HEI

APPENDIX AVAILABLE ON REQUEST

Special Report

Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality

Part II: Sensitivity Analyses

Appendix I. Random Effects Cox Models

Daniel Krewski, Richard T. Burnett, Mark S. Goldberg, Kristin Hoover, Jack Siemiatycki, Michael Jerrett, Michal Abrahamowicz, Warren H. White, and Others

---

Correspondence may be addressed to Dr Daniel Krewski, Professor of Epidemiology and Statistics, Department of Epidemiology and Community Medicine, Room 3229C, 451 Smyth Road, University of Ottawa, Ottawa Ontario K1H 8M5, Canada

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R824835 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

This document was reviewed by the HEI Health Review Committee but did not undergo the HEI scientific editing and production process.

UNIVERSITY OF OTTAWA

Faculty of Medicine **Faculty of Health Sciences**

Re-analysis of the Harvard Six-Cities Study and the American Cancer Society Study of Air Pollution and Mortality, Phase II: Sensitivity Analysis

Appendix A, B, C, D, E, F, G, H, and I

R. Samuel McLaughlin Center for Population Health Risk Assessment
Institute of Population Health
University of Ottawa

August, 2000

Appendix I

Random Effects Cox Models

I.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 recognized (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 this appendix is as follows. We introduce the random effects Cox model and its auxiliary random effects Poisson models in Sections I.2 and I.3, respectively. In Section I.4, we discuss the estimation of the

Replaced 12·1·00

nested random effects Cox models based on the orthodox BLUP approach to the auxiliary random effects Poisson models. The methods described in this section were used in fitting random effects Cox models to the American Cancer Society Study data (see Section I.5).

I.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 hazard functions for individuals 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 strata $s = 1, \ldots, a$. Clearly the (possibly censored) survival times 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.

We will focus on estimation and inference on three-level hierarchical Cox models with the following nested random effects structure. Suppose each 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). We introduce a class of models with nested random effects based on the class of Tweedie exponential dispersion model distributions denoted by $\mathrm{Tw}_r(\mu, \sigma^2)$, where $\mathrm{Tw}_r(\mu, \sigma^2)$ includes the normal $(r = 0)$, Poisson $(r = 1)$, gamma $(r = 2)$, compound Poisson $(1 < r < 2)$ and inverse Gaussian $(r = 3)$ distributions (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 dispersion model distribution, with

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

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 $U_{11}, \ldots, U_{mJ_m}$ are conditionally independent, and the conditional distribution of U_{ij}, given $\mathbf{U}_* = \mathbf{u}_*$, depends on u_i only which is

$$U_{ij}|\mathbf{U}_* = \mathbf{u}_* \sim \mathrm{Tw}_q(u_i, \omega^2), \tag{3}$$

Assumptions (1)-(3) together provide a full parametric specification of a nested random effects Cox model. To avoid non-positive random effects, we require $r \geq 2$ and $q \geq 2$. 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.

I.3 Auxiliary Random Effects Poisson Models

Let $\tau_{s1}, \ldots, \tau_{sq_s}$ denote the distinct uncensored event times in the sth stratum, with m_{sh} indicating the multiplicity of uncensored events happening 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 an uncensored event happens for individual (i,j,k) from the sth stratum at time τ_{sh} and 0 otherwise. Given the random effects $\mathbf{U} = \mathbf{u}$, Peto's version of the conditional partial likelihood (cf. Cox and Oakes 1984 p.103) is

$$p\ell(\boldsymbol{\beta}; \mathbf{Y}|\mathbf{u}) = \prod_{s=1}^{a} \prod_{h=1}^{q_s} \frac{\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij}^{Y^{(s)}_{ijk,h}} [\exp(\mathbf{x}_{ijk}^\top \boldsymbol{\beta})]^{Y^{(s)}_{ijk,h}} (m_{sh}!)}{[\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\mathbf{x}_{ijk}^\top \boldsymbol{\beta})]^{m_{sh}}}. \tag{4}$$

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

$$\begin{aligned} Y^{(s)}_{ijk,h} &\sim \text{Poisson}\left(u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})\right) \\ &= \text{Poisson}\left(\nu^{(s)}_{ijk,h}\right) \quad (i,j,k) \in \mathcal{R}(\tau_{sh}), \end{aligned} \tag{5}$$

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

$$\begin{aligned} \ell(\alpha, \boldsymbol{\beta}; \mathbf{Y}|\mathbf{u}) &= \prod_{s=1}^{a} \prod_{h=1}^{q} \prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \frac{[u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})]^{Y^{(s)}_{ijk,h}}}{\exp[u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})]} \\ &= \prod_{s=1}^{a} \prod_{h=1}^{q} \frac{\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij}^{Y^{(s)}_{ijk,h}} [\exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})]^{Y^{(s)}_{ijk,h}}}{\exp[\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\alpha_{sh} + \mathbf{x}_{ijk}^\top \boldsymbol{\beta})]}. \end{aligned} \tag{6}$$

We will show that the maximum conditional Poisson likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (6) are the maximum conditional partial likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (4).

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

$$\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} Y_{ijk,h}^{(s)} = m_{sh} \tag{7}$$

that

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

We therefore have

$$\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\hat{\alpha}_{sh} + \mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}}) = m_{sh}. \tag{9}$$

or

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

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

$$\begin{aligned}
\ell(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}; \mathbf{Y}|\mathbf{u}) &= \prod_{s=1}^{a}\prod_{h=1}^{q}\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \exp(-m_{sh}) u_{ij}^{Y_{ijk,h}^{(s)}} [\exp(\hat{\alpha}_{sh} + \mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})]^{Y_{ijk,h}^{(s)}} \\
&= \prod_{s=1}^{a}\prod_{h=1}^{q}\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \left\{\exp(-m_{sh}) u_{ij}^{Y_{ijk,h}^{(s)}} [\exp(\mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})]^{Y_{ijk,h}^{(s)}}\right\} \\
&\quad \times \left\{[\exp(\hat{\alpha}_{sh})]^{\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} Y_{ijk,h}^{(s)}}\right\} \\
&= \prod_{s=1}^{a}\prod_{h=1}^{q} \frac{\prod_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij}^{Y_{ijk,h}^{(s)}} [\exp(\mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})]^{Y_{ijk,h}^{(s)}} \exp(-m_{sh})}{[\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} u_{ij} \exp(\mathbf{x}_{ijk}^{\top}\hat{\boldsymbol{\beta}})]^{m_{sh}}} \\
&= \prod_{s=1}^{a}\left\{\prod_{h=1}^{q} \frac{m_{sh}^{m_{sh}} \exp(-m_{sh})}{m_{sh}!}\right\} p\ell(\hat{\boldsymbol{\beta}}; \mathbf{Y}|\mathbf{u}),
\end{aligned} \tag{11}$$

where the first and third equalities were obtained by using (9) and (7), and (10), respectively. Clearly the conditional partial likelihood and conditional Poisson likelihood share the same kernel at maximum conditional Poisson likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$.

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

$$p\ell(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{u}) = p\ell(\boldsymbol{\beta}; \mathbf{Y}|\mathbf{u}) f(\mathbf{u}). \tag{12}$$

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

$$\ell(\alpha, \beta; \mathbf{Y}, \mathbf{u}) = \ell(\alpha, \beta; \mathbf{Y}|\mathbf{u}) f(\mathbf{u}). \tag{13}$$

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 U does not depend on the regression parameter vector. Therefore we have

$$\ell(\hat{\alpha}, \hat{\beta}; \mathbf{Y}, \mathbf{U}) = \text{constant} \cdot p\ell(\hat{\beta}; \mathbf{Y}, \mathbf{U}). \tag{14}$$

This demonstrates that the maximum joint Poisson likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (6) are the maximum joint partial likelihood estimates for the regression parameter vector $\boldsymbol{\beta}$ from (4). 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 β 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 remainder of this appendix, 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 (5), (2) and (3).

I.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).

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 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

$$\widehat{\mathrm{U}} = \mathrm{E}(\mathrm{U}) + \mathrm{Cov}(\mathbf{U}, \mathbf{Y})\mathrm{Var}^{-1}(\mathbf{Y})\,(\mathbf{Y} - \mathrm{E}(\mathbf{Y})). \tag{15}$$

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 mode generally is neither linear nor unbiased. The orthodox BLUP of the random effects is a 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 U and their predictors are as follows:

$$\begin{aligned} c(i) &= \mathrm{E}(\widehat{U}_i - U_i)^2 \\ &= \frac{\sigma^2}{1+\sigma^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij}\mu_{ijk,h}^{(s)}}, \end{aligned} \tag{16}$$

where (i,j,k) in (16) runs over the risk set $\mathcal{R}(\tau_{sh})$ for fixed i. Here,

$$\begin{aligned} \mu_{ijk,h}^{(s)} &= \exp\left(\alpha_{sh} + \boldsymbol{\beta}^\top \mathbf{x}_{ijk}^{(s)}\right) \\ &= \exp\left((\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)\mathbf{x}_{ijk,h}^{(s)}\right) \\ &= \exp\left(\boldsymbol{\gamma}^\top \mathbf{x}_{ijk,h}^{(s)}\right), \end{aligned} \tag{17}$$

and, for fixed (i,j),

$$w_{ij} = 1 \Big/ \left(1 + \omega^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \mu_{ijk,h}^{(s)}\right), \tag{18}$$

where (i,j,k) in (18) runs over the risk set $\mathcal{R}(\tau_{sh})$. Similarly we have

$$\begin{aligned} c(ij) &= \mathrm{E}(\widehat{U}_{ij} - U_{ij})^2 \\ &= w_{ij}\left\{\omega^2 + c(i) w_{ij}\right\}, \end{aligned} \tag{19}$$

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

The cluster-specific random effects predictor can be expressed as

$$\begin{aligned} \widehat{U}_i &= \frac{1+\sigma^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij} Y_{ijk,h}^{(s)}}{1+\sigma^2 \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij}\mu_{ijk,h}^{(s)}} \\ &= c(i)\left(\frac{1}{\sigma^2} + \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij} Y_{ijk,h}^{(s)}\right), \end{aligned} \tag{20}$$

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

$$\hat{U}_{ij} = w_{ij}\hat{U}_i + \omega^2 w_{ij} \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} Y_{ijk,h}^{(s)}, \tag{21}$$

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

Using Chevyshev's inequality, it follows from (16) and (19) that we have the following consistency results (Ma 1999):

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

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

We also have

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

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

where n_{ij} is number of the induced observations $y_{ijk,h}^{(s)}$ contained in sub-cluster (i,j). The latter part of result (iv) holds if $\mu_{ijk,h}^{(s)}$s are bounded from zero. Results (i)-(iii) are usually referred as 'small dispersion asymptotics', whereas (iv) corresponds to large sample asymptotics. 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.

Estimation of Regression Parameters

Consider first the estimation for the regression parameters with the case of known dispersion parameters. Consideration of unknown dispersion parameters will be discussed in the 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 regression parameters $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$:

$$\psi(\gamma) = \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \mathbf{x}_{ijk,h}^{(s)} (Y_{ijk,h}^{(s)} - \hat{U}_{ij}\mu_{ijk,h}^{(s)}). \tag{22}$$

The solutions of $\psi(\gamma) = 0$ 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 $\psi(\gamma) = 0$, 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; its expression is given by

$$\begin{aligned} \mathrm{S}(\gamma) &= \sum_{i=1}^{m} c(i)\mathbf{e}_i\mathbf{e}_i^\top + \sum_{i=1}^{m}\sum_{j=1}^{J_i} \omega^2 w_{ij}\mathbf{f}_{ij}\mathbf{f}_{ij}^\top \\ &\quad - \sum_{s=1}^{a}\sum_{h=1}^{q_s}\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \mu_{ijk,h}^{(s)}\mathbf{x}_{ijk,h}^{(s)}(\mathbf{x}_{ijk,h}^{(s)})^\top, \end{aligned} \tag{23}$$

where

$$\mathbf{e}_i = \left(\sum_{s=1}^{a}\sum_{h=1}^{q_s}\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} w_{ij}\mu_{ijk,h}^{(s)}\mathbf{x}_{ijk,h}^{(s)}\right), \tag{24}$$

and

$$\mathbf{f}_{ij} = \left(\sum_{s=1}^{a}\sum_{h=1}^{q_s}\sum_{(i,j,k)\in\mathcal{R}(\tau_{sh})} \mu_{ijk,h}^{(s)}\mathbf{x}_{ijk,h}^{(s)}\right). \tag{25}$$

Here, index (i, j, k) runs over the risk set $\mathcal{R}(\tau_{sh})$ for fixed i in (24) and for fixed (i, j) in (25), respectively, and (i, j, k) runs freely over the risk set $\mathcal{R}(\tau_{sh})$ in the last term of (23). The resulting algorithm gives the following updated value for $\boldsymbol{\gamma}$:

$$\boldsymbol{\gamma}^* = \boldsymbol{\gamma} - \mathbf{S}^{-1}(\boldsymbol{\gamma})\boldsymbol{\psi}(\boldsymbol{\gamma}). \tag{26}$$

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 $-\mathrm{S}^{-1}(\gamma)$. For fixed m, the asymptotic covariance has the same expression $-\mathrm{S}^{-1}(\gamma)$ for large cluster sizes for the auxiliary model with only one level of gamma distributed random effects (Lee and Nelder 1996). In general, the estimating function $\boldsymbol{\psi}(\boldsymbol{\gamma})$ has been shown to be optimal in the sense that it attains the minimum asymptotic covariance for estimate $\hat{\boldsymbol{\gamma}}$ 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 partial likelihood for standard Cox model. Expression (23) shows that the asymptotic variance for regression parameter estimates based on 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 $\widehat{\boldsymbol{\beta}}_{(1)}$.

Estimation of Dispersion Parameters

We now discuss the situation in which the dispersion parameters are unknown. Inspired from generalized linear models, we adopt the following adjusted Pearson estimators for the dispersion parameters.

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} \left\{ (\hat{U}_i - 1)^2 + c(i) \right\}. \tag{27}$$

The first term in (27) is the Pearson estimator, with the second term being a bias correction term.

$$\hat{\omega}^2 = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{J_i} \sum_{j=1}^{J_i} \left\{ (\hat{U}_{ij} - \hat{U}_i)^2 + c(ij) + c(i) - 2c(i)w_{ij} \right\}. \tag{28}$$

The first term in (28) 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).

In fact, our orthodox BLUP approach depends on the 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 misspecification of the random effects distributions (Ma 1999), and thus covers non-Tweedie random effects such as log-normal random effects.

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 for $\hat{U}_i$ 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 subcluster (i, j) divided by the average of all responses, respectively. The initial dispersion parameter estimates are calculated from 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.

I.5 Applications

The methods described in the previous section can be applied to the American Cancer Society Study data. Because of the possibility of clustering effects at the city level, we included a random effect for each city in our re-analysis of these data. The corresponding random effects Cox regression model assumes that, given random effects, the hazard functions for individuals are conditionally independent with the hazard function for individual j from city i given by

$$h_{ij}^{(s)}(t) = h_0^{(s)}(t) u_i \exp(\boldsymbol{\beta}^\top \mathbf{x}_{ij}^{(s)}), \tag{29}$$

where the cohort is stratified on the basis of five year age groups, race and sex for the American Cancer Society Study ($s = 1, \ldots, 96$). The City-specific random effects are assumed to follow Tweedie distributions with mean unity and variance σ^2, with $r = 2$ corresponding to a gamma distribution. The Peto-Breslow approximation (Cox and Oakes 1984) for tied survival times was used in all analyses.

The results of fitting the random effects Cox regression model to the American Cancer Society Study data is shown in Table 50. In the ACS Study, the random effects Cox model provides estimates of relative risk with associated confidence limits similar to those based on the two-stage model, thereby confirming the accuracy of the simpler two-stage approach.

I.6 Summary

In this appendix, we introduced a class of random effects Cox models which accommodates a wide range of nested random effects distributions based on Tweedie exponential dispersion model distributions. 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 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. This new methodology is directly applicable to the analysis of survival data with one or more levels of clustering, of the type found in the American Cancer Society Study of air pollution and mortality.