BIOMETRICS 57, 96–102
March 2001

Multilevel Models for Survival Analysis with Random Effects

Kelvin K. W. Yau

Department of Management Sciences, City University of Hong Kong,
Tat Chee Avenue, Kowloon, Hong Kong
email: mskyau@cityu.edu.hk

SUMMARY. A method for modeling survival data with multilevel clustering is described. The Cox partial likelihood is incorporated into the generalized linear mixed model (GLMM) methodology. Parameter estimation is achieved by maximizing a log likelihood analogous to the likelihood associated with the best linear unbiased prediction (BLUP) at the initial step of estimation and is extended to obtain residual maximum likelihood (REML) estimators of the variance component. Estimating equations for a three-level hierarchical survival model are developed in detail, and such a model is applied to analyze a set of chronic granulomatous disease (CGD) data on recurrent infections as an illustration with both hospital and patient effects being considered as random. Only the latter gives a significant contribution. A simulation study is carried out to evaluate the performance of the REML estimators. Further extension of the estimation procedure to models with an arbitrary number of levels is also discussed.

KEY WORDS: Chronic granulomatous disease; Generalized linear mixed model; Multilevel model; Random effects; Residual maximum likelihood; Survival analysis.

1. Introduction

Many kinds of data have a hierarchical or clustered structure. Such data hierarchies may be present naturally in observational studies or may be due to the design of the experiment in experimental studies. In particular, for the analysis of survival data, when the response variable is the time to occurrence of a certain event, such data hierarchies may also appear. In a chronic granulomatous disease (CGD) study (see Section 3 for details), the effectiveness of a new treatment (γ-IFN) in reducing the rate of serious infections is investigated. Since each patient may experience multiple failure events, a commonly considered random effect survival model is the two-level model that assumes an independent identically distributed frailty term for each patient. Such a frailty term represents the variation due to the heterogeneity of patients. However, as each patient belongs to one of the 13 hospitals, the variation may possibly be due to a random hospital effect as well. In this case, a three-level survival model should be considered. According to the structure of the data hierarchy, infections are defined as level 1 units, patients as level 2 units, and hospitals as level 3 units. The purpose of the study is, therefore, to look for the effect of γ-IFN in reducing the rate of infection as well as estimate the variance(s) of the clustered random effects. Ignoring such random cluster effects may result in overlooking the importance of certain cluster effects and call into question the validity of traditional statistical techniques used for studying data relationships (Goldstein, 1995).

Early work on frailty models for survival data by Hougaard (1984, 1986a,b) considered the parametric modeling of the heterogeneity between individuals in the population using the gamma, inverse Gaussian, and positive stable distributions. Various approaches have been proposed in recent years to model survival data with two-level clustering (see Clayton, 1991; Gray, 1992; Klein, 1992; Nielsen et al., 1992; McGilchrist, 1993; Lin, 1994). The aim of the current work is to generalize the two-level model of McGilchrist (1993) to survival models with an arbitrary number of levels based on the generalized linear mixed model (GLMM) methodology.

The GLMM method starts with the construction of a log likelihood analogous to the likelihood associated with the best linear unbiased prediction (BLUP) of Henderson (1975) based on the Cox partial likelihood. Estimation of regression parameters and conditionally fixed random effects are achieved by maximizing this log likelihood at the initial step of estimation. Residual maximum likelihood (REML) estimating equations of variance component parameters are obtained following the spirit of Thompson (1980) and Fellner (1986, 1987). In the context of survival analysis, for the proportional hazards regression model, the partial likelihood approach of Cox (1972, 1975) to estimate the covariate effects on failure rates has been widely used in biomedical applications. One important feature of the partial likelihood approach is that the baseline hazard function is not involved in the estimation of regression parameters and hence can be unspecified. Moreover, under mild conditions, the maximum partial likelihood estimation of the covariate effects is unbiased and asymptotically normally distributed, with variances being estimated by the information matrix. The GLMM method, as indicated by McGilchrist (1994), has an advantage in that it preserves the cancellation property of the baseline hazard function when constructing

the partial likelihood and avoids the possible complications when integrating out the random terms of the marginal likelihood expression for the analysis of multivariate survival data. Application of this approach for the analysis of multivariate survival data with different variance structures can be found in McGilchrist (1993) and Yau and McGilchrist (1998).

This section is followed by an outline of the developments of the estimation and inference procedure for a three-level hierarchical survival model using GLMM. REML estimating equations of variance component parameters and expressions of asymptotic variances are derived. In Section 3, as an illustration, such a model is applied to the chronic granulomatous disease (CGD) data on recurrent infections of Fleming and Harrington (1991), with infection observations, patients, and hospitals being considered as the three levels. A simulation study is carried out in Section 4 to evaluate the performance of the REML estimators with varied combinations of fixed effects and variance component parameters. The censoring pattern is allowed to vary from 30 to 60% censoring. Further extension of the estimation procedure to models with an arbitrary number of levels is described in Section 5. A concluding discussion is given in the final section.

2. Three-Level Hierarchical Survival Model

For the CGD study, multiple failure time data are collected for the enrolled patients. Each patient is enrolled with one among several hospitals. The failure time observations, patients, and hospitals are defined, respectively, as level 1, level 2, and level 3 units. Let T_{ijk} be the observable failure/censoring time for the kth failure time observation of the jth patient in the ith hospital. Under Cox's proportional hazards model (Cox, 1972), the hazard function is given by

$$h(t;i,j,k) = \lambda(t)\exp(\eta_{ijk}), \qquad \eta_{ijk} = x'_{ijk}\beta + U_{ij},$$

where

$i = 1,2,\ldots,b$; $j = 1,2,\ldots,m_i$; $k = 1,2,\ldots,n_{ij}$;
b = number of hospitals; m_i = number of patients in the ith hospital;
n_{ij} = random number of multiple failure time observations for the jth patient in the ith hospital;
$\Sigma_{i=1}^{b} m_i = M$ = number of patients;
$\Sigma_{j=1}^{m_i} n_{ij} = n_i$ = total number of observations in the ith hospital;
$\Sigma_{i=1}^{b} n_i = N$ = total number of observations.

x'_{ijk} is a vector of risk variables corresponding to the kth observation for the jth patient of the ith hospital, β is a vector of fixed effect parameters, and U_{ij} is the unobservable random effect. The random effect $U_{ij} = E_i + F_{ij}$. Let $e = [E_1 \;\; E_2 \;\; \cdots \;\; E_b]'$, $f_i = [F_{i1} \;\; F_{i2} \;\; \cdots \;\; F_{in_i}]'$, and $f = [f_1' \;\; f_2' \;\; \cdots \;\; f_b']'$. The distributions of e and f are taken, respectively, to be multivariate $\mathrm{N}(\mathbf{0}, \theta_2 \boldsymbol{I}_b)$ and $\mathrm{N}(0, \theta_1 \boldsymbol{I}_M)$, where $\boldsymbol{I}_b$ and $\boldsymbol{I}_M$ are identity matrices with dimensions being specified by the subscripts. Furthermore, let $u_i = [U_{i1} \;\; U_{i2} \;\; \cdots \;\; U_{im_i}]'$ and $u = [u_1' \;\; u_2' \;\; \cdots \;\; u_b']'$. u then follows a multivariate normal distribution with mean zero and variance Ω, where $\Omega = \theta_1 \boldsymbol{I}_M + \theta_2 \boldsymbol{W}$ and

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{J}_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_{m_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{J}_{m_b} \end{bmatrix}_{M\times M},$$

with $\boldsymbol{J}_{m_i}$ an $m_i \times m_i$ matrix of ones and $i = 1,2,\ldots,b$. Now θ_1 and θ_2 represent the variances of random effects with respect to patients and hospitals. For this three-level hierarchical survival model with log-normal frailty, it is more convenient to use the following parameterization. Let $\theta = \theta_1$ and $\phi = \theta_2/\theta_1$. Ω can be rewritten as

$$\Omega = \theta \boldsymbol{A} = \theta \begin{pmatrix} \boldsymbol{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{A}_b \end{pmatrix},$$

where $\boldsymbol{A}_i = \boldsymbol{I}_{m_i} + \phi \boldsymbol{J}_{m_i}$. Simplification gives the following exact relationships: $\partial \boldsymbol{A}_i/\partial\phi = \boldsymbol{J}_{m_i}$, $\boldsymbol{A}_i^{-1} = \boldsymbol{I}_{m_i} - [\phi/(1 + m_i\phi)]\boldsymbol{J}_{m_i}$, and $\partial \boldsymbol{A}_i^{-1}/\partial\phi = [-1/(1+m_i\phi)^2]\boldsymbol{J}_{m_i}$, which are the basic elements for subsequent development.

Following the work of McGilchrist (1994), estimators of β and u are found by maximizing the BLUP likelihood at the initial step and extending it to obtain REML estimators of the parameters in the variance component. The BLUP likelihood is the sum of the two components l_1 and l_2, where l_1 is the partial log likelihood of failure times, with u treated as a fixed effect term, and l_2 is the log-probability density function of u. The construction of risk sets, which is essential for partial likelihood, follows a time scale corresponding to a renewal process (Therneau and Hamilton, 1997). For each of the N failure time observations, a triplet of variables (T, D, η) corresponding to our definition of multiple failure times are given by

$$T_r = \begin{cases} \text{time to failure/censoring} \\ \quad \text{since the entry of study} & \text{for the first episode,} \\ \text{time to failure/censoring} \\ \quad \text{since the latest infection} & \text{for the other episodes,} \end{cases}$$

$$D_r = \begin{cases} 0, & \text{censored observations,} \\ 1, & \text{otherwise,} \end{cases}$$

η_r = the linear predictor of the rth observation $= x_r'\beta + U_r$, $r = 1,2,\ldots,N$. Assuming no tied observations and that the failure times have been arranged so that $T_1 < T_2 < \cdots < T_N$, we have

$$l_1 = \sum_{r=1}^{N} D_r \left\{ \eta_r - \ln \sum_{s=r}^{N} \exp(\eta_s) \right\}.$$

Here we have assumed that there are no ties and left truncation is not acceptable. Tied observations are not a problem in the subsequent application. For the handling of tied observations, the standard approach of Peto (1972) and Breslow (1974) also applies. An alternative approach involves Monte Carlo methods to find maximum likelihood estimates in the presence of tied observations (see Sinha, Tanner, and Hall, 1994). When we use time since entry in all cases but with truncation at the latest event, the partial likelihood can be modified to account for delayed entry into the risk set (Klein and

Moeschberger, 1997). The second component of the BLUP likelihood is given by

$$l_2 = -(1/2)\left\{M\ln(2\pi\theta) + \ln|\boldsymbol{A}| + \theta^{-1}u'\boldsymbol{A}^{-1}u\right\}.$$

Let $\eta = [\eta_1 \quad \eta_2 \quad \cdots \quad \eta_N]'$. We have $\eta = \boldsymbol{X}\beta + \boldsymbol{Z}u$, where $\boldsymbol{X}, \boldsymbol{Z}$ are the design matrices of β, u, respectively, after the reordering is made consistent with the ordering of the failure/censoring times. For given initial values β_0 and u_0, the Newton–Raphson iterating procedure (McGilchrist, 1994; McGilchrist and Yau, 1995) for estimating β, u is given by the following iteration equation:

$$\begin{bmatrix}\hat{\beta}\\ \hat{u}\end{bmatrix} = \begin{bmatrix}\beta_0\\ u_0\end{bmatrix} + \boldsymbol{V}^{-1}[\boldsymbol{X} \quad \boldsymbol{Z}]'(\partial l_1/\partial\eta) - \boldsymbol{V}^{-1}\begin{bmatrix}\boldsymbol{0}\\ \Omega^{-1}u_0\end{bmatrix}, \tag{1}$$

where

$$\boldsymbol{V} = \begin{bmatrix}\boldsymbol{X}'\\ \boldsymbol{Z}'\end{bmatrix}\boldsymbol{B}[\boldsymbol{X} \quad \boldsymbol{Z}] + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \Omega^{-1}\end{bmatrix}$$

and $\boldsymbol{B} = -\partial^2 l_1/\partial\eta\partial\eta'$.

The method of computation of $(\partial l_1/\partial\eta)$ and $\boldsymbol{B}$ can be found in McGilchrist (1993). The matrix $\boldsymbol{V}^{-1}$ is partitioned conformally to $\beta \mid u$ as

$$\boldsymbol{V}^{-1} = \begin{bmatrix}\boldsymbol{P} & \boldsymbol{R}\\ \boldsymbol{R}' & \boldsymbol{T}\end{bmatrix}$$

and $\boldsymbol{T}$ is partitioned conformally to $u_1 \mid u_2 \mid \cdots \mid u_b$ as $\boldsymbol{T} = \text{diag}[\boldsymbol{T}_1 \quad \boldsymbol{T}_2 \quad \cdots \quad \boldsymbol{T}_b]$. The estimated variance of $\hat{\beta}$ is given by $\boldsymbol{P}$. The approximate REML estimator of θ is obtained by solving the equation of the first-order derivative of the REML log likelihood with respect to θ (McGilchrist and Yau, 1995), viz.

$$\begin{aligned}\hat{\theta} &= M^{-1}\left\{\text{tr}\left(\boldsymbol{A}^{-1}\boldsymbol{T}\right) + \hat{u}'\boldsymbol{A}^{-1}\hat{u}\right\}\\ &= M^{-1}\left(\text{tr}\,\boldsymbol{T} + \hat{u}'\hat{u} - \sum_{i=1}^{b}\frac{\phi}{1+m_i\phi}S_i\right),\end{aligned} \tag{2}$$

where

$$S_i = \left(\sum_{i=1}^{b}U_{ij}\right)^2 + \mathbf{1}_i'\boldsymbol{T}_i\mathbf{1}_i$$

and $\mathbf{1}_i$ is a vector of ones with dimension m_i.

Differentiating the REML log likelihood with respect to ϕ, we obtain an estimating equation for ϕ (McGilchrist and Yau, 1995), viz.

$$\text{tr}\left(\boldsymbol{A}^{-1}\frac{\partial\boldsymbol{A}}{\partial\phi}\right) + \hat{\theta}^{-1}\hat{u}'\left(\frac{\partial\boldsymbol{A}'}{\partial\phi}\right)\hat{u} + \hat{\theta}^{-1}\text{tr}\left(\boldsymbol{T}\frac{\partial\boldsymbol{A}^{-1}}{\partial\phi}\right) = 0.$$

Further simplification gives

$$\hat{\theta}\sum_{i=1}^{b}\left(\frac{m_i}{1+m_i\phi}\right) - \sum_{i=1}^{b}\left\{\frac{1}{(1+m_i\phi)^2}S_i\right\} = 0.$$

Letting

$$f(\phi) = \hat{\theta}\sum_{i=1}^{b}\left(\frac{m_i}{1+m_i\phi}\right) - \sum_{i=1}^{b}\left\{\frac{1}{(1+m_i\phi)^2}S_i\right\},$$

we have

$$f'(\phi) = \hat{\theta}\sum_{i=1}^{b}\frac{-m_i^2}{(1+m_i\phi)^2} + \sum_{i=1}^{b}\left\{\frac{2m_i}{(1+m_i\phi)^3}S_i\right\}.$$

The REML estimation of ϕ is accomplished by a Newton–Raphson iterative procedure,

$$\phi_{p+1} = \phi_p - \frac{f(\phi_p)}{f'(\phi_p)}, \tag{3}$$

where ϕ_{p+1} is an updated estimate of ϕ, with the final estimate being the limit for p going to infinity.

Note that the expressions $f(\phi)$ and its first-order derivative $f'(\phi)$ can be simplified when all m_i's are equal. For the analysis of the CGD data on recurrent infections, however, we need this more general case since the number of failure episodes among patients may be different.

Parameter estimation can be done by iterative application of equations (1), (2), and (3). Specifically, our iterative scheme is

Step 1. Given initial values, use equation (1) to update $\hat{\beta}, \hat{u}$ until convergence.
Step 2. Update θ by equation (2) and ϕ by equation (3).
Step 3. Using updated estimates of θ and ϕ, repeat steps 1 and 2 until convergence.
Step 4. Find the standard errors of $\hat{\beta}, \hat{\theta}$, and $\hat{\phi}$.

Asymptotic variances of the estimators in the variance component are obtained from the inverse of the REML information matrix (McGilchrist and Yau, 1995) as follows: Let $\boldsymbol{K}_1 = \theta^{-1}\boldsymbol{T}\boldsymbol{A}^{-1}$, $\boldsymbol{K}_2 = \theta^{-1}\boldsymbol{T}(\partial\boldsymbol{A}^{-1}/\partial\phi)$, $\boldsymbol{K}_3 = \boldsymbol{A}(\partial\boldsymbol{A}^{-1}/\partial\phi)$, and

$$\text{var}\begin{bmatrix}\hat{\theta}\\ \hat{\phi}\end{bmatrix} = 2\begin{bmatrix}a_{11} & a_{12}\\ a_{12} & a_{22}\end{bmatrix}^{-1},$$

where $a_{11} = \theta^{-2}\{\text{tr}(\boldsymbol{I}_N - \boldsymbol{K}_1)^2\}$, $a_{12} = -\theta^{-1}\text{tr}\{(\boldsymbol{I}_N - \boldsymbol{K}_1)^2\boldsymbol{K}_3\}$, and $a_{22} = \text{tr}(\boldsymbol{K}_2 - \boldsymbol{K}_3)^2$. Asymptotic variances of $\hat{\theta}_1$ and $\hat{\theta}_2$ are obtained using the delta method.

Note that only the REML estimators are presented here. In fact, maximum likelihood (ML) estimators can be obtained accordingly with $\boldsymbol{T}$ being replaced by $[\boldsymbol{Z}'\boldsymbol{B}\boldsymbol{Z} + \Omega^{-1}]^{-1}$. However, simulation results of McGilchrist (1993, 1994) have shown that the REML estimation procedure gives less biased estimators of the variance component parameters in the GLMM and is preferable.

3. Application to the CGD Data

The CGD data set in Fleming and Harrington (1991) consists of a placebo-controlled randomized trial of gamma interferon (γ-IFN) in chronic granulomatous disease. The study was conducted by the International CGD Cooperative Study Group in the late 1980s. In this study, 128 patients from 13 hospitals were followed for about 1 year. The number of patients in a hospital ranges from 4 to 26. Out of the 63 patients in the treatment group, 14 patients experienced at least one infection and a total of 20 infection observations were recorded. In the placebo group, 30 patients experienced at least one infection and a total of 56 infection observations were recorded. The aim of the trial was to investigate the effectiveness of

γ-IFN in reducing the rate of serious infections in CGD patients. A detailed description of this study and the CGD data set can be found in Fleming and Harrington (1991).

Applying the three-level survival model by considering that hospitals and patients are both random, the treatment effect (γ-IFN) is estimated to be $\beta = -1.069$, with an estimated standard error 0.320. This shows that the γ-IFN significantly reduces the rate of serious infection for CGD patients. The 95% confidence interval for the hazard ratio is given by [0.183, 0.643]. The two variance parameters θ_1 and θ_2 (standard errors in brackets) are, respectively, estimated by 0.758 (0.330) and 0.025 (0.118), giving a total variation due to random effects of 0.783 on the log frailty scale. Over 95% of this variation is due to the heterogeneity between patients. The random hospital effects only contribute a small portion in explaining the overall variability. In our model, the covariance between random effects is given by the following expression:

$$\operatorname{cov}(U_{ij}, U_{yz}) = \begin{cases} 0, & i \neq y, \\ \theta_2, & i = y,\ j \neq z, \\ \theta_1 + \theta_2, & i = y,\ j = z. \end{cases}$$

For any two patients in the same hospital, the within-hospital correlation is estimated by $\hat{\theta}_2/(\hat{\theta}_1 + \hat{\theta}_2) = 0.032$.

When considering patient effects as the only random component, McGilchrist's (1993) model is used and classified as a two-level model in the current framework. The treatment effect (standard error in brackets) is -1.063 (0.321), and the estimated variance of the random effects is 0.787. The results of the treatment effect and the total residual variation of the two models agree. The likelihood ratio statistic for testing the null hypothesis $\theta_2 = 0$ is 1.04, which is not significant at a 5% level ($x^2_{1,0.05} = 3.84$), indicates that a two-level model is adequate. One should note that such a hypothesis is on the boundary of the parameter set. As an additional remark, given that there are very few multiple responses for any patient, the present analysis agrees closely with the fitting of the Cox model to the time until the first infection: the treatment effect is -1.094 with a standard error of 0.335 (Lin, 1994).

4. Simulations

A simulation study of the three-level hierarchical survival model is performed to evaluate the performance of the estimators. In every simulated data set, there are 30 patients and each patient has three multiple failure time observations. In addition, each patient is assigned to 1 of 10 clusters (hospitals), with each cluster consisting of three patients, giving $10 \times 3 \times 3 = 90$ observations in each simulated data set. Note that this design is different from the design of the CGD application. In the application, any events happening in a fixed time period are studied, compared with three observations in the simulation. The baseline hazard is taken to be $\lambda(t) = 0.1$, and for each combination of i and j, components of x_{ijk} are randomly selected as zero or one and such value is common over k. Since censoring usually occurs for multiple failure time data only at the last observations, different percentages of censoring are considered for these last observations and data are simulated with 30 and 60% censoring on a per patient basis.

Every simulation has 500 identically generated data sets. Both the number of regression parameters and the values of the variance component parameters are allowed to vary. The dimension of β is chosen to be one or four with true parameter values (0.5) and (0.5, -0.5, 0.8, -0.8), respectively. θ_1 and θ_2 are chosen to be one or two, giving four sets of combinations for simulation.

Results of the simulation are given in Table 1. REML estimates of the regression coefficient β are never appreciably biased. The two variance component parameter estimates are slightly biased. SE_1 represents the average of the standard error of estimates from the 500 simulations. SE_2 gives the standard error of the 500 simulated estimates. In all cases, good agreement between SE_1 and SE_2 for regression coefficients is observed, indicating that the reported standard error is, on average, accurate. The standard errors of the variance component parameter estimates, however, are slightly overestimated. The MSE frailty is defined as the average over simulations of the mean square error in the estimation of frailties, which compares simulated random effects (u) with estimated random effects ($\hat{u}$). This quantity increases when either the dimension of β increases or the true values of the variance parameters increase. When the percentage of censored observations changes, there is, in general, no noticeable effect on the bias of regression coefficient estimates.

In order to investigate the effect of increased sample size on estimator performance, further simulation study, with 30% censoring, using 240 ($= 15 \times 4 \times 4$) simulated observations in each simulated data set is considered. Results of the simulation are given in Table 2. When compared with Table 1, part a, it is found that estimator performance, especially the variance component parameters, improves substantially in most cases when the sample size increases. Moreover, as expected, the values of SE_1, SE_2, and MSE frailties decrease with increased sample size, giving more precise estimates of random effects and standard errors.

5. Generalization to $(K + 1)$-Level Model

It may be possible that the observations are described by a hierarchy of clusters with more than three levels. As a generalization, we consider a hierarchical survival model with $(K+1)$ levels of clustering, where K is the number of nested random components.

We adopt the notations of Section 2, with $\Omega = \operatorname{var}(u)$ being defined by $(\theta_1 \boldsymbol{W}_1 + \theta_2 \boldsymbol{W}_2 + \cdots + \theta_k \boldsymbol{W}_k)$ in l_2 and equation (1), where $\boldsymbol{W}_i$ is a block diagonal matrix with each block on the diagonal consisting of a matrix of ones, $i = 1, 2, \ldots, K$. The blocks are partitioned conformally to the $(i + 1)$th level clustering. For given θ_i's, equation (1) gives the iterative estimates of β and u. Differentiating the REML log likelihood with respect to θ_i's, we obtain the following estimating equations (McGilchrist and Yau, 1995):

$$\operatorname{tr}(\Omega - \boldsymbol{T} - uu')\left(\Omega^{-1}\boldsymbol{W}_i\Omega^{-1}\right) = 0, \qquad i = 1, 2, \ldots, K. \quad (4)$$

In general, estimation of θ_i's can proceed using a numerical method to locate the roots of equation (4). For a two-level survival model, $\operatorname{var}(u) = \theta \boldsymbol{I}_M$, equation (4) becomes $\operatorname{tr}(\theta \boldsymbol{I}_M - \boldsymbol{T} - uu') = 0$ and hence $\hat{\theta} = M^{-1}\operatorname{tr}(\boldsymbol{T} + uu')$, which mirrors the results obtained by McGilchrist (1993).

6. Discussion

In this paper, multilevel survival models are formulated by extending Cox's partial likelihood in the GLMM framework

Table 1
Estimated biases and standard errors of REML estimators for three-level hierarchical survival model[a]

Parameter	True value	Average bias (SE)	SE_1	SE_2	MSE frailties
(a) Censoring Percentage = 30%					
Simulation 1					
β	0.5	−0.015 (0.022)	0.518	0.488	
θ_1	1.0	0.120 (0.018)	0.500	0.395	
θ_2	1.0	−0.049 (0.027)	0.729	0.614	
					0.539
Simulation 2					
β	0.5	−0.010 (0.028)	0.648	0.635	
θ_1	2.0	−0.024 (0.031)	0.782	0.693	
θ_2	2.0	−0.168 (0.044)	1.297	0.980	
					0.780
Simulation 3					
β_1	0.5	0.029 (0.025)	0.555	0.549	
β_2	−0.5	−0.028 (0.024)	0.557	0.537	
β_3	0.8	0.027 (0.025)	0.562	0.550	
β_4	−0.8	−0.009 (0.025)	0.561	0.569	
θ_1	1.0	0.126 (0.015)	0.536	0.342	
θ_2	1.0	0.075 (0.031)	0.828	0.697	
					0.650
Simulation 4					
β_1	0.5	0.033 (0.031)	0.708	0.701	
β_2	−0.5	−0.003 (0.032)	0.711	0.711	
β_3	0.8	−0.004 (0.033)	0.712	0.735	
β_4	−0.8	−0.011 (0.031)	0.717	0.701	
θ_1	2.0	0.098 (0.030)	0.880	0.660	
θ_2	2.0	−0.034 (0.052)	1.451	1.165	
					1.010
(b) Censoring Percentage = 60%					
Simulation 1					
β	0.5	−0.023 (0.021)	0.506	0.469	
θ_1	1.0	0.016 (0.014)	0.484	0.314	
θ_2	1.0	−0.103 (0.025)	0.695	0.564	
					0.567
Simulation 2					
β	0.5	−0.006 (0.028)	0.651	0.618	
θ_1	2.0	−0.080 (0.027)	0.784	0.601	
θ_2	2.0	−0.246 (0.047)	1.264	1.050	
					0.803
Simulation 3					
β_1	0.5	−0.039 (0.024)	0.543	0.528	
β_2	−0.5	0.006 (0.026)	0.550	0.576	
β_3	0.8	0.002 (0.025)	0.546	0.563	
β_4	−0.8	0.030 (0.025)	0.552	0.561	
θ_1	1.0	0.043 (0.015)	0.527	0.327	
θ_2	1.0	−0.089 (0.029)	0.747	0.643	
					0.654
Simulation 4					
β_1	0.5	0.012 (0.031)	0.712	0.695	
β_2	−0.5	−0.050 (0.031)	0.710	0.698	
β_3	0.8	0.012 (0.033)	0.717	0.737	
β_4	−0.8	0.016 (0.031)	0.716	0.698	
θ_1	2.0	0.038 (0.026)	0.879	0.580	
θ_2	2.0	−0.152 (0.049)	1.396	1.106	
					1.033

[a] SE_1, average of standard error of estimates; SE_2, standard error of estimates over simulations; MSE frailties, mean square error in estimation of frailties.

Table 2
Estimated biases and standard errors of REML estimators for three-level hierarchical survival model (with increased sample size); censoring percentage = 30%[a]

Parameter	True value	Average bias (SE)	SE_1	SE_2	MSE frailties
Simulation 1					
β	0.5	0.027 (0.015)	0.327	0.344	
θ_1	1.0	−0.003 (0.012)	0.275	0.279	
θ_2	1.0	0.076 (0.021)	0.537	0.474	
					0.359
Simulation 2					
β	0.5	−0.001 (0.019)	0.429	0.423	
θ_1	2.0	−0.089 (0.023)	0.471	0.509	
θ_2	2.0	0.054 (0.036)	0.999	0.806	
					0.495
Simulation 3					
β_1	0.5	0.005 (0.016)	0.339	0.362	
β_2	−0.5	0.005 (0.015)	0.340	0.345	
β_3	0.8	0.007 (0.015)	0.343	0.339	
β_4	−0.8	−0.018 (0.016)	0.342	0.352	
θ_1	1.0	0.026 (0.013)	0.290	0.289	
θ_2	1.0	0.035 (0.020)	0.533	0.452	
					0.427
Simulation 4					
β_1	0.5	−0.033 (0.021)	0.449	0.468	
β_2	−0.5	0.030 (0.020)	0.448	0.448	
β_3	0.8	−0.012 (0.021)	0.450	0.469	
β_4	−0.8	0.014 (0.021)	0.450	0.469	
θ_1	2.0	−0.006 (0.024)	0.504	0.540	
θ_2	2.0	−0.002 (0.036)	1.001	0.794	
					0.628

[a] SE_1, average of standard error of estimates; SE_2, standard error of estimates over simulations; MSE frailties, mean square error in estimation of frailties.

to analyze survival data with nested random effects. Implementable formulas for estimating both the regression and variance component parameters are derived for a three-level hierarchical model. In addition, the estimation procedure for models with an arbitrary number of levels is outlined and is shown to be a generalization of the two-level model considered by McGilchrist (1993). The current approach, on the one hand, preserves the cancellation property of Cox's method in which the baseline hazard function is not involved in the estimation of regression coefficients and can be unspecified. On the other hand, it circumvents the high-dimensional integration that is involved in computing the marginal likelihood for survival analysis with random effects. Our simulation study has shown that the estimators for regression coefficients are unbiased and the performance of the variance component estimator improves as the sample size increases.

The GLMM method is a combination of the linear mixed model and the generalized linear model and allows one or more normally distributed random components to be added to the usual fixed effects in the linear predictor. Early work on the development of the GLMM using a penalized quasi-likelihood approach has been given by Schall (1991), Breslow and Clayton (1993), and McGilchrist (1994). In this study, we follow McGilchrist's approach of GLMM modeling in which only a computable expression of the likelihood that can be expressed in terms of the linear predictor is required, which enables the application of the method to frailty models in survival analysis. Although we have only considered hierarchical models in this paper, the method in fact can apply to non-hierarchical models. As outlined in McGilchrist (1994), such GLMM formulation can apply to frailty models in survival analysis with any correlation structure in the random component.

The GLMM method has been generalized by Lee and Nelder (1996) and was named the hierarchical generalized linear model (HGLM). It allows other families of distribution of the random effects. Under appropriate conditions, the fixed effect estimators are shown to be asymptotically equivalent to those obtained from the use of marginal likelihood and hence inherits asymptotic consistency, efficiency, and normality of the GLMM. The random effect estimates are shown to be asymptotically best unbiased predictors.

Acknowledgements

The author is grateful to Professor C. A. McGilchrist for his valuable comments and suggestions. The author would like to thank the editor, an associate editor, and four referees for helpful comments on earlier versions of this paper. This work was supported in part by grants from the Research Grants Council of Hong Kong.

Résumé

Une méthode pour modéliser des données de survie regroupées sur plusieurs niveaux est décrite. La vraisemblance partielle de Cox est incorporée dans le modèle linéaire généralisé mixte (GLMM). L'estimation des paramètres est obtenue par maximisation d'une log-vraisemblance analogue à la vraisemblance associée à la meilleure prédiction linéaire non biaisée (BLUP) à l'étape initiale de l'estimation et est étendue pour obtenir des estimations du maximum de la vraisemblance résiduelle (REML) de la variance. Les équations pour trouver les estimations dans un modèle de survie hiérarchique à 3 niveaux sont développées en détail; ce modèle est appliqué à l'analyse de données sur des infections récentes dans la granulomatose chronique où les effets hôpital et patient sont supposés aléatoires. Seul l'effet patient apporte une contribution significative au modèle. Une étude de simulation est effectuée pour évaluer la performance des estimateurs REML. La procédure d'estimation généralisée à des modèles avec un nombre arbitraire de niveaux est aussi discutée.

References

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 generalised linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Clayton, D. G. (1991). A Monte Carlo method for Baynesian inference in frailty models. *Biometrics* **47,** 467–485.

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

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics* **28,** 51–60.

Fellner, W. H. (1987). Sparse matrices, and the estimation variance components by likelihood methods. *Communications in Statistics—Simulation and Computation* **16,** 439–463.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* New York: Wiley.

Goldstein, H. (1995). *Multilevel Statistical Models.* London: Arnold.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association* **87,** 942–951.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31,** 423–447.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71,** 75–83.

Hougaard, P. (1986a). A class of multivariate failure time distributions. *Biometrika* **73,** 671–678.

Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73,** 387–396.

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

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis Techniques for Censored and Truncated Data.* New York: Springer-Verlag.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B* **58,** 619–678.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* **13,** 2233–2247.

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

McGilchrist, C. A. (1994). Estimation in generalised mixed models. *Journal of the Royal Statistical Society, Series B* **56,** 61–69.

McGilchrist, C. A. and Yau, K. K. W. (1995). The derivation of BLUP, ML, REML estimation methods for generalised linear mixed models. *Communications in Statistics—Theory and Methods* **24,** 2963–2980.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19,** 25–44.

Peto, R. (1972). Contribution to the discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B* **34,** 205–207.

Schall, R. (1991). Estimation in generalised linear mixed models with random effects. *Biometrika* **78,** 719–727.

Sinha, D., Tanner, M. A., and Hall, W. J. (1994). Maximization of the marginal likelihood of grouped survival data. *Biometrika* **81,** 53–60.

Therneau, T. M. and Hamilton, S. A. (1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine* **16,** 2029–2047.

Thompson, R. (1980). Maximum likelihood estimation of variance components. *Math. Operations—forsch. Statist., Ser. Statistics* **11,** 545–561.

Yau, K. K. W. and McGilchrist, C. A. (1998). ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine* **17,** 1201–1213.

Received October 1997. *Revised May* 2000.
Accepted June 2000.