Random Effects Cox Models: A Poisson Modelling Approach

Renjun Ma[0,1], Daniel Krewski[1,2] and Richard T. Burnett[1,3]

[1]Faculty of Medicine, University of Ottawa, Ottawa, Canada, K1H 8M5

[2] School of Mathematics and Statistics, Carleton University, Ottawa, Canada, K1S 5B6

[3] Environmental Health Directorate, Health Canada, Ottawa, Canada, K1A 0L2

February 1, 2000

Abstract

We propose a Poisson modelling approach to random effects Cox proportional hazards models. Specifically we describe methods of statistical inference for a class of random effects Cox models which accommodate a wide range of nested random effects distributions. The orthodox BLUP approach to random effects Poisson modeling techniques enables us to study this new class of models as a single class, rather than as a collection of unrelated models. The explicit expressions for the random effects given by our approach facilitate incorporation of relatively large number of random effects. An important feature of this approach is that the principal results depend only on the first and second moments of the unobserved random effects. The application of proposed methods is illustrated through the re-analysis of data on the time to failure (tumour onset) in an animal carcinogenesis experiment previously reported by Mantel and Ciminera (1979).

Key words: Cox model; BLUP; estimating equation; frailty; generalized linear models; random effects; Tweedie exponential dispersion model

[0]Email address: renjun@zeus.med.uottawa.ca

1 Introduction

Although the incorporation of random effects into Cox models has gained increasing attention in analyses of event history data, these models pose considerable theoretical difficulties in the development of estimation and inference procedures (Clayton 1991). Until recently, previous research in this area has focussed mainly on survival models with one level of random effects (Sastry 1997; Sargent 1998). The frequentist approaches to nested frailty survival models have usually been restricted to piecewise constant baseline hazard functions and specific random effects distributions (Sastry 1997). On the other hand, Bayesian approaches to nested random effects Cox models are computationally intensive, and the assessment of convergence of computational techniques such as the Gibbs sampler remains an area of debate (Glifford 1993; Smith and Roberts 1993; Sargent 1998). Flexible frailty models that can be fit with reasonable computational effort are therefore needed.

Considerable progress has been made in recent years in the area of random effects generalized linear models (Breslow and Clayton 1993; Lee and Nelder 1996; Ma 1999). The connection between the Cox and Poisson regression models has long been recongnized (Whitehead 1980). In this paper, we show that random effects methods developed for use with generalized linear models can be applied by characterizing the random effects Cox model as a random effects Poisson regression model. Our approach deals with an unspecified baseline hazard function and a wide range of random effects distributions. Our approach can also handle ties and stratification in the same way as in the standard Cox model. Further, our explicit expressions for the random effects facilitate incorporation of relatively large numbers of random effects.

The organization of the paper is as follows. We introduce the random effects Cox model and its auxiliary random effects Poisson models in Sections 2 and 3, respectively. In Section 4, we discuss the estimation of the nested random effects Cox models based on the orthodox BLUP approach to the auxiliary random effects Poisson models. An illustrative example involving animal carcinogensis data is presented in Section 5, and potential extensions of the models are discussed in Section 6.

2 Random Effects Cox Model

In this section, we consider a Cox model with two levels of random effects. Suppose that the cohort of interest is stratified on the basis of one or more relevant covariates. Let the hazard function for individual (i, j, k) from stratum $s = 1, 2, \ldots, a$ at time t be denoted by $h_{ijk}^{(s)}(t)$. Given the random effects, we assume that the individual hazard functions are conditionally independent with

$$h_{ijk}^{(s)}(t) = h_0^{(s)}(t) u_{ij} \exp(\boldsymbol{\beta}^\top \mathbf{x}_{ijk}^{(s)}). \tag{1}$$

Here, $u_{ij} > 0$ are random effects, or frailties, shared by all individuals within the same group, and $h_0^{(s)}(t)$ is the baseline hazard function for stratum s. Clearly the survival times (either failed or censored) within the same group are correlated. The random effects are traditionally assumed not to depend on the regression parameter $\boldsymbol{\beta}$. Without loss of generality, we assume that the design matrix is of full rank.

Here, we will focus on three-level hierarchical Cox models with the following nested random effects structure. Suppose the cohort is composed of m independent clusters indexed by i. Within each cluster i, there are J_i correlated sub-clusters indexed by (i, j). Further, within each sub-cluster (i, j) there are n_{ij} individuals whose survival times are given by (1). One such hierarchy example was presented by Sastry (1997) where the children were clustered at both community and family levels.

We introduce a class of models with nested random effects based on the class of Tweedie exponential dispersion model distributions denoted by $\text{Tw}_r(\mu, \sigma^2)$, where $\text{Tw}_r(\mu, \sigma^2)$ includes the normal ($r = 0$), Poisson ($r = 1$), gamma ($r = 2$), compound Poisson ($1 < r < 2$) and inverse Gaussian ($r = 3$) distributions as special cases (Jørgensen, 1997). More specifically, we assume that the cluster level random effects $u_1, \ldots, u_m$ are independently identically distributed random effects following the Tweedie distribution, with

$$U_1, \ldots, U_m \;\sim\; \text{Tw}_r(1, \sigma^2). \tag{2}$$

We further assume that, given the cluster level random effects $\mathbf{U}_* = \mathbf{u}_* = (u_1, \ldots, u_m)$, the sub-cluster level random effects $U_{11}, ..., U_{mJ_m}$ are conditionally independent, and that the conditional distribution of U_{ij}, given $\mathbf{U}_* = \mathbf{u}_*$, depends on u_i only, and is given by

$$U_{ij} | U_i = u_i \;\sim\; \text{Tw}_q(u_i, \omega^2), \tag{3}$$

Assumptions (1)-(3) together provide a full specification of a nested random effects Cox model. To avoid non-positive random effects, we require $r \geq 2$ and $q \geq 2$. Here, the multiplicative sub-cluster random effect u_{ij} represents the effect of the (i, j)th sub-cluster on the individual relative risk due to the fixed effect $\boldsymbol{\beta}$. Under these assumptions, the hazard function in (1) can be rewritten as

$$h_{ijk}^{(s)}(t) = h_0^{(s)}(t) u_i v_{ij} \exp(\boldsymbol{\beta}^\top \mathbf{x}_{ijk}^{(s)}), \tag{4}$$

where $V_{ij} = U_{ij}/U_i$. It can be easily verified that $\text{E}(V_{ij}) = 1$ and $\text{Cov}[U_i, V_{ij}] = 0$. In the literature, V_{ij} and U_i are usually assumed to be independent, with V_{ij} referred to as sub-cluster random effect instead of u_{ij}.

A Cox model with one level of random effects is obtained as a special case of the Cox model with two levels of random effects by setting $\omega^2 = 0$ and $J_i = 1$ for all i.

3 Auxiliary Random Effects Poisson Models

Let $\tau_{s1}, \ldots, \tau_{sq_s}$ denote the distinct failure times in the sth stratum, with m_{sh} indicating the multiplicity of failures occurring at time τ_{sh} $(s = 1, \ldots, a)$. The risk set at time τ_{sh} is a subset of stratum s, $\mathcal{R}(\tau_{sh}) = \{(i,j,k) : \; t_{ijk} \geq \tau_{sh}\}$, where t_{ijk} is the observed survival time for individual (i,j,k) from the sth stratum. In addition, let $Y^{(s)}_{ijk,h}$ be 1 if a failure occurs for individual (i,j,k) from the sth stratum at time τ_{sh} and 0 otherwise. Let $\mathbf{Y}$ and $\mathbf{U}$ denote the vectors of $Y^{(s)}_{ijk,h}$ and the random effects U_{ij}, respectively. Given the random effects $\mathbf{U} = \mathbf{u}$, Peto's version of the conditional partial likelihood (cf. Cox and Oakes 1984 p.103) is

$$p\ell(\boldsymbol{\beta}; \mathbf{Y}|\mathbf{u}) = \prod_{s=1}^{a} \prod_{h=1}^{q_s} \frac{\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij}^{Y^{(s)}_{ijk,h}} \left\{\exp(\mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right\}^{Y^{(s)}_{ijk,h}} (m_{sh}!)}{\left\{\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right\}^{m_{sh}}}. \qquad (5)$$

We now define an auxiliary random effects Poisson regression model. Assume that the components of $\mathbf{Y}$ are conditionally independent, given random effects $\mathbf{U} = \mathbf{u}$, with

$$\begin{aligned} Y^{(s)}_{ijk,h} &\sim \text{Poisson}\left(u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right) \\ &= \text{Poisson}\left(\nu^{(s)}_{ijk,h}\right) \;\; (i,j,k) \in \mathcal{R}(\tau_{sh}), \end{aligned} \qquad (6)$$

where $\nu^{(s)}_{ijk,h} = u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})$. Given the random effects, the conditional likelihood for the random effects Poisson model is

$$\begin{aligned} \ell(\alpha, \boldsymbol{\beta}; \mathbf{Y}|\mathbf{u}) &= \prod_{s=1}^{a} \prod_{h=1}^{q} \prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \frac{\left\{u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right\}^{Y^{(s)}_{ijk,h}}}{\exp\left\{u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right\}} \\ &= \prod_{s=1}^{a} \prod_{h=1}^{q} \frac{\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij}^{Y^{(s)}_{ijk,h}} \left\{\exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right\}^{Y^{(s)}_{ijk,h}}}{\exp\left\{\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right\}}. \end{aligned} \qquad (7)$$

We will show that the maximum conditional Poisson likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (7) are the maximum conditional partial likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (5).

Consider the maximum likelihood estimates for $\nu^{(s)}_{ijk,h}$, denoted by $\hat{\nu}^{(s)}_{ijk,h}$, based on the conditional Poisson likelihood. Since $Y^{(s)}_{ijk,h}$ $(i,j,k) \in \mathcal{R}(\tau_{sh})$ are independent for $(i,j,k) \in \mathcal{R}(\tau_{sh})$ given the random effects, it follows from the relation

$$\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} Y^{(s)}_{ijk,h} = m_{sh} \tag{8}$$

that

$$\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \hat{\nu}^{(s)}_{ijk,h} = m_{sh}. \tag{9}$$

We therefore have

$$\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\hat{\alpha}_{sh} + \mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}}) = m_{sh}. \tag{10}$$

or

$$\exp(\hat{\alpha}_{sh}) = \frac{m_{sh}}{\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})}. \tag{11}$$

At its maximum $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$, the conditional Poisson likelihood for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is

$$\begin{aligned}
\ell(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}; \mathbf{Y}|\mathbf{u}) &= \prod_{s=1}^{a}\prod_{h=1}^{q}\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \exp(-m_{sh}) u_{ij}^{Y^{(s)}_{ijk,h}} \left\{\exp(\hat{\alpha}_{sh} + \mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})\right\}^{Y^{(s)}_{ijk,h}} \\
&= \prod_{s=1}^{a}\prod_{h=1}^{q}\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \left[\exp(-m_{sh}) u_{ij}^{Y^{(s)}_{ijk,h}} \left\{\exp(\mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})\right\}^{Y^{(s)}_{ijk,h}}\right] \\
&\quad \times \left[\left\{\exp(\hat{\alpha}_{sh})\right\}^{\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} Y^{(s)}_{ijk,h}}\right] \\
&= \prod_{s=1}^{a}\prod_{h=1}^{q} \frac{\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij}^{Y^{(s)}_{ijk,h}} \left\{\exp(\mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})\right\}^{Y^{(s)}_{ijk,h}} \exp(-m_{sh})}{\left\{\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})\right\}^{m_{sh}}} \\
&= \prod_{s=1}^{a}\left\{\prod_{h=1}^{q} \frac{m_{sh}^{m_{sh}} \exp(-m_{sh})}{m_{sh}!}\right\} p\ell(\hat{\boldsymbol{\beta}}; \mathbf{Y}|\mathbf{u}),
\end{aligned}$$

where the first and third equalities are obtained using (10) and (8), and (11), respectively. Clearly the conditional partial likelihood and conditional Poisson likelihood share the same kernel at the maximum conditional Poisson likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$.

Let $f(\mathbf{U};\xi)$ be the density function of $\mathbf{U}$ with parameter ξ. The joint partial likelihood of the regression parameter $\boldsymbol{\beta}$ given the data and the random effects is

$$p\ell(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{U}) = p\ell(\boldsymbol{\beta}; \mathbf{Y}|\mathbf{U}) f(\mathbf{U};\xi).$$

The joint likelihood of the regression parameter $\boldsymbol{\beta}$ given the data and the random effects for the auxiliary random effects Poisson regression model is

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{Y}, \mathbf{U}) = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{Y}|\mathbf{U}) f(\mathbf{U};\xi).$$

To obtain the regression parameter estimates, given the data and the random effects, maximizing the joint (partial) likelihood is equivalent to maximizing the conditional (partial) likelihood since the random effects vector $\mathbf{U}$ does not depend on the regression parameter vector. Therefore we have

$$\ell(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}; \mathbf{Y}, \mathbf{U}) = \text{constant} \cdot p\ell(\hat{\boldsymbol{\beta}}; \mathbf{Y}, \mathbf{U}).$$

This demonstrates that the maximum joint Poisson likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (7) are the maximum joint partial likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (5). We may therefore make inferences on the random effects Cox models by fitting random effects Poisson models.

The random effects are unobserved, and thus have to be predicted. Algorithms for fitting random effects models usually iterate between updating random effects and updating parameter estimates until convergence is achieved. Given the predicted random effects, the estimates of the regression parameter $\boldsymbol{\beta}$ for the auxiliary models are also the regression parameter estimates for the corresponding random effects Cox models. We therefore approximate the random effects using the consistent random effects predictors for the auxiliary models.

In the reminder of this paper, we will focus on the nested random effects Cox models specified by (1), (2) and (3) via fitting the auxiliary nested random effects Poisson models specified by (6), (2) and (3).

4 Orthodox BLUP Approach to Auxiliary Models

In this section, we discuss estimation of the auxiliary nested random effects Poisson models based on the orthodox BLUP approach to nested random effects Poisson models (Ma 1999).

4.1 Prediction of Random Effects

We will predict the random effects by the best linear unbiased predictor of $\mathbf{U}$ given $\mathbf{Y}$ in the literal sense (cf. Brockwell and Davis 1991 p.64). More specifically, letting $\mathbf{U}$ and $\mathbf{Y}$ be random vectors with finite second moments, the best linear unbiased predictor of $\mathbf{U}$ given $\mathbf{Y}$ is given by

$$\widehat{\mathbf{U}} = \mathrm{E}(\mathbf{U}) + \mathrm{Cov}(\mathbf{U}, \mathbf{Y})\mathrm{Var}^{-1}(\mathbf{Y})\left(\mathbf{Y} - \mathrm{E}(\mathbf{Y})\right).$$

We call $\widehat{\mathbf{U}}$ the orthodox BLUP of the random effects since the mode of the conditional density of the random effects given the data is also referred to as BLUP in the literature (McGilchrist 1993), although this modal predictor is neither linear nor unbiased in general. The orthodox BLUP of the random effects is the linear unbiased predictor of $\mathbf{U}$ given $\mathbf{Y}$ which minimizes the mean square distance between the random effects $\mathbf{U}$ and their predictor within the class of linear functions of $\mathbf{Y}$.

Explicit expressions for the mean square distances between the components of the random effects $\mathbf{U}$ and their predictors are as follows:

$$\begin{aligned} c_i &= \mathrm{E}(\widehat{U}_i - U_i)^2 \\ &= \frac{\sigma^2}{1 + \sigma^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij}\mu_{ijk,h}^{(s)}}, \end{aligned} \tag{12}$$

where (i,j,k) runs over the risk set $\mathcal{R}(\tau_{sh})$ for fixed i. Here,

$$\begin{aligned} \mu_{ijk,h}^{(s)} &= \exp\left(\alpha_{sh} + \boldsymbol{\beta}^\top \mathbf{x}_{ijk}^{(s)})\right) \\ &= \exp\left((\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)\mathbf{x}_{ijk,h}^{(s)}\right) \\ &= \exp\left(\boldsymbol{\gamma}^\top \mathbf{x}_{ijk,h}^{(s)}\right), \end{aligned}$$

and, for fixed (i,j),

$$w_{ij} = \left(1 + \omega^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \mu_{ijk,h}^{(s)}\right)^{-1},$$

where (i,j,k) runs over the risk set $\mathcal{R}(\tau_{sh})$. Similarly, we have

$$\begin{aligned} c_{ij} &= \mathrm{E}(\widehat{U}_{ij} - U_{ij})^2 \\ &= w_{ij}\left\{\omega^2 + c_i w_{ij}\right\}. \end{aligned} \tag{13}$$

The cluster random effects predictor can be expressed as

$$\begin{aligned} \widehat{U}_i &= \frac{1 + \sigma^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij} Y_{ijk,h}^{(s)}}{1 + \sigma^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij}\mu_{ijk,h}^{(s)}} \\ &= c_i \left(\frac{1}{\sigma^2} + \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij} Y_{ijk,h}^{(s)}\right), \end{aligned}$$

where (i,j,k) runs over the risk set $\mathcal{R}(\tau_{sh})$ for any given i. The sub-cluster random effects predictors are

$$\widehat{U}_{ij} = w_{ij}\hat{U}_i + \omega^2 w_{ij} \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} Y_{ijk,h}^{(s)},$$

where (i,j,k) runs over the risk set $\mathcal{R}(\tau_{sh})$ for any given (i,j).

Using Chebyshev's inequality, it follows from (12) and (13) that we have the following consistency results in terms of convergence in probability (Ma 1999):

(i) $\hat{U}_i \longrightarrow U_i$ as $\sigma^2 \to \mathbf{0}$;

(ii) $\hat{U}_{ij} \longrightarrow U_{ij}$ as $\omega^2 + \sigma^2 \to \mathbf{0}$.

(iii) $\hat{U}_{ij} \longrightarrow U_{ij}$ as $\min_{jksh}(\mu^{(s)}_{ijk,h}) \to \infty$.

Results (i)-(iii) are usually referred to as 'small dispersion asymptotics'. Let n_{ij} be the number of the induced observations $y^{(s)}_{ijk,h}$ contained in sub-cluster (i,j). We also have the following large sample asymptotics if $\min_{jk}(\mu_{ijk}) \geq c\log(\min_j(n_{ij}))/\min_j(n_{ij})$ for a positive constant c. That is, the only restriction is that μ_{ijk} should not tend to zero too quickly.

(iv) $\hat{U}_i \longrightarrow U_i$ as $J_i \to \infty$ and $\hat{U}_{ij} \xrightarrow{P} U_{ij}$ as $\min_j(n_{ij}) \to \infty$.

The magnitude of the n_{ij} depends not only on the number of individuals in sub-cluster (i,j), but also on the number of the failures in each individual's stratum. In other words, the greater the number of subjects, especially those with complete survival histories, the better we are able to predict the random effects.

4.2 Estimation of Regression Parameters

Consider first estimation of the regression parameters in the case of known dispersion parameters. Estimation of the unknown dispersion parameters will be discussed in next section.

Differentiating the joint likelihood of the auxiliary model for the data and random effects yields the joint score function. Replacing the random effects with their predictors, we have an unbiased estimating function for the regression parameters $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$:

$$\boldsymbol{\psi}(\boldsymbol{\gamma}) = \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \mathbf{x}^{(s)}_{ijk,h}(Y^{(s)}_{ijk,h} - \widehat{U}_{ij}\mu^{(s)}_{ijk,h}).$$

The solutions of $\boldsymbol{\psi}(\boldsymbol{\gamma}) = \mathbf{0}$ provide estimates of the regression parameters. The Newton scoring algorithm introduced by Jørgensen et al. (1995) can be used to solve this estimating equation.

The Newton scoring algorithm is defined as the Newton algorithm applied to the equation $\boldsymbol{\psi}(\boldsymbol{\gamma}) = \mathbf{0}$, but with the derivative of $\boldsymbol{\psi}(\boldsymbol{\gamma})$ replaced by its expectation. This expectation, denoted by $\mathbf{S}(\boldsymbol{\gamma})$, is called the sensitivity matrix:

$$\mathbf{S}(\boldsymbol{\gamma}) \;=\; \sum_{i=1}^{m} c_i \mathbf{e}_i \mathbf{e}_i^\top + \sum_{i=1}^{m}\sum_{j=1}^{J_i} \omega^2 w_{ij} \mathbf{f}_{ij}\mathbf{f}_{ij}^\top$$

$$- \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k) \in \mathcal{R}(\tau_{sh})} \mu_{ijk,h}^{(s)} \mathbf{x}_{ijk,h}^{(s)} (\mathbf{x}_{ijk,h}^{(s)})^\top, \tag{14}$$

where

$$\mathbf{e}_i = \left(\sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k) \in \mathcal{R}(\tau_{sh})} w_{ij} \mu_{ijk,h}^{(s)} \mathbf{x}_{ijk,h}^{(s)} \right), \tag{15}$$

and

$$\mathbf{f}_{ij} = \left(\sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k) \in \mathcal{R}(\tau_{sh})} \mu_{ijk,h}^{(s)} \mathbf{x}_{ijk,h}^{(s)} \right). \tag{16}$$

Here, the index (i, j, k) runs over the risk set $\mathcal{R}(\tau_{sh})$ for fixed i in (15) and for fixed (i, j) in (16), respectively, and (i, j, k) runs freely over the risk set $\mathcal{R}(\tau_{sh})$ in the last term of (14). The resulting algorithm gives the following updated value for $\boldsymbol{\gamma}$:

$$\boldsymbol{\gamma}^* = \boldsymbol{\gamma} - \mathbf{S}^{-1}(\boldsymbol{\gamma}) \boldsymbol{\psi}(\boldsymbol{\gamma}).$$

The sensitivity matrix multiplied by -1 has been shown to be the Godambe information matrix for the nested random effects Poisson model (Ma 1999). That is, the sensitivity matrix plays a role in the Newton scoring algorithm similar to that of the Fisher information matrix in the Fisher scoring algorithm.

Under mild regularity conditions, the solutions of $\boldsymbol{\psi}(\boldsymbol{\gamma}) = \mathbf{0}$, denoted by $\hat{\boldsymbol{\gamma}}$, have been shown to be consistent as $m \to \infty$ with the asymptotic covariance given by $-\mathbf{S}^{-1}(\boldsymbol{\gamma})$. The estimating function $\boldsymbol{\psi}(\boldsymbol{\gamma})$ has also been shown to be optimal in the sense that it attains the minimum asymptotic covariance for the estimator $\hat{\boldsymbol{\gamma}}$ among a certain class of linear functions of $\mathbf{Y}$ (Ma 1999). When there are no random effects, the sensitivity matrix becomes the negative Fisher information matrix derived from the partial likelihood for the standard Cox model. Expression (14) shows that the asymptotic variance for regression parameter estimates based on the standard Cox model is smaller than that based on the random effects Cox model if the regression parameter estimates are identical for both models.

An analogue of Wald's test is available for testing the hypothesis H_0 : $\boldsymbol{\beta}_{(1)} = \mathbf{0}$, where $\boldsymbol{\beta}_{(1)}$ is a sub-vector of $\boldsymbol{\beta}$. The test statistic is:

$$W = \widehat{\boldsymbol{\beta}}_{(1)}^\top \left\{ \mathbf{J}^{11}(\widehat{\boldsymbol{\gamma}}) \right\}^{-1} \widehat{\boldsymbol{\beta}}_{(1)},$$

where $\mathbf{J}^{11}(\widehat{\boldsymbol{\gamma}})$ is the block of the asymptotic covariance matrix of $\widehat{\boldsymbol{\gamma}}$ corresponding to $\boldsymbol{\beta}_{(1)}$. Asymptotically, this statistic follows a $\chi^2(k)$-distribution, where k is the size of the sub-vector $\widehat{\boldsymbol{\beta}}_{(1)}$.

4.3 Estimation of Dispersion Parameters

We now discuss the situation in which the dispersion parameters are unknown. In analogy with generalized linear models, we adopt the following adjusted Pearson estimator for the dispersion parameter σ^2:

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} \left\{ (\hat{U}_i - 1)^2 + c_i \right\}.$$

The first term is the Pearson estimator, with the second term being a bias correction term. The corresponding adjusted Pearson estimator for ω^2 is:

$$\hat{\omega}^2 = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{J_i} \sum_{j=1}^{J_i} \left\{ (\widehat{U}_{ij} - \widehat{U}_i)^2 + c_{ij} + c_i - 2c_i w_{ij} \right\}.$$

Again, the first term is the Pearson estimator, whereas the remaining terms are bias correction terms. These dispersion parameter estimates can also be shown to be consistent as $m \to \infty$ (Ma 1999). Unlike most other approaches in the literature, our asymptotic variance of the regression parameter estimator is not affected by the variability in the dispersion parameter estimators.

In fact, this orthodox BLUP approach depends on the random effects only via the first and second moments of the sub-cluster random effects. It has been shown to be robust, to a certain extent, against mispecification of the random effects distributions (Ma 1999), and thus covers non-Tweedie random effects such as log-normal random effects.

4.4 Computational Procedures

Initial values for the regression parameters are taken as the regression parameter estimates obtained from standard Poisson regression techniques assuming independent responses. Initial random effects predictions $\hat{U}_i$ and $\hat{U}_{ij}$ are given by the average of the responses within cluster i divided by the average of all responses and the average of the responses within sub-cluster (i, j) divided by the average of all responses, respectively. The initial dispersion parameter estimates are calculated from the adjusted Pearson estimators, omitting the bias-correction terms.

The algorithm then iterates between updating the regression parameter estimates via the Newton scoring algorithm, updating random effect predictors via the orthodox BLUP, and updating dispersion parameter estimates via the adjusted Pearson estimators.

5 An Illustrative Example

We illustrate the application of our approach to the random effects Cox model using data from an animal carcinogenesis experiment originally reported by Mantel and Ciminera (1979). This experiment involved 50 sets of three female weanling rats selected from within the same litter, with one animal assigned to a treatment group exposed to a putative carcinogen, and the remaining two serving as litter-matched controls. The time to tumour occurrence or censoring was recorded to the nearest week for each of the 150 animals employed in this study. This experiment thus involved a single binary covariate with values of 0 and 1 indicating assignment to the control or treated group, respectively.

Because of the possibility of intra-litter correlation (Gart et al. 1986), we included a random effect for each litter. The corresponding Cox regression model assumes that, given the random effects, the hazard functions for individuals are conditionally independent, with the hazard function for individual j from litter i given by

$$h_{ij}(t) = h_0(t) u_i \exp(x_{ij}\beta),$$

where x_{ij} is the indicator variable, reflecting exposure to the test agent. The litter random effect u_i are assumed to follow independent and identical Tweedie distributions with unity mean and dispersion parameter σ^2 described in (2).

Parameter estimates for both the standard and random effects Cox models are shown in Table 1 where the Peto-Breslow approximation (Cox and Oakes 1984) for tied failure times was used in both analyses. The estimates of the regression parameter β associated with the treatment effect are comparable under both models, as are the standard errors of these estimates. Based on the ratio of these estimates to their respective standard errors, the treatment effect is significant under both models.

Table 1: Parameter estimates for the animal carcinogenesis data.

	Parameter Estimates	
Cox Model	$\hat{\beta} \pm$ SE	σ^2
Standard	0.898 ± 0.317	-
Random effects	0.902 ± 0.312	0.293

Scatter plot of the litter random effects is shown in Figures 1. These 50 litters were labelled as $1, 3, \ldots, 99$ by Mantel and Ciminera (1979) and are re-numbered as $1, 2, \ldots, 50$ here for convenience. Litters 3, 21, 22, 25 and 37 demonstrated the lowest litter-specific relative risks, whereas litter 13 had the highest (Figure 1). Figure 2 shows that the litter random effects match the number of tumour occurrences in the corresponding litter; the higher the litter-specific relative risk, the higher the litter tumour occurrence. The one exception is litter 13, which had a higher litter-specific relative risk than litter 32, although litter 32 was the only litter with tumours occurring in all three littermates. Examination of the data revealed that all three rats in litter 13 had exceptionally low tumour onset times (Figure 3).

Figure 1, 2 and 3 are approximately here.

6 Discussion

In this paper, we have introduced a Poisson modelling approach to random effects Cox models. We have specifically focussed on Cox models with two levels of nested random effects. We may consider models with more than two levels of random effects. For such models, our method remains valid with (i, j, k) replaced by higher dimensional indices. The proposed Poisson modelling approach can also be extended to random effects Cox models with time dependent covariates in the following way. Suppose that all covariates assume constant values between two distinct failure times, as reflected by the corresponding step functions for the cumulative failure times. The incorporation of such time dependent covariates can be simply achieved by replacing $\mathbf{x}_{ijk}^{(s)}$ by $\mathbf{x}_{ijk}^{(s)}(t) = \mathbf{x}_{ijk}(\tau_{sh})$ for $Y_{ijk,h}^{(s)}$ in the model.

For the Cox model with one level of random effects ($J_i = 1$, $\omega^2 = 0$, with $u_{ij} = u_i$), the random effects have been previously characterized by gamma (Clayton 1991), positive stable (Hougaard 1986a, 1986b) and log-normal (McGilchrist 1993) distributions. Our framework effectively covers the gamma, log-normal and inverse Gaussian distributed random effects.

Our Poisson approach is not limited to Cox models with the nested random effects structures. Taking $u_{ij} = v_i v_j$ for balanced designs will lead to crossed random effects. For Cox models with only time dependent subject frailties $u_i(t)$ for each subject i, we can employ the techniques developed for Poisson models with an AR(p) structure on the latent variable $u_i(t)$. Since the distinct failure times are not equally spaced, a specific time series structure for time dependent frailties may not be appropriate. In the Cox model specified by (1)-(3), taking the second level random effects $u_{it} = u_i(t)$ as conditional on the subject random effect u_i, where t represents the distinct failure times in the stratum of the ith subject, we have correlated time dependent frailties for each subject.

Acknowledgements

This research was supported by the Health Effects Insitute and by a grant A8664 from the Natural Sciences and Engineering Research Council of Canada to D. Krewski.

References

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized lin ear mixed model. *Journal of American Statististical Association* **88**, 9-25.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods* 2nd ed. New York: Springer-Verlag.

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

Cox, D.R. and Oakes, D. (1984) *Analysis of Survival Data* New York: Chapman and Hall.

Gart, J., Krewski, D., Lee, P., Tarone, R. and Wahrendorf, J. (1986) *Statistical Methods in Cancer Research, Vol.III: The Design and Analysis of Long-Term Animal Experiments.* Lyon: International Agency for Research on Cancer.

Glifford, P. (1993). Discussion on the meeting on the Gibbs sampler and other Markov chain Monte Carlo methods. *Journal of Royal Statistical Society Ser. B* **55**, 53-54.

Hougaard, P. (1986a) Survival models for heterogeneous population derived from stable distributions. *Biometrika* **73**, 387-396.

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

Jørgensen, B. (1997). *The Theory of Dispersion Models.* London: Chapm an and Hall.

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

Ma, R. (1999) An Orthodox BLUP Approach to Generalized Linear Mixed Models. Ph.D. Thesis. Department of Statistics, The University of British Columbia.

Mantel, N. and Ciminera, J.L. (1979) Mantel-Haenszel analysis of littermatched time-to-response data with modifications for recovery of interlitter information. *Cancer Research* **37**, 3863-3868.

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

Sastry, N. (1997) A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of American Statistical Association***92**, 426-435.

Sargent, D.J. (1998) A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics* **54**, 1486-1497.

Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of Royal Statistical Society Ser. B* **55**, 3-23.

Whitehead, J. (1980) Fitting Cox's regression model to survival data using GLIM. *Applied Statistics* **29**, 268-275.