Journal of Statistical Planning and Inference 112 (2003) 43–62

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Empirical Bayes estimation smoothing of relative risks in disease mapping

Jane L. Meza*

University of Nebraska Medical Center, Omaha, NE 68198-4350, USA

Abstract

A common index of disease incidence and mortality is the standardized mortality ratio (SMR). The SMR is a reliable measure of relative risk for large geographical regions such as countries or states, but may be unreliable for small areas such as counties. This paper reviews several empirical Bayes methods for producing smoothed estimates of the SMR as well as the conditional autoregressive procedure which accounts for spatial correlation. A multi-level Poisson model with covariates is developed, and estimating functions are used to estimate model parameters as in Lahiri and Maiti (University of Nebraska Technical Report, 2000). A hybrid of parametric bootstrap and delta methods is used to estimate the MSE. The proposed measure captures all sources of uncertainty in approximating the MSE of the proposed empirical Bayes estimator of the SMR.

0. Introduction

Disease mapping is a method used by epidemiologists, medical demographers and biostatisticians to understand the geographical distribution of a disease. Disease maps may be useful for government agencies to allocate resources or identify hazards related to disease such as incidence of leukemia near nuclear installations. Faster computers have led to advances in disease mapping, allowing for more complex models and estimation methods, larger data sets and improved graphics.

Let θ_i be the unknown relative risk for region i with probability density function $f(\theta_i)$. The standardized mortality ratio (SMR) is a common measure of relative risk. For example, Ishibashi et al. (1999) recently used the SMR to study the death rate and causes of death for patients with ulcerative colitis in Fukuoka, Japan. The SMR

* Tel.: +1-402-559-4112; fax: +1-402-559-7259.
E-mail address: jmeza@unmc.edu (J.L. Meza).

0378-3758/03/$ - see front matter
PII: S0378-3758(02)00322-1

is defined as $\hat{\theta}_i = O_i/e_i$ where O_i is the observed number of deaths and e_i is the expected number of deaths for region i. The SMR, however, has limitations. The variance of the SMR is θ_i/e_i, and is large in regions where the expected value is small (small population size) and small for regions where the expected value is large (large population size). This makes interpretation of the SMR difficult (Lawson et al., 2000). For example, a SMR of 1 is obtained when the observed number of deaths and the expected number of deaths are each 500 or when the observed number of deaths and the expected number of deaths are each 5. Further, in regions where there are no observed deaths, the SMR is zero, regardless of the population size.

Much work has focused on borrowing information from other geographic areas via modeling and using empirical Bayes methods, thus reducing the total mean-square error. Clayton and Kaldor (1987), Marshall (1991a), Lahiri and Maiti (2000) and others used empirical Bayes methods to smooth the SMR. Here we review these methods and adapt the method used by Lahiri and Maiti (2000) to propose a new empirical Bayes estimator of the SMR. Lahiri and Maiti (2000) estimated the model parameters by solving a system of optimal estimating equations. We extend their Gamma model to include covariates via a regression model along with estimating functions (EF) to estimate the model parameters and address the problem of measuring the uncertainty of the proposed empirical Bayes estimator. See Christiansen and Morris (1997) and Waller et al. (1997a) for other methods involving the Gamma model with auxiliary data. Most work on measuring the uncertainty of the empirical Bayes estimator has focused on normal linear models. Morris (1983), Prasad and Rao (1990), Singh et al. (1998) and Butar and Lahiri (2003) all examine normal linear models. The models here do not assume normality and hence these methods do not apply. Readers interested in hierarchical Bayes are referred to Ghosh et al. (1998), Waller et al. (1997b) and Bernardinelli and Montomoli (1992).

In Section 1, the empirical Bayes estimator is described. Section 2 discusses the Gamma model and estimation of the model parameters. Section 3 extends the Gamma model considered in Section 2 to include auxiliary data and discusses estimation of the model parameters. A regression model along with estimating functions is proposed to estimate the model parameters. In Section 4, the Log-Normal model, an alternative to the Gamma model, and estimation of model parameters is discussed. Section 5 considers extension of the Log-Normal model to include covariates. The measure of uncertainty of the empirical Bayes estimator in these models is addressed in Section 6. Section 7 discusses spatial correlation. Finally, in Section 8 an example is given using Nebraska prostate cancer data previously considered by Cowles et al. (1999).

1. Empirical Bayes estimation of relative risks

Let O_{il} and N_{il} be the observed number of deaths from disease and the population size for the lth age group in the ith region ($i = 1,\dots,m$; $l = 1,\dots,L$), respectively. Consider the multiplicative model $\widehat{O_{il}} = \theta_i \xi_l N_{il}$, where θ_i is the effect due to the ith region ($i = 1,\dots,m$) and ξ_l is the effect due to the lth age group ($l = 1,\dots,L$). The L age groups are used to determine the expected counts $\xi_l N_{il}$. The ξ_l's are assumed to be known from an external source. Clayton and Kaldor (1987) discuss estimation of ξ_l.

Consider estimation of θ_i, the relative risk for the ith region. The SMR is defined as $\hat{\theta}_i = O_i/e_i$ where $O_i = \sum_{l=1}^{L} O_{il}$ is the number of observed deaths for region i and $e_i = \sum_{l=1}^{L} \xi_l N_{il}$ is the expected number of deaths in region i ($i=1,\ldots,m$). Conditional on θ_i, the O_i's are assumed to be independent Poisson random variables with $E(O_i|\theta_i) = e_i\theta_i$.

Suppose that θ_i has prior density function $f(\theta_i)$ with $E_\theta(\theta_i) = \mu_i$ and $\mathrm{Var}_\theta(\theta_i) = \sigma_i^2$. Under the squared error loss function, the best linear Bayes estimator of θ_i (Ericson, 1969) is given by

$$\hat{\theta}_i^{\mathrm{B}} = (1 - B_i)\hat{\theta}_i + B_i\mu_i, \tag{1}$$

where $B_i = \mu_i/(e_i\sigma_i^2 + \mu_i)$ for $i = 1,\ldots,m$. The empirical Bayes estimate replaces the unknown μ_i and σ_i^2 by their estimates $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ and is given by

$$\hat{\theta}_i^{\mathrm{EB}} = (1 - \hat{B}_i)\hat{\theta}_i + \hat{B}_i\hat{\mu}_i, \tag{2}$$

where $\hat{B}_i = \hat{\mu}_i/(e_i\hat{\sigma}_i^2 + \hat{\mu}_i)$ for $i = 1,\ldots,m$. Next, several choices for $f(\theta_i)$ and estimates for σ_i^2 and μ_i are discussed.

2. Gamma model

This section presents three methods from the literature for estimating θ_i assuming the Gamma model. One advantage of this model is the reduction from $2m$ ($\mu_1,\ldots,\mu_m$, $\sigma_1^2,\ldots,\sigma_m^2$) to two parameters μ and σ^2. We begin as before, but assume a Gamma prior on θ_i.

I. Conditional on θ_i's, the O_i's are independent Poisson random variables with means $e_i\theta_i$ ($i = 1,\ldots,m$).

II. A priori, θ_i's are independently distributed as Gamma random variables with shape parameter α and scale parameter β. That is $E(\theta_i) = \alpha/\beta = \mu$ and $\mathrm{Var}(\theta_i) = \alpha/\beta^2 = \sigma^2$.

Under the squared error loss function and the Gamma model, the Bayes estimator of θ_i is given by

$$\hat{\theta}_i^{\mathrm{B}} = \frac{O_i + \alpha}{e_i + \beta} = (1 - B_i)\hat{\theta}_i + B_i\mu, \tag{3}$$

where $B_i = \beta/(e_i + \beta)$ for $i = 1,\ldots,m$. In practice, α and β are unknown and are replaced by their estimates $\hat{\alpha}$ and $\hat{\beta}$ resulting in the empirical Bayes estimate

$$\hat{\theta}_i^{\mathrm{EB}} = (1 - \hat{B}_i)\hat{\theta}_i + \hat{B}_i\hat{\mu}, \tag{4}$$

where $\hat{B}_i = \hat{\beta}/(e_i + \hat{\beta})$ for $i = 1,\ldots,m$ and $\hat{\mu} = \hat{\alpha}/\hat{\beta}$. The next section discusses various methods for estimating α and β.

2.1. Maximum likelihood estimation

Clayton and Kaldor (1987) used the maximum likelihood (ML) method to estimate the model parameters α and β. Since the distribution of $O_i|\theta_i$ is Poisson ($e_i\theta_i$) and the distribution of θ_i is Gamma (α,β), the distribution of O_i is negative binomial

$(\alpha, \beta/(e_i+\beta))$. Thus, integrating out θ_i, the log likelihood for α and β can be expressed as

$$L(\alpha,\beta)=\sum_{i=1}^{m}\left\{\log\frac{\Gamma(O_i+\alpha)}{\Gamma(\alpha)}+\alpha\log(\beta)-(O_i+\alpha)\log(e_i+\beta)\right\}. \tag{5}$$

Equating the partial derivatives of $L(\alpha,\beta)$ with respect to α and β to zero yields the following equations which can be solved using a variety of iterative methods (Newton–Raphson, etc.) to produce estimates of α and β (Clayton and Kaldor, 1987):

$$0=\sum_{i=1}^{m}\sum_{j=0}^{O_i-1}\left(\frac{1}{\hat\alpha+j}\right)+m\log\hat\beta-\sum_{i=1}^{m}\log(e_i+\hat\beta), \tag{6}$$

$$\frac{\hat\alpha}{\hat\beta}=\frac{1}{m}\sum_{i=1}^{m}\frac{O_i+\hat\alpha}{e_i+\hat\beta}=\frac{1}{m}\sum_{i=1}^{m}\hat\theta_i^{\mathrm{EB}}. \tag{7}$$

When the observed number of deaths O_i is zero, the quantity $\sum_{j=0}^{-1} 1/(\hat\alpha+j)$ in Eq. (6) is assumed to be zero.

Clayton and Kaldor (1987) developed an alternate estimation method which is computationally simpler than the maximum likelihood method.

2.2. Alternate method of estimation

Clayton and Kaldor (1987) also proposed a method combining moment and ML estimators using Eq. (7) plus

$$\frac{\hat\alpha}{\hat\beta^2}=\frac{1}{m-1}\sum_{i=1}^{m}\left(1+\frac{\hat\beta}{e_i}\right)\left(\hat\theta_i^{\mathrm{EB}}-\frac{\hat\alpha}{\hat\beta}\right)^2. \tag{8}$$

The new equation is obtained by equating the Pearsonian chi-square from the two ML equations in Section 2.1 with their asymptotic expectation. The model parameters α and β can be estimated by solving these two equations recursively following the steps below (Clayton and Kaldor, 1987).

Step 1: Start with initial estimates $\hat\alpha$ and $\hat\beta$ of α and β.

Step 2: Calculate $\hat\theta_i^{\mathrm{EB}}=(1-\hat B_i)\hat\theta_i+\hat B_i\hat\mu$ where $\hat B_i=\hat\beta/(e_i+\hat\beta)$ and $\hat\mu=\hat\alpha/\hat\beta$.

Step 3: Calculate

$$\frac{1}{m}\sum_{i=1}^{m}\hat\theta_i^{\mathrm{EB}}=C_0, \tag{9}$$

$$\frac{1}{m-1}\sum_{i=1}^{m}\left(1+\frac{\hat\beta}{e_i}\right)\left(\hat\theta_i^{\mathrm{EB}}-\frac{\hat\alpha}{\hat\beta}\right)^2=D_0. \tag{10}$$

Step 4: The new estimates of α and β are $\hat\beta=C_0/D_0$ and $\hat\alpha=C_0\hat\beta$.

Repeat steps 2–4, calculating $\hat{\theta}_i^{\mathrm{EB}}$ with the updated estimates of α and β until convergence.

2.3. Method of moments estimation

Another method of estimating α and β is the method of moments. Marshall (1991a) proposed to estimate $\alpha/\beta = \mu$ by

$$\hat{\mu} = \frac{\sum_{i=1}^{m} \hat{\theta}_i e_i}{\sum_{i=1}^{m} e_i} \tag{11}$$

and estimate $\alpha/\beta^2 = \sigma^2$ by

$$\hat{\sigma}^2 = s^2 - \frac{\hat{\mu}}{\frac{1}{m}\sum_{i=1}^{m} e_i}, \quad \text{where } s^2 = \frac{\sum_{i=1}^{m} e_i(\hat{\theta}_i - \hat{\mu})^2}{\sum_{i=1}^{m} e_i}. \tag{12}$$

If the value of $\hat{\sigma}^2$ is negative, it will be truncated at zero. From these two equations we obtain the estimates of α and β as $\hat{\alpha} = \hat{\mu}^2/\hat{\sigma}^2$ and $\hat{\beta} = \hat{\mu}/\hat{\sigma}^2$.

The ML and alternate methods in the previous sections require estimation using iterative methods, whereas the method of moments estimate has a direct solution. Another iterative method using estimating equations proposed by Lahiri and Maiti (2000) is considered next.

2.4. Estimating functions

Lahiri and Maiti (2000) used EF to estimate α and β. In general, estimating functions combine the least-squares and maximum likelihood estimation (MLE) methods. Consider a simple example in which $y_1, \ldots, y_n$ are i.i.d with $E(y_i) = \mu$ for $i = 1, \ldots, n$. We can use the estimating function $g = \sum_{i=1}^{n}(y_i - \mu)b_i$ to obtain an estimate of μ in the following manner. Suppose that $\sum_{i=1}^{n} b_i = c$. The variance of g is minimized when $b_i = c/n$ for $i = 1, \ldots, n$. Equating g to zero gives $\hat{\mu} = \bar{y}$. The b_i's, however, need not be constant and can be differentiable functions of μ. For additional information on estimating equations for Poisson models, see Godambe (1991).

Lahiri and Maiti (2000) proposed the following EF which can be solved using iterative methods:

$$\sum_{i=1}^{m} a_i^{\mathrm{opt}}(\varphi)\left(\hat{\theta}_i - \frac{\alpha}{\beta}\right) = 0, \tag{13}$$

$$\sum_{i=1}^{m} b_i^{\mathrm{opt}}(\varphi)\left\{K_i^{-1}\left(\hat{\theta}_i - \frac{\alpha}{\beta}\right)^2 - 1\right\} = 0, \tag{14}$$

where $a_i^{\mathrm{opt}} = K_i^{-1}/\sum_{i=1}^{m} K_i^{-1}$, $b_i^{\mathrm{opt}} = h_i^{-1}/\sum_{i=1}^{m} h_i^{-1}$, $K_i = \alpha(\beta + e_i)/\beta^2 e_i$, $h_i = (1 + 3\mu + 3\sigma^2(e_i + 2/\mu) + 6e_i\sigma^4/\mu^2)/(\mu + \sigma^2 e_i) - 1$, $\mu = \alpha/\beta$, $\sigma^2 = \alpha/\beta^2$ and $\varphi = (\alpha, \beta)'$.

2.5. Comparison of estimation methods

The method of moments is computationally simpler than the other methods considered. The ML, alternate and EF approaches require use of iterative methods.

Marshall (1991a) compared the ML, alternate, and method of moments estimates and found little difference between the three.

3. Extension of Gamma model using auxiliary data

The Gamma model can be extended to incorporate covariate information and estimated using EF:

I. Conditional on θ_i's, the O_i's are independent Poisson random variables with means $e_i\theta_i$, $(i=1,\dots,m)$.

II. A priori, θ_i's are independently distributed as Gamma random variables with shape parameter α and scale parameter β_i, $(i=1,\dots,m)$. Let $E(\theta_i)=\alpha/\beta_i=\mu_i$ and suppose there are covariates $x_i=(x_{i1},\dots,x_{ip})'$ such that $\mu_i=x_i'b$.

Under the squared error loss function and the Gamma model, the Bayes estimator of θ_i is given by (3) with $B_i=(\alpha/x_i'b)/(e_i+\alpha/x_i'b)$ and $\mu_i=x_i'b$ for $i=1,\dots,m$. The empirical Bayes estimate of θ_i is given in (4) with $\hat{B}_i=(\hat{\alpha}/x_i'\hat{b})/(e_i+\hat{\alpha}/x_i'\hat{b})$ for $i=1,\dots,m$ and $\hat{\mu}_i=x_i'\hat{b}$.

EF as in Lahiri and Maiti (2000) can be used to estimate $\varphi=(\alpha,b_1,\dots,b_p)'$. Define the functions f and g_k by

$$f(\varphi;O)=\sum_{i=1}^{m}a_i(\varphi)(\hat{\theta}_i-\mu_i), \tag{15}$$

$$g_k(\varphi;O)=\sum_{i=1}^{m}c_{ki}(\varphi)[\kappa_i^{-1}(\hat{\theta}_i-\mu_i)^2-1]\left[\frac{\partial\mu_i}{\partial b_k}\right] \tag{16}$$

for $k=1,\dots,p$ where $O=(\hat{\theta}_1,\dots,\hat{\theta}_m)$, $\mu_i=x_i'b$, $\kappa_i=\mathrm{Var}(\hat{\theta}_i)=x_i'b(1+e_ix_i'b/\alpha)/e_i$, and $a_i(\varphi)$ and $c_{ki}(\varphi)$ are constants to be determined optimally. Note that $E[f(\varphi;O)]=0$ and $E[g_k(\varphi;O)]=0$ for $k=1,\dots,p$ and for all φ, where the expectation is taken with respect to the extended Gamma model.

We need to find $a_i(\varphi)$ for $i=1,\dots,m$ such that $\mathrm{Var}[f(\varphi;O)]$ is minimum subject to $\sum_{i=1}^m a_i(\varphi)=1$, where the variance is taken with respect to the extended Gamma model. Following Lahiri and Maiti (2000)

$$a_i^{\mathrm{opt}}=\frac{\kappa_i^{-1}}{\sum_{i=1}^m\kappa_i^{-1}}. \tag{17}$$

Similarly, we need to find $c_{ki}(\varphi)$ for $k=1,\dots,p$ such that $\mathrm{Var}[g_k(\varphi;O)]$ is minimum subject to $\sum_{i=1}^m[(\partial\mu_i/\partial b_k)c_{ki}(\varphi)]=1$.

By Theorem A.1 (Appendix A)

$$c_{ki}^{\mathrm{opt}}(\varphi)=\frac{[(\partial\mu_i/\partial b_k)h_i]^{-1}}{\sum_{i=1}^m h_i^{-1}}, \tag{18}$$

where $h_i=\mathrm{Var}(\hat{\theta}_i-\mu_i)^2/\kappa_i=(1+3\mu_i+3\sigma_i^2(e_i+2/\mu_i)+6e_i\sigma_i^4/\mu_i^2)/(\mu_i+\sigma_i^2e_i)-1$, $\mu_i=x_i'b$ and $\sigma_i^2=(x_i'b)^2/\alpha$ for $i=1,\dots,m$.

Estimate $\varphi = (\alpha, b_1, \ldots, b_p)'$ by $\hat{\varphi} = (\hat{\alpha}, \hat{b}_1, \ldots, \hat{b}_p)'$ where $\hat{\alpha}, \hat{b}_1, \ldots, \hat{b}_p$ are found by solving

$$\sum_{i=1}^{m} a_i^{\text{opt}}(\varphi)(\hat{\theta}_i - \mu_i) = 0, \tag{19}$$

$$\sum_{i=1}^{m} c_{ki}^{\text{opt}}(\varphi)[\kappa_i^{-1}(\hat{\theta}_i - \mu_i)^2 - 1]\left[\frac{\partial \mu_i}{\partial b_k}\right] = 0 \tag{20}$$

for $k = 1, \ldots, p$ using iterative methods.

4. Log-Normal model

Let $\zeta_i = \log(\theta_i)$ and for simplicity, assume that ζ_i are i.i.d. $N(\mu, \sigma^2)$. Under the Log-Normal model, the empirical Bayes estimate $E(\theta_i | O_i)$ does not have a closed form, and will be approximated. To obtain the empirical Bayes estimate in a closed form, Clayton and Kaldor (1987) assumed the Poisson likelihood for ζ_i given O_i to be quadratic. They considered a bias-corrected version using $\hat{\zeta}_i = \log[(O_i + 0.5)/e_i]$ (since the choice of $\hat{\zeta}_i = \log(O_i/e_i)$ is undefined for $O_i = 0$) and found the empirical Bayes estimator of ζ_i to be

$$\hat{\zeta}_i^{\text{EB}} = \frac{\hat{\mu} + \hat{\sigma}^2(O_i + 0.5)\tilde{\zeta}_i - 0.5\hat{\sigma}^2}{1 + \hat{\sigma}^2(O_i + 0.5)}, \tag{21}$$

which can be written as

$$\hat{\zeta}_i^{\text{EB}} = (1 - \hat{\gamma}_i)\hat{\zeta}_i + \hat{\gamma}_i\hat{\mu} - 0.5\hat{\gamma}_i\hat{\sigma}^2, \tag{22}$$

where $\hat{\gamma}_i = [1 + \hat{\sigma}^2(O_i + 0.5)]^{-1}$. Hence, θ_i can then be estimated as $\exp(\hat{\zeta}_i^{\text{EB}})$.

Clayton and Kaldor (1987) used the EM algorithm to obtain estimates of μ and σ^2 by cycling between (21) (or equivalently (22)) and (23) and (24) until convergence.

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} \hat{\zeta}_i^{\text{EB}}, \tag{23}$$

$$\hat{\sigma}^2 = \frac{1}{m}\left(\hat{\sigma}^2\sum_{i=1}^{m}[1 + \hat{\sigma}^2(O_i + 0.5)]^{-1} + \sum_{i=1}^{m}(\hat{\zeta}_i^{\text{EB}} - \hat{\mu})^2\right). \tag{24}$$

5. Extension of Log-Normal model using auxiliary data

The Log-Normal model can be extended to include auxiliary information. Let $\zeta_i = \log(\theta_i)$ and take $\hat{\zeta}_i = \log[(O_i + 0.5)/e_i]$. Suppose there are covariates $x_i = (x_{i1}, \ldots, x_{ip})'$ such that $\zeta_i = x_i'b + \varepsilon_i$ where ε_i is $N(0, \sigma^2)$. The empirical Bayes estimator of ζ_i is given by (23) with $\hat{\mu}_i = x_i'\hat{b}$. Next, two possible methods of estimating $\varphi = (\sigma^2, b_1, \ldots, b_p)'$ are discussed.

5.1. Method of simulated moments estimation

There are p equations of the form

$$\sum_{i=1}^{m} x_{ij}\hat{\zeta}_i = \sum_{i=1}^{m} x_{ij}E_\varphi(\zeta_i) \quad \text{for } 1 \leqslant j \leqslant p \tag{25}$$

and an additional equation of the form

$$\sum_{i=1}^{m}\sum_{j\neq i} \hat{\zeta}_i\hat{\zeta}_j = \sum_{i=1}^{m}\sum_{j\neq i} E_\varphi(\zeta_i\zeta_j) \tag{26}$$

yielding a total of $p+1$ equations. Note that in (25) and (26), the left hand side is equated to its expectation and that $E_\varphi(\zeta_i)$ and $E_\varphi(\zeta_i\zeta_j)$ are unknown.

Now

$$E_\varphi(\zeta_i) = E_\varphi(x_i'b + \varepsilon_i) = E_\varphi(x_i'b + \sigma z), \tag{27}$$

where z is a standard normal random variable. Accordingly, $E_\varphi(\zeta_i)$ can be estimated as in Jiang (1998) by generating R independent standard normal random deviates. In other words, estimate $E_\varphi(\zeta_i)$ by

$$\frac{1}{R}\sum_{r=1}^{R}(x_i'b + \sigma z_r), \tag{28}$$

where z_r is a standard normal deviate for $r = 1,\ldots,R$.

Similarly, $E_\varphi(\zeta_i\zeta_j) = E_\varphi(x_i'b + \sigma z)(x_j'b + \sigma z)$ can be estimated by

$$\frac{1}{R}\sum_{r=1}^{R}(x_i'b + \sigma z_r)(x_j'b + \sigma z_r). \tag{29}$$

These $p+1$ equations can be solved using the Newton–Raphson procedure using initial estimates for σ^2 and $b_1,\ldots,b_p$. See Jiang (1998) for more details.

5.2. Maximum likelihood estimation

Breslow (1984) modeled extra-poisson variation in log-linear models and outlined the ML method for estimating $\varphi = (\sigma^2, b_1, \ldots, b_p)'$. Assuming the observed values O_i are large, $\zeta_i = \log(\theta_i)$ is approximately $N(x_i'b, \sigma^2 + \tau_i^2)$, where $\tau_i^2 = 1/E(O_i)$. We will estimate τ_i^2 by $\hat{\tau}_i^2 = 1/O_i$. As in Breslow (1984)

Step 1: Find the weighted least-squares solution to

$$\hat{\zeta}_i = x_i'b + \varepsilon_i, \tag{30}$$

where $E(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$ and the weights are given by $w_i = (\hat{\sigma}^2 + \hat{\tau}_i^2)^{-1}$.

An initial estimate of σ^2 can be taken as 0. If the residual sum of squares from the regression is approximately equal to its degrees of freedom, then the procedure stops. If not, proceed to step 2.

Step 2: Assuming that the w_i are the inverse variances, then

$$\sum_{i=1}^{m} \frac{(\hat{\zeta}_i - x_i'\hat{b})^2}{(\sigma^2 + \tau_i^2)} = m - p \tag{31}$$

in expectation, where $\hat{\zeta}_i = \log(O_i/e_i)$. An estimate of σ^2 can be found from this equation by solving recursively for σ^2 to obtain

$$\sigma^2 = (m-p)^{-1} \sum_{i=1}^{m} \frac{(\hat{\zeta}_i - x_i'\hat{b})^2}{1 + (\sigma^2 O_i)^{-1}}. \tag{32}$$

At the first iteration of step 2, a nonzero estimate of σ^2 is needed for Eq. (32). Thus, at the first iteration of step 2, estimate σ^2 by

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{m} O_i(y_i - x_i'\hat{b})^2 - (m-p)}{\sum_{i=1}^{m} O_i(1 - O_i q_i)}, \tag{33}$$

where q_i are the diagonal elements of $X(X'WX)^{-1}X'$. If the w_i's do not vary widely, the quantity $\sum_{i=1}^{m} O_i(1 - O_i q_i)$ can be estimated by $((m-p)/m)\sum_{i=1}^{m} O_i$.

Step 3: Update the estimates of the weights $w_i = (\hat{\sigma}^2 + \hat{\tau}_i^2)^{-1}$.

Repeat steps 1–3 until convergence. See Breslow (1984) for details.

6. Measures of uncertainty

Up to this point, we have only discussed the point estimators. Now, we consider estimating the measure of uncertainty for all estimators.

6.1. Bootstrap approach

The measure of uncertainty for the methods discussed can be estimated using bootstrap calculations. The calculations are discussed here using the Gamma model, but can be altered to accommodate the other models. The variance can be estimated in the following manner:

Step 1: Simulate B counts O_i^* as independent Poisson random variables with means $e_i\theta_i$.

Step 2: For each bootstrap sample, calculate θ_i^*, where θ_i^* is an estimate of θ_i.

Step 3: The sample variance of the B bootstrap samples is the bootstrap estimator of the variance.

6.2. Details for the extended Gamma model using estimating functions

A naive measure of uncertainty of $\hat{\theta}_i^{\text{EB}}$ is

$$\text{Var}(\theta_i | O_i, \hat{\varphi}) = \frac{O_i + \hat{\alpha}}{(e_i + \hat{\alpha}/x_i'\hat{b})^2} \tag{34}$$

for $i = 1, \ldots, m$. This underestimates the uncertainty of $\hat{\theta}_i^{\text{EB}}$ due to the estimation of $\varphi = (\alpha, b_1, \ldots, b_p)'$, especially for small regions.

Laird and Louis (1987) and Lahiri and Maiti (2000) used a parametric bootstrap method to define a measure of uncertainty for a normal model and the Gamma model, respectively. Under this framework, the following bootstrap model is proposed.

Let O_i^* be the sample count for county i generated from the parametric bootstrap procedure.

I. Conditional on the θ_i^*'s, the O_i^*'s are independent Poisson random variables with means $e_i\theta_i^*$ for $i=1,\dots,m$.

II. A priori, θ_i^*'s are independently distributed as Gamma random variables with shape parameter $\hat{\alpha}$ and scale parameter $\hat{\beta}_i$, $(i=1,\dots,m)$. That is $E(\theta_i^*)=\hat{\alpha}/\hat{\beta}_i=\hat{\mu}_i$ and $\hat{\mu}_i=x_i'\hat{b}$ where $\hat{b}$ is the vector of estimated regression parameters, $\hat{b}=(\hat{b}_1,\dots,\hat{b}_p)'$, $x_i=(x_{i1},\dots,x_{ip})'$. Let $\hat{\varphi}^*=(\hat{\alpha}^*,\hat{b}_1^*,\dots,\hat{b}_p^*)'$. That is, $\hat{\varphi}^*$ is found by replacing O_i by O_i^*, $(i=1,\dots,m)$. The Laird–Louis method gives the following measure of uncertainty for $\hat{\theta}_i^{\mathrm{EB}}$:

$$\mathrm{Var}_i^{\mathrm{LL}}=E_*[\mathrm{Var}(\theta_i|O_i,\hat{\varphi}^*)]+\mathrm{Var}_*[E(\theta_i|O_i,\hat{\varphi}^*)], \tag{35}$$

where E_* and Var_* are the expectation and variance with respect to the above model.

Under the Gamma model, Lahiri and Maiti (2000) found that $\mathrm{Var}_i^{\mathrm{LL}}$ may be smaller than the naive measure $\mathrm{Var}(\theta_i|O_i,\hat{\varphi})$ for some small regions and developed an alternate measure of uncertainty of $\hat{\theta}_i^{\mathrm{EB}}$.

As in Lahiri and Maiti (2000), begin by considering the integrated Bayes risk of $\hat{\theta}_i^{\mathrm{EB}}$, defined as $r(\hat{\theta}_i^{\mathrm{EB}})=E(\hat{\theta}_i^{\mathrm{EB}}-\theta_i)^2$ where the expectation is with respect to the extended Gamma model. A measure of uncertainty of $\hat{\theta}_i^{\mathrm{EB}}$ can be found by estimating $r(\hat{\theta}_i^{\mathrm{EB}})$. It can easily be shown that

$$r(\hat{\theta}_i^{\mathrm{EB}})=M_{1i}(\varphi)+M_{2i}(\varphi), \tag{36}$$

where $M_{1i}(\varphi)=E(\hat{\theta}_i^{\mathrm{B}}-\theta_i)^2=E[(O_i+\alpha)/(e_i+\alpha/x_i'b)^2]$ and $M_{2i}(\varphi)=E(\hat{\theta}_i^{\mathrm{EB}}-\hat{\theta}_i^{\mathrm{B}})^2$. The term $M_{2i}(\varphi)$ can be estimated by the first two terms in the Taylor series expansion of $\hat{\theta}_i^{\mathrm{EB}}$ around φ to obtain

$$M_{2i}(\varphi)\doteq E[\mathrm{tr}\{(\nabla\hat{\theta}_i^{\mathrm{B}})(\nabla\hat{\theta}_i^{\mathrm{B}})'(\hat{\varphi}-\varphi)(\hat{\varphi}-\varphi)'\}], \tag{37}$$

where $\nabla\hat{\theta}_i^{\mathrm{B}}=[(\partial/\partial\alpha)\hat{\theta}_i^{\mathrm{B}},(\partial/\partial b_1)\hat{\theta}_i^{\mathrm{B}},\dots,(\partial/\partial b_p)\hat{\theta}_i^{\mathrm{B}}]$ and the approximation of $M_{2i}(\varphi)$ is of order $\mathrm{O}(m^{-1})$. $M_{1i}(\hat{\varphi})$ can be approximated by examining the first three terms in the Taylor Series expansion of $M_{1i}(\hat{\varphi})$ around φ to obtain

$$E[M_{1i}(\hat{\varphi})]\doteq M_{1i}(\varphi)+G_{1i}(\varphi)+G_{2i}(\varphi), \tag{38}$$

where $G_{1i}(\varphi)=[\nabla M_{1i}(\varphi)]E(\hat{\varphi}-\varphi)$, $\nabla M_{1i}(\varphi)=[(\partial/\partial\alpha)M_{1i}(\varphi),(\partial/\partial b_1)M_{1i}(\varphi),\dots,(\partial/\partial b_p)M_{1i}(\varphi)]$, $G_{2i}(\varphi)=\frac{1}{2}\mathrm{tr}[H_i(\varphi)\Sigma(\varphi)]$, $\Sigma(\varphi)=E(\hat{\varphi}-\varphi)(\hat{\varphi}-\varphi)'$ and

$$H_i(\varphi)=\begin{bmatrix}\frac{\partial^2}{\partial\alpha^2}M_{1i}(\varphi) & \frac{\partial^2}{\partial\alpha\partial b_1}M_{1i}(\varphi) & \cdots & \frac{\partial^2}{\partial\alpha\partial b_p}M_{1i}(\varphi)\\ & \frac{\partial^2}{\partial b_1^2}M_{1i}(\varphi) & \cdots & \frac{\partial^2}{\partial b_1\partial b_p}M_{1i}(\varphi)\\ & & \ddots & \vdots\\ & & & \frac{\partial^2}{\partial b_p^2}M_{1i}(\varphi)\end{bmatrix}. \tag{39}$$

Since the bias of $M_{1i}(\hat{\varphi})$ involves a term of order $O(m^{-1})$, the bias cannot be ignored. Estimate $M_{1i}(\varphi)$ by

$$m_{1i}(\hat{\varphi}) = (O_i + \hat{\alpha})/(e_i + \hat{\alpha}/x_i'\hat{b})^2 + G_{1i}(\hat{\varphi}) + G_{2i}(\hat{\varphi}), \tag{40}$$

where $G_{1i}(\hat{\varphi}) = [\nabla M_{i1}(\varphi)]|_{\hat{\varphi}}\ E_*(\hat{\varphi}^* - \hat{\varphi})$, $G_{2i}(\hat{\varphi}) = \frac{1}{2}\mathrm{tr}[H_i(\hat{\varphi})\Sigma(\hat{\varphi})]$ and $\Sigma(\hat{\varphi}) = E_*(\hat{\varphi}^* - \hat{\varphi})(\hat{\varphi}^* - \hat{\varphi})'$. Estimate $M_{2i}(\varphi)$ by

$$m_{2i}(\hat{\varphi}) = \mathrm{tr}[(\nabla\hat{\theta}_i^{\mathrm{EB}})(\nabla\hat{\theta}_i^{\mathrm{EB}})'\Sigma(\hat{\varphi})], \tag{41}$$

where $\nabla\hat{\theta}_i^{\mathrm{EB}} = [(\partial/\partial\alpha)\nabla\hat{\theta}_i^{\mathrm{EB}}, (\partial/\partial b_1)\nabla\hat{\theta}_i^{\mathrm{EB}}, \ldots, (\partial/\partial b_p)\nabla\hat{\theta}_i^{\mathrm{EB}}]|_{\hat{\varphi}} = [(e_i - O_i/x_i'\hat{b})/(e_i + \hat{\alpha}/x_i'\hat{b})^2, (O_i + \hat{\alpha})\hat{\alpha}x_{i1}/(x_i'\hat{b}e_i + \hat{\alpha})^2, \ldots, (O_i + \hat{\alpha})\hat{\alpha}x_{ip}/(x_i'\hat{b}e_i + \hat{\alpha})^2]$.

A measure of uncertainty that accounts for the uncertainty due to estimation of φ is

$$\mathrm{Var}_i^P = m_{1i}(\hat{\varphi}) + m_{2i}(\hat{\varphi}) \tag{42}$$

which is correct up to the order $o(m)^{-1}$

The values $E_*(\hat{\varphi}^* - \hat{\varphi})$ and $\Sigma(\hat{\varphi})$ and can be estimated by bootstrap Monte Carlo simulation in the following manner:

Step 1: Generate $B=1000$ parametric bootstrap samples of size m, $(O_i^*,\ i=1,\ldots,m)$.

Step 2: For each bootstrap sample, calculate estimates of the parameters to obtain $\hat{\varphi}^* = (\hat{\alpha}^*, \hat{b}_1^*, \ldots, \hat{b}_p^*)'$.

Step 3: For each bootstrap sample, calculate $(\hat{\varphi}^* - \hat{\varphi})$.

Step 4: Combine the bootstrap estimates to estimate $E_*(\hat{\varphi}^* - \hat{\varphi})$ and $\Sigma(\hat{\varphi})$.

7. Spatial correlation

It is possible that the relative risks (or log relative risks) are correlated. It is well known that disease rates vary by geography. Much work has focused on the case when the correlation depends on geographical proximity. Marshall (1991b) gives an extensive review of methods for analysis of spatial patterns of disease.

A primary concern is identifying clusters of disease. A cluster of disease is a focus of high (or low) incidence (Marshall, 1991b). A cluster which can be explained by examining the age distribution of the area of concern is not as interesting as areas for which the clustering is unexplained. Clustering may occur due to an elevated risk in the cluster so that individuals within the cluster are at independently higher risk than individuals outside the cluster. Clustering may also occur due to spatial interaction of a disease with a high rate of transmission. Clusters may also occur by chance, for example from random effects. The Log-Normal model is an example of a model which takes into account the random effects by using extra-Poisson variation.

Several methods to account for geographic proximity in the Bayesian approach exist. One infrequently used option is to apply a prior which is position dependent. A second option is to model the neighboring areas as being correlated. Alternatively, spatial auto-models are used to model the spatial dependence of the prior. In this method, the mean is modeled conditional on the mean of its neighbors (Marshall, 1991b). One such method is the spatial conditional autoregression (CAR) model proposed by Clayton and Kaldor (1987). For more information on spatial autoregression procedures, see

Besag (1974), Cook and Pocock (1983), Clayton and Kaldor (1987), Mollie and Richardson (1991), Marshall (1991b) and Cressie (1992).

7.1. Estimation under the conditional autoregression model

Suppose that the log relative risks ζ_i are correlated by geographic proximity. Assume the Log-Normal model and let ζ be the vector of log relative risks where $E(\zeta)=\mu$ and $\mathrm{Var}(\zeta)=\Sigma$. The simplest form of the Log-Normal model is when the ζ_i's are i.i.d., in which case $\mu_i=c$ and $\Sigma=\sigma^2 I$ for all $i=1,\dots,m$. Conditional autoregression is defined in following manner:

$$E(\zeta_i|\zeta_j, j\neq i)=\mu_i+\rho\sum_j W_{ij}(\zeta_j-\mu_j), \tag{43}$$

$$\mathrm{Var}(\zeta_i|\zeta_j, j\neq i)=\sigma^2. \tag{44}$$

The matrix W is known as the adjacency matrix of the map, where $W_{ij}=1$ if area i and area j are adjacent and $W_{ij}=0$ otherwise. Besag (1974) showed that $E(\zeta)=\mu$ and $\mathrm{Var}(\zeta)=\Sigma=\sigma^2(I-\rho W)^{-1}$ assuming the conditional distributions $\zeta_i|\zeta_j,\ j\neq i$ are normal.

The Bayes estimate of θ is $\exp(b)$ where b is the mean of the posterior density of ζ given O

$$b=[\Sigma^{-1}-\psi''(\hat\zeta)]^{-1}[\Sigma^{-1}\mu-\psi''(\hat\zeta)\hat\zeta+\psi'(\hat\zeta)]. \tag{45}$$

Here, ψ is the Poisson likelihood for $\zeta|O$ which is assumed to be quadratic. $\psi'(\hat\zeta)$ and $\psi''(\hat\zeta)$ are the first and second derivatives of ψ with respect to ζ. The empirical Bayes estimate of θ is $\exp(\hat b)$, where μ and Σ are replaced by their MLEs $\hat\mu$ and $\hat\Sigma$.

The EM algorithm can be used to obtain MLEs of μ and Σ as follows.

E-step: Using the current (or initial) estimates of μ,σ^2 and ρ, calculate

$$Q=-\frac{m}{2}\log(2\pi)+\sum_{i=1}^{m}\frac{1}{2}\log(1-\rho\lambda_i)-\frac{m}{2}\log(\sigma^2)$$
$$-\frac{1}{2\sigma^2}[\mathrm{tr}(S(1-\rho W))+(b-\mu)'(1-\rho W)(b-\mu)], \tag{46}$$

where λ_i are the eigenvalues of the adjacency matrix W and $S=[\Sigma^{-1}-\psi''(\hat\zeta)]^{-1}$.

M-step: Maximize Q with respect to μ and Σ to update their estimates.

$$\hat\mu=b, \tag{47}$$

$$\hat\sigma^2=\frac{1}{m}[\mathrm{tr}(S(1-\rho W))+(b-\hat\mu)'(1-\rho W)(b-\hat\mu)]. \tag{48}$$

To estimate ρ, maximize $g(\rho)=\sum_{i=1}^{m}[\log(1-\rho\lambda_i)]-m\log(\hat\sigma^2)$. Clayton and Kaldor (1987) used a NAG search subroutine to estimate ρ; Mollie and Richardson (1991) used the Newton–Raphson algorithm. An application of the CAR procedure can be found in Cowles et al. (1999).

7.2. Measures of uncertainty for the conditional autoregression model

Mollie and Richardson (1991) provide details on estimating the uncertainty using the conditional expected information of the prior of ζ. Cressie (1992) used an asymptotic expansion as in Prasad and Rao (1990). Alternatively, the bootstrap approach discussed in Section 6.1 may also be used.

8. Example

In this section, the methods discussed are applied to the data set analyzed by Cowles et al. (1999). The data, compiled by the Nebraska Cancer Registry for 1990–1991, consist of prostate cancer incidence records gathered from each Nebraska hospital as well as Nebraska residents diagnosed and/or treated at hospitals in Colorado, Missouri, Wyoming, Iowa and South Dakota as well as patients diagnosed and/or treated at Nebraska outpatient facilities. The US Census Bureau estimates of average population per square mile for 1990 were used as a covariate.

Table 1 displays estimates of incidence of prostate cancer for combined years 1990 and 1991 for the 93 Nebraska counties. Table 2 summarizes the estimates. The ML method for the Log-Normal model is computationally simpler than the method of simulated moments, so only the ML method is displayed here.

The SMRs range from 0 (Arthur and Grant counties) to 2.60 (Thomas county). In general, the SMRs are highly variable when the expected value is small and the observed value is non-zero. The SMR and its estimated standard deviation are zero when the observed value is zero.

All of the empirical Bayes methods produce non-zero estimates. In general, when the expected count is large (for example Douglas county), the empirical Bayes estimate is similar to the SMR. When the expected count is small (for example Banner county), the empirical Bayes methods result in an estimate smoothed toward the overall mean.

For the Gamma model, the ML estimates exhibited less smoothing and were more variable than the other estimates. This may be due to the sensitivity of the iterative procedure to the choice of the initial estimates of model parameters. The alternate (ALT) and EF estimation methods produce similar estimates. The ALT, method of moments (MOM) and EF estimation methods have smaller variability than the ML method, with smallest variability in the MOM estimate.

In the extended Gamma model using auxiliary data, the method of EF using covariates (EFC), had larger variability than the ALT, MOM and EF methods for the Gamma model without covariates. One possible explanation of the performance of the EFC estimate is the choice of auxiliary variable. The selection of an auxiliary variable thought to be more highly associated with incidence of prostate cancer may produce EFC estimates with smaller variability.

The EM method for the Log-Normal model produces estimates similar to the ALT, MOM and EF estimates in the Gamma model. In the Log-Normal model, the EM method has smaller variability than the ML method. Again, the choice of auxiliary variable may produce ML estimates with smaller variability.

Table 1
Estimators of Nebraska prostate cancer rates, 1990–1991

County	O	SMR	SD (SMR)	Gamma model						Log-Normal model	
				ML	ALT	MOM	SD(MOM)	EF	EFC	EM	ML
Thomas	4	2.60	1.30	2.18	1.06	1.11	0.07	1.06	1.61	1.13	1.42
Dodge	113	1.83	0.17	1.83	1.63	1.62	0.12	1.63	1.71	1.65	1.77
Hooker	4	1.73	0.87	1.64	1.02	1.07	0.15	1.02	1.40	1.05	1.21
Boone	24	1.58	0.32	1.57	1.23	1.24	0.19	1.23	1.48	1.25	1.41
Holt	37	1.49	0.25	1.49	1.25	1.26	0.13	1.25	1.45	1.27	1.39
McPherson	73	1.48	0.17	1.48	1.34	1.34	0.08	1.34	1.46	1.34	1.43
Deuel	8	1.44	0.51	1.43	1.05	1.09	0.16	1.05	1.32	1.07	1.20
Kimball	12	1.40	0.40	1.40	1.08	1.12	0.17	1.08	1.33	1.10	1.22
Platte	60	1.40	0.18	1.40	1.26	1.27	0.14	1.26	1.33	1.27	1.35
Saunders	47	1.37	0.20	1.37	1.22	1.23	0.07	1.22	1.33	1.22	1.31
Kearney	17	1.33	0.32	1.33	1.09	1.12	0.17	1.09	1.27	1.10	1.21
Antelope	22	1.32	0.28	1.33	1.12	1.14	0.18	1.12	1.28	1.13	1.23
Logan	2	1.30	0.92	1.32	0.96	1.02	0.11	0.96	1.21	0.98	1.04
Burt	24	1.28	0.26	1.28	1.11	1.13	0.16	1.11	1.24	1.11	1.20
Gage	58	1.26	0.16	1.26	1.16	1.18	0.06	1.16	1.23	1.17	1.22
Adams	61	1.24	0.16	1.24	1.16	1.17	0.13	1.16	1.18	1.16	1.21
Franklin	14	1.24	0.33	1.25	1.05	1.08	0.07	1.05	1.21	1.06	1.14
Clay	18	1.16	0.27	1.17	1.03	1.07	0.14	1.03	1.14	1.04	1.09
Lancaster	273	1.15	0.07	1.15	1.13	1.14	0.07	1.13	1.19	1.13	1.14
Stanton	10	1.13	0.36	1.15	0.99	1.04	0.07	0.99	1.11	1.00	1.05
Cuming	25	1.09	0.22	1.10	1.02	1.05	0.15	1.02	1.09	1.02	1.06
Douglas	519	1.07	0.05	1.07	1.07	1.07	0.16	1.07	1.08	1.07	1.07
Garfield	6	1.07	0.43	1.10	0.96	1.02	0.10	0.96	1.10	0.97	1.00
Furnas	16	1.04	0.26	1.05	0.98	1.02	0.13	0.98	1.05	0.98	1.01
Scotts Bluff	66	1.03	0.13	1.04	1.01	1.03	0.16	1.01	1.01	1.01	1.03
Pierce	16	1.00	0.25	1.02	0.96	1.00	0.14	0.96	1.01	0.97	0.98
Harlan	10	0.99	0.31	1.02	0.95	1.00	0.12	0.95	1.03	0.96	0.97
Merrick	16	0.99	0.25	1.01	0.96	1.00	0.13	0.96	1.00	0.96	0.98
Custer	29	0.98	0.18	0.99	0.96	0.99	0.13	0.96	1.00	0.96	0.97
Lincoln	52	0.98	0.14	0.98	0.96	0.98	0.11	0.96	0.98	0.97	0.97
Saline	26	0.98	0.19	0.99	0.96	0.99	0.16	0.96	0.98	0.96	0.97
Johnson	11	0.96	0.29	0.98	0.94	0.99	0.14	0.94	0.99	0.95	0.95
Knox	24	0.96	0.20	0.97	0.95	0.98	0.13	0.95	0.98	0.95	0.95
Colfax	19	0.95	0.22	0.96	0.94	0.98	0.13	0.94	0.95	0.94	0.95
Pawnee	10	0.95	0.30	0.98	0.94	0.99	0.13	0.94	1.00	0.95	0.95
Sioux	3	0.95	0.55	1.04	0.93	1.00	0.10	0.93	1.08	0.95	0.95
Nuckolls	14	0.94	0.25	0.96	0.93	0.98	0.15	0.93	0.97	0.94	0.94
Box Butte	17	0.92	0.22	0.94	0.93	0.97	0.13	0.93	0.95	0.93	0.93
Valley	12	0.92	0.27	0.95	0.92	0.97	0.09	0.93	0.96	0.94	0.93
York	24	0.92	0.19	0.94	0.93	0.96	0.15	0.93	0.93	0.93	0.93
Cass	28	0.90	0.17	0.91	0.91	0.94	0.12	0.91	0.89	0.92	0.91
Webster	11	0.90	0.27	0.93	0.92	0.97	0.14	0.92	0.95	0.93	0.92
Dawson	33	0.87	0.15	0.88	0.89	0.92	0.11	0.89	0.89	0.90	0.89
Wayne	12	0.86	0.25	0.89	0.90	0.95	0.11	0.90	0.89	0.91	0.89

Table 1 (*continued*)

County	O	SMR	SD (SMR)	Gamma model						Log-Normal model	
				ML	ALT	MOM	SD(MOM)	EF	EFC	EM	ML
Otoe	26	0.85	0.17	0.87	0.88	0.92	0.13	0.88	0.87	0.89	0.87
Polk	12	0.84	0.24	0.87	0.89	0.94	0.14	0.89	0.89	0.90	0.87
Brown	7	0.83	0.31	0.88	0.90	0.96	0.13	0.90	0.93	0.91	0.88
Sarpy	47	0.82	0.12	0.82	0.84	0.87	0.13	0.84	0.91	0.85	0.83
Washington	21	0.82	0.18	0.84	0.87	0.91	0.10	0.87	0.82	0.88	0.85
Dakota	17	0.80	0.19	0.82	0.86	0.90	0.13	0.86	0.76	0.88	0.84
Fillmore	11	0.80	0.24	0.83	0.87	0.93	0.07	0.87	0.86	0.89	0.85
Keith	13	0.80	0.22	0.83	0.87	0.92	0.13	0.87	0.86	0.89	0.85
Cheyenne	14	0.79	0.21	0.82	0.86	0.91	0.12	0.86	0.85	0.88	0.84
Blaine	1	0.77	0.77	1.00	0.92	0.99	0.08	0.92	1.07	0.93	0.92
Cedar	16	0.76	0.19	0.78	0.84	0.89	0.12	0.84	0.81	0.86	0.81
Dawes	11	0.76	0.23	0.79	0.85	0.91	0.12	0.85	0.84	0.87	0.82
Cherry	9	0.75	0.25	0.79	0.86	0.91	0.12	0.86	0.85	0.88	0.82
Dixon	10	0.75	0.24	0.78	0.85	0.91	0.13	0.85	0.82	0.87	0.82
Thayer	14	0.75	0.20	0.77	0.84	0.89	0.15	0.84	0.81	0.86	0.81
Buffalo	35	0.74	0.12	0.75	0.79	0.82	0.10	0.79	0.75	0.81	0.77
Hall	56	0.74	0.10	0.75	0.78	0.80	0.12	0.78	0.70	0.79	0.76
Hamilton	11	0.72	0.22	0.75	0.83	0.89	0.10	0.83	0.79	0.86	0.80
Madison	1	0.71	0.71	0.95	0.91	0.99	0.14	0.91	0.69	0.93	0.91
Thurston	8	0.71	0.25	0.76	0.85	0.91	0.14	0.85	0.80	0.87	0.81
Rock	3	0.70	0.41	0.81	0.88	0.96	0.14	0.89	0.92	0.91	0.86
Jefferson	15	0.68	0.18	0.71	0.79	0.84	0.12	0.79	0.74	0.82	0.75
Perkins	5	0.66	0.30	0.73	0.85	0.92	0.12	0.85	0.82	0.88	0.81
Richardson	16	0.66	0.16	0.68	0.77	0.82	0.11	0.77	0.71	0.80	0.73
Sheridan	10	0.65	0.21	0.69	0.80	0.86	0.07	0.80	0.75	0.83	0.76
Loup	1	0.64	0.64	0.89	0.90	0.98	0.07	0.91	1.02	0.92	0.89
Dundy	4	0.63	0.31	0.71	0.85	0.92	0.03	0.85	0.82	0.88	0.81
Gosper	3	0.63	0.37	0.74	0.87	0.94	0.15	0.87	0.86	0.89	0.83
Howard	8	0.62	0.22	0.66	0.80	0.86	0.12	0.80	0.73	0.84	0.75
Phelps	11	0.61	0.18	0.64	0.77	0.83	0.11	0.77	0.69	0.81	0.72
Greeley	4	0.58	0.29	0.66	0.83	0.90	0.12	0.83	0.78	0.87	0.78
Seward	15	0.58	0.15	0.60	0.72	0.77	0.13	0.72	0.63	0.77	0.67
Banner	1	0.57	0.57	0.83	0.90	0.97	0.07	0.90	0.98	0.92	0.87
Nemaha	9	0.57	0.19	0.60	0.76	0.82	0.11	0.76	0.66	0.80	0.70
Red Willow	12	0.57	0.17	0.60	0.74	0.79	0.12	0.74	0.65	0.78	0.68
Chase	5	0.55	0.25	0.62	0.80	0.87	0.11	0.80	0.72	0.84	0.74
Wheeler	1	0.53	0.53	0.79	0.89	0.97	0.12	0.89	0.96	0.91	0.86
Nance	5	0.52	0.23	0.59	0.79	0.86	0.10	0.79	0.69	0.83	0.73
Butler	9	0.47	0.16	0.51	0.69	0.76	0.10	0.70	0.58	0.76	0.63
Morrill	5	0.44	0.20	0.50	0.74	0.81	0.10	0.74	0.61	0.80	0.67
Hayes	1	0.39	0.39	0.63	0.86	0.94	0.12	0.86	0.85	0.89	0.82
Key Paha	1	0.37	0.37	0.60	0.86	0.94	0.07	0.86	0.84	0.89	0.81
Hitchcock	3	0.35	0.20	0.44	0.74	0.82	0.15	0.74	0.59	0.82	0.67
Frontier	2	0.32	0.23	0.44	0.77	0.85	0.18	0.77	0.63	0.84	0.71
Garden	1	0.15	0.15	0.28	0.72	0.81	0.12	0.72	0.52	0.84	0.69

Table 1 (*continued*)

County	O	SMR	SD (SMR)	Gamma model						Log-Normal model	
				ML	ALT	MOM	SD(MOM)	EF	EFC	EM	ML
Boyd	1	0.13	0.13	0.24	0.68	0.77	0.06	0.68	0.45	0.83	0.66
Sherman	1	0.11	0.11	0.21	0.65	0.74	0.11	0.65	0.40	0.82	0.65
Arthur	0	0.00	0.00	0.60	0.88	0.96	0.03	0.88	0.92	0.91	0.84
Grant	0	0.00	0.00	0.52	0.86	0.95	0.11	0.87	0.87	0.90	0.83

O: observed values for 1990 and 1991 combined, SMR: SMR for 1990 and 1991 combined, SD(SMR): estimated standard deviation of the SMR, ML: maximum likelihood method (2.1), ALT: alternate Method (2.2), MOM: method of moments (2.3), SD(MOM): boostrap estimated standard deviation, MOM method, EF: estimating functions (2.4), EFC: Estimating Functions using covariates (3), EM: EM algorithm for Log-Normal model (4).

Table 2
Summary of estimators

Estimator	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Mean	Standard deviation
SMR	0.00	0.64	0.83	1.04	2.60	0.87	0.40
Gamma model							
ML	0.21	0.73	0.88	1.05	2.18	0.92	0.33
ALT	0.65	0.84	0.90	0.96	1.63	0.93	0.15
MOM	0.74	0.90	0.96	1.02	1.62	0.97	0.14
EF	0.65	0.84	0.90	0.96	1.63	0.93	0.15
EFC	0.40	0.80	0.92	1.08	1.71	0.95	0.25
Log-Normal model							
EM	0.76	0.87	0.91	0.98	1.65	0.95	0.14
ML	0.63	0.81	0.89	1.01	1.77	0.94	0.21

Clayton and Kaldor (1987) examined lip cancer in Scotland and found the CAR model produced similar estimates to the Log-Normal model except for counties in which the observed value was small and the SMR was dramatically different than the surrounding counties. A similar pattern was seen in prostate cancer incidence in Nebraska. For example, the SMR for Madison county was 0.71 based on 1 observed case of prostate cancer, much lower than its neighboring counties of Boone (1.58), Platte (1.40), Antelope (1.32), Stanton (1.13) and Pierce (1.00). The estimated SMR under the Log-Normal model (0.93) was lower than the estimate under the CAR model (1.15) (Cowles et al., 1999).

The measure of uncertainty for the MOM estimates was estimated using the bootstrap approach described in Section 6.1. The standard deviation using the bootstrap approach (SD(MOM)) was more stable (range: 0.03, 0.19) than the estimated standard deviation of the SMR (SD(SMR)) (range: 0, 1.30). The value of SD(MOM) was less than SD(SMR) except when the SMR is zero, in which case SD(SMR) is incorrectly estimated to be 0.

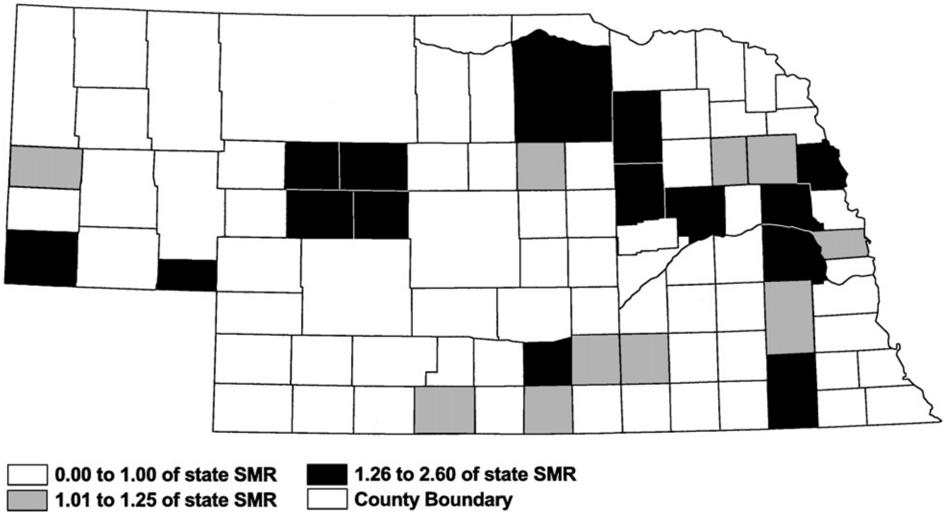

Fig. 1. Nebraska prostate cancer incidence, 1990–1991, SMR.

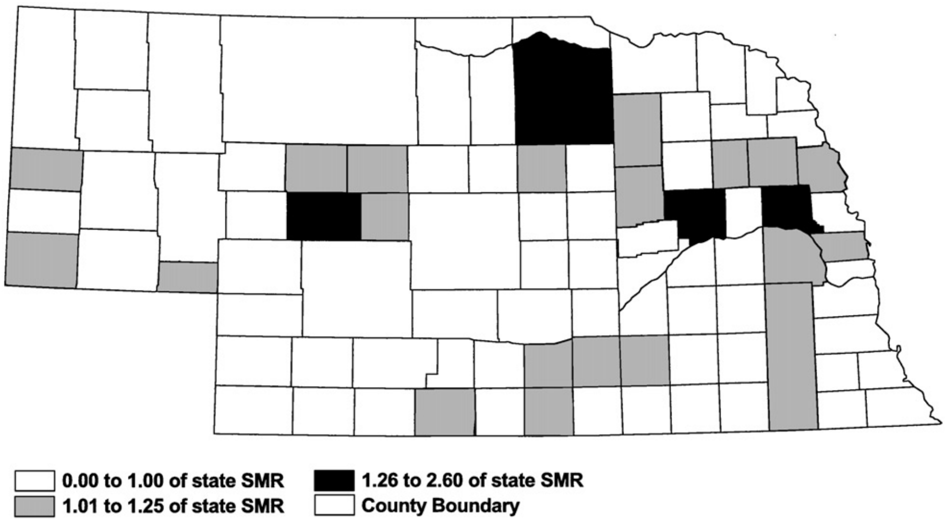

Fig. 2. Nebraska Prostate Cancer Incidence, 1990–1991 empirical Bayes smoothing SMR using MOM.

Fig. 1 displays the ratio of the county SMR to the state SMR for the incidence of prostate cancer in Nebraska for years 1990–1991. Counties with an SMR higher than the state SMR are shaded. There were 15 counties with an SMR over 1.25 of the state SMR and 10 counties with an SMR between 1.01 and 1.25 of the state SMR. Fig. 2 displays the ratio of the empirical Bayes MOM estimate to the state SMR. After smoothing the SMR, there were 4 counties (Dodge, Holt, McPherson and Platte) with smoothed SMRs over 1.25 of the state SMR and 21 counties with smoothed SMRs between 1.01 and 1.25 of the state SMR. Note from Figs. 1 and 2 that counties with higher incidence rates are often contiguous. Cowles et al. (1999) applied the CAR procedure and found there were 2 counties (Dodge and McPherson) with smoothed SMRs over 1.25 of the state SMR after accounting for spatial correlation.

Appendix A

Theorem A.1. (a) *The optimal choice for* $a_i(\varphi)$, *denoted by* $a_i^{\text{opt}}(\varphi)$, *which minimizes* $\text{Var}[f(\varphi;O)]$ *subject to* $\sum_{i=1}^m a_i(\varphi)=1$ *is given by*

$$a_i^{\text{opt}} = \frac{\kappa_i^{-1}}{\sum_{i=1}^m \kappa_i^{-1}}$$

for $i=1,\ldots,m$.

(b) *The optimal choice for* $c_{ki}(\varphi)$, *denoted by* $c_{ki}^{\text{opt}}(\varphi)$, *which minimizes* $\text{Var}[g_k(\varphi;O)]$ *subject to* $\sum_{i=1}^m [(\partial\mu_i/\partial b_k)c_{ki}(\varphi)]=1$ *is given by*

$$c_{ki}^{\text{opt}}(\varphi) = \frac{[(\partial\mu_i/\partial b_k)h_i]^{-1}}{\sum_{i=1}^m h_i^{-1}},$$

where $h_i=(1+3\mu_i+3\sigma_i^2(e_i+2/\mu_i)+6e_i\sigma_i^2/\mu_i^2)/(\mu_i+\sigma_i^2e_i)-1$, $\mu_i=x_i'b$ *and* $\sigma_i^2=(x_i'b)^2/\alpha$ *for* $i=1,\ldots,m$.

The following Lemma due to Lahiri and Maiti (2000) is needed to prove Theorem A.1.

Lemma A.1. *The optimal choice of* a_i $(i=1,\ldots,m)$ *which minimizes* $\sum_{i=1}^m a_i^2V_i$ *subject to* $\sum_{i=1}^m a_i=1$ *is given by*

$$a_i = \frac{V_i^{-1}}{\sum_{i=1}^m V_i^{-1}}.$$

Proof. The Lemma follows from the fact that

$$\sum_{i=1}^m a_i^2V_i = \sum_{i=1}^m V_i\left(a_i - \frac{V_i^{-1}}{\sum_{i=1}^m V_i^{-1}}\right)^2 + \left(\sum_{i=1}^m V_i^{-1}\right)^{-1}$$

since $\sum_{i=1}^m a_i=1$. □

Proof of Theorem A.1. First note that unconditionally, O_i follows a negative binomial distribution with mean $e_i\alpha/\beta_i$ and variance $e_i\alpha/\beta + e_i^2\alpha/\beta_i^2$. Thus $\text{Var}(\hat{\theta}_i)=\kappa_i$ and $\text{Var}(\hat{\theta}_i-\mu_i)^2/\kappa_i=h_i$ using the formula for moments of a negative binomial distribution.

Theorem A.1 (a) follows from Lemma A.1 by noting that $\mathrm{Var}[f(\varphi;O)]=\sum_{i=1}^m a_i^2(\varphi)\kappa_i$. Since $\mathrm{Var}[g_k(\varphi;O)] = \sum_{i=1}^m [(\partial\mu_i/\partial b_k)c_{ki}]^2 h_i$, by Lemma A.1 the optimal choice of $[(\partial\mu_i/\partial b_k)c_{ki}]$ which minimizes $\mathrm{Var}[g_k(\varphi;O)]$ subject to $\sum_{i=1}^m [(\partial\mu_i/\partial b_k)c_{ki}(\varphi)] = 1$ is $(\partial\mu_i/\partial b_k)c_{ki}^{\mathrm{opt}}(\varphi) = h_i^{-1}/\sum_{i=1}^m h_i^{-1}$ and thus

$$c_{ki}^{\mathrm{opt}}(\varphi) = \frac{[(\partial\mu_i/\partial b_k)h_i]^{-1}}{\sum_{i=1}^m h_i^{-1}}. \qquad \square$$

References

Bernardinelli, L., Montomoli, C., 1992. Empirical Bayes vs. fully Bayesian analysis of geographical variation in disease risk. Statist. Med. 11, 983–1007.

Besag, J.E., 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). J. Roy. Statist. Soc. B 36, 192–236.

Breslow, N.E., 1984. Extra-Poisson variation in log-linear models. Appl. Statist. 33, 38–44.

Butar, F.B., Lahiri, P., 2003. On measures of uncertainty of empirical Bayes small-area estimators. J. Statist. Plann. Inference 112, this issue.

Christiansen, C.L., Morris, C.N., 1997. Hierarchical Poisson regression modeling. J. Amer. Statist. Assoc. 92, 618–632.

Clayton, D., Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43, 671–681.

Cook, D.G., Pocock, S.J., 1983. Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. Biometrics 39, 361–371.

Cowles, M.K., Anderson, J., Mueller, K.J., Ullrich, F., 1999. Prostate cancer incidence rates, by County in Nebraska, 1990–1991. Nebraska Health Data Rep. 2 (4), 1–10.

Cressie, N., 1992. Smoothing regional maps using empirical bayes predictors. Geog. Anal. 24 (1), 75–95.

Ericson, W.A., 1969. A note on the posterior mean of a population mean. J. Roy. Statist. Soc. B 31, 332–334.

Ghosh, M., Natarajan, K., Stroud, T.W.F., Carlin, B.P., 1998. Generalized linear models for small-area estimation. J. Amer. Statist. Assoc. 93, 273–282.

Godambe, V.P., 1991. Estimating Functions. Oxford University Press, New York.

Ishibashi, N., Hirota, Y., Ikeda, M., Hirohata, T., 1999. Internat. J. Epi. 28, 609–613.

Jiang, J., 1998. Consistent estimators in generalized linear mixed models. J. Amer. Statist. Assoc. 93, 720–729.

Laird, N.M., Louis, T.A., 1987. Empirical Bayes confidence intervals based on bootstrap samples. J. Amer. Statist. Assoc. 82, 739–750.

Lahiri, P., Maiti, T., 2000. Empirical Bayes estimation of relative risks in disease mapping. University of Nebraska Technical Report.

Lawson, A.B., Biggeri, A.B., Boehning, D., Lesaffre, E., Viel, J.F., Clark, A., Schlattmann, P., Divino, F., 2000. Disease mapping models: an empirical evaluation. Statist. Med. 19, 2217–2241.

Marshall, R.J., 1991a. Mapping disease and mortality rates using empirical Bayes estimators. Appl. Statist. 40, 283–294.

Marshall, R.J., 1991b. A review of methods for the statistical analysis of spatial patterns of disease. J. Roy. Statist. Soc. A 154, 421–441.

Mollie, A., Richardson, S., 1991. Empirical Bayes estimates of cancer mortality rates using spatial models. Statist. Med. 10, 95–112.

Morris, C.N., 1983. Parametric empirical Bayes inference: theory and applications. J. Amer. Statist. Assoc. 78, 47–59.

Prasad, N.G.N., Rao, J.N.K., 1990. The estimation of mean squared errors of small area estimators. J. Amer. Statist. Assoc. 85, 163–171.

Singh, A.C., Stukel, D.M., Pfeffermann, D., 1998. Bayesian versus frequentist measures of error in small area estimation. J. Roy. Statist. Soc. B 60, 377–396.

Waller, L.A., Carlin, B.P., Xia, H., Gelfand, A.E., 1997a. Hierarchical spatio-temporal mapping of disease rates. J. Amer. Statist. Assoc. 92, 607–617.

Waller, L.A., Carlin, B.P., Xia, H., 1997b. Structuring correlation within hierarchical spatio-temporal models for disease rates. In: Gregoire, T.G., Brillinger, D.R., Diggle, P.J., Russek-Cohen, E., Warren, W.G., Wolfinger, R.D. (Eds.), Modeling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions, Lecture Notes in Statistics, Vol. 122. Springer, New York, pp. 309–320.