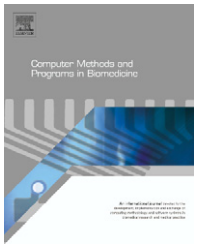
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE XXX (2007) XXX–XXX

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Multilevel survival modelling of recurrent urinary tract infections

Kui Wang[a,b], *Kelvin K.W. Yau*[b], *Andy H. Lee*[a,*], *Geoffrey J. McLachlan*[c]

[a] *School of Public Health, Curtin University of Technology, GPO Box U 1987, Perth, WA 6845, Australia*
[b] *Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China*
[c] *Department of Mathematics and Institute for Molecular Bioscience, University of Queensland, St. Lucia, Qld 4072, Australia*

ARTICLE INFO

Article history:
Received 28 October 2006
Received in revised form
16 January 2007
Accepted 31 May 2007

Keywords:
Multilevel modelling
Random effects
Residual maximum likelihood
Survival analysis
Urinary tract infections

ABSTRACT

A multilevel survival frailty model is presented for analyzing clustered and recurrent urinary tract infections among elderly women residing in aged-care institutions. At the subject level, serial dependence is expected between recurrent events recorded on the same individual. At the cluster level, correlations of observations within the same institution are present due to the inherent residential environment and hierarchical setting. Two random components are therefore incorporated explicitly within the survival frailty model to account for the simultaneous heterogeneity and autoregressive structure. A Splus computer program is developed for the estimation of fixed effect and variance component parameters.

1. Introduction

Survival frailty models are commonly used to analyse survival data in different health and biomedical settings, by assuming gamma and log-normal distributions for the random effects [1–3]. Alternatively, random effects Cox models can be defined by specifying the first and second moments of the frailty distribution [4]. The advantage of applying log-normal frailty model is its flexibility on the correlation structure for the failure time data, while keeping the interpretation of regression coefficients meaningful. For example, in order to handle time dependent correlated frailties, Yau and McGilchrist [5] proposed a log-normal frailty model incorporating an autoregressive correlation structure for the frailty term.

Multilevel models [6] are also available for handling nested survival data. A multilevel frailty model with two nested random effects has been developed, in which the random effects follow a gamma distribution [7]. For log-normal frailty, similar multilevel models [8] were considered following the generalised linear mixed modelling approach [2]. Zhang and Steele [9] proposed a semi-parametric multilevel survival model, with a non-linear effect for the continuous covariate and a linear effect for categorical covariate in the log-hazard function. Recently, Ha and Lee [10] used multilevel mixed linear models to analyse censored survival data. An application of multilevel frailty modelling of clustered grouped survival data can be found in [11] where the MCMC method is used for parameter estimation.

Our modelling of multilevel survival data is motivated by a longitudinal study of recurrent urinary tract infections sustained by a group of elderly women residing in aged-care institutions. At the facility/cluster level, all subjects from the same institution share a common random institution effect. At the subject level, repeated measurements (recurrent times)

* *Corresponding author*. Tel.: +61 8 9266 4180; fax: +61 8 9266 2958.
E-mail address: Andy.Lee@curtin.edu.au (A.H. Lee).
0169-2607/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2007.05.013

from the same individual are expected to be correlated. An autoregressive covariance structure is thus specified as part of the variance component for the recurrent times.

2. Recurrent urinary tract infections

Urinary tract infection (UTI) is one of the most common bacterial infections in women, and one in four of these women will develop a recurrence. Between 10% and 20% of women aged 60 years and over are affected by asymptomatic infection or bacteriuria [12]. Various risk factors predispose women of different age groups to recurrence [13]. The prevalence of recurrent UTI also increases for women living in nursing homes [14].

A retrospective cohort study was conducted in 2003 to determine the risk factors associated with recurrent UTI among elderly women in residential aged-care facilities [15]. Eligibility criteria for the subjects were defined to be female residents aged 60 years or above with an institutionalisation period of at least 6 months. A total of 201 subjects satisfying the selection criteria were recruited from six randomly selected aged-care institutions in Perth, Western Australia. Women residing in the same institution were likely to be correlated in terms of contracting UTI because of their exposure to the same environment [15].

It was found that 93 of the 201 women experienced at least one UTI episode during the 2 years follow-up period. For this subgroup of women, the outcome variable was taken to be the duration between successive UTI episodes. In addition to age (in years), available covariates were binary variables indicating the presence or absence of diabetes mellitus, stroke history, history of prior UTI, urinary incontinence, hysterectomy, faecal incontinence, immuno-compromised, and anatomical abnormalities of the urinary tract. Information on these variables was retrieved from records or medication charts held at each institution. The variables were chosen because they are either established or postulated risk factors for recurrent UTI [13].

3. Multilevel survival frailty model with autocorrelation

For the modelling of clustered recurrent times, let T denotes the duration between successive recurrent events or the time to end of study, with D being the associated indicator of event (1) or censor (0). Suppose T_{ijk} is the observed kth recurrent time of the jth individual nested within the ith institution, with $k = 1, 2, \ldots, n_{ij}$, $j = 1, 2, \ldots, m_i$, $i = 1, 2, \ldots, b$. Here, n_{ij} is the number of repeated observations on subject j; m_i is the number of subjects within institution i; and b is the number of randomly selected institutions. There are altogether $\sum_{i=1}^{b} m_i = M$ subjects, $\sum_{j=1}^{m_i} n_{ij} = n_i$ observations within the ith institution, and $\sum_{i=1}^{b} n_i = N$ observations in total. In this three-level hierarchical setting, conditional on unobservable institution random effect u_i and subject frailties v_{ijk}, observations (T_{ijk}, D_{ijk}) are assumed to be independent. Following the survival frailty approach [2,5,8], the proportional hazard function may be written as:

$$h(t; i, j, k) = \lambda(t)\exp(\eta_{ijk}), \quad \eta_{ijk} = x'_{ijk}\beta + u_i + v_{ijk},$$

where $\lambda(t)$ is the underlying baseline hazard, x'_{ijk} is a covariate vector corresponding to t_{ijk}, and β is the associated vector of regression coefficients. Let $u = (u_1, \ldots, u_b)'$ and $v = (v_{111}, v_{112}, \ldots, v_{211}, v_{212}, \ldots, v_{b11}, v_{b12}, \ldots)'$. The linear predictor can be expressed as:

$$\eta = X\beta + Z_1 u + Z_2 v.$$

Without loss of generality, we assume u to be normally distributed, $N(0, \sigma^2 I_b)$, independent of v. To further account for the time dependent correlated frailties, a first-order autoregressive correlation structure is adopted for the subject random effects [5,16], so that v follows a $N(0, \theta A(\phi))$ distribution, where $A = \mathrm{diag}(A_{11}, A_{12}, \ldots, A_{b1}, \ldots, A_{bm_b})$ is a block diagonal matrix with:

$$A_{ij}(\phi) = \frac{1}{1-\phi^2}\begin{pmatrix} 1 & \phi & \cdots & \phi^{n_{ij}-1} \\ \phi & 1 & \cdots & \phi^{n_{ij}-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n_{ij}-1} & \phi^{n_{ij}-2} & \cdots & 1 \end{pmatrix}.$$

The following expressions can be derived as:

$$A_{ij}^{-1} = (1+\phi^2)I_{ij} - \phi J_{ij} - \phi^2 K_{ij} \quad \text{and} \quad \mathrm{trace}\left(\frac{\partial A_{ij}^{-1}}{\partial \phi} A_{ij}\right) = -\frac{2\phi}{1-\phi^2},$$

where I_{ij}, J_{ij} and K_{ij} are $n_{ij} \times n_{ij}$ matrices; I_{ij} is the identity matrix; J_{ij} has its sub-diagonal entries ones and zeros elsewhere; K_{ij} takes on the value 1 at the first and last element of its principal diagonal and zeros elsewhere. To simplify notation, I, J and K represent the respective block diagonal matrix with element I_{ij}, J_{ij} and K_{ij}, respectively.

The best linear unbiased prediction (BLUP) log-likelihood is the sum of two components $l = l_1 + l_2$, where l_1 is the logarithm of the partial likelihood of recurrent times conditional on u and v, and l_2 is the logarithm of the probability density function of u and v, namely:

$$l_2 = -\frac{1}{2}(b\log(2\pi\sigma^2) + \sigma^{-2}u'u) - \frac{1}{2}(N\log(2\pi\theta) + \log|A| + \theta^{-1}v'A^{-1}v).$$

From now onwards, we use i as the index of observations. By sorting the recurrent event/censoring times T_i in ascending order, we have $\eta_i = x'_i\beta + z'_{1i}u + z'_{2i}v$, where x'_i is the vector of fixed covariates, while $z'_{1i}u$ and $z'_{2i}v$ return the value of u or v for the ith observation. For the above log-normal frailty model with u and v conditionally fixed:

$$l_1 = \sum_{i=1}^{N} D_i \left[\eta_i - \log\sum_{j=i}^{N}\exp(\eta_j)\right].$$

When the variance parameters σ^2, ϕ and θ are held fixed, the estimates of β, u and v are given by the Newton–Raphson

Please cite this article in press as: K. Wang et al., Multilevel survival modelling of recurrent urinary tract infections, Comput. Methods Programs Biomed. (2007), doi:10.1016/j.cmpb.2007.05.013

equation:

$$\begin{bmatrix}\tilde{\beta}\\ \tilde{u}\\ \tilde{v}\end{bmatrix} = \begin{bmatrix}\beta_0\\ u_0\\ v_0\end{bmatrix} + V^{-1}\begin{bmatrix}\partial l/\partial\beta\\ \partial l/\partial u\\ \partial l/\partial v\end{bmatrix},$$

where

$$\frac{\partial l}{\partial\beta} = X'\frac{\partial l_1}{\partial\eta}, \qquad \frac{\partial l}{\partial u} = Z_1'\frac{\partial l_1}{\partial\eta} - \sigma^{-2}u, \qquad \frac{\partial l}{\partial v} = Z_2'\frac{\partial l_1}{\partial\eta} - \theta^{-1}A^{-1}v,$$

$$\text{and } V = \begin{bmatrix}X'\\ Z_1'\\ Z_2'\end{bmatrix}\left(-\frac{\partial^2 l_1}{\partial\eta\partial\eta'}\right)\begin{bmatrix}X & Z_1 & Z_2\end{bmatrix} + \begin{pmatrix}0 & 0 & 0\\ 0 & \sigma^{-2}I_b & 0\\ 0 & 0 & \theta^{-1}A^{-1}\end{pmatrix}.$$

The method of computing $\partial l_1/\partial\eta$ and $\partial l_1/\partial\eta\partial\eta'$ can be found in Ref. [2]. The estimation procedure is iterative. Once estimates for β, u and v are obtained, then σ^2, ϕ and θ can be replaced by their REML estimates [17] derived below, which in turn are used to update the estimating equation for (β, u, v). The iterative cycle continues until all parameter estimates converge.

To obtain REML estimates for the variance components, let the block matrix V^{-1} partitioned conformally to β, u and v as:

$$V^{-1} = \begin{pmatrix}V_{11} & V_{12} & V_{13}\\ V_{21} & V_{22} & V_{23}\\ V_{31} & V_{32} & V_{33}\end{pmatrix}$$

The REML estimator of the first variance component is given by:

$$\hat{\sigma}^2 = \frac{[\text{trace}(V_{22}) + u'u]}{b}.$$

For the second variance component, the REML estimating equations for θ and ϕ are:

$$\hat{\theta} = N^{-1}[\text{trace}(A^{-1}(V_{33} + vv'))],$$

$$\text{trace}\left[\frac{\partial A^{-1}}{\partial\phi}A\right] = \hat{\theta}^{-1}\left[\text{trace}\left((V_{33} + vv')\frac{\partial A^{-1}}{\partial\phi}\right)\right].$$

The simplification is analogous to that for the log-normal survival model with correlated frailty [5], viz.:

$$\hat{\theta} = N^{-1}[(1+\phi^2)L_1 - 2\phi L_2 - \phi^2 L_3],$$

where $L_1 = \text{trace}(V_{33} + vv')$, $L_2 = (1/2)\,\text{trace}[J(V_{33} + vv')]$, and $L_3 = \text{trace}[K(V_{33} + vv')]$.

Estimation of the correlation parameter ϕ requires solving the cubic equation:

$$f(\phi) = C_1\phi^3 + C_2\phi^2 + C_3\phi + C_4 = 0,$$

where $C_1=(N-M)(L_1-L_3)$, $C_2=(2M-N)L_2$, $C_3=NL_3-(N+M)L_1$, and $C_4=NL_2$.

Standard numerical algorithms such as Newton–Raphson may be used to solve for $\hat{\phi}$.

The matrix V_{11} provides the asymptotic variance and covariance of the regression coefficients $\hat{\beta}$. Furthermore, asymptotic variances of the variance component estimators are obtained from the inverse of the REML information matrix [17], as follows:

$$\text{var}\begin{pmatrix}\hat{\sigma}^2\\ \hat{\theta}\\ \hat{\phi}\end{pmatrix} = 2\begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{12} & a_{22} & a_{23}\\ a_{13} & a_{23} & a_{33}\end{pmatrix}^{-1},$$

where

$$a_{11} = \sigma^{-4}\,\text{trace}(I_b - \sigma^{-2}V_{22})^2,$$

$$a_{12} = \sigma^{-4}\theta^{-2}\,\text{trace}(V_{23}A^{-1}V_{32}),$$

$$a_{13} = -\sigma^{-4}\theta^{-1}\,\text{trace}\left(V_{23}\frac{\partial A^{-1}}{\partial\phi}V_{32}\right),$$

$$a_{22} = \theta^{-2}\,\text{trace}(I_N - K_1)^2,$$

$$a_{23} = -\theta^{-1}\,\text{trace}[(I_N - K_1)^2K_3],$$

$$a_{33} = \text{trace}(K_2 - K_3)^2;$$

in which $K_1=\theta^{-1}V_{33}A^{-1}$, $K_2=\theta^{-1}V_{33}(\partial A^{-1}/\partial\varphi)$ and $K_3=A(\partial A^{-1}/\partial\varphi)$.

Note that the survival frailty model with autocorrelation [5] can be formulated as a special case of this multilevel survival frailty model with autocorrelation by setting the first vector of random effects $u=0$. Specifically, l_1 and l_2 can be modified as follows:

$$l_1 = \sum_{i=1}^{N} D_i\left[\eta_i - \log\sum_{j=i}^{N}\exp(\eta_j)\right],$$

$$l_2 = -\frac{1}{2}(N\log(2\pi\theta) + \log|A| + \theta^{-1}v'A^{-1}v)$$

where $\eta_i = x_i'\beta + z_{2i}'v$.

Estimation of β and v can be achieved via the Newton–Raphson equation:

$$\begin{bmatrix}\tilde{\beta}\\ \tilde{v}\end{bmatrix} = \begin{bmatrix}\beta_0\\ v_0\end{bmatrix} + V^{-1}\begin{bmatrix}\partial l/\partial\beta\\ \partial l/\partial v\end{bmatrix},$$

$$\text{where } V = \begin{bmatrix}X'\\ Z_2'\end{bmatrix}\left(-\frac{\partial^2 l_1}{\partial\eta\partial\eta'}\right)\begin{bmatrix}X & Z_2\end{bmatrix} + \begin{pmatrix}0 & 0\\ 0 & \theta^{-1}A^{-1}\end{pmatrix}.$$

Table 1 – Frequency distribution of recurrent UTI of 93 elderly women

Number of recurrent UTI	Number of women
0	31
1	23
2	10
3	9
4	5
5	5
6	5
7	2
8	0
9	2
≥10	1

Please cite this article in press as: K. Wang et al., Multilevel survival modelling of recurrent urinary tract infections, Comput. Methods Programs Biomed. (2007), doi:10.1016/j.cmpb.2007.05.013

Table 2 – Results from fitting multilevel survival and survival frailty models with autocorrelation to the recurrent UTI data

Risk factor	Multilevel survival model with autocorrelation		Survival frailty model with autocorrelation	
	Estimate (S.E.)	Hazard ratio	Estimate (S.E.)	Hazard ratio
Age	−0.005 (0.014)	0.995	−0.007 (0.014)	0.993
Diabetes	−0.064 (0.294)	0.938	−0.121 (0.312)	0.886
Stroke history	0.074 (0.245)	1.077	0.071 (0.264)	1.074
Prior UTI	0.972 (0.253)*	2.643	0.877 (0.253)*	2.404
Urinary incontinence	−0.017 (0.242)	0.983	0.154 (0.252)	1.166
Hysterectomy	0.370 (0.688)	1.448	−0.198 (0.651)	0.820
Faecal incontinence	0.879 (0.402)*	2.408	0.660 (0.434)	1.935
Immuno-compromised	0.250 (0.475)	1.284	0.201 (0.516)	1.223
Anatomical abnormalities	0.590 (0.304)	1.804	0.636 (0.325)	1.889
Institution	$\sigma^2 = 0.265$ (0.242)			
Subject	$\theta = 0.619$ (0.219)*, $\phi = 0.444$ (0.218)*		$\theta = 0.619$ (0.230)*, $\phi = 0.575$ (0.170)*	

* p-value < 0.05.

Letting the block matrix V^{-1} partitioned conformally to β and υ as:

$$V^{-1} = \begin{pmatrix} V_{11} & V_{13} \\ V_{31} & V_{33} \end{pmatrix}.$$

The matrix V_{11} provides the asymptotic variance and covariance of the regression coefficients $\hat{\beta}$. The REML estimator of the variance component is given by:

$$\hat{\theta} = N^{-1}[\text{trace}(A^{-1}(V_{33} + \upsilon\upsilon'))],$$

$$\text{trace}\left[\frac{\partial A^{-1}}{\partial \phi} A\right] = \hat{\theta}^{-1}\left[\text{trace}\left((V_{33} + \upsilon\upsilon')\frac{\partial A^{-1}}{\partial \phi}\right)\right].$$

4. Application

For the recurrent UTI study, there are altogether $N = 285$ observations from $M = 93$ elderly women nested within the $b = 6$ randomly selected institutions. The frequency distribution of recurrent UTI is given in Table 1. One third of the cohort had no recurrence during the study period. Descriptive statistics of the covariates were taken at baseline in view of the repeated measures recorded for each individual. The average age of the cohort was 85.8 (S.D. 8.4) years. About 16% of them were diabetic, 28% experienced a stroke and 34.4% had a history of prior UTI. A large proportion of women (46.2%) suffered from urinary incontinence, yet only 5.4% had faecal incontinence symptoms. The data also revealed that 13% of the cohort had anatomical abnormalities of the urinary tract, 5.4% had undergone hysterectomy but only 4.3% were immuno-compromised.

With the nine available covariates, results from fitting multilevel survival and survival frailty models with autocorrelation to the recurrent UTI data are presented in Table 2. It appears that the hazard rate of recurrent UTI is significantly associated with the subject's history of prior UTI. For both models, a positive coefficient is obtained for this risk factor, implying an increased hazard of UTI recurrence if the woman experienced UTI before (adjusted hazard ratio being 2.643 and 2.404, respectively). However, an additional significant risk factor, namely, faecal incontinence, is evident under the multilevel survival model, with an adjusted hazard ratio of 2.408. Both models demonstrate significant autocorrelation and subject frailty. In the multilevel setting, a mild institution effect is found, which explains the similarity between the two sets of results. By accommodating institutional effect in the multilevel model, the association between faecal incontinence and risk of recurrent UTI becomes more apparent.

5. Conclusion

In this study a multilevel survival frailty model is proposed for analysing clustered recurrent times. It may be considered as an extension of the log-normal frailty model for time dependent correlated frailties [5]. An REML approach is adopted to estimate the parameters of the fixed and random components. An application to the recurrent UTI study demonstrates the usefulness of the method in analysing hierarchical survival data. In particular, an additional risk factor for recurrent UTI is identified, which may have important clinical implications to control this common disease for elderly women.

There are several potential extensions of the current model. One direction is to investigate other correlation structures of random effects at each level. Another extension is to generalise the methodology to more than three levels. Because of the flexibility of the log-normal frailty model in specifying the correlation structure of the random components, extensions to other multilevel settings are feasible by modifying the estimation procedure outlined in Section 3, findings of which will be reported elsewhere.

Acknowledgements

The authors are grateful to the reviewer for helpful comments. This research was supported by grants from the Australian Research Council and the Research Grants Council of Hong Kong. The model fitting procedure has been implemented as a Splus computer program available from http://fbstaff.cityu.edu.hk/mskyau/.

Please cite this article in press as: K. Wang et al., Multilevel survival modelling of recurrent urinary tract infections, Comput. Methods Programs Biomed. (2007), doi:10.1016/j.cmpb.2007.05.013

REFERENCES

[1] C.A. McGilchrist, C.W. Aisbett, Regression with frailty in survival analysis, Biometrics 47 (1991) 461–466.
[2] C.A. McGilchrist, REML estimation for survival models with frailty, Biometrics 49 (1993) 221–225.
[3] I.D. Ha, Y. Lee, J.K. Song, Hierarchical likelihood approach for frailty models, Biometrika 88 (2001) 233–243.
[4] R. Ma, D. Krewski, R.T. Burnett, Random effects Cox models: a Poisson modelling approach, Biometrika 90 (2003) 157–169.
[5] K.K.W. Yau, C.A. McGilchrist, ML and REML estimation in survival analysis with time dependent correlated frailty, Stat. Med. 17 (1998) 1201–1213.
[6] H. Goldstein, W. Browne, J. Rasbash, Multilevel modelling of medical data, Stat. Med. 21 (2002) 3291–3315.
[7] N. Sastry, A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil, J. Am. Stat. Assoc. 92 (1997) 426–435.
[8] K.K.W. Yau, Multilevel models for survival analysis with random effects, Biometrics 57 (2001) 96–102.
[9] W. Zhang, F. Steele, A semiparametric multilevel survival model, Appl. Stat. 53 (2004) 387–404.
[10] I.D. Ha, Y. Lee, Multilevel mixed linear models for survival data, Lifetime Data Anal. 11 (2005) 131–142.
[11] M.C.M. Wong, K.F. Lam, E.C.M. Lo, Multilevel modelling of clustered grouped survival data using Cox regression model: an application to ART dental restorations, Stat. Med. 25 (2006) 447–457.
[12] J. Harper, Managing UTI in adults, Practitioner 244 (2000) 464–471.
[13] A.V. Franco, Recurrent urinary tract infections, best practice and research, Clin. Obstet. Gynaecol. 19 (2005) 861–873.
[14] L.E. Nicolle, Urinary tract infection in geriatric and institutionalized patients, Curr. Opin. Urol. 12 (2002) 51–55.
[15] L. Xiang, A.H. Lee, K.K.W. Yau, G.J. McLachlan, A score test for zero-inflation in correlated count data, Stat. Med. 25 (2006) 1660–1671.
[16] K.K.W. Yau, A.H. Lee, P.J.W. Carrivick, Modeling zero-inflated count series with application to occupational health, Comput. Methods Programs Biomed. 74 (2004) 47–52.
[17] C.A. McGilchrist, K.K.W. Yau, The derivation of BLUP, ML, REML estimation methods for generalized linear mixed models, Commun. Stat.: Theory Methods 24 (1995) 2963–2980.

Please cite this article in press as: K. Wang et al., Multilevel survival modelling of recurrent urinary tract infections, Comput. Methods Programs Biomed. (2007), doi:10.1016/j.cmpb.2007.05.013