ .

7 Simulation

The proposed h-likelihood method can be viewed as a modification of the Markov Chain Monte Carlo EM (MCEM) method of Huang and Wolfe [4]. Thus we first compare the two methods through simulation studies. Huang and Wolf [4] considered a joint cause-specific frailty model with a shared log-frailty v_i for two event types; $k = 1$ for failure (T) and $k = 2$ for informative censoring (C), i.e.

$$\lambda_{ij}^T(t|v_i) = \lambda_{01}(t) \exp(\eta_{ij1}) \text{ and } \lambda_{ij}^C(t|v_i) = \lambda_{02}(t) \exp(\eta_{ij2}),$$

where $\lambda_{0k}(\cdot)$ is the unknown baseline hazard function for each event type k , and

$$\eta_{ij1} = X_{ij}^T \beta_1 + v_i \text{ and } \eta_{ij2} = X_{ij}^T \beta_2 + \alpha v_i.$$

Here v_i 's follow $N(0, \sigma^2)$, and under a competing risks setting, λ_{ij}^T and λ_{ij}^C would be equivalent to λ_{ij1} for events of interest and λ_{ij2} for competing events such as drop-outs [4] or deaths prior to events of interest, respectively. This model is a special case of model (1) with $v_{i1} = v_i$ and $v_{i2} = \alpha v_i$, which can have a different impact on λ_{ij1} and λ_{ij2} via parameter α [4,5]; if $\alpha > 0$ a subject with higher frailty will experience an earlier competing event, whereas if $\alpha < 0$ the competing event will be more likely delayed for a subject with higher frailty.

The h-likelihood-based inference for the shared joint model above is straightforward simply by replacing ℓ_{ijk} and ℓ_i in (6) with

$$\ell_{ijk} = \ell_{ijk}(\beta_k, \lambda_{0k}, \alpha; t_{ij}, \delta_{ijk} | v_i) = \delta_{ijk} (\log \lambda_{0k}(t_{ij}) + \eta_{ijk}) - \Lambda_{0k}(t_{ij}) \exp(\eta_{ijk})$$

and

$$\ell_i = \ell_i(\sigma^2; v_i) = -\log(2\pi\sigma^2)/2 - v_i^2/(2\sigma^2).$$

Here the dispersion parameters are $\theta = (\alpha, \sigma^2)^T$. Specifically, to eliminate the nuisance parameters λ_{0k} we first use the corresponding profile h-likelihood h_p in (10). Then we use h_p to estimate the fixed and random effects (β, v) , and the adjusted profile h-likelihood $h_A(\theta)$ in (19) to estimate the dispersion parameters $\theta = (\alpha, \sigma^2)^T$. For our comparison, we adopt the same simulation scenarios as in Huang and Wolf [4]. The results are summarized in Table 1. Here, we list the simulation results by Huang and Wolfe [4] as presented in Liu and Huang [5]. Overall, the results from the h-likelihood (HL) are comparable to those from the MCEM method, even though the HL method produces slightly larger biases for σ^2 than the MCEM does.

Next, data for the cause-specific hazard frailty model assuming a bivariate normal distribution is generated using a technique similar to Beyersmann *et al.* [23]. Let there be two event types, Type I and Type II as well as independent censoring. Samples sizes of $N=100$ and $N=200$, where $(n, n_i) = (50, 2), (50, 4)$ and $(100, 2)$, are considered. Data were generated with two covariates (X_{ij1}, X_{ij2}) , where X_{ij1} follows a standard normal distribution and X_{ij2} is a Bernoulli random variable with probability 0.5. The random effects are bivariate normal with

$$\begin{pmatrix} v_{i1} \\ v_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right) \quad (25)$$

where $\theta = (\sigma_{11}, \sigma_{22}, \sigma_{12}) = (1, 1, 0.5)$. The conditional cause-specific hazard rates for each event type are,

$$\lambda_{ij1}(t|x_{ij}, v_{i1})=0.5\exp(0.6x_{ij1}-0.4x_{ij2}+v_{i1}), \quad (26)$$

$$\lambda_{ij2}(t|x_{ij}, v_{i2})=2\exp(-0.3x_{ij1}-0.7x_{ij2}+v_{i2}). \quad (27)$$

That is, $\beta_1 = (\beta_{11}, \beta_{12}) = (0.6, -0.4)$ and $\beta_2 = (\beta_{21}, \beta_{22}) = (-0.3, 0.7)$. Under this scenario, approximately 65% of the events are Type I and 35% are Type II when there is no censoring. Censoring times are generated from a Uniform(0, c) distribution where the value of c is empirically selected to achieve the approximate right censoring rate, 0% and 30%. With 30% censoring, the proportions of Type I and Type II events are about 45% and 25%, respectively. For each scenario, 1000 datasets were generated.

Estimators of the Type I effects (β_{11}, β_{12}) perform well with no censoring and 30% censoring (Table 2). As the sample size increases the estimators become less biased. The estimated standard errors (SE) of the regression coefficients tend to underestimate the empirical standard deviation (SD) which is the estimate of true $\{\text{var}(\hat{\beta})\}^{1/2}$; thus most of the coverage probabilities are less than 95%. This is also a known problem when using the penalized partial likelihood approach for frailty models [18], particularly when the cluster size is as small as $n_i = 2$. Again increasing the sample size gives more accurate estimates of the SEs and reduces the SDs of the estimates and the MSE's, as expected.

Estimates of the variance components in Table 3 are more biased compared to the estimates of the corresponding regression coefficients presented in Table 2, in particular, the estimated variance for the Type II random effects σ_{22} . This was expected since there were fewer Type II events in the simulation scenarios considered. Increasing the number of clusters or the cluster size improves the estimation of this parameter. The variations of estimators (SD, SE and MSE) are larger for the censoring case, as expected.

For 0% censoring, 98% of the samples converged when $n = 50$ and $n_i = 2$ and 99% converged when $n = 50$ and $n_i = 4$. For all of the remaining scenarios, all of the samples converged. Samples that failed to converge were not included in Tables 2 and 3.

In addition, with 1000 replications we carried out simulation studies with a negative correlation $\sigma_{12} = -0.5$, i.e. $\rho = -0.5$, between Type I and Type II events, under the same setting used in Tables 2 and 3. This setting gave the rates of Type I events, Type II events, and the censoring proportion as approximately 48%, 27%, and 25%, respectively. The results are summarized in Table 4, which indicates that the trends of the parameter estimates are similar to those seen in Table 2 and Table 3. Furthermore, we have also conducted a simulation study when the rate of Type II events is higher than the rate of Type I events. For this case, we simply exchanged the two constant baseline hazards 0.5 and 2 in the simulation models (26) and (27), which gave the rates of Types I events, Type II events, and the censoring proportion of approximately 14%, 60% and 26%, respectively. The results are presented in Table 5, showing that the biases in the estimated variance component σ_{11} for Type I random effects v_{i1} are larger than one for σ_{22} . This might imply that the fewer

number of Type I events in this case (the fewer number of Type II events in Tables 2, 3 and 4) might have introduced the biases in estimation of the variance components, but the results improve as the cluster size or the number of clusters increases.

8 Application

The B-14 phase III breast cancer clinical trial conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP) was a randomized double-blinded multi-center trial comparing tamoxifen to placebo following surgery in patients who had negative axillary lymph nodes and estrogen receptor positive breast cancer. The study concluded that patients treated with tamoxifen had a significantly better outcome than those treated with placebo [24].

In this section the cause-specific hazard frailty model is used to estimate the effect of tamoxifen on different types of failures where patients experienced multiple events with competing risks. This analysis will use a high risk subset of patients from the B-14 study with a tumor size greater than 2.5 centimeters. In this subset there are 731 women (371 placebo and 360 tamoxifen) who are eligible with follow-up. The median age for women on either placebo or treatment was 55 years. Multiple types of treatment failure are possible; local, regional, or distant recurrence of the original cancer as well as a new second primary cancer or death because patients were followed as long as they did not withdraw their consents.

For the purpose of this analysis the types of failures will be divided into three event types. The first type (Type I) is a local or regional recurrence, the second type (Type II) is a new second primary cancer in the contralateral breast and the third type (Type III) is a distant recurrence, other new second primary cancer or death. We assume that these three types of events compete against each other because once a recurrence or second primary occurs, then non-protocol therapies are often administered after the event, which would prohibit an accurate assessment of the effect of the treatment solely on that particular event type under consideration.

A cause-specific hazard frailty model for age and treatment was fitted assuming both univariate and trivariate normal distributions with considering correlation structures among random effects. The regression coefficients and estimated variance of the random effects assuming a univariate normal distribution with only one random effect per subject are in the upper left-hand side of Table 6. Adjusted for age, the relative risk of a Type I event for an individual on tamoxifen compared to the same individual being on placebo is 0.48 with a 95% confidence interval of (0.31, 0.74). The estimated variance of the random effects is 1.895, suggesting a fairly heterogeneous group of subjects. Ignoring the correlation between event types, fitting the standard frailty model for each event type by treating other events as independent censoring [10] results in lower estimates of the treatment effect (the upper right-hand side of Table 6); fitting this naive model is equivalent to fitting a cause-specific hazard frailty model with an independent assumption among three random effects per subject (one random effect per event type).

Figure 1 shows the predicted cumulative incidence curves of a Type I event for a 55 year old woman from the cause-specific univariate frailty model, which was calculated by modifying Cheng *et al.* (1998) [25]. To be brief, for a given set of the covariates x_0 and a known frailty value v_0 , the cumulative incidence function for type k events can be predicted from

$$\hat{F}_k(t|x_0, v_0) = \sum_{t_{ij} \leq t} \hat{S}(t_{ij}|x_0, v_0) \hat{\lambda}_k(t_{ij}|x_0, v_0),$$

where $\hat{S}(t_{ij}|x_0, v_0) = \exp\{-\sum_k \hat{\Lambda}_k(t_{ij}|x_0, v_0)\}$, $\hat{\Lambda}_k(t_{ij}|x_0, v_0) = \hat{\Lambda}_{0k}(t_{ij}) \exp(x_0^T \hat{\beta}_k + v_0)$,

$$\hat{\Lambda}_{0k}(t_{ij}) = \sum_{r: t_{(kr)} \leq t_{ij}} \frac{\delta_{ijk}}{\sum_{ij \in \mathcal{R}_{kr}} \exp(X_{ij}^T \hat{\beta}_k + \hat{v}_i)},$$

and hence

$$\hat{\lambda}_k(t_{ij}|x_0, v_0) = \frac{\delta_{ijk} \exp(x_0^T \hat{\beta}_k + v_0)}{\sum_{ij \in \mathcal{R}_{kr}} \exp(X_{ij}^T \hat{\beta}_k + \hat{v}_i)}.$$

The incidence of a Type I event increases faster for the placebo group compared to the tamoxifen group. Ten years after surgery an average women on tamoxifen has a 7% chance of a local or regional recurrence while a women on placebo has a 13% chance. This probability increases for women at high risk (75th percentile of the estimated frailty distribution) and decreases for women at low risk (25th percentile of the estimated frailty distribution).

The estimated 25th and 75th percentiles of the random effects are respectively -1.07 and 0.82 . In Figure 2, the estimated cause-specific random effects are larger in general for subjects who had an event early and decrease for later event times. Thus those subjects who had an event early are more frail than those who survived longer, as expected. The estimated cause-specific random effects for every subject who did not have an event is less than 0, which is reasonable since there is no evidence from the observed data that these subjects should be at higher risk than an average person.

Additionally, the cause-specific hazard frailty model was fit assuming a trivariate normal distribution with one random effect per event type (three random effects per subject) by using an exchangeable correlation structure. The estimated regression coefficients along with their standard errors and confidence intervals as well as the estimated variance components are given under the trivariate case in Table 6. The estimated treatment effects for each event type are less than the corresponding estimates for the univariate case in Table 6, but the patterns were similar; patients on tamoxifen had a significantly lower risk of a Type I event compared to patients on placebo. Tamoxifen did not significantly lower the risk for other event types. The estimated variance of the random effects for each event type are

all similar ranging from 0.706 to 0.789. There is also a strong positive correlation between the cause-specific random effects, indicating that patients who experienced a local or regional recurrence will also be at a greater risk for developing a second primary cancer in the contralateral breast as well as any of the Type III events. This is because patients who have a larger random effect for a Type I event would also tend to have a larger random effect for Type II and Type III events, and larger random effects would increase the risk of failure for an individual or a cluster.

Assuming a trivariate normal distribution for the frailties, the AIC was 6980.2, while it was 6967.8 for the univariate case and 6993.0 for the independent case (i.e. the standard frailty model with three independent frailties), indicating that the univariate model provides a better fit in this example.

9 Discussion

The h-likelihood provides a new estimation procedure for fitting the cause-specific hazard models with multivariate frailties for competing risks data, which is computationally efficient and directly produces the estimates of standard errors for the parameters from the observed information matrix required for Newton-Raphson method [13]. Simulation results demonstrate that the h-likelihood approach performs well, producing reasonable estimates even for small cluster sizes.

A drawback of the h-likelihood is that it might be difficult to be implemented because of the numerous derivatives that need to be calculated. Once the derivatives are calculated, however, the analysis is computationally efficient, whereas the EM algorithm will always be computationally intensive. The h-likelihood procedure presented here is mainly focused on estimating regression coefficients while accounting for the correlation between event types. There are other methods such as Cheng *et al.* [26] that maybe more appropriate when the main objective of the analysis is the association between competing events, not the effect of risk factors.

The present work only considered the lognormal frailty distribution. It may be also interesting to consider other distributions when competing risks are present, in particular the compound Poisson frailty distribution. This distribution is unique in that it allows for a subgroup of zero frailty, a subgroup where no one experiences the event. An advantage of the compound Poisson distribution would be that it has a closed form of the Laplace transform, which would make calculation of the marginal survival function simpler.

Two broad classes of models for analyzing the competing risks data have been developed based on the Cox's proportional hazards model; (i) model the cause-specific hazard of each event type separately by Prentice *et al.* [28] and (ii) model the subhazard (i.e. the hazard function of a subdistribution) for the event of interest by Fine and Gray [8]. The cause-specific hazard model associates the covariates with the cause-specific hazard function, whereas the subhazard model directly associates the covariates with the cumulative probability of cause-specific events over time, i.e. the cumulative incidence function (CIF). Recently, these two modelling approaches have been extended for clustered competing risks

data by using frailties. The cause-specific frailty model (1) can take into account the correlation between events of interest and competing events via frailties, while the subhazard frailty models [6,7,15] do not by assuming that the frailty effects on both types of events are independent. Therefore, the cause-specific frailty model (1) would be more appropriate when a dependency between both types of events or informative censoring is present.

Ha *et al.* [15] have shown via a simulation study that the likelihood ratio test from the h-likelihood for testing the null hypothesis of $H_0 : \sigma_{ii} = 0$ when the single variance component is on the boundary of the parameter space performs well under the subdistribution hazard frailty model. However, in the present case, the application of the mixture chi-squares distributions [20,21] might be challenging due to the number of parameters that needs be dealt with, especially with the trivariate case presented in Section 8, which could be an interesting future work.

Acknowledgments

Dr. Jeong's research was supported in part by National Institute of Health (NIH) grants 5-U10-CA69974-09 and 5-U10-CA69651-11. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology, Korea (No. 2010-0021165).

References

1. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of family tendency in chronic disease incidence. *Biometrika*. 1978; 65:141–151.
2. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society: Series B*. 1972; 34:182–220.
3. Kalbfleisch, J., Prentice, RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons, Inc; 1980.
4. Huang X, Wolfe R. A frailty model for informative censoring. *Biometrics*. 2002; 58:510–520. [PubMed: 12229985]
5. Liu L, Huang X. The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine*. 2008; 27:2665–2683. [PubMed: 17910008]
6. Katsahian S, Resche-Rigon M, Chevret S, Porcher R. Analysing multicentre competing risks data with a mixed proportional hazards model for the subdistribution. *Statistics in Medicine*. 2006; 25:4267–4278. [PubMed: 16960919]
7. Katsahian S, Boudreau C. Estimating and testing for center effects in competing risks. *Statistics in Medicine*. 2011; 30:1608–1617. [PubMed: 21341296]
9. Zhou B, Latouche A, Rocha V, Fine J. Competing risks regression for stratified data. *Biometrics*. 2011; 67:661–670. [PubMed: 21155744]
8. Fine J, Gray R. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94:496–509.
10. Gorfine M, Hsu L. Frailty-based competing risks model for multivariate survival data. *Biometrics*. 2011; 67:415–426. [PubMed: 20707868]
11. Xue X, Brookmeyer R. Bivariate frailty model for the analysis of multivariate survival Time. *Lifetime Data Analysis*. 1996; 2:227–289.
12. Lee Y, Nelder JA. Hierarchical Generalized Linear Models (with Discussion). *Journal of the Royal Statistical Society: Series B*. 1996; 58:619–678.
13. Ha I, Lee Y, Song J. Hierarchical likelihood approach for frailty models. *Biometrika*. 2001; 88:233–243.
14. Ha I, Lee Y. Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*. 2003; 12:663–681.

15. Ha I, Christian NJ, Jeong J, Park J, Lee Y. Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Statistical Methods in Medical Research*. 2013; doi: 10.1177/0962280214526193
16. Johansen S. An extension of Cox's regression model. *International Statistical Review*. 1983; 51:165–174.
17. Breslow NE. Covariance analysis of censored survival data. *Biometrics*. 1974; 30:89–99. [PubMed: 4813387]
18. Therneau TM, Grambsch PM, Pankratz VS. Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*. 2003; 12:156–175.
19. Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*. 2000; 56:1016–1022. [PubMed: 11129456]
20. Noh M, Ha I, Lee Y. Dispersion frailty models and HGLMs. *Statistics in Medicine*. 2006; 25:1341–1354. [PubMed: 16217839]
21. 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.
22. Ha I, Lee Y, MacKenzie G. Model selection for multi-component frailty models. *Statistics in Medicine*. 2007; 26:4790–4807. [PubMed: 17476647]
23. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Statistics in Medicine*. 2009; 28:956–971. [PubMed: 19125387]
24. Fisher B, Costantino JP, Redmond C, et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen receptor-positive tumors. *New England Journal of Medicine*. 1989; 320:479–484. [PubMed: 2644532]
25. Cheng S, Fine J, Wei L. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*. 1998; 54:219–228. [PubMed: 9544517]
26. Cheng Y, Fine J, Kosorok M. Nonparametric association analysis of exchangeable clustered competing risks data. *Biometrics*. 2009; 65:385–393. [PubMed: 18549422]
27. Zhou B, Fine J, Latouche A, Labopin M. Competing risks regression for clustered data. *Biostatistics*. 2012; 13:371–383. [PubMed: 22045910]
28. Prentice R, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978; 34:541–554. [PubMed: 373811]
29. Lee, Y., Nelder, JA., Pawitan, Y. *Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood*. Chapman & Hall; 2006.

Appendix: Derivation of the gradient vector and elements for the information matrix for the adjusted profile likelihood

Let $\tau = (\beta, \nu)$ and $\hat{\tau} = \hat{\tau}(\theta) = (\hat{\beta}(\theta), \hat{\nu}(\theta))$. First calculate the derivatives in (21). Since $h_p/\nu/\hat{\tau} = 0$, the total derivative of the first term $\hat{h}_p/\theta_q$ is

$$\begin{aligned}
\frac{\partial \hat{h}_p}{\partial \theta_q} &= \frac{\partial h_p}{\partial \theta_q} \Big|_{\tau=\hat{\tau}} + \left(\frac{\partial h_p}{\partial v} \Big|_{\tau=\hat{\tau}} \right) \left(\frac{\partial \hat{v}}{\partial \theta_q} \right) \\
&= \frac{\partial h_p}{\partial \theta_q} \Big|_{\tau=\hat{\tau}} \\
&= \sum_{i=1}^n \frac{\partial l_i(\theta; \hat{v}_i)}{\partial \theta_q} \\
&= \sum_{i=1}^n -\frac{1}{2} \text{trace} \left(\Sigma^{-1} \Sigma'_q \right) + \frac{1}{2} \hat{v}_i^T \left(\Sigma^{-1} \Sigma'_q \Sigma^{-1} \right) \hat{v}_i,
\end{aligned}$$

where $\Sigma'_q = \partial \Sigma / \partial \theta_q$.

The derivative of the second term $\hat{H} / \theta_q$ in (21) is more complicated as

$$\frac{\partial \hat{H}}{\partial \theta_q} = \frac{\partial H}{\partial \theta_q} \Big|_{\tau=\hat{\tau}} + \left(\frac{\partial H}{\partial v} \Big|_{\tau=\hat{\tau}} \right) \left(\frac{\partial \hat{v}}{\partial \theta_q} \right).$$

The term $\hat{v} / \theta_q$ is calculated following Lee *et al.* [29]. From h_p , given θ_p , let $\hat{v}(\theta_q)$ be the solution to $g(\theta_q) = h_p' / v|_{\tau=\hat{\tau}} = 0$. Then,

$$\frac{\partial g(\theta_q)}{\partial \theta_q} = \frac{\partial^2 h_p}{\partial v \partial \theta_q} \Big|_{\tau=\hat{\tau}} + \left(\frac{\partial^2 h_p}{\partial v^2} \Big|_{\tau=\hat{\tau}} \right) \left(\frac{\partial \hat{v}}{\partial \theta_q} \right) = 0.$$

Solving for $\hat{v} / \theta_q$ gives a $2n \times 1$ vector,

$$\begin{aligned}
\frac{\partial \hat{v}}{\partial \theta_q} &= \left(-\frac{\partial^2 h_p}{\partial v^2} \Big|_{\tau=\hat{\tau}} \right)^{-1} \left(\frac{\partial^2 h_p}{\partial v \partial \theta_q} \Big|_{\tau=\hat{\tau}} \right) \\
&= \left(\mathbf{Z}^T \hat{\mathbf{W}} \mathbf{Z} + \mathbf{Q} \right)^{-1} \left(\left[\left(\Sigma^{-1} \Sigma'_q \Sigma^{-1} \right) \otimes I_n \right] \hat{v} \right),
\end{aligned}$$

where $\hat{\mathbf{W}}$ is $\mathbf{W}$ evaluated at $(\hat{\beta}, \hat{v}, \theta)$, that is, when $W_k = W_k(\hat{\beta}_k, \hat{v}_k, \theta) = \hat{W}_k$. Now since $\mathbf{X}$ and $\mathbf{Z}$ are constant matrices that have no dependence on θ , it follows that the total derivative of $\hat{H} / \theta_q$ is,

$$\frac{\partial \hat{H}}{\partial \theta_q} = \begin{pmatrix} \mathbf{X}^T \hat{\mathbf{W}}'_q \mathbf{X} & \mathbf{X}^T \hat{\mathbf{W}}'_q \mathbf{Z} \\ \mathbf{Z}^T \hat{\mathbf{W}}'_q \mathbf{X} & \mathbf{Z}^T \hat{\mathbf{W}}'_q \mathbf{Z} + \mathbf{Q}'_q \end{pmatrix},$$

where $\hat{\mathbf{W}}'_q = \partial \hat{\mathbf{W}} / \partial \theta_q$ and $Q'_q = \partial Q / \partial \theta_q$. Based on the structure of $\mathbf{W}$, $\hat{\mathbf{W}}'_q$ is found by evaluating $\hat{\mathbf{W}}'_k / \theta_q$. Since $\hat{\mathbf{W}}_k$ does not depend on θ , the total derivative of $\hat{\mathbf{W}}_k$ is

$$\hat{\mathbf{W}}'_{kq} = \frac{\partial \hat{\mathbf{W}}_k}{\partial \theta_q} = \frac{\partial W_k}{\partial \theta_q} \Big|_{\tau=\hat{\tau}} + \left(\frac{\partial W_k}{\partial v_k} \Big|_{\tau=\hat{\tau}} \right) \left(\frac{\partial \hat{v}_k}{\partial \theta_q} \right) = \left(\frac{\partial W_k}{\partial v_k} \Big|_{\tau=\hat{\tau}} \right) \left(\frac{\partial \hat{v}_k}{\partial \theta_q} \right).$$

The derivative $\hat{\mathbf{W}}_k / v_k$ is found by differentiating $W_k(\beta_k, v_k) = M_k - N_k R_k C_k (R_k N_k)^T$ with respect to v_k . Given the structure of v defined earlier, $\hat{v}_1 / \theta_q$ is the first n elements of the vector $\hat{v} / \theta_q$ and $\hat{v}_2 / \theta_q$ are the last n elements. Since Q does not depend on v , the total derivative is not needed to find Q / θ_q so,

$$Q'_q = \frac{\partial Q}{\partial \theta_q} = \left(-\sum^{-1} \sum'_q \sum^{-1} \right) \otimes I_n. \quad (28)$$

Using (28) there is a slightly simpler expression for $\partial \hat{v} / \partial \theta_q = -(\mathbf{Z}^T \hat{\mathbf{W}} \mathbf{Z} + Q)^{-1} \cdot \{Q'_q \hat{v}\}$. The next step is to calculate the terms in the observed information (22). First,

$$\begin{aligned} -\frac{\partial^2 \hat{h}_p}{\partial \theta_q \partial \theta_s} &= -\frac{\partial^2 h_p}{\partial \theta_q \partial \theta_s} \Big|_{\tau=\hat{\tau}} - \left(\frac{\partial^2 h_p}{\partial v \partial \theta_s} \Big|_{\tau=\hat{\tau}} \right) \left(\frac{\partial \hat{v}}{\partial \theta_s} \right) \\ &= \sum_{i=1}^n -\frac{\partial^2 l_i(\theta; \hat{v}_i)}{\partial \theta_q \partial \theta_s} - (Q'_q \hat{v}) \left(\frac{\partial \hat{v}}{\partial \theta_s} \right) \\ &= \sum_{i=1}^n \left(\frac{1}{2} \text{trace} \left(\sum^{-1} \sum'_s \sum^{-1} \sum'_q \sum^{-1} \right) + \frac{1}{2} \hat{v}_i^T \left(\sum^{-1} \sum'_s \sum^{-1} \sum'_q \sum^{-1} \right) \hat{v}_i + \frac{1}{2} \hat{v}_i^T \left(\sum^{-1} \sum'_q \sum^{-1} \sum'_s \sum^{-1} \right) \hat{v}_i \right) - \end{aligned}$$

The last term needed to calculate (22) is

$$\frac{\partial^2 \hat{H}}{\partial \theta_q \partial \theta_s} = \begin{pmatrix} \mathbf{X}^T \hat{\mathbf{W}}''_{qs} \mathbf{X} & \mathbf{X}^T \hat{\mathbf{W}}''_{qs} \mathbf{Z} \\ \mathbf{Z}^T \hat{\mathbf{W}}''_{qs} \mathbf{X} & \mathbf{Z}^T \hat{\mathbf{W}}''_{qs} \mathbf{Z} + Q''_{qs} \end{pmatrix},$$

where

$$Q''_{qs} = \frac{\partial^2 Q}{\partial \theta_q \partial \theta_s} = \left(\sum^{-1} \sum'_s \sum^{-1} \sum'_q \sum^{-1} + \sum^{-1} \sum'_q \sum^{-1} \sum'_s \sum^{-1} \right) \otimes I_n$$

and $\hat{\mathbf{W}}''_{qs} = \partial \hat{\mathbf{W}}'_q / \partial \theta_s$. Like earlier, $\hat{\mathbf{W}}''_{qs}$ is found by finding $\partial^2 \hat{\mathbf{W}}_k / \theta_q \theta_s$ for $k = 1, 2$,

$$\hat{W}_{kqs}'' = \frac{\partial^2 \hat{W}_k}{\partial \theta_q \partial \theta_s} = \left[\left(\frac{\partial^2 W_k}{\partial v_k^2} \Big|_{\tau=\hat{\tau}} \right) \frac{\partial \hat{v}_k}{\partial \theta_q} \right] \frac{\partial \hat{v}_k}{\partial \theta_s} + \left(\frac{\partial W_k}{\partial v_k} \Big|_{\tau=\hat{\tau}} \right) \frac{\partial^2 \hat{v}_k}{\partial \theta_q \partial \theta_s},$$

where

$$\begin{aligned} \frac{\partial^2 \hat{v}}{\partial \theta_q \partial \theta_s} &= (\mathbf{Z}^T \hat{\mathbf{W}} \mathbf{Z} + \mathbf{Q})^{-1} \left(\mathbf{Z}^T \hat{\mathbf{W}}'_s \mathbf{Z} + \mathbf{Q}'_s \right) (\mathbf{Z}^T \hat{\mathbf{W}} \mathbf{Z} + \mathbf{Q})^{-1} \mathbf{Q}'_q \hat{v} - (\mathbf{Z}^T \hat{\mathbf{W}} \mathbf{Z} + \mathbf{Q})^{-1} \left[\mathbf{Q}''_{qs} \hat{v} + \mathbf{Q}'_q \frac{\partial \hat{v}}{\partial \theta_s} \right] \\ &= -(\mathbf{Z}^T \hat{\mathbf{W}} \mathbf{Z} + \mathbf{Q})^{-1} \left[\left(\mathbf{Z}^T \hat{\mathbf{W}}'_s \mathbf{Z} + \mathbf{Q}'_s \right) \frac{\partial \hat{v}}{\partial \theta_q} + \mathbf{Q}''_{qs} \hat{v} + \mathbf{Q}'_q \frac{\partial \hat{v}}{\partial \theta_s} \right] \end{aligned}$$

(29)

is a $2n \times 1$ vector and $\frac{\partial^2 \hat{v}_k}{\partial \theta_q \partial \theta_s}$ is the first n elements of (29) if $k = 1$ and the second n elements if $k = 2$. The term $\frac{\partial^2 W_k}{\partial v_k^2}$ is found by twice differentiating $W_k(\beta_k, v_k) = \mathbf{M}_k - \mathbf{N}_k \mathbf{R}_k \mathbf{C}_k (\mathbf{R}_k \mathbf{N}_k)^T$ with respect to v_k .

Figure 1.

Predicted cumulative incidence of Type I events, for an average subject $V = 0$, high risk subject $V = 0.82$ (75th percentile), and low risk subject $V = -1.07$ (25th percentile).

Figure 2. Estimated cause-specific frailties versus the first observed event time for each subject; boxplot on the right hand side is the distribution of the estimated cause-specific random effects.

Comparison of the two methods (HL and MCEM); cause-specific hazard frailty model - shared case: 40 clusters with 5 subjects in each cluster ($n = 40$, $n_i = 5$); type I ("T")= 45%, type II ("C")=35% informative dropout, censoring=20% administrative censoring

Table 1

Parameter	HL				MCEM			
	Mean	SD	SE	CP	Mean	SD	SE	CP
$\alpha = 1$								
$10\beta_{\text{age}}^T = 1.0$	1.006	0.247	0.234	0.932	1.023	0.251	0.256	0.952
$\beta_{\text{tr}}^T = -1.4$	-1.427	0.288	0.276	0.952	-1.423	0.288	0.291	0.954
$10\beta_{\text{age}}^C = 2.0$	2.010	0.284	0.275	0.956	2.012	0.311	0.302	0.942
$\beta_{\text{tr}}^C = 1.2$	1.224	0.316	0.314	0.958	1.225	0.322	0.331	0.960
$\sigma^2 = 1.0$	1.029	0.433	-	-	0.977	0.405	-	-
$\alpha = 1.0$	1.031	0.305	-	-	1.027	0.291	-	-
$\alpha = -1$								
$10\beta_{\text{age}}^T = 1.0$	1.018	0.242	0.242	0.952	0.996	0.251	0.243	0.940
$\beta_{\text{tr}}^T = -1.4$	-1.425	0.287	0.271	0.944	-1.415	0.276	0.289	0.954
$10\beta_{\text{age}}^C = 2.0$	2.014	0.309	0.299	0.948	2.052	0.305	0.308	0.956
$\beta_{\text{tr}}^C = 1.2$	1.211	0.300	0.291	0.944	1.199	0.306	0.312	0.952
$\sigma^2 = 1.0$	1.042	0.466	-	-	1.001	0.458	-	-
$\alpha = -1.0$	-1.015	0.291	-	-	-1.034	0.337	-	-

HL; simulation results from h-likelihood method

MCEM; simulation results from MCEM method in [4]

Mean and SD; mean and standard deviation of estimates from 500 iterations

SE, mean of the estimated standard errors; %Bias, percent bias

MSE, mean square error; CP, coverage probability.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Simulation results for β ; cause-specific hazard frailty model - bivariate case; type I = 65%, type II = 35% with no censoring; type I = 45%, type II = 25% with 30% censoring

Censoring	Sample Size	Parameter	Mean	SD	SE	%Bias	MSE	CP	
0%	$n = 50, n_f = 2$	β_{11}	0.615	0.194	0.178	2.5	0.038	93.6	
		β_{12}	-0.407	0.348	0.330	1.7	0.121	93.9	
		β_{21}	-0.308	0.250	0.231	2.7	0.063	94.6	
		β_{22}	0.693	0.467	0.445	-1.0	0.218	94.1	
	$n = 50, n_f = 4$	β_{11}	0.606	0.120	0.116	1.0	0.014	93.7	
		β_{12}	-0.408	0.221	0.215	2.0	0.049	93.0	
		β_{21}	-0.304	0.162	0.152	1.3	0.026	93.8	
		β_{22}	0.712	0.303	0.292	1.7	0.092	94.7	
	30%	$n = 100, n_f = 2$	β_{11}	0.588	0.128	0.123	-2.0	0.017	93.5
			β_{12}	-0.390	0.222	0.227	-2.5	0.049	95.5
			β_{21}	-0.293	0.167	0.156	-2.3	0.028	93.3
			β_{22}	0.695	0.310	0.302	-0.7	0.096	95.3
$n = 50, n_f = 2$		β_{11}	0.614	0.218	0.200	2.3	0.048	93.9	
		β_{12}	-0.432	0.380	0.378	8.0	0.145	93.7	
		β_{21}	-0.320	0.303	0.272	6.7	0.092	93.2	
		β_{22}	0.773	0.564	0.537	10.4	0.323	95.3	
$n = 50, n_f = 4$		β_{11}	0.599	0.142	0.131	-0.2	0.020	93.1	
		β_{12}	-0.404	0.256	0.248	1.0	0.066	94.6	
		β_{21}	-0.303	0.186	0.177	1.0	0.035	94.6	
		β_{22}	0.723	0.356	0.349	3.3	0.127	95.2	
$n = 100, n_f = 2$	β_{11}	0.585	0.142	0.136	-2.5	0.020	94.6		
	β_{12}	-0.390	0.263	0.259	-2.5	0.069	94.5		
	β_{21}	-0.303	0.189	0.181	0.3	0.036	94.9		
	β_{22}	0.687	0.358	0.355	-1.9	0.128	94.8		

Mean and SD; mean and standard deviation of estimates from 1000 iterations

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

SE, mean of estimated standard errors; %Bias, percent bias; MSE, mean square error
CP; coverage probability

Table 3

Simulation results for θ , cause-specific hazard frailty model - bivariate case; type I = 65%, type II = 35% with no censoring; type I = 45%, type II = 25% with 30% censoring

Censoring	Sample Size	Parameter	Mean	SD	%Bias	MSE
0%	$n = 50, n_i = 2$	σ_{11}	1.084	0.642	8.4	0.419
		σ_{22}	1.270	0.962	27.0	0.998
		σ_{12}	0.524	0.605	4.8	0.367
		ρ	0.399	0.395	-20.2	0.166
	$n = 50, n_i = 4$	σ_{11}	1.020	0.377	2.0	0.143
		σ_{22}	1.086	0.501	8.6	0.258
		σ_{12}	0.506	0.338	1.2	0.114
		ρ	0.483	0.270	-3.4	0.073
	$n = 100, n_i = 2$	σ_{11}	0.957	0.402	-4.3	0.163
		σ_{22}	1.005	0.518	0.5	0.268
		σ_{12}	0.485	0.355	-3.0	0.126
		ρ	0.479	0.287	-4.2	0.083
30%	$n = 50, n_i = 2$	σ_{11}	1.150	0.744	15.0	0.576
		σ_{22}	1.430	1.497	43.0	2.426
		σ_{12}	0.450	0.701	10.0	0.494
		ρ	0.349	0.441	-30.2	0.217
	$n = 50, n_i = 4$	σ_{11}	1.006	0.417	0.6	0.174
		σ_{22}	1.084	0.561	8.4	0.322
		σ_{12}	0.495	0.384	-1.0	0.147
		ρ	0.473	0.275	-5.4	0.076
	$n = 100, n_i = 2$	σ_{11}	0.955	0.424	-4.5	0.182
		σ_{22}	1.020	0.543	2.0	0.295
		σ_{12}	0.474	0.396	-5.2	0.157
		ρ	0.470	0.290	-6.0	0.085

Mean and SD, mean and standard deviation of estimates from 1000 iterations

% Bias, percent bias; MSE, mean square error.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Negative correlation; simulation results for β and θ with $\sigma_{12} = -0.5$; cause-specific hazard frailty model - bivariate case; type I = 48%, type II = 27% with 25% censoring

Censoring	Sample Size	Parameter	Mean	SD	SE	%Bias	MSE	CP
25%	$n = 50, n_I = 2$	β_{11}	0.617	0.222	0.201	2.8	0.050	93.5
		β_{12}	-0.425	0.391	0.378	6.3	0.154	94.6
		β_{21}	-0.319	0.277	0.257	6.3	0.077	94.4
		β_{22}	0.758	0.555	0.510	8.3	0.311	93.4
	$n = 50, n_I = 4$	β_{11}	0.606	0.144	0.131	1.0	0.021	93.0
		β_{12}	-0.389	0.247	0.246	-2.8	0.061	96.7
		β_{21}	-0.312	0.184	0.175	4.0	0.034	93.2
		β_{22}	0.719	0.354	0.338	2.7	0.126	93.4
	$n = 100, n_I = 2$	β_{11}	0.586	0.152	0.135	-2.3	0.023	93.4
		β_{12}	-0.388	0.259	0.257	-3.0	0.067	95.4
		β_{21}	-0.286	0.184	0.172	4.7	0.034	93.2
		β_{22}	0.705	0.336	0.339	0.7	0.113	96.6
	$n = 50, n_I = 2$	σ_{11}	1.256	0.855	-	25.6	0.797	-
		σ_{22}	1.462	1.517	-	46.2	2.515	-
		σ_{12}	-0.535	0.731	-	7.0	0.536	-
		ρ	-0.364	0.401	-	-27.2	0.179	-
	$n = 50, n_I = 4$	σ_{11}	1.052	0.432	-	5.2	0.189	-
		σ_{22}	1.074	0.597	-	7.4	0.362	-
		σ_{12}	-0.524	0.367	-	4.8	0.135	-
		ρ	-0.476	0.315	-	-4.8	0.100	-
	$n = 100, n_I = 2$	σ_{11}	0.947	0.407	-	-5.3	0.168	-
		σ_{22}	1.019	0.569	-	1.9	0.324	-
		σ_{12}	-0.503	0.428	-	0.6	0.183	-
		ρ	-0.467	0.355	-	-6.6	0.127	-

Table 5

Proportions of event types reversed; simulation results for β and θ , cause-specific hazard frailty model - bivariate case; type I = 14%, type II = 60% with 26% censoring

Censoring	Sample Size	Parameter	Mean	SD	SE	%Bias	MSE	CP
26%	$n = 50, n_i = 2$	β_{11}	0.627	0.399	0.368	4.5	0.160	96.5
		β_{12}	-0.466	0.764	0.737	16.5	0.588	96.5
		β_{21}	-0.296	0.173	0.172	-1.3	0.029	95.4
		β_{22}	0.716	0.349	0.337	2.3	0.122	94.2
	$n = 50, n_i = 4$	β_{11}	0.607	0.246	0.235	1.2	0.061	94.9
		β_{12}	-0.412	0.489	0.465	3.0	0.239	95.2
		β_{21}	-0.302	0.118	0.113	0.7	0.014	94.1
		β_{22}	0.699	0.224	0.221	-0.1	0.052	94.3
	$n = 100, n_i = 2$	β_{11}	0.588	0.244	0.237	-2.0	0.060	94.2
		β_{12}	-0.396	0.486	0.469	-1.0	0.236	95.6
		β_{21}	-0.294	0.125	0.117	-2.0	0.016	93.8
		β_{22}	0.690	0.243	0.231	-1.4	0.006	93.6
	$n = 50, n_i = 2$	σ_{11}	1.438	1.247	-	43.8	1.745	-
		σ_{22}	1.122	0.810	-	12.2	0.671	-
		σ_{12}	0.521	0.713	-	4.2	0.509	-
		ρ	0.371	0.417	-	-25.8	0.191	-
	$n = 50, n_i = 4$	σ_{11}	1.137	0.887	-	13.7	0.806	-
		σ_{22}	1.037	0.399	-	3.7	0.161	-
		σ_{12}	0.510	0.423	-	2.0	0.179	-
		ρ	0.514	0.389	-	2.8	0.152	-
	$n = 100, n_i = 2$	σ_{11}	1.176	1.097	-	17.6	1.234	-
		σ_{22}	0.943	0.398	-	-9.7	0.168	-
		σ_{12}	0.491	0.502	-	-1.8	0.252	-
		ρ	0.515	0.369	-	3.0	0.136	-

Table 6

Estimates of the cause-specific hazard frailty model—univariate, independent and trivariate cases.

Event Type	Effect	Univariate case With correlation		Independent case Ignoring correlation	
		Estimate (SE)	95% CI	Estimate (SE)	95% CI
Type I	Age	-0.015 (0.010)	(-0.036, 0.005)	-0.017 (0.009)	(-0.035, 0.001)
	Treatment	-0.742 (0.226)	(-1.186, -0.299)	-0.633 (0.199)	(-1.023, -0.243)
Type II	Age	-0.002 (0.014)	(-0.028, 0.025)	-0.001 (0.013)	(-0.025, 0.024)
	Treatment	-0.179 (0.273)	(-0.716, 0.357)	-0.041 (0.250)	(-0.532, 0.449)
Type III	Age	0.016 (0.007)	(0.001, 0.031)	0.017 (0.005)	(0.007, 0.027)
	Treatment	-0.262 (0.149)	(-0.556, 0.032)	-0.137 (0.102)	(-0.336, 0.063)
Random Effect	Variance	$\sigma_s^2=1.895$			
			$\sigma_I^2=0.116, \sigma_{II}^2=0.001, \sigma_{III}^2=0.001$		
Event Type	Effect	Trivariate case		95% CI	
		Estimate	SE		
Type I	Age	-0.017	0.010	(-0.036, 0.002)	
	Treatment	-0.684	0.211	(-1.057, -0.271)	
Type II	Age	-0.002	0.013	(-0.027, 0.024)	
	Treatment	-0.111	0.260	(-0.621, 0.399)	
Type III	Age	0.016	0.006	(0.004, 0.028)	
	Treatment	-0.203	0.125	(-0.448, 0.042)	
Type I	Variance	0.789			
Type II	Variance	0.706			
Type III	Variance	0.771			
Type I, Type II	Correlation	0.886			
Type I, Type III	Correlation	0.848			
Type II, Type III	Correlation	0.897			

CI, confidence interval; SE, standard error σ_s^2 , common variance for all event types $\sigma_I^2, \sigma_{II}^2, \sigma_{III}^2$; independent variance for each event type