BIOMETRICS 56, 1016–1022
December 2000

Estimation of Multivariate Frailty Models Using Penalized Partial Likelihood

Samuli Ripatti[1,2,*] and Juni Palmgren[2,3]

[1]Rolf Nevanlinna Institute, University of Helsinki, P.O. Box 4, FIN-00014 Helsinki, Finland
[2]Department of Mathematical Statistics, Stockholm University, SE-10691 Stockholm, Sweden
[3]Department of Medical Epidemiology, Box 281, Karolinska Institutet, SE-17177 Stockholm, Sweden
*email: samuli.ripatti@rni.helsinki.fi

SUMMARY. There exists a growing literature on the estimation of gamma distributed multiplicative shared frailty models. There is, however, often a need to model more complicated frailty structures, but attempts to extend gamma frailties run into complications. Motivated by hip replacement data with a more complicated dependence structure, we propose a model based on multiplicative frailties with a multivariate log-normal joint distribution. We give a justification and an estimation procedure for this generally structured frailty model, which is a generalization of the one presented by McGilchrist (1993, *Biometrics* **49**, 221–225). The estimation is based on Laplace approximation of the likelihood function. This leads to estimating equations based on a penalized fixed effects partial likelihood, where the marginal distribution of the frailty terms determines the penalty term. The tuning parameters of the penalty function, i.e., the frailty variances, are estimated by maximizing an approximate profile likelihood. The performance of the approximation is evaluated by simulation, and the frailty model is fitted to the hip replacement data.

KEY WORDS: Dependent failure times; Frailty; Multivariate survival models; Penalized partial likelihood; Proportional hazards model; Random effects model; Variance components.

1. Introduction

Survival models have been extensively used in medical research during the last 25 years. Especially the Cox proportional hazards model has been widely applied to describe the effect of exposure and other covariates on patient's survival. One of the basic assumptions in the Cox model is independence of survival times from one observational unit to another given the current time and observed values of covariates. However, often all relevant covariates have not been observed and this assumption is not plausible. Units may be clustered in a way that introduces association of survival probabilities within a cluster. If, e.g., the sampling unit is a patient, groups of patients may have unobserved genetic or environmental factors in common. Similarities within groups may also arise through the sampling scheme by which the patients have been selected. Alternatively, if the focus is on several events for the same person, then an individual patient may form a group and one would expect within-individual correlation between events. This type of data is often called recurrent event data. Ignoring such dependence in the analysis affects the estimation of the relation between hazards, causing convergence of the hazards and therefore nonproportionality (Hougaard, 1991). The precision of the parameter estimates in the model is also overestimated (Wei, Lin, and Weissfeld, 1989).

Multivariate survival models can take into account within-cluster dependencies. A large bulk of the research to date focuses on random effects survival models or frailty models, where the event times are assumed to be independent conditional on unobserved frailty terms. In a proportional hazards frailty model, it is assumed that conditional unit-specific hazards are proportional over time. The frailties are unobserved random variables assumed to follow a probability distribution, the shape of which is described with a few parameters. The shared frailty model with gamma distributed frailties acting multiplicatively on hazard has been studied in several articles (Vaupel, Manton, and Stallard, 1979; Clayton and Cuzick, 1985; Klein, 1992; Andersen et al., 1993). Hougaard (1995) discussed the impact of different frailty distributions in a shared frailty model.

In order to describe more complicated dependencies, the shared frailty model needs to be extended. One may, e.g., wish to incorporate dependencies that reflect genetic structures between members in a family, or one may wish to use a hierarchical nested variance component model. Another limitation of the shared frailty model is the restriction to only positively correlated survival times. Petersen, Andersen, and Gill (1996) and Korsgaard and Andersen (1998) used a multiplicative gamma-distributed frailty model, where frailties were divided into additive independent gamma-distributed components. This model, however, results in a complicated likelihood. An alternative approach would be to follow Breslow and Clayton (1993), who used penalized quasi-likelihood estimation for generalized linear mixed models with normally

distributed random effects. We use a similar approximate inference procedure for frailty models by assuming multivariate multiplicative frailties drawn from a multivariate log-normal distribution. We thus get a flexible specification of frailty variance components, and we can handle negative dependencies. Our procedure generalizes the results of Therneau and Grambsch (2000), who showed an exact connection between the shared gamma frailty model and a penalized likelihood procedure. Our paper links together the likelihood approximation of Breslow and Clayton (1993) and the penalized likelihood theory (Green, 1987) and uses this to derive a generalization of the shared frailty model estimation of McGilchrist (1993).

The organization of the article is as follows. Section 2 motivates the theoretical problem through a hip replacement dataset. Section 3 describes the general frailty model. In Section 4.1, the use of penalized partial likelihood in the estimation of fixed and random effects is motivated, and the estimation of variance components is described in Section 4.2. A simulation study to investigate properties of the estimation procedure is presented in Section 5 along with an application to a hip fraction dataset. Finally, some weaknesses and suggestions for further study are given in Section 6.

2. Total Hip Endoprosthesis Data

Turula (1989) collected data on all primary total hip endoprostheses that had been inserted from 1967 to 1986 at the Orthopaedic Hospital of the Invalid Foundation in Finland. A total of 2976 patients who had undergone 3562 consecutive primary total hip athroplasties were followed up to the end of year 1988.

To investigate the factors affecting the survival of the hip endoprostheses, patient-, hip-, and operation-specific covariates were collected. Over 20% of the patients received bilateral implants and 30% of the patients had more than one operation, the maximum number of operations per patient being eight during the follow-up. Bilateral implants and possible renewal implants in one patient induce clustering in the dataset, which has to be taken into account in the modeling procedure. The dependencies between renewals may also differ for left and right hips, which motivates the need for a more flexible frailty structure than a simple shared frailty model.

3. The Frailty Model

Let T_i for unit $i = 1, \ldots, n$ denote the event time, C_i the censoring time, $U_i = \min(T_i, C_i)$, and $\delta_i = I_{\{T_i \leq C_i\}}$. Given a q-vector of random effects, or frailties $\boldsymbol{b}$, the event times are assumed independent with the conditional hazard function for unit i of the proportional hazards (Cox, 1972) form,

$$\lambda_i(t \mid \boldsymbol{b}) = \lambda_0(t) \exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}), \tag{1}$$

with $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ vectors of explanatory variables. The model is completed by assuming that jointly $\boldsymbol{b} \sim p(\boldsymbol{b}; \boldsymbol{D})$, with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{\theta})$, with $\boldsymbol{\theta}$ denoting a vector of unknown parameters.

The model includes as a special case the gamma-Cox shared frailty model (Andersen et al., 1993), in which the elements of $\boldsymbol{b}$ are independent and follow $\log(\text{gamma}(1/\theta, 1/\theta))$. Model (1) can be extended to include time–varying covariates or frailties or the baseline hazard can be modeled parametrically, but these extensions are not considered here.

4. Approximate Likelihood

4.1 *Penalized Partial Likelihood*

If conditionally on $\boldsymbol{b}$ the censoring is independent and noninformative also of $\boldsymbol{b}$, then the likelihood for model (1) in terms of the parameters $(\lambda_0(t), \boldsymbol{\beta}, \boldsymbol{\theta})$ is

$$\begin{aligned} L(\lambda_0(t), \boldsymbol{\beta}, \boldsymbol{\theta}) &= \int \prod_{i=1}^{n} \lambda_i(t \mid \boldsymbol{b})^{\delta_i} S_i(t \mid \boldsymbol{b})\, p(\boldsymbol{b}; \boldsymbol{D}(\boldsymbol{\theta}))\, d\boldsymbol{b} \\ &= \int \prod_{i=1}^{n} [\lambda_0(t) \exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b})]^{\delta_i} \\ &\qquad \times \exp[-\Lambda_0(t) \exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b})] \\ &\qquad \times p(\boldsymbol{b}; \boldsymbol{D}(\boldsymbol{\theta}))\, d\boldsymbol{b}, \end{aligned} \tag{2}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ and the unobserved frailties are integrated out.

We restrict $\boldsymbol{b}$ to follow a multivariate normal distribution, but the derived likelihood approximations can be easily adapted to other frailty distributions as well. With normally distributed frailties, the integral becomes difficult to solve. We therefore follow Breslow and Clayton (1993) in their approach for the generalized linear mixed model and use a Laplace approximation for the integral in (2). As shown in Appendix A, this leads to the approximate marginal log likelihood

$$\begin{aligned} & l(\lambda_0(t), \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &\quad \approx -\frac{1}{2} \log |\boldsymbol{D}(\boldsymbol{\theta})| \\ &\qquad - \frac{1}{2} \log \left| \sum_{i=1}^{n} \Lambda_0(t) \exp\left(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\tilde{\boldsymbol{b}}\right) \boldsymbol{Z}_i\boldsymbol{Z}_i' - \boldsymbol{D}(\boldsymbol{\theta})^{-1} \right| \\ &\qquad + \sum_{i=1}^{n} \delta_i \left[\log(\lambda_0(t)) + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\tilde{\boldsymbol{b}}\right] \\ &\qquad\qquad - \Lambda_0(t) \exp\left(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\tilde{\boldsymbol{b}}\right) - \frac{1}{2}\tilde{\boldsymbol{b}}' \boldsymbol{D}(\boldsymbol{\theta})^{-1} \tilde{\boldsymbol{b}}. \end{aligned} \tag{3}$$

If both $\boldsymbol{\theta}$ were known and $\boldsymbol{b}$ were considered a fixed effects parameter, then the second line in (3) would be a penalized log likelihood (Green, 1987), where $-(1/2)\boldsymbol{b}'\boldsymbol{D}(\boldsymbol{\theta})^{-1}\boldsymbol{b}$ is the penalty term penalizing for extreme values of $\boldsymbol{b}$. Since the combined third and fourth lines are the full likelihood for a Cox model with $\boldsymbol{b}$ as another set of parameters and a penalty term, it turns out that it can be maximized using penalized fixed effects partial likelihood (PPL),

$$\begin{aligned} & \sum_{i=1}^{n} \delta_i \left((\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}) - \log \sum_{j \in R(t_i)} \exp(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{b}) \right) \\ &\qquad - \frac{1}{2}\boldsymbol{b}' D(\boldsymbol{\theta})^{-1} \boldsymbol{b} \end{aligned} \tag{4}$$

as shown in Appendix B. If we ignore the terms on the first two lines of (3) and use the PPL to estimate $(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\boldsymbol{b}}(\boldsymbol{\theta}))$, then the estimation procedure becomes simple, but some information may be lost. The characteristics of the information loss are evaluated in a simulation study in Section 5.

For given $\boldsymbol{\theta}$, the estimating equations based on the first partial derivatives of the PPL are, for $\boldsymbol{\beta}$,

$$\sum_{i=1}^{n} \delta_i \left[\boldsymbol{X}_i - \frac{\boldsymbol{X}_i \exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b})}{\sum_{j \in R(t_i)} \exp(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{b})} \right] = 0, \quad (5)$$

and for $\boldsymbol{b}$,

$$\sum_{i=1}^{n} \delta_i \left[\boldsymbol{Z}_i - \frac{\boldsymbol{Z}_i \exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b})}{\sum_{j \in R(t_i)} \exp(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{b})} \right] - \boldsymbol{D}(\boldsymbol{\theta})^{-1}\boldsymbol{b} = 0, \quad (6)$$

and $(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\boldsymbol{b}}(\boldsymbol{\theta}))$ can be found by alternating between solving (5) and (6). Note that equation (5) can be solved with standard Cox regression software using estimated values of frailties as the offset term. The inverse of the minus second partial derivative matrix can be used as an approximate covariance matrix for $\hat{\boldsymbol{\beta}}$. However, because the uncertainty in the estimation of $\boldsymbol{\theta}$ is ignored, these estimates of covariance are naive and underestimate the true variance of $\hat{\boldsymbol{\beta}}$. The seriousness of the underestimation of $\text{var}(\hat{\boldsymbol{\beta}})$ is studied in the simulation study in Section 5.

4.2 *Variance Component Estimation*

If we assign the maximized values $(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\boldsymbol{b}}(\boldsymbol{\theta}))$ of the PPL into (3), we get an approximate profile likelihood function for $\boldsymbol{\theta}$,

$$l(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) \approx -\frac{1}{2}\log|\boldsymbol{D}(\boldsymbol{\theta})| - \frac{1}{2}\log\left|K''(\hat{\boldsymbol{b}})\right| - \frac{1}{2}\hat{\boldsymbol{b}}'\boldsymbol{D}(\boldsymbol{\theta})^{-1}\hat{\boldsymbol{b}}, \quad (7)$$

where $K''(\hat{\boldsymbol{b}})$ is derived in Appendix A, given $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}; \boldsymbol{b} = \hat{\boldsymbol{b}}$.

Now we propose the use of $K''_{PPL}(\hat{\boldsymbol{b}}) = (\partial^2 PPL)/(\partial \boldsymbol{b} \partial \boldsymbol{b}')$ instead of $K''(\hat{\boldsymbol{b}})$ and motivate this choice by the former performing better in the simulations and it being a side product of the previous step in the algorithm.

After differentiation and some simplification, an estimating equation for $\boldsymbol{\theta}$ is

$$-\frac{1}{2}\left[\text{tr}\left(\boldsymbol{D}^{-1}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}}\right) + \text{tr}\left(K''_{PPL}(\hat{\boldsymbol{b}})^{-1}\frac{\partial \boldsymbol{D}^{-1}}{\partial \boldsymbol{\theta}}\right) - \hat{\boldsymbol{b}}'\boldsymbol{D}^{-1}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}}\boldsymbol{D}^{-1}\hat{\boldsymbol{b}}\right] = 0. \quad (8)$$

The corresponding Fisher information matrix, derived by differentiating (7) twice and taking the expectation with respect to $\boldsymbol{b}$, is

$$\mathcal{J} = \frac{1}{2}\left[\text{tr}\left(\boldsymbol{D}^{-1}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}}\boldsymbol{D}^{-1}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}} + \boldsymbol{D}^{-1}\frac{\partial^2 \boldsymbol{D}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) + \text{tr}\left(K''_{PPL}(\hat{\boldsymbol{b}})^{-1}\frac{\partial \boldsymbol{D}^{-1}}{\partial \boldsymbol{\theta}}K''_{PPL}(\hat{\boldsymbol{b}})^{-1}\frac{\partial \boldsymbol{D}^{-1}}{\partial \boldsymbol{\theta}} - K''_{PPL}(\hat{\boldsymbol{b}})^{-1}\frac{\partial^2 \boldsymbol{D}^{-1}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right)\right].$$

Note that, in a shared frailty model with i.i.d. frailty terms, $\boldsymbol{D}(\boldsymbol{\theta})$ is a diagonal $(q \times q)$ matrix with elements θ, and the solution to (8) is of the simple form

$$\hat{\theta} = \frac{\hat{\boldsymbol{b}}'\hat{\boldsymbol{b}} + \text{tr}(K''_{PPL}(\hat{\boldsymbol{b}})^{-1})}{q}. \quad (9)$$

The variance for $\hat{\theta}$ in the shared frailty model is then

$$\text{var}(\hat{\theta}) = 2\hat{\theta}^2\left[q + \frac{1}{\hat{\theta}^2}\text{tr}\left(K''_{PPL}(\hat{\boldsymbol{b}})^{-1}K''_{PPL}(\hat{\boldsymbol{b}})^{-1}\right) - \frac{2}{\hat{\theta}}\text{tr}\left(K''_{PPL}(\hat{\boldsymbol{b}})^{-1}\right)\right]^{-1}.$$

The estimate and its variance equal the ones presented by McGilchrist (1993).

4.3 *Computation*

The maximization of the approximate likelihood (3) is done in two steps. First, an initial value for $\boldsymbol{\theta}$ is guessed and (5) and (6) are solved using the Newton–Raphson technique for fixed $\boldsymbol{\theta}$. Then $\boldsymbol{\beta}$ and $\boldsymbol{b}$ are fixed at the values obtained and (8) is solved to find a new value for $\boldsymbol{\theta}$. Then (5) and (6) are solved again and the two steps are iterated until convergence. After they converge, the cumulative baseline hazard can be estimated using Breslow's estimator (Breslow, 1974).

5. Simulation and Application

5.1 *Simulations*

The performance of the likelihood approximation is evaluated in three sets of simulations: (i) a shared frailty model without fixed effects, (ii) a shared frailty model with fixed covariates, and (iii) a hierarchical frailty model.

(i) The first simulation study looks at the performance of frailty variance estimates in the absence of fixed effects. Bivariate datasets were created from a proportional hazards model $t_{ij} = \exp(-\boldsymbol{b}_i)\epsilon_{ij}$, $i = 1, \ldots, 100$, $j = 1, 2$, with $\epsilon_{ij} \sim \exp(0.1)$. Frailties were drawn from a normal distribution with a fixed variance. Three different frailty variances were used, $\theta = 0.2, 0.5$, or 1, and 500 datasets were created for each value of θ.

Results from the simulation can be seen in Table 1. The frailty variances are slightly overestimated. The relative bias is biggest for the smallest variance value. The estimated standard errors for the variances are clearly underestimated.

(ii) For the shared frailty model with three fixed covariates, two types of datasets were created from a proportional hazards model $t_{ij} = \exp(-\boldsymbol{X}_{ij}\beta - \boldsymbol{b}_i)\epsilon_{ij}$, $i = 1, \ldots, s$, $j = 1, \ldots, t$, with $\epsilon_{ij} \sim \exp(0.1)$. The first type had $s = 100$ and $t = 2$ and the second type had $s = 50$ and $t = 4$. The fixed covariates were $\boldsymbol{X}_{ij} = (X_{ij1}, X_{i2}, X_{i3})'$, with $X_{ij1} \sim \text{N}(0, 1)$ that varies within clusters, $X_{i2} \sim \text{N}(0, 1)$ constant within clusters but varies from one cluster to another, and X_{i3} a binary treatment allocation covariate on the cluster level, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3) = (1, -0.7, 0.5)$. Three different values for the frailty variance were used, i.e., $\theta = 0.5, 1$, or 2, and 20% of the event times were independently right censored. Again, 500 datasets were created for each θ value.

Results for the shared frailty model simulations are presented in Tables 2 and 3. Table 2 shows results from the simulation of the bivariate cluster model and Table 3 from the simulation with four event times per cluster. In the shared frailty model simulations, the parameters $(\hat{\boldsymbol{\beta}}, \hat{\theta})$ show

Table 1

Estimated frailty variances and their estimated and empirical standard errors in 500 simulations based on a bivariate shared frailty model with frailty variances $\theta = 0.2$, 0.5, or 1

Parameter	Correct value	Mean	Estimated SE	Empirical SE
$\hat{\theta}$	0.2	0.223	0.0913	0.1852
$\hat{\theta}$	0.5	0.532	0.1512	0.2745
$\hat{\theta}$	1.0	1.073	0.2323	0.4083

slight bias toward zero. The bias for $\hat{\beta}$ is connected to the underestimation of the variance component θ. In bivariate simulations, the larger the frailty variance, the bigger the bias in its estimation, and this is reflected in the slight underestimation of $\hat{\beta}$. The bias diminishes considerably in Table 3 when the number of replicates per cluster grows to four. The frailty variance is well estimated, and due to this, the fixed effects are also well estimated.

The estimated standard errors for $\hat{\beta}$ are slightly underestimated, as expected since the variation in the estimation of θ is ignored. This is reflected in the rate by which the 95% confidence interval covers the true parameter value. These hit rates are presented in the last column of Tables 2 and 3. In Table 2, they vary from 83.8 to 95.0% and in Table 3 between 90.4 and 95.0%. The estimated standard error for the frailty variance estimate $\hat{\theta}$ is substantially underestimated, with the value of about 75% of the right level.

We also calculated correlations between the parameter estimates $(\hat{\beta}, \hat{\theta})$ calculated over the simulations and found that the association is largest between the frailty variance estimate and the fixed effect $\hat{\beta}_1$ varying within clusters. The standard error for $\hat{\beta}_1$ was also most underestimated.

Table 2

Estimated parameters and their estimated and empirical standard errors and hits on the 95% confidence intervals in 500 simulations based on a bivariate shared frailty model with frailty variances $\theta = 0.5$, 1, or 2

Parameter	Correct value	Mean	Estimated SE	Empirical SE	Hits (%)
		$\theta = 0.5$			
$\hat{\beta}_1$	1	0.987	0.118	0.144	89.0
$\hat{\beta}_2$	−0.7	−0.689	0.123	0.133	92.2
$\hat{\beta}_3$	0.5	0.481	0.178	0.176	95.0
$\hat{\theta}$	0.5	0.538	0.176	0.267	79.4
		$\theta = 1$			
$\hat{\beta}_1$	1	0.954	0.122	0.139	90.4
$\hat{\beta}_2$	−0.7	−0.668	0.140	0.153	90.8
$\hat{\beta}_3$	0.5	0.471	0.181	0.186	94.0
$\hat{\theta}$	1	0.962	0.243	0.371	79.4
		$\theta = 2$			
$\hat{\beta}_1$	1	0.921	0.126	0.149	83.8
$\hat{\beta}_2$	−0.7	−0.643	0.167	0.182	89.8
$\hat{\beta}_3$	0.5	0.469	0.185	0.188	95.0
$\hat{\theta}$	2	1.766	0.355	0.598	47.0

Table 3

Estimated parameters and their estimated and empirical standard errors and hits on the 95% confidence intervals in 500 simulations based on a shared frailty model with four repetitions per cluster and with frailty variances $\theta = 0.5$, 1, or 2

Parameter	Correct value	Mean	Estimated SE	Empirical SE	Hits (%)
		$\theta = 0.5$			
$\hat{\beta}_1$	1	0.976	0.113	0.127	91.2
$\hat{\beta}_2$	−0.7	−0.673	0.139	0.145	91.0
$\hat{\beta}_3$	0.5	0.476	0.174	0.171	95.0
$\hat{\theta}$	0.5	0.493	0.168	0.237	86.4
		$\theta = 1$			
$\hat{\beta}_1$	1	0.978	0.116	0.118	93.2
$\hat{\beta}_2$	−0.7	−0.690	0.173	0.179	93.6
$\hat{\beta}_3$	0.5	0.491	0.177	0.188	92.2
$\hat{\theta}$	1	0.978	0.268	0.346	84.4
		$\theta = 2$			
$\hat{\beta}_1$	1	0.964	0.117	0.128	90.4
$\hat{\beta}_2$	−0.7	−0.666	0.224	0.232	93.8
$\hat{\beta}_3$	0.5	0.483	0.178	0.186	93.4
$\hat{\theta}$	2	1.928	0.462	0.600	84.8

Additionally, we performed a set of shared frailty simulations with 10 event times per cluster with 50 clusters and the results were quite similar to Table 3 with only four repetitions per cluster, indicating that little additional bias reduction for fixed effects parameters or for their standard error estimates is achieved by large cluster sizes. However, the estimated standard errors for $\hat{\theta}$ were closer to the correct values.

(iii) For the hierarchical frailty model, each dataset has 50 independent clusters that contain two subclusters (call them left and right) and each subcluster has two (potentially censored) event times. In total, each dataset again has 200 event or censoring times. Here $\boldsymbol{b}_1$ is a vector of frailties on the cluster level, $\boldsymbol{b}_2$ is a vector of frailties in the left subcluster and $\boldsymbol{b}_3$ in the right subcluster. Thus, $\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ and each $\boldsymbol{b}_j, j = 1, 2, 3$ is a 1×50 vector with elements that are drawn independently from $\mathrm{N}(0, \theta_j)$.

Table 4

Estimated parameters and their estimated and empirical standard errors in 500 simulations based on a hierarchical frailty model

Parameter	Correct value	Mean	Estimated SE	Empirical SE
$\hat{\beta}_1$	1	0.921	0.116	0.139
$\hat{\beta}_2$	−0.7	−0.691	0.131	0.157
$\hat{\beta}_3$	0.5	0.511	0.235	0.249
$\hat{\theta}_1$	0.2	0.093	0.082	0.156
$\hat{\theta}_2$	0.5	0.401	0.209	0.352
$\hat{\theta}_3$	1	0.865	0.312	0.457

Table 5
Model fits to the hip replacement data

Variable	$\hat{\beta}$ (SE) Model 1	Model 2	Model 3
Age	−0.0405 (0.0237)	−0.0439 (0.0258)	−0.0420 (0.0243)
Weight	0.0093 (0.0094)	0.0108 (0.0103)	0.0093 (0.0096)
Pay class	0.0257 (0.1977)	0.0892 (0.2152)	0.0280 (0.2034)
Operation score	0.6706 (0.1690)	0.6349 (0.1776)	0.6628 (0.1715)
Type 2	−0.4177 (0.2586)	−0.4379 (0.2841)	−0.4283 (0.2664)
Type 3	−0.8475 (0.3519)	−0.9278 (0.3775)	−0.8844 (0.3596)
Subject level frailty variance $\hat{\theta}_1$		0.6464 (0.2901)	0.0557 (0.3532)
Left hip frailty variance $\hat{\theta}_2$			0.1222 (0.2835)
Right hip frailty variance $\hat{\theta}_3$			0.1410 (0.283)

This model leads to the joint distribution of random effects for two clustered event times, one from the left subcluster and one from the right subcluster, with

$$\begin{pmatrix} b_1 + b_2 \\ b_1 + b_3 \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta_1 + \theta_2 & \theta_1 \\ \theta_1 & \theta_1 + \theta_3 \end{bmatrix} \right).$$

Censoring, covariates, and parameters related to the event time were similar to those in the shared frailty simulations. The frailty variances were set to be $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (0.2, 0.5, 1)$. This leads to a correlation of 0.218 between the frailties.

The simulation results are given in Table 4. The overall results are similar to those in the shared frailty simulations. The estimates of $\boldsymbol{\beta}$ are very close to the correct ones. Their standard errors are again underestimated, but the difference between the empirical standard errors and the estimated ones are small, showing similar underestimation as in the shared frailty simulations. The variance components are also estimated fairly well.

5.2 *Hip Endoprosthesis Data*

The estimation procedure is illustrated using a subset of the hip fraction data collected at the Orthopaedic Hospital of the Invalid Foundation in Finland and described in Section 2. The data of all primary total hip endoprostheses inserted from 1967 to 1986 were followed up until the end of 1988. Here a subset of 562 male patients who had undergone 826 consecutive primary total hip arthroplasties is used. Failure time for a prosthesis is defined as the time from the operation to the replacement or removal of the prosthesis. Death of a patient is treated as censoring.

Two patient-specific covariates (age and weight at the time of operation) and three operation/hip-specific covariates (indicator of pay class, prosthesis type, and an operation score measuring the difficulty of the operation) are used in the model. Two different frailty structures are fitted: a shared frailty model with patient-specific independent frailties and a hierarchical frailty model with independent frailties on the patient level and separately on the right and left hips, allowing different variance components within left and right hip prostheses of a patient. The estimates from these models are compared with a standard Cox model.

Results from these models are presented in Table 5. Model 1 is a Cox model ignoring the dependencies in the data, model 2 is a shared frailty model with i.i.d. frailties for each patient, and model 3 is a hierarchical frailty model allowing different frailties for left and right hips on top of the patient-specific frailties when there was more than one operation per hip side.

In model 1, age at operation, operation score, and prosthesis type seem to affect the survival. Age has a positive effect on the survival of a prosthesis, while the more difficult the operation, the shorter the survival. When patient level frailties are added in model 2, the effects seem to grow except for the operation score. The frailty variance is estimated to be 0.65. In model 3, the hip side-specific frailties are added. The estimates for the hipside-specific frailty variances are very close to each other (0.12 and 0.14 for the left- and right-side variation, respectively). The patient-level frailty variance θ_1 is estimated to be very small, if not zero.

A noteworthy feature of the fitted models is that the more flexible variance component structure in model 3 gives estimates closer to those in model 1. The tight model structure for the dependence in the shared frailty model 2 may well induce model misspecification bias in the fixed effects estimates.

6. Discussion

The estimation in this general frailty model framework is done by using a simple two-step procedure where the fixed effects and the individual frailty terms are estimated keeping the frailty variance parameter fixed. This procedure leads to simple estimating equations but results in an underestimation of the estimated variances of the fixed effects parameters because the variance estimation does not take into account the variability in the estimation of $\hat{\boldsymbol{\theta}}$. This is a problem known in the mixed model literature and is also seen in the simulations presented here. One solution to correct this is via (semi)parametric bootstrap (Laird and Louis, 1985), but then additional assumptions concerning, e.g., the censoring have to be introduced.

Acknowledgements

This work was inspired by Terry Therneau (personal communications) and started while the first author visited the Department of Biostatistics at the University of Copenhagen. Per Kragh Andersen and Jørgen Holm Petersen gave helpful comments at early stages of the work. We would also like to thank an associate editor for clarifying comments. The research was supported by the Academy of Finland, Yrjö Jahnsson Foundation and the Nordic Academy for Advanced Study.

Résumé

Il existe un nombre croissant d'articles dans la littérature sur l'estimation des modèles multiplicatifs de distribution gamma ayant une fragilité partagée. Cependant, modéliser des structures de fragilité plus complexe est souvent nécessaire, mais nous sommes confrontés à des difficultés lorsque l'on essaie d'étendre les fragilités gamma. Notre travail a été motivé par des données concernant des données de prothèses de hanches ayant une structure de dépendance plus compliquée, et nous proposons un modèle basé sur des fragilités multiplicatives avec une distribution log-normal jointe. Nous donnons une justification ainsi qu'une procédure d'estimation pour ce modèle de fragilité structuré de façon générale, qui est une généralisation du modèle proposé par McGilchrist (1993, *Biometrics* **49,** 221–225). L'estimation est basée sur une approximation de Laplace de la fonction de vraisemblance. Cette procédure mène à l'estimation d'équations basée sur une vraisemblance pénalisée partielle d'effets fixes, où la distribution marginale des termes de fragilité déterminent le terme de pénalité. Les paramètres de réglages de la fonction de pénalité, c'est à dire les variances des fragilités, sont estimés par une maximisation de la vraisemblance approximative de "profile". La performance de l'approximation est évaluée par simulation et le modèle de fragilité est appliqué aux données de prothèses de hanches.

References

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30,** 89–99.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association* **88,** 9–25.

Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A* **148,** 82–117.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression model. *International Statistical Review* **55,** 245–259.

Hougaard, P. (1991). Modelling heterogeneity in survival data. *Journal of Applied Probability* **28,** 695–701.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis* **1,** 255–273.

Johansen, S. (1983). An extension of Cox's regression model. *International Statistical Review* **51,** 158–262.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48,** 795–806.

Korsgaard, I. R. and Andersen, A. H. (1998). The additive genetic gamma frailty model. *Scandinavian Journal of Statistics* **25,** 255–269.

Laird, N. M. and Louis, T. A. (1985). Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association* **82,** 739–757.

McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics* **49,** 221–225.

Petersen, J., Andersen, P., and Gill, R. (1996). Variance components models for survival data. *Statistica Neerlandica* **50,** 193–211.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model.* New York: Springer-Verlag.

Turula, K. B. (1989). Factors affecting the survival of the total hip endoprosthesis (in Finnish). Ph.D. thesis, Department of Public Health, University of Helsinki.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16,** 439–454.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Received February 1999. *Revised May* 2000.
Accepted May 2000.

Appendix A

Derivation of the Approximate Marginal Likelihood

We follow Breslow and Clayton (1993) in their approximation for the generalized linear mixed model. Laplace's method for integral approximation allows the marginal log likelihood to be approximated by

$$\begin{aligned} l(\lambda_0(t), \boldsymbol{\beta}, \boldsymbol{\theta}) &= \log(L(\lambda_0(t), \boldsymbol{\beta}, \boldsymbol{\theta})) \\ &\approx -\frac{1}{2}\log|\boldsymbol{D}(\boldsymbol{\theta})| - \frac{1}{2}\log\left|K''(\tilde{\boldsymbol{b}})\right| - K(\tilde{\boldsymbol{b}}), \quad (10) \end{aligned}$$

where

$$\begin{aligned} K(\tilde{\boldsymbol{b}}) = -\Bigg[&\sum_{i=1}^{n} \delta_i \left[\log(\lambda_0(t)) + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\tilde{\boldsymbol{b}}\right] \\ &- \Lambda_0(t)\exp\left(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\tilde{\boldsymbol{b}}\right) - \frac{1}{2}\tilde{\boldsymbol{b}}' D(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{b}}\Bigg], \end{aligned}$$

and $\tilde{\boldsymbol{b}} = \tilde{\boldsymbol{b}}(\boldsymbol{\beta}, \boldsymbol{\theta})$ denotes the solution to the partial derivatives of $K(\boldsymbol{b})$ with respect to $\boldsymbol{b}$, i.e.,

$$\begin{aligned} K'(\tilde{\boldsymbol{b}}) &= \frac{\partial K(\tilde{\boldsymbol{b}})}{\partial \tilde{\boldsymbol{b}}} \\ &= -\sum_{i=1}^{n} \boldsymbol{Z}_i \left[\delta_i - \Lambda_0(t)\exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\tilde{\boldsymbol{b}})\right] - \boldsymbol{D}(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{b}} \\ &= 0. \quad (11) \end{aligned}$$

The set of second partial derivatives of $K(\boldsymbol{b})$ with respect to $\boldsymbol{b}$ is denoted $K''(\boldsymbol{b})$ and has the form

$$K''(\tilde{\boldsymbol{b}}) = \sum_{i=1}^{n} \Lambda_0(t) \exp\left(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\tilde{\boldsymbol{b}}\right) \boldsymbol{Z}_i\boldsymbol{Z}_i' + \boldsymbol{D}(\boldsymbol{\theta})^{-1}. \quad (12)$$

Appendix B

From Penalized Likelihood to Penalized Partial Likelihood

The logarithm of the penalized likelihood defined in the last two lines of (3) can be factored into two parts, i.e.,

$$\begin{aligned} L_{PEN} &= \sum_{i=1}^{n} \delta_i [\log(\lambda_0(t)) + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}] \\ &\qquad - \Lambda_0(t) \exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}) - \frac{1}{2}\boldsymbol{b}'D(\boldsymbol{\theta})\boldsymbol{b} \\ &= \sum_{i=1}^{n} \delta_i \Bigg[\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b} \\ &\qquad - \log \sum_{j \in R(t_i)} \exp(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{b}) \Bigg] - \frac{1}{2}\boldsymbol{b}'D(\boldsymbol{\theta})\boldsymbol{b} \\ &\quad + \sum_{i=1}^{n} \delta_i \Bigg[\log(\lambda_0(t)) \\ &\qquad + \log \sum_{j \in R(t_i)} \exp(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{b}) \Bigg] \\ &\qquad - \Lambda_0(t) \exp(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}) \\ &= PPL + h(\lambda_0(t), \boldsymbol{\beta}, \boldsymbol{b}). \end{aligned}$$

If the second factor is maximized with respect to the discretized baseline hazard with one event per time interval while keeping $(\boldsymbol{\beta}, \boldsymbol{b})$ fixed and the resulting Nelson–Aalen estimator for $\Lambda_0(t)$ is plugged into the penalized likelihood, then the penalized fixed effects partial likelihood on the first line is a likelihood profile (Johansen, 1983) for $(\boldsymbol{\beta}, \boldsymbol{b})$ if $\boldsymbol{b}$ is considered a fixed effect. With fixed $\boldsymbol{\theta}$, PPL is maximized to obtain $(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\boldsymbol{b}}(\boldsymbol{\theta}))$, which also maximizes L_{PEN}.