5-8-2006

Disease Mapping and Spatial Regression with Count Data

Jon Wakefield
University of Washington, jonno@u.washington.edu

Disease Mapping and Spatial Regression with Count Data

Jon Wakefield

Departments of Statistics and Biostatistics, Box 357232, University of Washington, Seattle, Washington 98195-7232, U.S.A.

Summary. In this paper we provide critical reviews of methods suggested for the analysis of aggregate count data in the context of disease mapping and spatial regression. We introduce a new method for picking prior distributions, and propose a number of refinements of previously-used models. We also consider ecological bias, mutual standardization, and choice of both spatial model and prior specification. We analyse male lip cancer incidence data collected in Scotland over the period 1975–1980, and outline a number of problems with previous analyses of these data. A number of recommendations are provided. In disease mapping studies, hierarchical models can provide robust estimation of area-level risk parameters, though care is required in the choice of covariate model, and it is important to assess the sensitivity of estimates to the spatial model chosen, and to the prior specifications on the variance parameters. Spatial ecological regression is a far more hazardous enterprise for two reasons. First, there is always the possibility of ecological bias, and this can only be alleviated via the inclusion of individual-level data. For the Scottish data we show that the previously used mean model has limited interpretation from an individual perspective. Second, when residual spatial dependence is modelled, and if the exposure has spatial structure, then estimates of exposure association parameters will change when compared with those obtained from the independence across space model, and the data alone cannot choose the form and extent of spatial correlation that is appropriate.

Keywords: Bayesian Methods; Ecological Bias; Ecological Correlation Studies; Hierarchical Models; Prior Distributions; Spatial Epidemiology; Standardization.

1. Introduction

In this paper we consider the analysis of population and health counts, aggregated over a set of disjoint geographical areas; recent reviews of methods for spatial epidemiological data in general may be found in Lawson et al. (1999), Elliott et al. (2000) and Waller and Gotway (2004). We critically review a number of approaches for the analysis of spatially aggregated count data, propose a number of refinements to currently-used models, and describe a procedure for prior choice. We consider two distinct aims: *disease mapping* to obtain relative risk estimates for each study area, and *spatial regression* to estimate the association between relative risk and potential risk factors. In general, counts in areas that are geographically close will display residual spatial dependence; *residual* here acknowledges that known confounders have been included in the analysis model. In a disease mapping context this dependence may be exploited in estimation of risk summaries, by smoothing across "neighbouring" areas. In a regression context, the dependence must be acknowledged since conventional statistical analysis techniques are inappropriate for dependent data.

Disease mapping has a long history in epidemiology (Walter, 2000) as part of the classic triad of person/place/time. A number of statistical reviews are available, see for example Smans and Esteve (1992), Clayton and Bernardinelli (1992), Mollié (1996) and Wakefield et al. (2000). There are numerous examples of both cancer atlases, see for example, Kemp et al. (1985) and Devesa et al. (1999), and mapping studies for

specific cancer sites; for example, Toledano et al. (2001) report spatio-temporal trends for testicular cancer, and Jarup et al. (2002) carry out mapping for prostate cancer. Similarly, numerous ecological correlation studies have been reported. For example, the contribution of environmental factors to cancer risk have been summarised by Boffetta and Nyberg (2003), who report evidence from ecological studies including three that considered mesothelioma and lung cancer in relation to asbestos, eight that investigated lung cancer and proximity to various industries that produce air pollution, and six which assessed the association between nitrates in drinking water and stomach cancer. Exposures may be directly measured in air, water or soil, or be indirect surrogates such as distance from a point source of risk such as an incinerator (e.g. Elliott et al., 1996) or a foundry (e.g. Lawson and Williams, 1994), or a line source such as a road.

As motivating example we examine incidence rates of lip cancer in males in 56 counties of Scotland, registered in 1975–1980. These data were originally reported by Kemp et al. (1985). The data consist of the observed and expected number of cases (based on the age population in each county), a covariate measuring the proportion of the population engaged in agriculture, fishing, or forestry (AFF) (exposure to sunlight is a risk factor for lip cancer, and the AFF variable is related to exposure to sunlight), and the standardised morbidity ratio (SMR), which is the ratio of the observed to expected cases. The AFF variable was read from Figure 3.6 of Kemp et al. (1985) by Clayton and Bernardinelli (1992) and takes one of only 6 values. The data include the centroids of each area under the Great Britain National Grid projection system. This is a *conformal* projection that preserves local shape when moving from three-dimensional to two-dimensional coordinates for the purposes of mapping, Waller and Gotway (2004) provide details on this and other projections. The data, along with a figure displaying the labelled centroids of each county, are available as supplementary material at http://www.biostatistics.oxfordjournals.org.

The structure of this paper is as follows. In Sections 2 and 3 we describe and critique models for disease mapping and spatial regression, respectively, illustrating their use with the Scottish data. In Section 4 we analyse a more comprehensive version of these data, obtained from the original source, and we conclude with a discussion in Section 5.

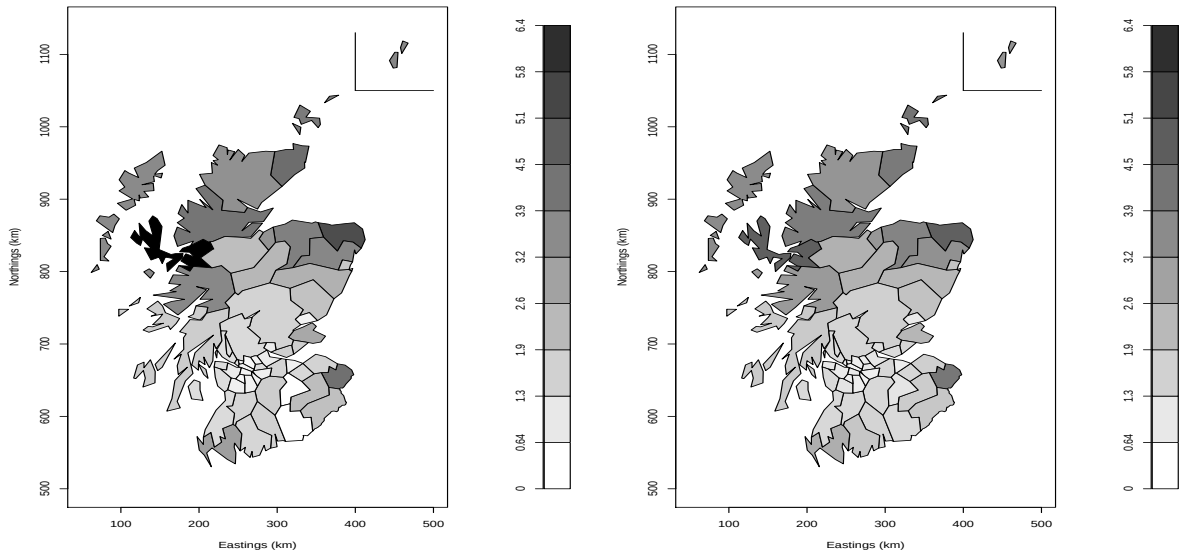
2. Disease Mapping

2.1. Motivation

Disease mapping is often carried out to investigate the geographical distribution of disease burden. Area specific estimates of risk may inform public health resource allocation by estimating the disease burden in specific areas, and the informal comparison of risk maps with exposure maps may provide clues to etiology/generate hypotheses. An additional use is to provide a context within which specific studies may be placed; for example, surveillance will be greatly helped if we have a knowledge of the variability in residual spatial risk, and the nature of that variability (spatial versus non-spatial), in order to know the "null" distribution, that is, the distribution in the absence of a "hot spot". In a similar vein, regression will be aided if we have a "prior" on the magnitude and forms of the non-spatial and spatial background variability.

2.2. Drawbacks of Simple Approaches

We assume that the study region is partitioned into n non-overlapping, areas. Though the total burden of disease is of interest, control for confounding allows the *residual* geographical distribution of risk to be



(a) SMR estimates

(b) Smoothed estimates

Fig. 1. Raw and smoothed estimates in 56 counties of Scotland. On this and subsequent maps, the Shetland Isles (county 8) has been moved south by 100km in order to use space more efficiently

investigated and modelled. Consider a confounder with K strata and let N_{ik} and Y_{ik} be the population size and number of cases in area i, stratum k, $i = 1, ..., n, k = 1, ..., K$. A starting model is $\mathrm{E}[Y_{ik}|p_{ik}] = N_{ik}p_{ik}$, where p_{ik} is the probability of disease in area i and confounder stratum k. For small-area studies in particular, the estimation of $n \times K$ probabilities is not feasible, and for a rare disease it is usual to assume that the effect of being in area i is to multiplicatively change "reference" stratum specific risks, p_k, by a constant, i.e.

$$p_{ik} = \theta_i \times p_k. \tag{1}$$

The assumption $Y_{ik}|p_{ik} \sim \mathrm{Poisson}(N_{ik}p_{ik})$ leads to

$$Y_i|\theta_i \sim \mathrm{Poisson}(E_i\theta_i) \tag{2}$$

where $E_i = \sum_{k=1}^{K} N_{ik}p_k$ are the *expected numbers* of cases in area i, based on the confounder-specific populations. This procedure is known as indirect standardization.

The SMR is given by $\mathrm{SMR}_i = Y_i/E_i$, and is an estimate of the relative risk associated with area i, and corresponds to the maximum likelihood estimator (MLE) of θ_i in model (2), $i = 1, ..., n$. Figure 1(a) shows the SMRs for the Scottish lip cancer data, and indicates a large spread with an increasing trend in the south-north direction. The variance of the estimator is $\mathrm{var}(\mathrm{SMR}_i) = \mathrm{SMR}_i/E_i$, which will be large if E_i is small. For the Scottish data the expected numbers are highly variable, with range 1.1–88.7. This variability suggests that the extreme SMRs may be based on small expected numbers, many of the large, sparsely-populated rural areas in the north have high SMRs.

Maps showing p-values of exceedence of 1 are even less informative than maps of SMRs. Although they account for sample size they do not show the extent of the risk, and areas with large populations may provide statistically significant SMRs, even for small exceedences of 1.

The above considerations led to methods being developed to *smooth* the SMRs using random effects models that use the data from the totality of areas to provide more reliable estimates in each of the constituent areas. We first describe models that do not use spatial information (Tsutakawa et al., 1985; Manton et al., 1989) before turning to models that allow both spatial and non-spatial variability (Clayton and Kaldor, 1987; Besag et al., 1991).

2.3. Non-Spatial Models

We begin by describing a simple Poisson-Gamma two-stage model that offers analytic tractability and ease of estimation, and is useful for exploratory analyses, for example, to decide on the form of the area-level risk-exposure model. At the first stage assume the likelihood is given by

$$Y_i|\theta_i, \boldsymbol{\beta} \sim_{ind} \text{Poisson}\left(E_i\mu_i\theta_i\right), \tag{3}$$

where $\mu_i = \mu(\boldsymbol{x}_i, \boldsymbol{\beta})$ describes a regression model in area-level covariates $\boldsymbol{x}_i$. At the second stage assume that across the map the deviations of the relative risks from the mean, μ_i, are modelled by

$$\theta_i|\alpha \sim_{iid} \text{Ga}(\alpha, \alpha), \tag{4}$$

a gamma distribution with mean 1, and variance $1/\alpha$. The marginal distribution of $Y_i|\boldsymbol{\beta}, \alpha$ is negative binomial with mean and variance

$$E[Y_i|\boldsymbol{\beta}, \alpha] = E_i\mu_i, \quad \text{var}(Y_i|\boldsymbol{\beta}, \alpha) = \text{E}[Y_i|\boldsymbol{\beta}, \alpha](1 + \text{E}[Y_i|\boldsymbol{\beta}, \alpha]/\alpha), \tag{5}$$

so that the variance increases as a quadratic function of the mean, and the scale parameter α can accommodate "overdispersion". This form is substantively more reasonable than the naive Poisson model; it is important to consider excess-Poisson variability resulting from unmeasured confounders, data anomalies in numerator and denominator, and model misspecification (Wakefield and Elliott, 1999, provide a discussion of these aspects).

A fully Bayesian approach to inference would consider the posterior distribution $p(\theta_1, ..., \theta_n, \boldsymbol{\beta}, \alpha|\boldsymbol{y})$ and carry out inference via the marginal distributions $p(\theta_i|\boldsymbol{y})$. Unfortunately the latter are unavailable in closed form, but an *empirical Bayes* approach obtains estimates $\widehat{\boldsymbol{\beta}}, \widehat{\alpha}$ and then proceeds as if these are known, i.e. considers $p(\theta_1, ..., \theta_n|\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \boldsymbol{y})$. Estimates $\widehat{\boldsymbol{\beta}}, \widehat{\alpha}$ usually correspond to the MLEs from $\prod_{i=1}^n \Pr(Y_i|\boldsymbol{\beta}, \alpha)$. If the aim is to gain clues to unexplained variability, θ_i may be examined; here we report the relative risk, $\text{RR}_i = \theta_i\mu_i$, where relative is with respect to E_i. Using Bayes theorem the conditional posterior is $\theta_i|\boldsymbol{y}, \widehat{\boldsymbol{\beta}}, \widehat{\alpha} \sim \text{Ga}(\widehat{\alpha} + y_i, \widehat{\alpha} + E_i\widehat{\mu}_i)$, yielding empirical Bayes estimates

$$\widehat{RR}_i = \text{E}[\text{RR}_i] \times (1 - w_i) + \text{SMR}_i \times w_i, \tag{6}$$

a weighted combination of the estimate $\text{E}[\text{RR}_i] = \widehat{\mu}_i \times \text{E}[\theta_i] = \widehat{\mu}_i$, and the SMR in area i. The *weight*

$$w_i = \frac{E_i\widehat{\mu}_i}{\widehat{\alpha} + E_i\widehat{\mu}_i}. \tag{7}$$

on the observed SMR increases as E_i increases so for areas with large populations the data dominate. If α is large then the random effects have a tight spread, and there is more shrinkage since SMRs that are far from unity are inconsistent with the total collection of estimates. This behavior illustrates both the potential benefits and hazards of smoothing; the estimates will be less variable than the SMRs, but an outlying

estimate that is not based on a large expected number, will be shrunk, and we may miss an important excess. Conlon and Louis (1999) provide a discussion of the inherent bias due to shrinkage of random effects estimators.

The above model is subtly different from the alternative $Y_i|\mathrm{RR}_i \sim_{ind} \mathrm{Poisson}(E_i \times \mathrm{RR}_i)$ with $\mathrm{RR}_i \sim_{ind} \mathrm{Ga}(\alpha^\star\mu_i, \alpha^\star)$ which has mean $\mathrm{E}[\mathrm{RR}_i] = \mu_i$ and variance $\mathrm{var}(\mathrm{RR}_i) = \mu_i/\alpha^\star$. The subtlety is that this model implies first two moments of $\mathrm{E}[Y_i|\boldsymbol{\beta}, \alpha^\star] = E_i\mu_i$ and $\mathrm{var}(Y_i|\boldsymbol{\beta}, \alpha^\star) = \mu_i(1+E_i/\alpha^\star)$, so that the mean coincides with (5), but the variance differs. For this model empirical Bayes estimates differ from (6) and are given by

$$\widehat{RR}_i = \widehat{\mu}_i \times \frac{\widehat{\alpha}^\star}{\widehat{\alpha}^\star + E_i} + \frac{y_i}{E_i} \times \frac{E_i}{\widehat{\alpha}^\star + E_i} = \mathrm{E}[\mathrm{RR}_i] \times (1 - w_i^\star) + \mathrm{SMR}_i \times w_i^\star$$

where $w_i^\star = E_i/(\widehat{\alpha}^\star + E_i)$. One would be concerned if the two models gave significantly different estimates, revealing a general issue: there are multiple choices for the manner in which random effects may be incorporated, and the data will often be insufficiently numerous to decide between competing options.

A Poisson-lognormal non-spatial random effect model is given by:

$$Y_i|\beta, V_i \sim_{ind} \mathrm{Poisson}(E_i\mu_i \mathrm{e}^{V_i}), \quad V_i \sim_{iid} N(0, \sigma_v^2) \tag{8}$$

where V_i are area-specific random effects that capture the residual unexplained log relative risk in area i, $i = 1, ..., n$. The marginal distribution of this model is not available in closed form though the variance agrees with (5); the addition of spatial random effects is straightforward, however. Empirical Bayes is not so convenient for this model, and so we resort to a fully Bayesian approach, for which prior distributions are required.

2.4. Prior Choice for Non-Spatial Model

For a rare disease, a log-linear link is a natural choice: $\log \mu(\boldsymbol{x}_i, \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}$, where x_{ij} is the value of the j-th covariate in area i. For regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_J)$, an improper prior $p(\boldsymbol{\beta}) \propto 1$ may often be used, but such a choice may lead to an improper posterior (an example with a linear link is given in Section 3.8). If there are a large numbers of covariates, or there is high dependence amongst the elements of $\boldsymbol{x}$, then more informative priors will be beneficial. In this case it is convenient to specify lognormal priors for positive parameters $\exp(\beta_j)$, since one may specify two quantiles of the distribution, and directly solve for the two parameters of the lognormal. Denote by $\mathrm{LN}(\mu, \sigma)$ the lognormal distribution for a generic parameter θ with $\mathrm{E}[\log\theta] = \mu$ and $\mathrm{var}(\log\theta) = \sigma^2$, and let θ_1 and θ_2 be the q_1 and q_2 quantiles of this prior. Then it is straightforward to show that

$$\mu = \log(\theta_1)\left(\frac{z_{q_2}}{z_{q_2} - z_{q_1}}\right) - \log(\theta_2)\left(\frac{z_{q_1}}{z_{q_2} - z_{q_1}}\right), \ \sigma = \frac{\log(\theta_1) - \log(\theta_2)}{z_{q_1} - z_{q_2}}. \tag{9}$$

As an example, suppose that for the ecological relative risk e^{β_1} we believe there is a 50% chance that the relative risk is less than 1, and a 95% chance that it is less than 5; with $q_1 = 0.5, \theta_1 = 1.0$ and $q_2 = 0.95, \theta_2 = 5.0$, we obtain $\mu = 0$ and $\sigma = \log 5/1.645 = 0.98$.

It is not straightforward to specify a prior for σ_v, which represents the standard deviation of the log residual relative risks, a difficult parameter to interpret epidemiologically. The choice of a gamma distribution, $\mathrm{Ga}(a, b)$, for the precision $\tau_v = 1/\sigma_v^2$, is convenient since it produces a marginal distribution for the residual relative risks in closed form. Specifically the two-stage model

$$V_i|\sigma_v \sim_{iid} N(0, \sigma_v^2), \quad \tau_v = \sigma_v^{-2} \quad \sim \quad \mathrm{Ga}(a, b)$$

produces a marginal distribution for V_i which is $t_{2a}(0, b/a)$, a Student's t distribution with $2a$ degrees of freedom, location zero, and scale b/a; this is equivalent to the residual relative risks following a log t distribution. To determine a and b we specify the range $\exp(\pm R)$ within which the residual relative risks lie with probability q, and use the relationship $\pm t^{2a}_{q/2}\sqrt{b/a} = \pm R$, where t^{2a}_q is the q-th quantile of a Student t random variable with $2a$ degrees of freedom, to give $b = R^2 a/(t^{2a}_{q/2})^2$. For example, if we assume *a priori* that the residual relative risks follow a log Student t distribution with 2 degrees of freedom, with 95% of these risks falling in the interval (0.5,2.0), we obtain the prior, $\tau_v \sim \text{Ga}(1, 0.0260)$, an exponential distribution. In terms of σ_v this results in (2.5%, 97.5%) quantiles of (0.084,1.01) with posterior median 0.19.

It is important to assess whether the prior allows all reasonable levels of variability in the residual relative risks, in particular small values should not be excluded. As pointed out by Kelsall and Wakefield (1999) the prior Ga(0.001,0.001), which has previously been suggested, should be avoided for this reason; this choice corresponds to relative risks which follow a log t distribution with 0.002 degrees of freedom.

2.5. Non-Spatial Analysis of the Scottish Lip Cancer Data

Figure 2 shows relative risk estimates from a variety of models, with the SMRs on the left, referenced as position 0. At position 1 the empirical Bayes estimates obtained without the use of the covariate AFF are displayed. The weights on the SMR, (7), range between 0.45 and 0.99, with median 0.83. For these data the residual variability is large, from (4) the standard deviation of the random effects is $1/\sqrt{\alpha}$, and is estimated as 0.73, with 90% interval for residual relative risks (0.16,2.4).

In position 2 empirical Bayes estimates using a log-linear model in AFF, $\log \mu_i = \beta_0 + \beta_1 x_i$, are displayed. Four of the counties (4, 6, 14 and 32) have proportion in AFF equal to 0.24 (the highest value) and we see that the estimates for these counties are all moved upwards relative to the no covariate model (position 1) when the covariate is added to the model. The latter is worrying, and we see the reason in Figure 3; the log-linear model (dashed line) does not fit the data well for large values of AFF. This suggests that we use a more flexible model; exploratory work suggests the cubic form

$$\log \mu_i = \beta_0 + \beta_1(x_i - \overline{x}) + \beta_2(x_i - \overline{x})^2 + \beta_3(x_i - \overline{x})^3. \tag{10}$$

Figure 3 shows that this cubic model provides a better fit to the data (dotted line), and in particular flattens off for larger values of x. With the linear and cubic covariate models the standard deviation of the random effects are 0.58 and 0.53, respectively. We might expect the standard deviation to be reduced in size when we add an important covariate but this does not have to happen, for an explanation see Price et al. (1996). In position 3 of Figure 2 estimates under the cubic model are plotted, and we see that for counties 4, 6, 14 and 32 the estimates appear more reasonable. This illustrates the importance of deciding how much local smoothing is appropriate. A similar issue is relevant to the extent and nature of spatial smoothing.

We now report a fully Bayesian version of the normal model, (8), with log-linear cubic model (10) using Markov chain Monte Carlo (MCMC). The covariates are centered in (10) to reduce dependence in the posterior distribution, thereby reducing the dependence in the Markov chain. Flat priors were placed on $\beta_0, \beta_1, \beta_2, \beta_3$ and the previously-discussed gamma prior, $\text{Ga}(1, 0.0260)$, was assumed for σ_v^{-2}.

We see that the estimates under the empirical Bayes gamma and fully Bayesian lognormal model, at positions 3 and 4 respectively, each with cubic mean model, are very similar, illustrating that the most important aspect is not the inferential method or the choice of gamma or lognormal random effects, but a judicious choice of the covariate model. We define the random variables $\exp(\pm 1.96 \times \sigma_v)$ as the endpoints of a 95%

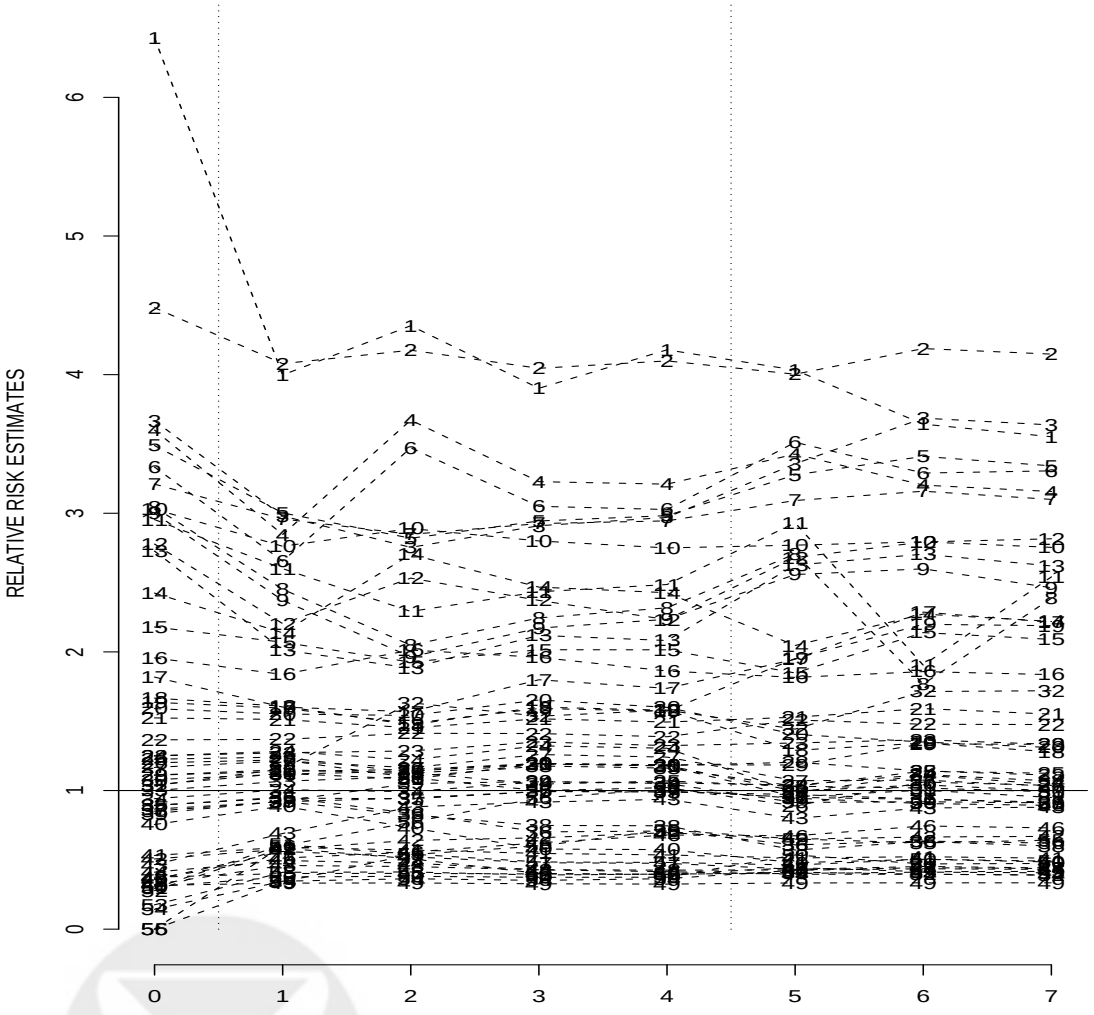


Fig. 2. Relative risk estimates for Scottish lip cancer data: 0 denote the SMRs; 1 the empirical Bayes non-spatial estimates without the use of AFF; 2 the empirical Bayes non-spatial estimates with a log-linear model in AFF; 3 the empirical Bayes non-spatial estimates with a log-linear cubic model in AFF; 4 the fully Bayes non-spatial estimates with a log-linear cubic model in AFF; 5 the joint model with a log-linear cubic model in AFF; 6 the initial ICAR model with a log-linear cubic model in AFF; 7 the refined ICAR model with a log-linear cubic model in AFF. Plotting symbol is county number.

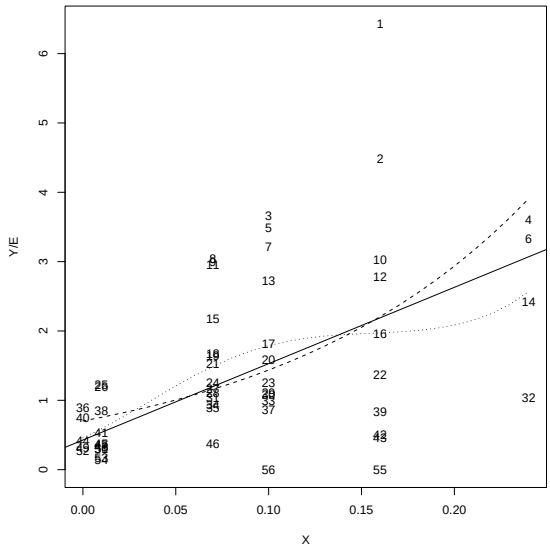


Fig. 3. Plot of Y/E versus proportion in AFF, x, with plotting symbol county number. Solid line corresponds to a model with identity link and linear in x; dashed line to a log link and linear in x; and dotted line to a log link and cubic in x.

interval for the residual relative risks. Posterior mean estimates of these endpoints are (0.35,2.96), showing that the posterior interval is considerably wider than the prior interval of (0.5,2.0). A 95% posterior interval for σ_v is (0.40,0.73) with median 0.55.

2.6. Spatial Models

In general we might expect residual relative risks in areas that are "close" to be more similar than in areas that are not "close", and we would like to exploit this information in order to provide more reliable relative risk estimates in each area. This is analogous to the use of a covariate x, in that areas with similar x values are likely to have similar relative risks. Unfortunately the modelling of spatial dependence is much more difficult since spatial location is acting as a surrogate for unobserved covariates; we need to choose an appropriate spatial model, but do not directly observe the covariates whose effect we are trying to mimic.

We first consider the model

$$Y_i|\boldsymbol{\beta},\boldsymbol{\gamma},U_i,V_i \sim_{ind} \text{Poisson}(E_i\mu_i \text{e}^{U_i+V_i}) \tag{11}$$

with

$$\log\mu_i = g(\boldsymbol{S}_i,\boldsymbol{\gamma}) + f(\boldsymbol{x}_i,\boldsymbol{\beta}), \tag{12}$$

where $\boldsymbol{S}_i = (S_{i1}, S_{i2})$ denotes spatial location, represented as the centroid of area i, and $g(\boldsymbol{S}_i,\boldsymbol{\gamma})$ is a regression model that we may include to capture large-scale spatial trend. The random effects $V_i|\sigma_v^2 \sim_{iid} N(0,\sigma_v^2)$ represent non-spatial contributions to the overdispersion, and U_i spatial contributions. We describe two forms for the latter.

We may assume that $\boldsymbol{U} = (U_1, ..., U_n)$ arise from a zero mean multivariate normal distribution with variances $\text{var}(U_i) = \sigma_u^2$ and correlations $\text{corr}(U_i, U_j) = \rho^{d_{ij}}$ where d_{ij} is the distance between the centroids of areas i and j, and ρ is a parameter that determines the extent of the correlation. This model is *isotropic* since it assumes that the correlation is the same in all spatial directions. We refer to this as the *joint* model, since we have specified the joint distribution for $\boldsymbol{U}$.

To define a *conditional* model we need to specify a rule for determining the "neighbours" of each area. A number of authors have taken areas i and j to be neighbours if they share a *common boundary*. This is reasonable if all regions are of similar size and arranged in a regular pattern (as is the case for pixels in image analysis where these models originated), but is not particularly attractive otherwise. Various other neighbourhood/weighting schemes are possible, for example Cressie and Chan (1989) assumed the neighbourhood structure was a function of the distance between area centroids. For area i we let ∂i denote the indices of the set of neighbours of area i. Besag et al. (1991) suggested a model that included a non-spatial and a spatial random effect and assigned the spatial random effects an intrinsic conditional autoregressive (ICAR) prior. Under this specification $U_i|U_j, j \in \partial_i \sim N\left(\overline{U}_i, \frac{\omega_u^2}{m_i}\right)$, where m_i is the number of neighbours of area i, and $\overline{U}_i$ is the mean of the spatial random effects of these neighbours. The parameter ω_u^2 is a conditional variance and its magnitude determines the amount of spatial variation. The variance parameters σ_v^2 and ω_u^2 are on different scales, σ_v is on the log relative risk scale while ω_u is on the log relative risk scale, *conditional* on $U_j, j \in \partial i$. Hence they are not comparable, in contrast to the joint model in which σ_u is on the same scale as σ_v. Notice that if ω_u^2 is "small" then although the residual is strongly dependent on the neighbouring value the overall contribution to the residual relative risk is small. This is a little counterintuitive but stems from spatial models having two aspects: the *extent* and *total amount* of spatial dependence, and in the ICAR model there is only a single parameter controlling both aspects. In the joint model the extent of spatial dependence is determined by ρ and the total amount by σ_u^2. A non-spatial random effect should always be included along with ICAR random effects since this model cannot take a limiting form that allows non-spatial variability; in the joint model with U_i only, this is achieved as $\rho \to 0$. If the majority of the variability is non-spatial, inference for the ICAR model might incorrectly suggest that spatial dependence was present. Leroux et al. (1999) showed via a simulation study that if the data were truly independent, a model with ICAR random effects and no non-spatial random effects produced a serious overestimation of ω_u^2, which led to very poor efficiency in the estimation of regression coefficients. In terms of implementation, both models require MCMC, but the conditional model needs far less computation than the joint model, for which $n \times n$ matrix inversions are typically necessary at each iteration.

Unfortunately, both the joint and conditional models suffer from a level of arbitrariness in their specification because the areas are not regular in shape or constant in size in a spatial epidemiological setting. For both models, normality of random effects can be replaced by other choices such as Laplacian and Student t distributions; see the discussion of Besag et al. (1991) and Best et al. (1999). The simple correlation structure described for the joint model can be extended to more complex forms, the Matérn class for example; see Matérn (1986), and the discussion of Diggle et al. (1998). For the Scottish data, in common with many applications, the data are spatially sparse, and little can be learnt about even a single parameter, hence we do not proceed to more complex forms here. Prior specification on the variances and spatial parameters also requires careful thought, as we discuss in the next section. Various other residual spatial models have been proposed, see for example, Cressie and Chan (1989), Besag et al. (1991), Clayton et al. (1993), Diggle et al. (1998), Leroux et al. (1999), Best et al. (2000), Mugglin et al. (2000), Knorr-Held and Raßer (2000), Kelsall and Wakefield (2002), Fernández and Green (2002), Christensen and Waagepetersen (2002) and Green and Richardson (2002). Richardson (2003) provides an excellent review of this literature. A number

of comparisons between spatial models have been carried out, see for example Lawson et al. (2000) and Best et al. (2005).

We have concentrated on Bayesian spatial models, but a number of frequentist approaches are possible, though have not been extensively investigated. Thurston et al. (2000) describe a negative binomial additive model, that potentially offers a useful alternative to the models described here; the negative binomial aspect would allow for overdispersion, while the generalised additive model allows flexible modelling of latitude and longitude to model non-small scale spatial variability. Recent work on generalised linear models with splines may be applicable also, see for example Lin and Zhang (1999) and Gu and Ma (2005). Allowing for small-scale residual spatial dependence in these models would be desirable, however.

2.7. Prior Choices for Spatial Models

Previously, priors have been specified for each of the variance components separately, but it is more practical to represent beliefs about the total variability. Proper priors are required for the parameters of the spatial model, see Berger et al. (2001) for a discussion in the context of the joint model.

For the joint model in which a multivariate normal distribution is assigned to $\boldsymbol{U}$, we have $V_i \sim_{iid} N(0, \sigma_v^2)$ and, independently, $U_i \sim_{iid} N(0, \sigma_u^2)$ so that the residual relative risk $\mathrm{e}^{V_i+U_i}$ is lognormal with parameters 0 and $\sigma_v^2 + \sigma_u^2$. If we specify inverse gamma priors for σ_v^2 and σ_u^2, the implied prior for $\sigma_v^2 + \sigma_u^2$ is not inverse gamma so that we cannot easily control the total residual relative risk. We write the total precision as $\tau_T = (\sigma_v^2 + \sigma_u^2)^{-1}$, and as in Section 2.4 specify $\tau_T \sim \mathrm{Ga}(a, b)$ so that marginally we have a log Student's t distribution for the total residual relative risks. We let $p = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$ represent the proportion of the total residual variation that is attributable to the spatial component, and assign a beta prior, $\mathrm{Be}(c, d)$, to p, and transform from (σ_T^2, p) to (σ_v^2, σ_u^2) via

$$\begin{aligned} \sigma_v^2 &= (1-p)\tau_T^{-1} = (1-p)(\sigma_v^2 + \sigma_u^2) \\ \sigma_u^2 &= p\tau_T^{-1} = p(\sigma_v^2 + \sigma_u^2). \end{aligned}$$

This prior allows us to control the amount of total residual variability, a quantity for which prior knowledge is available, and induces positive dependence in the joint prior for (σ_v^2, σ_u^2).

Rather than consider the parameter ρ, we specify a lognormal prior, using equations (9), for the distance at which the correlations fall to a half, $d_{1/2} = \log 2/\log \rho$. For example, if we believe there is a 5% chance that the correlation falls to a half in less than 4km, and a 95% chance that it falls to a half in less than 125km we obtain $d_{1/2} \sim \mathrm{LN}(3.11, 1.05^2)$.

Given its conditional interpretation, it is not straightforward to specify a prior for the ICAR parameter ω_u^2. Specifying an ICAR model for the spatial effects does not define a proper n-dimensional joint distribution, and none of the marginal distributions for U_i exist. Rather

$$p(\boldsymbol{U}|\omega_u^2) \propto (\omega_u^2)^{-(n-1)/2} \exp\left[-\frac{1}{2}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{U}\right] = (\omega_u^2)^{-(n-1)/2} \exp\left[-\frac{1}{2\omega_u^2}\sum_{i<j}(U_i - U_j)^2\right], \qquad (13)$$

where $\boldsymbol{Q}$ is the $n \times n$ matrix with, for $i \neq j$, $Q_{ij} = -1/\omega_u^2$, if areas i and j are neighbours and $Q_{ij} = 0$ otherwise, and $Q_{ii} = m_i/\omega_u^2$.

For prior specification we follow an approximate strategy and consider the $n-1$, random variables $\boldsymbol{Z} = (Z_1, ..., Z_{n-1})$ where $Z_i = U_i - U_n$, $i = 1, ..., n-1$. Hence $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{U}$, where $\boldsymbol{A} = [\boldsymbol{I}|\text{-}\mathbf{1}]$, $\boldsymbol{I}$ is the $(n-1)\times(n-1)$

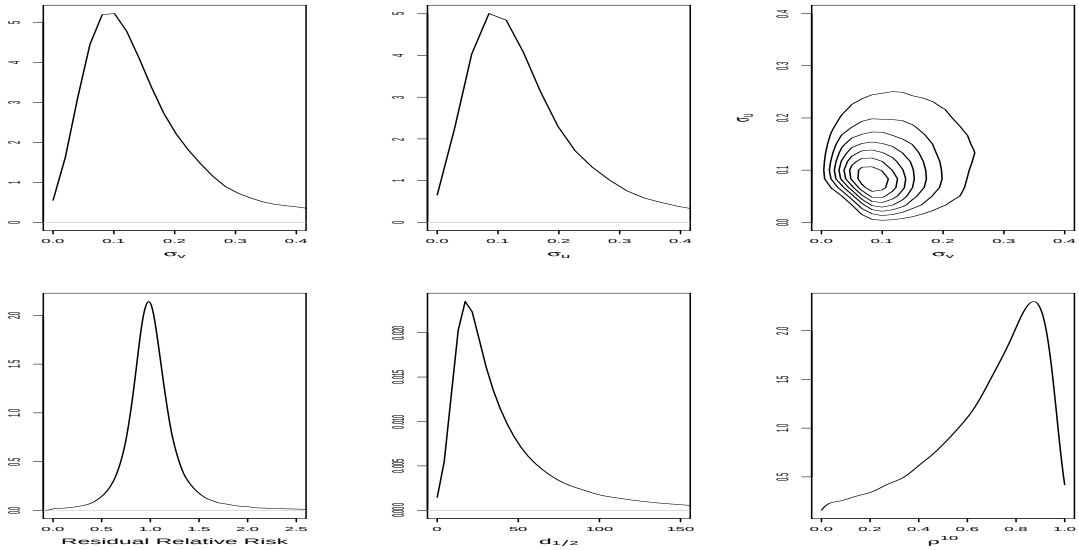


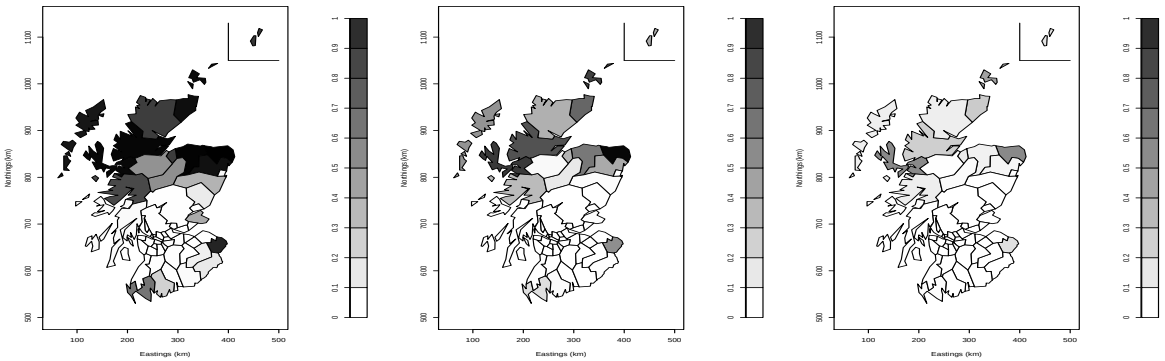
Fig. 4. Priors for the joint spatial model. First row: univariate and joint marginals for σ_v and σ_u. Second row: the residual relative risk $\exp(V_i + U_i)$ margin, the distance at which correlations fall to a half, $d_{1/2}$, and the correlation between areas whose centroids are 10km apart, ρ^{10}.

identity matrix, and **-1** is an $(n-1)\times 1$ vector of -1's. The joint distribution of $\boldsymbol{Z}$ exists, and is an $(n-1)$-dimensional normal distribution with mean zero and variance-covariance matrix $\overline{\boldsymbol{A}}\boldsymbol{Q}^{-1}\overline{\boldsymbol{A}}^{\mathrm{T}}$ with $\overline{\boldsymbol{A}} = [\boldsymbol{I}|\boldsymbol{0}]^{\mathrm{T}}$ a generalized inverse of $\boldsymbol{A}$, and $\boldsymbol{0}$ the $(n-1)\times 1$ vector of 0's; Besag and Kooperberg (1995) give further details, see Lemma 3.1 and Corollary 3.1. The marginal variance for Z_i is $\mathrm{var}(Z_i) = a_i\omega_u^2$, where the constants a_i are determined by the neighbourhood structure, and are the diagonal elements of $\overline{\boldsymbol{A}}\boldsymbol{Q}^{-1}\overline{\boldsymbol{A}}^{\mathrm{T}}$. We let $\overline{\sigma}_z^2 = \overline{a}\omega_u^2$ represent the average marginal variance, and specify a prior for $\overline{\sigma}_z^2$, which induces a prior for ω_u^2. Once the calibration between ω_u^2 and $\overline{\sigma}_z^2$ has been carried out we specify priors for $\tau_T = (\sigma_v^2 + \overline{\sigma}_z^2)^{-1}$ and p, as described for the joint model, and then take $\sigma_v^2 = (1-p)\tau_T^{-1}$ and $\omega_u^2 = p\tau_T^{-1}/\overline{a}$. This procedure is approximate in a number of ways, we have considered $\boldsymbol{Z}$ rather than $\boldsymbol{U}$, and U_i is not marginally normally distributed, but we have found it more useful than previously available prescriptions; see for example, Bernardinelli et al. (1995).

2.8. Spatial Models for the Scottish Lip Cancer Data

We assign improper flat priors to each element of $\boldsymbol{\beta}$, and for the joint spatial model assume that $\tau_T \sim \mathrm{Ga}(1, 0.0260)$, $p \sim \mathrm{Be}(1,1)$ and $d_{1/2} \sim \mathrm{LN}(3.11, 1.05^2)$. Figure 4 shows smoothed marginal densities based on samples from these priors, including induced quantities of interest such as the residual relative risk, $\exp(U_i + V_i)$, and ρ^{10}, the correlation at a distance of 10km. The induced dependence between σ_v and σ_u is apparent.

For the ICAR model the same priors were assumed and we set $\omega_u^2 = p\tau_T^{-1}/\overline{a}$ where $\overline{a} = 1.164$ for the Scottish geography with a common boundary neighbourhood scheme. This neighbourhood scheme is not particularly appealing for Scotland because of the irregularity of the areas. Following other authors (e.g. Thomas et al. 2000) we initially assume that the three islands, which have no common boundary neighbours, only have a non-spatial random effect.



(a) Threshold = 2 (b) Threshold = 3 (c) Threshold = 4

Fig. 5. Posterior probabilities of exceedance of different thresholds under the joint model.

Positions 5–7 of Figure 2 show estimates from spatial models, each with the cubic model in AFF. In the non-spatial model we have shrinkage to the overall regression model but for the spatial model we have, in addition, local smoothing so that estimates can move away from the regression model.

A striking feature of Figure 2 is the differences in the estimates for areas 8 and 11 under the joint spatial (position 5) and the ICAR (position 6) models. The explanation is that for the three islands without neighbours under the ICAR formulation, there are only non-spatial contributions to the relative risk. Table 1 reports posterior summaries for the parameters of the random effects distributions, and shows that the majority of the total variability is spatial for these data. Hence we see large shrinkage for the three islands since we are assuming a *common* non-spatial model across islands and non-islands, resulting in too much shrinkage for the islands. There are a number of possibilities for refining the model. One is to assume $V_i \sim_{iid} N(0, \tau_T^{-1})$ for the islands so that we have the same total variability as non-islands, but with all of this variability assumed to be non-spatial. Given our parameterization of the prior it is straightforward to fit this model. The resultant estimates are shown in position 7, and differ little from those in position 6, which is reassuring. This model may be useful in general circumstances in which there are areas which have no neighbours due to physical boundaries. Further possibilities include defining neighbours for the islands as the nearest points of the mainland (or the nearest island), or assuming a distinct non-spatial distribution for the islands, with only three islands this option is not feasible here, however.

Figure 1(b) shows relative risk estimates under the joint model; the smoothness compared to the SMRs in Figure 1(a) is apparent. Under a Bayesian sampling-based approach it is straightforward to carry out inference for functions of interest. As an illustration, Figure 5(a)–(c) show the posterior probabilities that the relative risk in each area exceeds the values 2, 3 and 4. We see a number of areas with high probabilities, suggesting that, in a serious investigation, these be examined more closely to discover the characteristics of the individuals, or health hazards that are present, in these areas to explain these excesses. Such plots are also useful for reflecting the uncertainties inherent in smoothed maps.

In Table 1 we examine the sensitivity of estimates of the non-spatial and spatial contributions of residual relative risk, by comparing the original gamma prior for τ_T to the alternative Ga$(1, 0.1399)$. This prior gives relative risks that follow a log student t distribution with 2 degrees of freedom, and fall within the range

Table 1. Sensitivity of spatial model parameters to prior choice for $\tau_T = (\sigma_u^2 + \sigma_v^2)^{-1}$, p is the proportion of the total variability that is spatial.

Spatial Model	Prior Specification	Posterior median			
		σ_v	σ_u	p	$d_{1/2}$ (km)
Joint	$\tau_T \sim \text{Ga}(1, 0.0260)$	0.23	0.48	0.82	79
Joint	$\tau_T \sim \text{Ga}(1, 0.1399)$	0.24	0.49	0.82	80
ICAR	$\tau_T \sim \text{Ga}(1, 0.0260)$	0.23	0.53	0.85	–
ICAR	$\tau_T \sim \text{Ga}(1, 0.1399)$	0.22	0.54	0.86	–

(0.2,5) with probability 0.95. We see that across these prior scenarios the majority of the residual variability, 82%–86%, is explained by the spatial component. Overall there is little sensitivity of the parameters in Table 1 to the priors considered, though the joint and ICAR models can give quite different estimates in particular areas. Interval estimates for $d_{1/2}$ are very wide, reflecting the lack of information on the strength of the residual spatial variability. For example, for the prior choice in row 1 of Table 1, a 95% interval for $d_{1/2}$ is (32km,243km).

2.9. Conclusions

The preferred model here would be that which includes a cubic model in AFF and a spatial component, since the association with AFF is strong and there is significant residual spatial dependence. A full analysis would examine the sensitivity of the relative risk estimates to the prior specifications. There is a large amount of residual variability for these data, which is not surprising since we have no information on risk factors such as smoking, alcohol and diet. Although it is important to consider models that include residual spatial dependence for small-area studies, empirical Bayes non-spatial models are very useful for exploratory purposes, particularly for choosing an appropriate mean model. Estimates from these models, along with the SMRs, provide baseline estimates which may be compared with those from spatial models.

3. Spatial Regression

Spatial regression differs from disease mapping in that the aim is to estimate the association between risk and covariates, rather than to provide area-specific relative risk estimates. This is a crucial distinction which has important implications for modelling both the mean function and the residual variability.

3.1. Drawbacks of Approaches Under Independence

In the usual implementation of regression models the standard errors are calculated under the assumption that the response data are independent, after control for known covariates. In a spatial context, and particularly when the areas are small, one would expect "residual" dependency between counts in areas that are geographically close, due to unmeasured risk factors, or errors in the data, that have spatial structure.

3.2. Non-Spatial Models

We begin by fitting models with no spatial dependency, in order to see the subsequent effect of including such dependency. A naive starting point is the Poisson regression model $Y_i|\boldsymbol{\beta} \sim \text{Poisson}(E_i\mu_i)$, where $\mu_i = \mu(\boldsymbol{x}_i, \boldsymbol{\beta})$, for $i = 1, ..., n$. As discussed in Section 2.3, this model will almost always be inappropriate for spatial count data, since the Poisson model does not have a variance parameter. An easy way of extending this model is to use quasi-likelihood (McCullagh and Nelder, 1989) and specify the first two moments:

$$\text{E}[Y_i|\boldsymbol{\beta}] = E_i \exp(\beta_0 + \beta_1 x_i), \quad \text{var}(Y_i|\boldsymbol{\beta}) = \kappa \times \text{E}[Y_i|\boldsymbol{\beta}], \tag{14}$$

κ represents the level of non-Poisson variability and fitting is straightforward with identical point estimates to the Poisson model, and standard errors multiplied by $\kappa^{1/2}$. This model does not assume a distribution for the data, and so is not helpful in the context of disease mapping where prediction is required, though Kriging approaches for non-Gaussian data may be useful. The parametric disease mapping models described earlier, in particular the Poisson models with gamma, (4), and lognormal, (8), random effects models may also be used in a regression setting.

3.3. Non-Spatial Regression Models for the Scottish Lip Cancer Data

The exposure variable AFF only takes one of six values since it was read from a map key, and hence AFF_i is measured with error and, in theory, an errors-in-variables model could be built, based on the widths of the cut-points. In common with the other authors who have examined these data, including Clayton and Kaldor (1987), Clayton and Bernardinelli (1992), Breslow and Clayton (1993), Yasui and Lele (1997), Leroux et al. (1999), Conlon and Louis (1999), Leroux (2000), Lee and Nelder (2001), Banerjee et al. (2004) and Waller and Gotway (2004)) we begin by assuming the log-linear model $\log \mu(x_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_i$. We provide a discussion of the appropriateness of this model in Section 3.7, where we also provide a more careful interpretation of the coefficients of the model; the ecological interpretation is that $\exp(\beta_1)$ is the multiplicative change in risk between an area with all individuals in AFF, and another area with no individuals in AFF.

Table 3 gives estimates from a number of regression models, in all cases $d_{1/2} \sim \text{LN}(3.11, 1.05^2)$. The simple Poisson regression model gives an estimate of 7.4 so that the area-level multiplicative difference in risk between areas with proportion in AFF 0.24 and 0 (the range of the observed data) is $\exp(7.4 \times 0.24) = 5.9$. Model (14) gives $\widehat{\kappa} = 4.9$, a considerable amount of overdispersion, giving a more than doubling in the standard error. The negative binomial parametric version of this model with first two moments (5) gives a slight reduction in the size of the coefficient and a very similar standard error.

Plots of AFF versus Eastings and Northings were examined, and there is spatial structure in the exposure, Figure 6, with a skewed U-shaped trend in the south-north direction, with low values in the heavily-populated urban areas in the Glasgow region, and a general increase with northings. Leroux et al. (1999) and Leroux (2000) include northings (latitude) in the log-linear model. When we include the Northing projection centroids in the model the AFF coefficient is reduced, because of the south-north trend in exposure. The inclusion of eastings gave a far smaller change since there is little east-west gradient in the exposure surface. The decision of whether to include a large-scale trend in the risk surface is a difficult one in a regression setting since exposures of interest will often have spatial structure and so estimates of relative risks of interest will in general change. Non-removal of the trend will attribute any such trend to the exposure estimate, and may also invalidate the assumption of stationarity of spatial models such as the joint specification. Ideally the choice will be based on epidemiological grounds; if it believed that the trend is due to plausible unmea-

Table 2. Estimates and standard errors for ecological regression coefficient, β_1, in log-linear model in AFF.

Model	Further specifications	Estimate	St. Err.
Poisson	–	7.4	0.60
Quasi-likelihood	–	7.4	1.3
Quasi-likelihood	Northings	5.5	1.2
Quasi-likelihood	Eastings and northings	5.6	1.2
Negative binomial	–	7.2	1.3
Poisson lognormal	$\sigma_v^{-2} \sim \text{Ga}(1, 0.0260)$	6.8	1.5
Poisson lognormal	$\sigma_v^{-2} \sim \text{Ga}(1, 0.0260)$, $\beta_1 \sim N(0, 4.21)$	6.1	1.4
Poisson Joint	$\tau_T \sim \text{Ga}(1, 0.0260)$	3.4	1.3
Poisson Joint	$\tau_T \sim \text{Ga}(1, 0.1399)$	3.4	1.3
Poisson Joint	$\tau_T \sim \text{Ga}(1, 0.0260)$	3.5	1.3
Poisson ICAR	$\tau_T \sim \text{Ga}(1, 0.0260)$	4.9	1.3
Poisson ICAR	$\tau_T \sim \text{Ga}(1, 0.1399)$	4.9	1.4

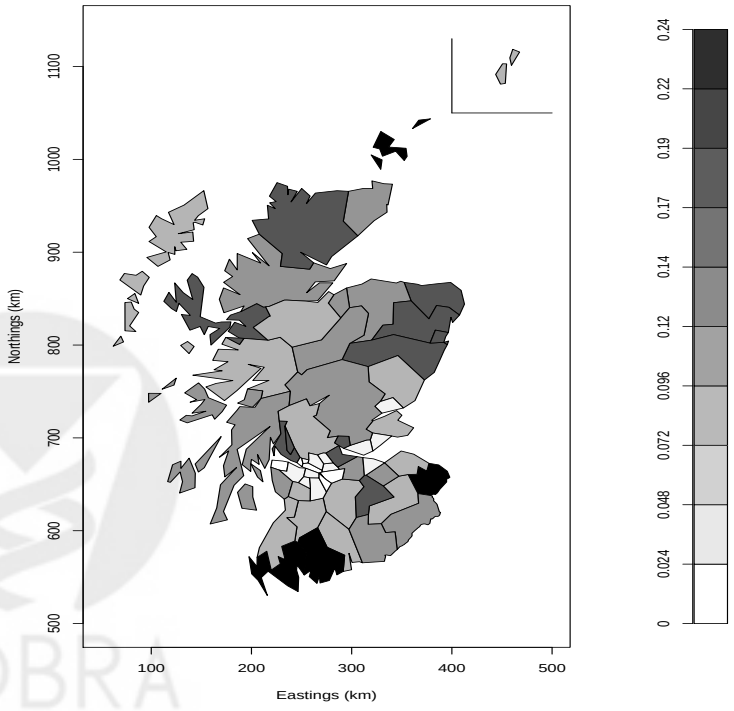


Fig. 6. Map of the exposure, percentage individuals employed in agriculture, fishing and farming (AFF).

sured factors such as socio-economic status then the trend should be included, though it is clearly preferable to have a direct measure of the variable responsible for the trend.

We now turn to the Poisson-lognormal model, (8). We choose improper flat priors on β_0 and β_1, and a Ga(1,0.0260) prior on σ_v^{-2}. This yields a posterior mean and standard deviation of 6.8 and 1.5, showing reasonable agreement with the comparable quasi-likelihood/negative binomial models. To illustrate the use of a proper prior on β_1, suppose we believe that the 50% and 95% points of the relative risk between areas with 10% and 0% in AFF are 1 and 2. This yields a normal prior for β_1 of N(0,4.21), and a posterior mean and standard deviation of 6.1 and 1.4, illustrating the effect of the prior in reducing both the estimate of the association, and the standard error.

3.4. Spatial Models

Unfortunately there are currently no simple ways of fitting frequentist fixed effects, non-linear models with spatially dependent residuals, and so we concentrate on random effects models, and the form given in (11) and (12), with no large-scale trend. It would be desirable to perform sandwich estimation in a spatial regression setting, but unfortunately the non-lattice nature of the data does not easily allow any concept of replication across space, as was used by Heagerty and Lumley (2000) in the case of lattice data.

3.5. Spatial Regression Models for the Scottish Lip Cancer Data

We specify identical priors for σ_v^2, σ_u^2, $d_{1/2}$ and ω_u^2 as in the disease mapping analyses. Improper flat priors were placed on β_0 and β_1.

Table 3 shows that the estimated relative risks are greatly attenuated under each of the spatial models. The explanation is spatial dependence in AFF, and in the disease counts. The choice of the level of spatial smoothing is a difficult one without knowing the true spatial dependence model. Here, the use of an informative prior distribution based on another study region might be useful. It is interesting to see that the standard error is reduced when spatial dependence is acknowledged, perhaps in conflict with intuition, Wakefield (2003) provides more discussion of this issue. Under the first joint model in Table 3 the posterior 5%, 50%, 95% quantiles of the distance at which correlations fall to a half are (35km,72km,199km); the prior 5%, 50%, 95% quantiles are (4km, 22km, 125km), again illustrating that there is very little information in the data on the range of the correlation, though the posterior is shifted to the right, relative to the prior. Spatial variation accounts for 77% of the total, with posterior means for σ_v and σ_u of 0.26 and 0.52.

3.6. Ecological Bias

There is a vast literature describing sources of ecological bias, see for example, Richardson et al. (1987), Piantadosi et al. (1988), Greenland and Morgenstern (1989), Greenland (1992), Greenland and Robins (1994), Morgenstern (1998) and Wakefield (2003). Ecological bias occurs because of within-area variability in exposures and confounders, and there are a number of distinct consequences of this variability; we discuss pure specification bias, confounding and standardization.

3.6.1. Pure Specification Bias

So-called pure specification bias arises because a nonlinear risk model changes its form under aggregation. To illustrate, we specify a model at the level of the individual and then aggregate to determine the implied ecological form. Let Y_{ij} denote the individual binary disease outcome with $Y_{ij} = 0/1$ representing non-case/case, and x_{ij} the exposure of individual j in area i, $i = 1, ..., n$, $j = 1, ..., N_i$. For simplicity we assume a univariate exposure and no confounders and let $p(\boldsymbol{\alpha}, x)$ denote the risk for an individual with exposure x as a function of parameters $\boldsymbol{\alpha}$. The individual outcome, Y_{ij}, is Bernoulli with probability of disease p_{ij}, written as $Y_{ij}|x_{ij} \sim_{ind} \text{Bern}(p_{ij})$. The implied aggregate (average) risk is

$$p_i = \frac{1}{N_i} \sum_{j=1}^{N_i} p_{ij} \tag{15}$$

where $p_{ij} = p(\boldsymbol{\alpha}, x_{ij})$; (15) clearly shows that to avoid ecological pure specification bias we require the average of the individual risks, rather than the risk associated with the average exposure, which would be used in a naive model. If N_i is large, then an alternative derivation is to assume that exposures x_{ij} are drawn independently from a distribution $f_i(x|\boldsymbol{\phi}_i)$ where $\boldsymbol{\phi}_i$ denote the parameters of this distribution. It then follows that, marginally, $Y_{ij}|\boldsymbol{\phi}_i$ is Bernoulli with probability of disease

$$p_i = \int p(\boldsymbol{\alpha}, x) f_i(x|\boldsymbol{\phi}_i)\, dx, \tag{16}$$

for a continuous exposure, and

$$p_i = \sum_{k=1}^{K} p(\boldsymbol{\alpha}, x_k) f_i(x_k|\boldsymbol{\phi}_i) \tag{17}$$

for a K-level discrete exposure. In a disease mapping context Knorr-Held and Besag (1998, p.2050) considered a discrete exposure with $f_i(x|\boldsymbol{\phi}_i)$ the distribution of a generalised Bernoulli random variable with $\boldsymbol{\phi}_i = (\upsilon_{i1}, ..., \upsilon_{iK})$ and υ_{ik} representing the probability of falling in exposure category k in area i. In this case we obtain $p_i = \sum_{k=1}^{K} p_{ik}\upsilon_{ik}$ as the risk associated with an individual randomly selected from area i and with no knowledge of their exposure. In an ecological regression context Richardson et al. (1987) and Plummer and Clayton (1996) considered (16) with $f_i(x|\boldsymbol{\phi}_i)$ a normal distribution with $\boldsymbol{\phi}_i = \{\overline{x}_i, s_i^2\}$. As an illustration of the problems that arise due to pure specification bias we present a simple example using the normal model. For a rare outcome, a common disease model is $p_{ij} = p(\boldsymbol{\alpha}, x_{ij}) = \text{e}^{\alpha_0 + \alpha_1 x_{ij}}$ and (16) assumes the closed-form

$$p_i = \exp(\alpha_0 + \alpha_1 \overline{x}_i + \alpha_1^2 s_i^2/2) \tag{18}$$

which may be compared with the naive ecological model $\text{e}^{\beta_0 + \beta_1 \overline{x}_i}$, where we have used β_0, β_1 in the ecological model to emphasise that in this model we are not estimating the individual-level parameters α_0, α_1. To gain intuition as to the extent of the bias we observe that in (18) the within-area variance s_i^2 is acting like a confounder, and there is no pure specification bias if the exposure is constant within each area, or if the variance is independent of the mean exposure in the area. The expression (18) also allows us to characterise the direction of bias. For example, if $\alpha_1 > 0$ and the within-area variance increases with the mean, then overestimation will occur. In general, there is no pure specification bias if the disease model is linear in x, or if all the moments of the within-area distribution of exposure are independent of the mean. Bias will also be small if α_1 is close to zero, since the risk model is then approximately linear.

3.6.2. Confounding

Between-area confounding is analogous to conventional confounding, since the area is the unit of analysis, and so the implications are relatively straightforward to understand. Within-area confounding is more complex. In an ecological study we need to control for the complete within-area *distribution* of exposures and confounders. We illustrate in the simplest situation in which we have a binary exposure, x_1, a binary confounder, x_2, and assume the individual-level risk model: $p(\boldsymbol{\alpha}, \boldsymbol{x}) = \exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2)$. Then Lasserre et al. (2000) show that the aggregate form is

$$p_i = x_{i00}\mathrm{e}^{\alpha_0} + x_{i10}\mathrm{e}^{\alpha_0+\alpha_1} + x_{i01}\mathrm{e}^{\alpha_0+\alpha_2} + x_{i11}\mathrm{e}^{\alpha_0+\alpha_1+\alpha_2+\alpha_3}, \tag{19}$$

where $x_{ix_1x_2}$ is the proportion of individuals in area i in exposure/confounder stratum x_1, x_2. Letting x_{i1+} and x_{i+1} represent the proportion of individuals in area i who have $x_1 = 1$ and $x_2 = 1$ respectively (the marginal prevalences) we may rewrite the average risk (19) as

$$p_i = (1 - x_{i1+} - x_{i+1} + x_{i11})\mathrm{e}^{\alpha_0} + (x_{i1+} - x_{i11})\mathrm{e}^{\alpha_0+\alpha_1} + (x_{i+1} - x_{i11})\mathrm{e}^{\alpha_0+\alpha_2} + x_{i11}\mathrm{e}^{\alpha_0+\alpha_1+\alpha_2+\alpha_3},$$

showing that the marginal prevalances alone, which ecological data will often consist of, are not sufficient to characterise the joint distribution, unless x_1 and x_2 are independent, in which case x_2 is not a within-area confounder.

3.6.3. Standardization

If the response is standardized with respect to confounders, then the exposure must be also; this is known as mutual standardization; Rosenbaum and Rubin (1984) provide a discussion in the context of direct standardization. To illustrate the problem in the context of ecological studies consider a continuous exposure x and suppose the individual level model is given by $p_k(\boldsymbol{\alpha}, \boldsymbol{\gamma}, x_{ik}) = \exp(\alpha_0 + \alpha_1 x_{ik} + \gamma_k)$, for $k = 1, ..., K$ stratum levels with associated relative risks e^{γ_k}, with $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_K)$ and $\gamma_1 = 0$; we will refer to the confounder as age. For an individual randomly selected in stratum k the aggregate risk is

$$p_{ik} = \exp(\alpha_0 + \gamma_k) \int \exp(\alpha_1 x) f_{ik}(x|\boldsymbol{\phi}_{ik})\ dx,$$

where $f_{ik}(x)$ is the distribution of the exposure in area i and confounder stratum k, with parameters $\boldsymbol{\phi}_{ik}$. If the population in area i stratum k is N_{ik}, with Y_{ik} cases, then summing over stratum:

$$\mathrm{E}[Y_i|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}_i] = \sum_k \left\{ N_{ik}\mathrm{e}^{\alpha_0+\gamma_k} \int \mathrm{e}^{\alpha_1 x} f_{ik}(x|\boldsymbol{\phi}_{ik})\ dx \right\}. \tag{20}$$

If we assume a common exposure distribution across stratum, $f_i(x|\boldsymbol{\phi}_i)$, so that age does not correspond to a within-area confounder, then (20) simplifies to

$$\mathrm{E}[Y_i|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}_i] = E_i\mathrm{e}^{\alpha_0} \int \mathrm{e}^{\alpha_1 x} f_i(x|\boldsymbol{\phi}_i)\ dx,$$

where $E_i = \sum_{k=1}^K N_{ik}\mathrm{e}^{\gamma_k}$. This mean is often used in conjuction with a Poisson model; we have standardized for the confounder, via indirect standardization, but for this to be valid we need to assume that the exposure is constant across age groups. The correct mean model is given by (20), and requires the exposure distribution by age, which will rarely be available.

The above discussion makes it clear that to prevent ecological bias we need individual-level data to control for the within-area distribution of confounders and exposures. Prentice and Sheppard (1995) describe a very powerful method for reducing ecological bias based on subsamples of individual exposure-confounder data, but not individual disease outcomes, within each area. Haneuse and Wakefield (2005,2006) describe a hybrid design in which case-control and ecological data are combined.

3.7. Ecological Bias in the Scottish Lip Cancer Data

We build a model from the level of the individual in order to aid interpretation of an ecological association. Let Y_{ij} be an indicator of lip cancer with $Y_{ij}|p_{ij} \sim \text{Bern}(p_{ij})$, and let the risk for individual j in area i, who is in age group $k_{ij} \in \{1, ..., K\}$ be given by

$$p_{ij} = \exp(\alpha_0 + \alpha_1 x_{ij} + \gamma_{k_{ij}}) = \{(1 - x_{ij})\mathrm{e}^{\alpha_0} + x_{ij}\mathrm{e}^{\alpha_0+\alpha_1}\}\mathrm{e}^{\gamma_{k_{ij}}}, \tag{21}$$

with x_{ij} an indicator of whether individual j in area i works in AFF; the second form illustrates that we have a linear model in the exposure x_{ij}. The parameters $\exp(\gamma_k)$ are the risks associated with being in age stratum k, $k = 1, ..., K$, while $\exp(\alpha_1)$ is the parameter of primary interest and is the multiplicative change in risk for an exposed individual when compared to an unexposed individual, which is assumed the same across all age groups so that there is no interaction between exposure and age.

We now consider the effect of aggregation. Suppose the number of individuals in stratum k who are unexposed (exposed) in area i is N_{i0k} (N_{i1k}), and Y_{i0k} (Y_{i1k}) of these are cases. For a rare disease: $Y_{ixk}|\boldsymbol{\alpha}, \boldsymbol{\gamma} \sim_{ind} \text{Poisson}(N_{ixk}\mathrm{e}^{\alpha_0+\alpha_1 x+\gamma_k})$, and so $Y_{i0k}+Y_{i1k}|\boldsymbol{\alpha}, \boldsymbol{\gamma} \sim_{ind} \text{Poisson}(N_{i0k}\mathrm{e}^{\alpha_0+\gamma_k}+N_{i1k}\mathrm{e}^{\alpha_0+\alpha_1+\gamma_k})$. We now assume that the proportions exposed, x_i, and unexposed, $1 - x_i$, in area i are constant across age groups, that is $N_{i0k} = (1 - x_i)N_{ik}$, $N_{i1k} = x_i N_{ik}$, where N_{ik} is the population in confounder stratum k in area i. Summing over k: $Y_i|\boldsymbol{\alpha}, \boldsymbol{\gamma} \sim_{ind} \text{Poisson}\left([1 - x_i]\mathrm{e}^{\alpha_0} \sum_{k=1}^{K} N_{ik}\mathrm{e}^{\gamma_k} + x_i\mathrm{e}^{\alpha_0+\alpha_1} \sum_{k=1}^{K} N_{ik}\mathrm{e}^{\gamma_k}\right)$ to give

$$Y_i|\boldsymbol{\alpha}, \boldsymbol{\gamma} \sim_{ind} \text{Poisson}\left(E_i(\boldsymbol{\gamma})\{[1 - x_i]\mathrm{e}^{\alpha_0} + x_i\mathrm{e}^{\alpha_0+\alpha_1}\}\right) \tag{22}$$

where $E_i(\boldsymbol{\gamma}) = \sum_{k=1}^{K} N_{ik}\mathrm{e}^{\gamma_k}$.

The addition of random effects to this model is straightforward; if we begin with the individual-level model $p_{ij} = \exp(\alpha_0 + \alpha_1 Z_{ij} + \gamma_{k_{ij}} + V_i + U_i)$, we obtain

$$Y_i|\alpha_0, \alpha_1, \gamma, V_i, U_i \sim_{ind} \text{Poisson}(E_i(\boldsymbol{\gamma})\{[1 - x_i]\mathrm{e}^{\alpha_0} + x_i\mathrm{e}^{\alpha_0+\alpha_1}\} \exp\{V_i + U_i\}). \tag{23}$$

The above model is an example of inference over a collapsed margin. Byers and Besag (2000) considered such a situation with a Poisson model, but were apparently unaware that the likelihood is available in closed form, and instead introduced auxiliary variables within an MCMC scheme.

From an individual-level perspective we see there are a number of problems with the analyses reported in Section 3.5, and those carried out previously. The most obvious is that model (22) is linear, and not log-linear, in form, because the original model, (21), is linear and a linear model is preserved under aggregation. The correct interpretation of the parameters in the log-linear model that was fitted in Section 3.5 is that $\exp(\beta_1)$ is the relative risk associated with the *contextual* effect of the proportion of individuals who are exposed to AFF in each area; i.e. it is not individual occupation in AFF that is relevant to individual risk, but the proportion of individuals in the area who are employed in AFF, a model that does not make substantive sense. The case of a binary covariate has been extensively studies in the social sciences literature, see Wakefield (2004) for further details.

Table 3. Estimates and standard errors for individual relative risk, $\exp(\alpha_1)$, in linear model.

Model	Relative risk	St. Err.
Quasi-likelihood	23	7.0
Poisson lognormal	21	7.5
Poisson Joint	6.4	3.3
Poisson ICAR	9.9	3.3

In order for model (22) to be relevant we also need to assume that the proportion in AFF is constant across age-groups, which is unlikely to hold, but unfortunately we do not have data on the proportion in AFF by age group. Finally, if the expected numbers are age-standardized using *a priori* internal standardization, rather than external standardization in which reference rates are based on data from another region, then the subsequent estimation of the AFF association will be distorted if age is a confounder, since some of the effect of AFF will already have been absorbed into the age effects. The disease and population counts are required by county and age stratum in order to fit a model in which age and AFF are simultaneously estimated.

3.8. Individual-Level Models for the Scottish Lip Cancer Data

Table 3 contains estimated relative risks under a number of different models. A quasi-likelihood version of (22) is given by

$$\mathrm{E}[Y_i|\alpha_0,\alpha_1] = x_{i1}\mathrm{e}^{\alpha_0} + x_{i2}\mathrm{e}^{\alpha_0+\alpha_1}, \quad \mathrm{var}(Y_i|\alpha_0,\alpha_1) = \kappa \times \mathrm{E}[Y_i|\alpha_0,\alpha_1],$$

where $x_{i1} = E_i(1-x_i)$ and $x_{i2} = E_i x_i$, a linear model. Fitting this model gave $\widehat{\kappa} = 4.15$ with an estimate of the relative risk of 23 with standard error 7.0. The fit corresponding to this model is shown as the solid line in Figure 3, and is certainly better than the log-linear model.

Turning to a Bayesian approach, perhaps surprisingly, the use of an improper flat prior for α_1 leads to an improper posterior. Consider the model $Y_i|\boldsymbol{\alpha} \sim_{ind} \mathrm{Poisson}(E_i\{[1-x_i]\mathrm{e}^{\alpha_0}+x_i\mathrm{e}^{\alpha_0+\alpha_1})$. Assigning an improper uniform prior to α_0 we integrate this parameter from the model to give

$$p(\alpha_1|\boldsymbol{y}) \propto \prod_{i=1}^{n}\left(\frac{E_i[(1-x_i)+x_i\mathrm{e}^{\alpha_1}]}{\sum_{i=1}^{n}E_i[(1-x_i)+x_i\mathrm{e}^{\alpha_1}]}\right)^{y_i} p(\alpha_1),$$

the likelihood contribution of which tends to the constant

$$\prod_{i=1}^{n}\left(\frac{E_i(1-x_i)}{\sum_{i=1}^{n}E_i(1-x_i)}\right)^{y_i} \tag{24}$$

as $\alpha_1 \to -\infty$, showing that a proper prior is required. The constant (24) is non-zero unless $x_i = 1$ in any area with $y_i \neq 0$. The reason for the impropriety is that $\alpha_1 = -\infty$ corresponds to a relative risk of zero, so that exposed individuals cannot get the disease, which is not inconsistent with the observed data unless all individuals in area i are exposed, $x_i = 1$, and $y_i \neq 0$ in that area, since then clearly the cases are due to exposure. A similar argument holds as $\alpha \to \infty$, with replacement of $1 - x_i$ by x_i, in (24) providing the limiting constant. Figure 7 illustrates this behavior for the Scottish lip cancer example, for which $x_i = 0$ in five areas.

In the Bayesian analyses we now describe next assumed that the relative risk was less than 1 with probability 0.5 and less than 50 with probability 0.95 to give the lognormal prior $\mathrm{LN}(0, 2.38^2)$ for the relative risk e^{α_1}.

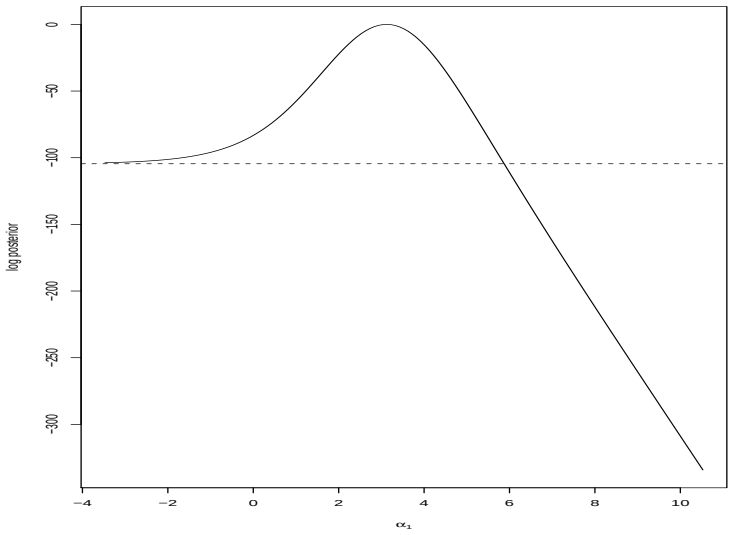


Fig. 7. Log likelihood for α_1 for the Scottish data; the horizontal line is the constant to which this function tends to as $\alpha_1 \to -\infty$.

Priors for the other parameters were the initial choices described above. The non-spatial Poisson-lognormal model, (8) gave posterior mean and standard deviation of 21 and 7.5, in close agreement with quasi-likelihood.

We fit joint and ICAR versions of (23) with the priors used previously. The posterior mean and standard deviation of the relative risk from the joint analysis are 6.4 and 3.3 while under the ICAR model the posterior mean and standard deviation are 9.9 and 3.3.

3.9. Conclusions

There are a number of problems with the spatial regressions described here. There are likely to be missing confounders as we have no information on lifestyle characteristics of the individuals in the areas such as diet, smoking and alcohol. *A priori* internal standardization was carried out which may distort the regression coefficient. Finally, the assumption of constant proportions of individuals in AFF across age groups is highly dubious. However, the coefficients observed are very large, and so we may conclude that there is an association between lip cancer incidence and occupations that lead to exposure to sunlight; placing a numerical value on the relative risk would be hazardous, however!

In general, an appropriate individual model should be aggregated to give the ecological model, so that the assumptions and aggregate data required for accurate inference can be clarified. The models representing spatial trends in risk, at both short and non-short scales, should also be carefully chosen, in order to decide on the manner in which spatial dependence is modelled.

4. Postscript: Analyses with Augmented Data

In previous sections we used the Scottish lip cancer data that are routinely available and have been analysed by numerous authors. We now analyse data provided by Michel Smans of the International Agency for

Research on Cancer. These data consist of disease counts, Y_{ik} and populations N_{ik} by county i and age group k, $i = 1, ..., 56$, $k = 1, ..., 10$. The first nine age strata were collapsed from the original 18 5-year strata due to sparsity of cases, so that the age categories are: [0,44), [45,49), ... , [80,85), 85+. Two county-level covariates were also provided: the proportion in AFF, and the proportion of all economically active households in Social class IV (partly skilled) and V (unskilled); these variables were obtained from General Register Office (1983). There were some small differences in the expected numbers constructed from the full data, when compared to the widely-used data, due to rounding errors. The new AFF x variable did not always agree with the widely-used data, even accounting for measurement error, apparently due to the mis-reading of Figure 3.6 in the cancer atlas. For example, four values of 0.16 in the widely-used data were actually 0.01.

With respect to disease mapping, the proportionality assumption (1) was assessed by splitting the data into two sets of age categories (< 65, ≥ 65) and calculating SMRs for each. This resulted in very similar SMRs in each age category across areas, so proportionality appears reasonable for these two age groups. This could be examined more formally using regression modelling with interactions between age and county.

Fitting a quasi-likelihood model with the original expected numbers, the new AFF variable and the naive log-linear ecological regression model gave an estimate of β_1 of 8.9 with standard error 1.15, compared to 7.4 with standard deviation 1.3 for the original data. Hence we see attenuation when using the mis-measured x variable. The reduced standard error was due to a reduction in κ of 4.9 to 3.8, perhaps due to induced model misspecification when the mis-measured x is used.

To assess whether *a priori* internal standardisation distorts the estimate of the AFF association we carry out simultaneous quasi-likelihood estimation of age and AFF using the naive mean model $\mathrm{E}[Y_{ik}|\boldsymbol{\beta}, \boldsymbol{\gamma}] = N_{ik} \exp(\beta_0 + \beta_1 x_i + \gamma_k)$, and $\mathrm{var}(Y_{ik}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \kappa \times \mathrm{E}[Y_{ik}|\boldsymbol{\beta}, \boldsymbol{\gamma}]$. The estimate of $\widehat{\beta}_1 = 7.5$ differs from the expected numbers estimate of 8.9, showing that age is a confounder, and bias due to a lack of mutual standardization.

We also analyzed the age-stratified data using the more appropriate individual model (22), and the new AFF variable with the joint spatial model, and the same priors as previously with flat priors on γ_k, $k = 2, ..., 10$. The posterior mean for the relative risk was 17 with standard deviation 6.3, and 95% interval (7.2,32), which is considerably larger than in Table 3.8. With an ICAR spatial model the corresponding figures were 19 and 6.0 with 95% interval (9.5,33). For each pair of analyses in Sections 3 and 4 the ICAR model gives less attenuated relative risk estimates than the joint model, perhaps because the smoothing is more local under the ICAR model, and so less of the south-north trend in risk is being absorbed into the spatial residuals.

The social class variable may be seen, at least potentially, as an ecological surrogate for lifestyle characteristics such as smoking, diet and alcohol consumption. Perhaps surprisingly, the inclusion of the social class variable did not lead to a significant change in the coefficient associated with AFF. This is only an attempt to assess between-area confounding, however. To examine within-area confounding would require the cross-classification of AFF and social class within each area, which is unavailable.

The final aspect of these data that we cannot access is within-area confounding due to age, since we do not have information on the proportion in AFF within each age stratum; if this were available then we would not be susceptible to ecological bias, at least due to this source, since we could completely characterise the within-area distribution of exposure and confounder.

5. Discussion

Throughout this paper we have stressed that though the models applied in disease mapping and spatial regression studies have similar features, the enterprises have very different aims, and modelling strategies should reflect this. Disease mapping is an exercise in prediction and therefore the form of the regression model can be very flexible, and need not reflect any causal mechanism. The use of the raw SMR is statistically valid (as E_i increases the SMR_i tends to the correct area-level relative risk), but residual spatial dependence is potentially useful since it may be exploited to smooth estimates in neighbouring areas. Area-level estimates should be carefully examined, however, to see that appropriate amounts of smoothing (in both regression and spatial models) have been used. Inappropriate smoothing may be reflected in unexpected changes in estimates when compared to the SMRs, as we saw with the Scottish data. Prior choice for the variance parameters is important, particularly if the number of areas is not large. In spatial regression the aim is to estimate causal parameters, and so the form of the mean model is of vital importance. Building an aggregate model from the level of the individual is an important step towards understanding potential sources of ecological bias. Valid inference in spatial regression also requires acknowledgment of residual spatial dependence. Careful modelling of residual spatial dependence, at both small and large scales, is required however, since regression coefficients of interest will often be sensitive to the form of the dependence assumed. Whether to include large-scale trends should be based on epidemiological considerations concerning likely unmeasured confounders. Short-scale dependence models should ideally be informed by priors based on previous studies of the disease under study. We have discussed prior choices in some detail, but we stress that each study must be considered separately, and we would not recommend uncritical use of the specific prior choices used here.

All fitting was carried out using the freely-available `R` and `WinBUGS` packages; code to implement each of the models described here may be found at `http://faculty.washington.edu/jonno/cv.html`. The `WinBUGS` software is a very flexible piece of freely-available computer software that uses MCMC algorithms to generate dependent samples from the posterior distribution of a user-specified model. The `GeoBUGS` software has a library of many common spatial models; see Thomas et al. (2000) for further details. Rue and Held (2005) provide details of alternative MCMC algorithms.

Multilevel models provide a means for modelling dependencies in data in order to provide appropriate standard errors and to allow smoothing, but they cannot control for confounding. As argued in Wakefield (2003), in spatial regression studies more effort should be placed on confounding/within-area modelling than spatial dependence, as the latter will be of secondary importance. This was illustrated in the Scottish lip cancer data, where previous analyses had investigated a variety of spatial models, but with an incorrect mean function. In principle, residual spatial dependence is a problem for individual-level studies also, though the collection of individual-level confounder variables is likely to reduce the extent of shared unmeasured variables, and hence residual dependence.

We have seen that for disease mapping great care is required in modelling both covariates and the large-term spatial trend. Splines are appealing in disease mapping studies, to model both covariates and spatial coordinates, as they provide flexible modelling and, at least for non-spatial models, can easily be incorporated into an empirical Bayes procedure. Computationally simple frequentist fitting is ideal for exploratory analyses, though incorporating residual spatial dependence is currently not straightforward.

Inferentially we have described frequentist methods for inference in non-spatial models, since these are computationally convenient, and Bayesian methods for spatial models; for spatial regression fitting an appropriate

mean model is more important than the choice of any particular inferential paradigm.

Acknowledgments

I would like to thank Carol Gotway for supplying the Great Britain National Grid coordinates for the Scottish data, Michel Smans for the original Scottish data, and the editor for detailed comments on the paper. This work was supported by grant R01 CA095994 from the National Institutes of Health.

References

Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, Florida: Chapman and Hall/CRC Press.

Berger, J.O., De Oliveira, V., and Sanso, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association 96*, 1361–1374.

Bernardinelli, L., Clayton, D., Montomoli, C., Ghislandi, M., and Songini, M. (1995). Bayesian estimates of disease maps: how important are priors. *Statistics in Medicine 14*, 2411–2431.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic auto-regressions. *Biometrika 82*, 733–746.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics 43*, 1–59.

Best, N.G., Ickstadt, K., and Wolpert, R.L. (2000). Ecological modelling of health and exposure data measured at disparate spatial scales. *Journal of the American Statistical Association 95*, 1076–1088.

Best, N., Richardson, S., and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research 14*, 35–59.

Best, N., Waller, L, Thomas, A., Conlon, E., and Arnold, R. (1999). Bayesian models for spatially correlated disease and exposure data. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds.), *Sixth Valencia international meeting on Bayesian statistics*, London, pp. 131–156. Oxford University Press.

Boffetta, P. and Nyberg, F. (2003). Contribution of environmental factors to cancer risk. *British Medical Journal 68*, 71–94.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*, 9–25.

Byers, S. and Besag, J. (2000). Inference on a collapsed margin in disease mapping. *Statistics in Medicine 19*, 2243–2249.

Christensen, O.F. and Waagepetersen, R. (2002). Bayesian prediction of spatial count data using generalised linear mixed models. *Biometrics 58*, 280–286.

Clayton, D.G. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risk. In P. Elliott, J. Cuzick, D. English, and R. Stern (Eds.), *Geographical and Environmental Epidemiology: Methods for Small-area Studies*, Oxford, pp. 205–20. Oxford University Press.

Clayton, D., Bernardinelli, L., and Montomoli, C. (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology 22*, 1193–1202.

Clayton, D.G. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics 43*, 671–682.

Conlon, E.M. and Louis, T.A. (1999). Addressing multiple goals in evaluating region-specific risk using Bayesian methods. In A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.F. Viel, and R. Bertollini (Eds.), *Disease Mapping and Risk Assessment for Public Health*, pp. 31–47. John Wiley.

Cressie, N. and Chan, N.H. (1989). Spatial modelling of regional variables. *Journal of the American Statistical Association 84*, 393–401.

Devesa, S.S., Grauman, D.J., Blot, W.J., Hoover, R.N., and Fraumeni, J.F. (1999). *Atlas of Cancer Mortality in the United States 1950–94*. NIH Publications No. 99–4564, National Institutes of Health.

Diggle, P.J., Tawn, J.A., and Moyeed, R.A. (1998). Model-based geostatistics (with discussion). *Applied Statistics 47*, 299–350.

Elliott, P., Shaddick, G., Kleinschmidt, I., Jolley, D., Walls, P., Beresford, J., and Grundy, C. (1996). Cancer incidence near municipal solid waste incinerators in Great Britain. *British Journal of Cancer 73*, 702–707.

Elliott, P., Wakefield, J.C., Best, N.G., and Briggs, D.J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press.

Fernández, C. and Green, P.J. (2002). Modeling spatially correlated data via mixtures: a Bayesian approach. *Journal of the Royal Statistical Society, Series B 64*, 805–826.

Green, P.J. and Richardson, S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association 97*, 1055–1070.

Greenland, S. (1992). Divergent biases in ecologic and individual level studies. *Statistics in Medicine 11*, 1209–1223.

Greenland, S. and Morgenstern, H. (1989). Ecological bias, confounding and effect modification. *International Journal of Epidemiology 18*, 269–274.

Greenland, S. and Robins, J. (1994). Ecological studies: biases, misconceptions and counterexamples. *American Journal of Epidemiology 139*, 747–760.

Gu, C. and Ma, P. (2005). Generalized non-parametric mixed-effects model: computation and smoothing parameter selection. *Journal of Computational and Graphical Statistics 14*, 485–504.

Haneuse, S. and Wakefied, J. (2005). The combination of ecological and case-control data. Under revision.

Haneuse, S. and Wakefied, J. (2006). Hierarchical models for combining ecological and case-control data. Under revision.

Heagerty, P.J. and Lumley, T. (2000). Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association 95*, 197–211.

Jarup, L., Toledano, M.B., Best, N., Wakefield, J., and Elliott, P. (2002). Geographical epidemiology of prostate cancer in Great Britain. *International Journal of Epidemiology 97*, 695–699.

Kelsall, J.E. and Wakefield, J.C. (1999). Discussion of "Bayesian models for spatially correlated disease and exposure data" by N. Best, L. Waller, A. Thomas, E. Conlon and R. Arnold. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds.), *Sixth Valencia international meeting on Bayesian statistics*, London. Oxford University Press.

Kelsall, J.E. and Wakefield, J.C. (2002). Modeling spatial variation in disease risk: a geostatistical approach. *Journal of the American Statistical Association 97*, 692–701.

Kemp, I., Boyle, P., Smans, M., and Muir, C. (1985). *Atlas of Cancer in Scotland, 1975–1980: Incidence and Epidemiologic Persepctive*. IARC Scientific Publication 72, Lyon, France, International Agency for Research on Cancer.

Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine 17*, 2045–60.

Knorr-Held, L. and Raßer, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics 56*, 13–21.

Lasserre, V., Guihenneuc-Jouyaux, C., and Richardson, S. (2000). Biases in ecological studies: utility of including within-area distribution of confounders. *Statistics in Medicine 19*, 45–59.

Lawson, A.B., Biggeri, A.B., Böhning, D., Lesaffre, E., Viel, J.F., and Bertollini, R. (1999). *Disease Mapping and Risk Assessment for Public Health*. New York: John Wiley and Sons.

Lawson, A.B., Biggeri, A.B., Böhning, D., Lesaffre, E., Viel, J.F., Clark, A., Schlattmann, P., and Divino, F. (2000). Disease mapping models: an empirical evaluation. *Statistics in Medicine 19*, 2217–2241.

Lawson, A. and Williams, F. (1994). Armadale: a case-study in environmental epidemiology. *Journal of the Royal Statistical Society, Series A 157*, 285–98.

Lee, Y. and Nelder, J.A. (2001). Modelling and analysing correlated non-normal data. *Statistical Modelling 1*, 3–16.

Leroux, B.G. (2000). Modeling spatial disease rates using maximum likelihood. *Statistics in Medicine 19*, 2321–2332.

Leroux, B.G., Lei, X., and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In M.E. Halloran and D.A Berry (Eds.), *Statistical Models in Epidemiology. the Environment and Clinical Trials*, pp. 179–192. New York: Springer.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by smoothing splines. *Journal of the Royal Statistical Society, Series B 61*, 381–400.

Manton, K.G., Woodbury, M.A., Stallard, E., Riggan, W.B., Creason, J.P., and Pellom, A.C. (1989). Empirical Bayes procedures for stabilizing maps of U.D. cancer mortality rates. *Journal of the American Statistical Association 84*, 637–50.

Matérn, B. (1986). *Spatial Variation, Second Edition*. Springer-Verlag, Berlin.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models, Second Edition*. London: Chapman and Hall.

Mollié, Annie (1996). Bayesian mapping of disease. In WalterR. Gilks, Sylvia Richardson, and DavidJ. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, New York, NY, USA, pp. 359–379. Chapman & Hall.

Morgenstern, H. (1998). Ecologic study. In P. Armitage and T. Colton (Eds.), *Encyclopedia of Biostatistics Vol. 2*, pp. 1255–76. New York: John Wiley and Sons.

Mugglin, A.S., Carlin, B.P., and Gelfand, A.E. (2000). Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association 95*, 877–887.

Office, GeneralRegister (1983). *Census 1981: Great Britain – Economic Activity, Part IV*. London, Office of Population Censuses and Surveys, HMSO.

Piantadosi, S., Byar, D.P., and Green, S.B. (1988). The ecological fallacy. *American Journal of Epidemiology 127*, 893–904.

Plummer, M. and Clayton, D. (1996). Estimation of population exposure. *Journal of the Royal Statistical Society, Series B 58*, 113–126.

Prentice, R.L. and Sheppard, L. (1995). Aggregate data studies of disease risk factors. *Biometrika 82*, 113–25.

Price, P.N., Nero, A.V., and Gelman, A. (1996). Bayesian prediction of mean indoor radon concentrations fo Minnesota counties. *Health Physics 71*, 922–936.

Richardson, S. (2003). Spatial models in epidemiological applications. In P.J. Green, N.L. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 237–259. Oxford: Oxford Statistical Science Series.

Richardson, S., Stucker, I., and Hémon, D. (1987). Comparison of relative risks obtained in ecological and individual studies: some methodological considerations. *International Journal of Epidemiology 16*, 111–20.

Rosenbaum, P.R. and Rubin, D.B. (1984). Difficulties with regression analyses of age-adjusted rates. *Biometrics 40*, 437–443.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Thoery and Application*. Chapman and Hall/CRC.

Smans, M. and Esteve, J. (1992). Practical approaches to disease mapping. In P. Elliott, J. Cuzick, D. English, and R. Stern (Eds.), *Geographical and environmental epidemiology: Methods for small-area studies*, Oxford, pp. 141–50. Oxford University Press.

Thomas, A., Best, N.G., Arnold, R.A., and Spiegelhalter, D.J. (2000). *GeoBUGS User Manual*. London: Imperial College of Science, Technology and Medicine.

Thurston, S.W., Wand, M.P., and Wiencke, J.K. (2000). Negative binomial additive models. *Biometrics 56*, 139–144.

Toledano, M., Jarup, L., Best, N., Wakefield, J.C., and Elliott, P. (2001). Spatial and temporal trends of testicular cancer in Great Britain. *British Journal of Cancer 84*, 1482–1487.

Tsutakawa, R.K., Shoop, G.L., and Marienfeld, C.J. (1985). Empirical Bayes estimation of cancer mortality rates. *Statistics in Medicine 4*, 201–212.

Wakefield, J. and Elliott, P. (1999). Issues in the statistical analysis of small area health data. *Statistics in Medicine 18*, 2377–2399.

Wakefield, J.C. (2003). Sensitivity analyses for ecological regression. *Biometrics 59*, 9–17.

Wakefield, J.C. (2004). Ecological inference for 2×2 tables (with discussion). *Journal of the Royal Statistical Society, Series A 167*, 385–445.

Wakefield, J.C., Best, N.G., and Waller, L.A. (2000). Bayesian approaches to disease mapping. In P. Elliott, J.C. Wakefield, N.G. Best, and D. Briggs (Eds.), *Spatial Epidemiology: Methods and Applications*, pp. 104–27. Oxford: Oxford University Press.

Waller, L.A. and Gotway, C.A. (2004). *Applied Spatial Statistics for Pubic Health Data*. Wiley.

Walter, S.D. (2000). Disease mapping: a historical perspective. In P. Elliott, J.C. Wakefield, N.G. Best, and D. Briggs (Eds.), *Spatial Epidemiology: Methods and Applications*, pp. 223–239. Oxford University Press.

Yasui, Y. and Lele, S. (1997). A regression method for spatial disease rates: an estimating function approach. *Journal of the American Statistical Association 92*, 21–32.