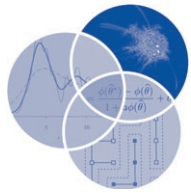
Overview



Hierarchical modeling in spatial epidemiology

Andrew B. Lawson*

This paper considers the basic concepts and methods used in hierarchical modeling for data arising in spatial epidemiology. Following discussion of basic statistical and epidemiological concepts relevant to small-area health studies, the paper reviews the different approaches to model formulation, parameter estimation, and also software resources.

How to cite this article:
WIREs Comput Stat 2014, 6:405–417. doi: 10.1002/wics.1315

Keywords: hierarchical; spatial; epidemiology; disease mapping; modeling; small area; health

INTRODUCTION

The analysis of small-area health data is the focus of spatial epidemiology. Spatial epidemiology can be thought of broadly as the analysis of health outcomes which are geo-referenced, i.e., any set of health events that have associated with them a location (either an address or an area within which the event took place) can fall under this category. Hierarchical modeling of such disease data is closely related to conventional multi-level modeling, but has some special features related to the spatial referencing of outcomes. In the following there will be an emphasis on hierarchical modeling and its toolkit. Statistical methods employed in the analysis of small-area heath data are diverse in their range and besides basic exploratory and descriptive methodology common to many subject areas, there is a need to employ particular spatial statistical methods which are designed for such data. The basic characteristic of data encountered in this application area is its *discrete* nature, whether in the form of spatial locations of cases of disease, or counts of disease within defined geographical regions. Hence, methods developed for continuous spatial processes, such as Kriging, are not directly applicable or only approximately valid.

*Correspondence to: lawsonab@musc.edu

Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

Conflict of interest: The author has declared no conflicts of interest for this article.

Often geographical hypotheses of interest in public health (PH) focus on whether the residential address of cases of disease yields insight into etiology of the disease or, in a PH application, whether adverse environmental health hazards exist locally within a region (as exemplified by local increases in disease risk). For example, in a study of the relationship between malaria endemicity and diabetes in Sardinia a strong negative relationship was found (Refs 1 and 2, ch. 9). This relation had a spatial expression, and the geographical distribution of malaria was important in generating explanatory models for the relation. In PH practice, it is of considerable importance to be able to assess whether localized areas which have larger than expected numbers of cases of disease are related to any underlying environmental cause. Here spatial evidence of a link between cases and a source is fundamental in the analysis. Evidence such as a decline in risk with distance from the *putative* source of hazard or elevation of risk in a preferred direction is important in this regard.

SOME GENERAL SPATIAL EPIDEMIOLOGICAL ISSUES

Before considering the study of the spatial distribution of disease, there are some fundamental epidemiological ideas that should be considered.

Relative Risk

Within any geographical area the local density of cases of disease can be studied. We often want to examine

this as it gives information about local variations in disease. If we have census tracts then the count of cases of a particular disease could be the data of interest. These crude counts of disease cannot be used on their own as the density of cases will be affected by the variation in the population of the area. This is true whether we observe case addresses (the residential address of a disease case) or the aggregated count of cases within small areas.

Hence, underlying the disease incidence is the variation in the population 'at risk' of the disease. This background population will vary in its composition (age, gender, susceptibility groups) and in its density with spatial location. Hence, this variation should be accounted for in any analysis of the disease occurrence. Clearly, if areas of high susceptibility (with frail population groups) coincide with areas of high disease occurrence then there is likely to be less interest in these areas (in terms of adverse disease presence) than areas where there is high disease occurrence and low number of susceptibles. Local occurrence of disease (counting of cases within areas) within short time spans (e.g., individual months or years) is termed *incidence*. Longer term accumulation of disease cases is often termed *prevalence*. Here the term incidence is used throughout. In general prevalence can be analyzed as for incidence.

To simplify discussion, initially, we will assume that we have a small administrative area (such as a census tract, postcode, zip code, county etc.) within which we observe the disease incidence. Often we want to compare the observed count of disease with what would have arisen from the underlying population. This will tell us if there is any excess disease risk in the local area. Let us assume there are $i = 1, \ldots, p$ tracts or small areas in a study area. Often a ratio of the observed count y_i in the ith tract to the expected count e_i derived from the background population is used to examine excess risk: the *relative risk* of a disease within the i th area can be estimated by $\frac{y_i}{e_i}$. This ratio represents the relative risk compared to that which the local population suggests should be seen in the area. Usually the count y_i will be available from government PH data sources and the expected count (or rate) is usually computed from known rates for the disease in population subgroups (broken by age and gender). This is known as *standardization*. The calculation of expected rates can be very important and different methods of calculation could lead to different conclusions about disease risk. Note that this relative risk definition implies a multiplicative model for risk. This is a common assumption in epidemiology.

Standardized Mortality/Morbidity/ Incidence Ratio (SMR or SIR)

The above relative risk ratio is commonly computed for certain types of data. The most common is where incident cases are involved and is called the *standardized incidence ratio* (SIR). Sometimes live cases are described by the term morbidity and so standardized mortality ratio (SMR) is sometimes used. This can be confusing as when deaths from a disease are recorded (mortality) the same acronym is applied (SMR). Of course, different expected rate calculations (denominators) would usually be used depending on whether incidence or mortality was to be considered.

Standardization

Expected rates in the small areas or tracts $\{e_i\}$ are calculated (estimated) from the local population structure. Usually an external (reference) standard population rate will be known and applied to the local population. This population is used as a guide to what should be present in terms of incidence of disease. For discussion of this issue, see Ref 3 For example, suppose that the national US rate for prostate cancer (PrCA) is to be used to standardize the rates in South Carolina counties. The rate for different population groups must be known. Hence we must know the rate for each age × gender group nationally and we also must know the population in these groups locally. Define the US rate for PrCA in the kth age group and jth gender group as e_{kj}.. Define the population in these groups in the ith area as p_{kji}.. Hence the expected rate in the ith area will be simply:

$$e_i = \sum_j \sum_k e_{kj} p_{kji}$$

That is, the numbers in each tract in different age × sex groups are multiplied with known rates for the disease for equivalent groups in a *standard* population. The standard population may be the *national* population (as above) or even the *study region* population. The study region population may be the most relevant if we want to study relative spatial differences across a study region. Also note that other standardizations could be used where covariates are used to standardize the rates.

The standardized ratio of either incidence, mortality, or morbidity is the relative risk ratio computed with standardized expected rates, as specified above:

$$\mathrm{SIR}_i = \frac{y_i}{e_i}$$

Figure 1 displays a SMR map for 26 census tracts in Falkirk central Scotland. The SMR map is often

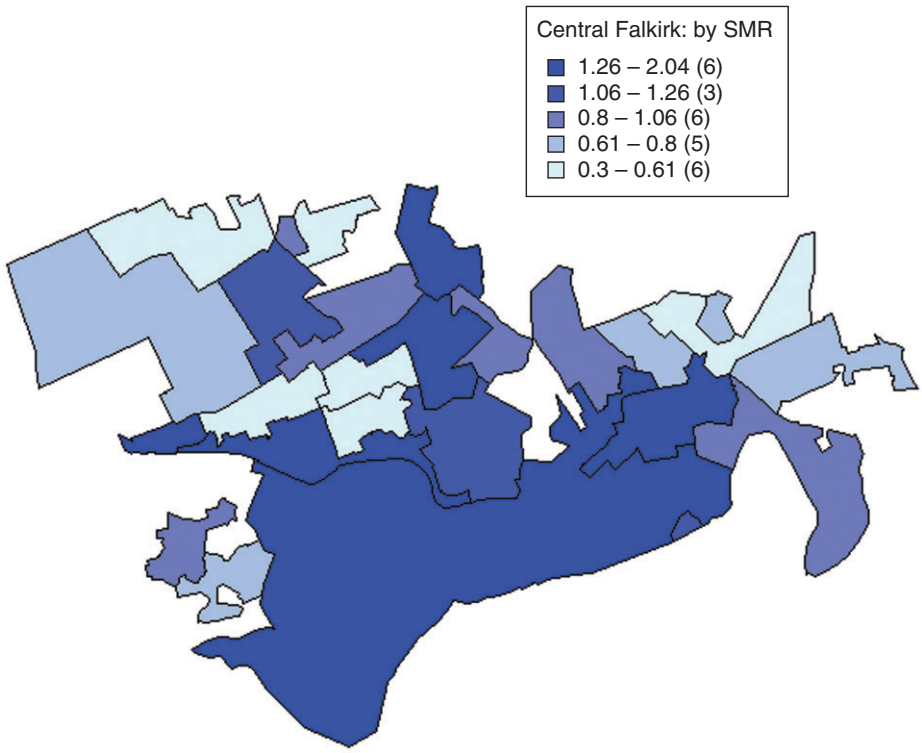


FIGURE 1 | Central Scotland: 26 census enumeration districts (EDs) in the center of the city of Falkirk. Respiratory cancer deaths [standardized mortality ratios (SMRs)] for the period 1976–1983. Scottish national rate used for standardization.

used by PH professionals to examine the distribution of disease risk. Areas of the map with SMRs greater than 2 or 3 (say) may be of concern. More formally, tests can be carried out to assess whether risk excesses are significant statistically (Ref 4, ch. 17 and Ref 5, ch. 5) Visual assessment is not adequate for this purpose. Also note that the SMR is one estimate of relative risk, and there are many other ways to estimate risk.

The Ecological and Atomistic Fallacy

Many mapping studies attempt to relate incidence of disease in regions with some other measurable *explanatory* variable relating to the etiology of the disease, e.g., we might want to examine the relation between the number of smokers in regions and the incidence of respiratory cancer in the same regions. This might be achieved by applying regression analysis to the disease incidence and explanatory variable. The relation between these variables is indirect in that we only have access to the average value for an area and not the individual level measures. Hence an average relationship can only be measured. There is no direct link between whether an individual smokes and whether they develop lung cancer.

The *ecological fallacy* arises when such regional average characteristics are ascribed to *individuals* within the region concerned. Any region-based analysis will suffer from this problem. It is known fact that in some extreme cases the relation between the covariate and the outcome is reversed when individual analysis is carried out. Hence, ecological analyses are sometimes viewed with caution. Of course, at the aggregate level the relation remains valid. The *atomistic fallacy* occurs when analysis is based on individuals, and the variability of individuals' response to disease is not accounted for in inference at the regional level. These, and other aggregation issues, are further discussed in Refs 6–8, and 9.

Confounders and Deprivation Indices

All disease maps contain the influence of variables affecting, or pertaining to, the local populations which are not accounted for in standardized rates or control diseases. We can try to allow for these effects in the following two ways:

- Include as many known explanatory variables in the expected rate or regression model to allow for these effects. These variables are called known *confounders*.
- Include the effect of unmeasured confounders via the use of *random effects*.

In the first case, the solution is to include in the study as many known variables that affect the outcome so that extra variation is explained. Of course it may not be feasible to include all know confounders simply due to (realistic) study limitations. To make allowance for unmeasured confounders (whether known or unknown) it is possible to admit random effects into any regression models. These are additional unobserved variates that will soak up extra variation of various kinds.

Often adverse disease incidence is known to be related to a range of poverty-related explanatory variables, e.g., unemployment, housing type, welfare status, car ownership. That is, we expect there to be measurable adverse risk in areas where these variates indicate low income and poverty. These variables are often available from national census. There has been some effort to combine such variables in composite measures known as *deprivation indices*.[10] In North America these are often termed urbanicity indices. Deprivation indices are now routinely available from government census data organizations and can be incorporated directly into a disease map as a covariate or as an offset term.

SOME SPATIAL STATISTICAL ISSUES

A fundamental feature of geo-referenced data available for analysis in PH applications is that it is usually discrete (either in the form of a point process or counting process), and the cases of concern arise from within a local human population which varies in spatial density and in susceptibility to the disease of interest. Hence any model or test procedure must make allowance for this background (nuisance) population effect. The background population effect can be allowed for in a variety of ways. For count data it is commonplace to obtain *expected* rates for the disease of interest based on the age–sex structure of the local population (see, e.g., Ref 11, ch. 3), and some crude estimates of local relative risk are often computed from the ratio of observed to expected counts (e.g., standardized mortality /incidence ratios: SMRs). In the following, we will focus on the modeling of small-area count data.

Count Data

Figure 2 displays a typical count data example: congenital death counts for South Carolina counties for the year 1990.

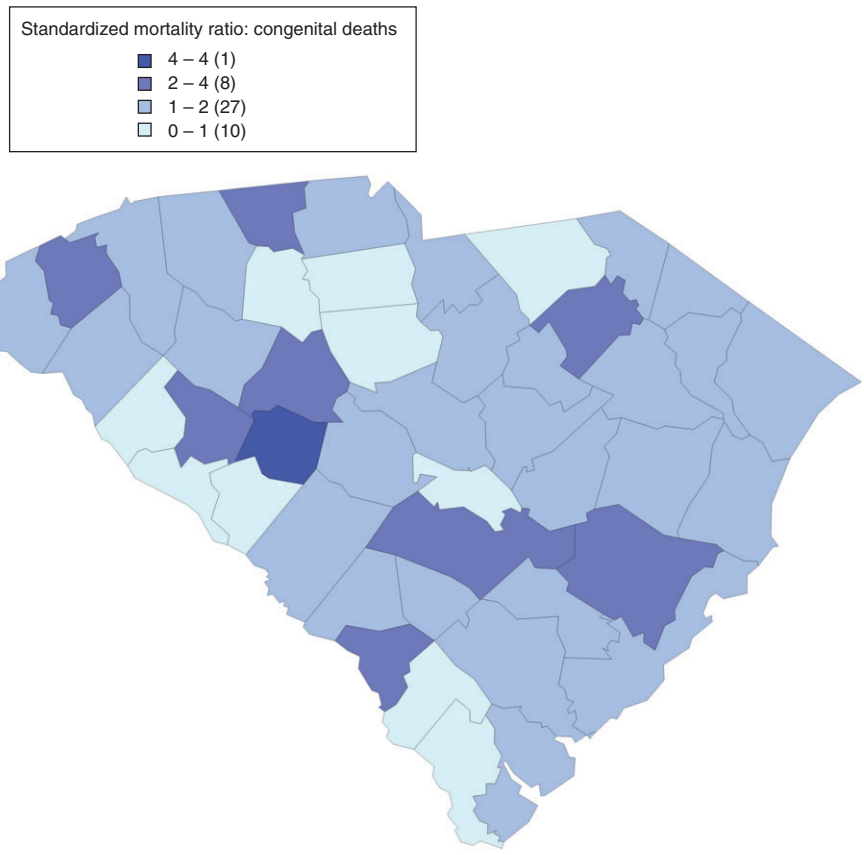


FIGURE 2 | South Carolina: congential deaths 1990 by county: standardized mortality ratio (SMR).

A considerable literature has developed concerning the analysis of count data in spatial epidemiology (e.g., see reviews in Refs 2, 11, 12, and 13).

The usual model adopted for the analysis of region counts $\{y_i, i=1, \ldots, p\}$ to be independent Poisson random variables with parameters $\{\lambda_i, i=1, \ldots, p\}$. Often the $\lambda_i s$ are assumed to be constant within areas. Usually the expected count is modeled as

$$E\left(y_i\right) = \lambda_i = e_i\theta_i, i = 1, ..., p.$$

Here the model assumes that the background expected rate (e_i) is modified by the relative risk (θ_i) in a multiplicative fashion. The relative risk is the fundamental modeling focus and most, if not all, models for risk directly model this parameter. Note that θ_i can lie on the positive real line, and can be considered to have a null value of 1, when the disease risk is simply an expression of the background effect (e_i). Values of $\theta_i < 1$ can also arise. Also note that the SIR/SMR is a simple (ML) estimator for θ_i under a saturated Poisson data model.

We express this basic data model as

$$y_i|e_i, \theta_i \sim \text{Poiss}\left(e_i\theta_i\right). \quad (1)$$

This implies that conditional on $\{\theta_i\}$ the counts are independent Poisson random variables, and so have a Poisson likelihood. Alternative forms of likelihood arise when the discrete outcome is observed in finite population (binomial likelihood) or as a binary variable (Bernoulli likelihood). In the former case then

$$y_i|n_i, p_i \sim \text{Bin}\left(n_i, p_i\right) \quad (2)$$

where n_i is the finite population in the ith area and p_i is the unit probability of disease. In the latter case a 0/1 variable would be defined to have

$$y_i|p_i \sim \text{Bern}\left(p_i\right). \quad (3)$$

Models for Relative Risk

Hierarchical models for risk can be now specified via functions of relative risk, in the Poisson case, or unit probability, in the binomial or Bernoulli case. For the Poisson case a log link is usually assumed whereby

$$\log\left(\theta_i\right) = \left(\text{linear}\right) \text{predictor}, \quad (4)$$

whereas in the binomial or Bernoulli case a logit link is often assumed. These links allow the predictor to take on a variety of values over a wide negative and positive range. Both these definitions with linear predictors represent special cases of generalized linear models. If the predictor includes random effects as well as covariates, then these are special cases of generalized linear mixed models. In general, then it can be specified that

$$E\left(y_i\right) = \mu_i$$
$$g\left(\mu_i\right) = \eta_i$$
$$\eta_i = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}\gamma$$

where $x_i^{\mathrm{T}}\beta$ is a linear predictor composed of x_i^{T}, the ith row of covariate design matrix x and β is a parameter vector, and $z_i^T\gamma$ is a linear predictor consisting of z_i^T, the ith row of random effects and γ is a vector of 0/1 indicators. This specification is quite general and assumes that the effects are additive and appear in linear combinations. For the Poisson example $E(y_i) = e_i\theta_i$ and $g(\mu_i) = \log(\mu_i) = \log(e_i) + \log(\theta_i)$ and

$$\log\left(\theta_i\right) = x_i^T\beta + z_i^T\gamma,$$

whereas for the binomial example $E(y_i) = n_i p_i$ and usually a logit link is assumed: $g^{-1}\left(\mu_i/n_i\right) = p_i = \frac{\exp\left(x_i^T\beta + z_i^T\gamma\right)}{1+\exp\left(x_i^T\beta + z_i^T\gamma\right)}$. When a binary outcome is observed then this simplifies to $E(y_i) = p_i$ and $\text{logit}\left(p_i\right) = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}\gamma$.

A variety of hierarchical models arise from particular specifications of $x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}\gamma$. Note that case-event data can be modeled by conditioning on the case–control realization and by using the binary labels (case/control) as the outcome (see, e.g., Ref 14, ch. 2). Hence, it is also possible to use the above formulations for case-event outcomes also (see Section on Models for Relative Risk) below.

SPECIAL CASES

Disease Mapping

In this area, the focus is on the processing of the disease map to extract random noise and smooth risk variation. Often applications in health services research require the production of an 'accurate' map of relative risks. Models for relative risk range from simple SMRs to posterior expected estimates from Bayesian models. In the count data situation, define the model for the observed counts as

$$y_i|e_i, \theta_i \sim \text{Poisson}\left(e_i\theta_i\right)$$
$$\log\theta_i = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}\gamma.$$

The simplest model assumes no linkages to covariates or random terms and the ML estimator of θ_i is the SMR, i.e., $\widehat{\theta}_i = y_i/e_i$. More often, and

more generally, $\log\theta_i$ is assumed to be equal to a linear predictor involving covariates and regression parameters ($x_i^T\beta$). The simplest model here would constitute a minimal model with

$$\log\theta_i = \beta_0 \quad (5)$$

$$\text{or } \mu_i = e_i \exp\left(\beta_0\right). \quad (6)$$

This intercept-only model describes overall (average) risk for all areas, and is of course not of primary interest as any spatial structure in risk is not modeled. In fact, if this model was true then there would be no need to carry out further modeling.

A first more realistic extension to this model is to consider random noise at the unit level (i), with a random intercept model. Disease variation is often considered to consist of random noise and so a simple random effect model could be used whereby:

$$\log\theta_i = \beta_0 + v_i.$$

This model assigns noise completely to one uncorrelated component at the unit level (v_i). It is convenient at this point to consider how this noise is specified. As $\{v_i\}$ is a random effect then it is required that its form be specified. In a non-Bayesian setting the mean and variance for v_i will usually be specified. Normally a prior specification of $E(v_i)=0$ with associated variance structure would be made. In a Bayesian setting this would usually be a prior distribution. A suitable choice could be

$$v_i \sim N\left(0, \tau_v^{-1}\right)$$

where τ_v is the precision of the Gaussian distribution. For a fully specified Bayesian hierarchical approach then the model would be of the form:

$$\begin{aligned} y_i|e_i,\theta_i &\sim \text{Poisson}\left(e_i\theta_i\right) \quad (7)\\ \log\theta_i &= \beta_0 + v_i,\\ \beta_0 &\sim N\left(0, \tau_0^{-1}\right)\\ v_i &\sim N\left(0, \tau_v^{-1}\right). \end{aligned}$$

At this stage there is a hierarchy of conditioning defining the model: $\theta_i \leftarrow \beta_0,\ v_i;\ \beta_0 \leftarrow \tau_0; v_i \leftarrow \tau_v$. This can be specified via conditional distributions: $[y_i|e_i,\theta_i]$, $[\beta_0|\tau_0]$, $[v_i|\tau_v]$. This hierarchical specification forms the basis for extended model ingredients. This model (7) can be described as a uncorrelated heterogeneity (UH) model. In the following we will assume a Bayesian paradigm for the hierarchical modeling described.

Within a model hierarchy a decision about the truncation point for random variation must be made. For example, should precisions (τ_0, τ_v) have (hyper) prior distributions or be fixed. For random effects it is generally recommended that precisions should be estimable and so should be given a prior distribution. For fixed effects, such as β_0, it is not essential to provide a prior distribution for precisions (e.g., τ_0). However, in mixed effect models, where there are random components included then there can be a need to estimate the precision of fixed effects. In mixed models it is recommended that all precisions for Gaussian distributions have hyper-prior specifications. Hence model (7) with added hyper-prior distributions could be

$$\begin{aligned} y_i|e_i,\theta_i &\sim \text{Poisson}\left(e_i\theta_i\right) \quad (8)\\ \log\theta_i &= \beta_0 + v_i,\\ \beta_0 &\sim N\left(0, \tau_0^{-1}\right)\\ v_i &\sim N\left(0, \tau_v^{-1}\right) \quad (9)\\ \tau_* &\sim Ga\left(a, b\right) \end{aligned}$$

where a gamma distribution with mean a/b is assumed. Usually a non-informative choice for these parameters is made such as (0.001,0.001) or (1,0.00001).

The UH model only ascribed random variation to uncorrelated noise. It is possible to add components to this specification to allow for other spatial effects. First, it may be the overall variation in risk is slowly varying over the spatial region of interest (Ref 14, ch. 6, Figure 6.8). In that case it is possible to make a simple extension by adding a regression on the coordinates to the centroids of the study areas. For example, $\log\theta_i = \beta_0 + \beta_1 x_i + \beta_2 y_i + v_i$, where (x_i, y_i) are the cartesian coordinates of these areas, with $\beta_* \sim N\left(0, \tau_*^{-1}\right)$. This could be termed a spatial trend model.

More commonly in spatial epidemiology, a random effect addressing prior spatial correlation in the data is included with the UH effect. This is often included to allow for 'clustering' of risk via spatial smoothing. The correlated effect can be termed a correlated heterogeneity (CH) effect, and this model assigns noise to two components (UH and CH). Both components are usually fitted to capture all the noise components thought to be present. This is often termed the Besag, York and Mollie (BYM)[15] or convolution model.

The logged risk model becomes:

$$\log\left(\theta_i\right) = \beta_0 + v_i + u_i.$$

To be able to estimate these components, prior distributions are assumed for each component.

Usually these consist of independent uncorrelated zero mean normal distribution for the UH effect:

$$v_i \sim N\left(0, \tau_v^{-1}\right),$$

where τ_v is a precision parameter and a spatial correlation prior distribution for the CH component. This could be chosen in a variety of ways. Commonly, a Markov random field (MRF) prior distribution is assumed for $\{u_i\}$. The intrinsic singular Gaussian distribution[15–18] is used where the conditional mean of the region effect is based only on a neighborhood of the region:

$$u_i|.....] \sim N\left(\overline{u}_{\delta_i}, \tau_u^{-1}/n_{\delta_i}\right)$$

where δ_i is a neighborhood of the ith area, and n_{δ_i} is the number of regions in the ith neighborhood, $\overline{u}_{\delta_i}$ is the mean of the neighboring u_j values (where $j \in \delta_i$) and τ_u is a precision parameter which controls the degree of smoothing. This is sometimes known as an ICAR prior distribution. An alternative to this specification is to assume a fully parameterized covariance and a multivariate normal distribution for CH:

$$\mathbf{u} \sim \mathbf{N}_p(\mathbf{0}, \Sigma)$$

where the elements of Σ are $\sigma_{ij} = \mathrm{cov}(u_i, u_j)$. These covariance elements can be parameterized with a distance-based form such as $\sigma_{ij} = \tau \exp\left(-\alpha d_{ij}^{\psi}\right)$. Here, d_{ij} is the distance between the ith and jth observations, α and ψ are distance and shape parameters, and at zero distance the instantaneous variance is τ. This is more heavily parameterized than the MRF model above and the model also requires the inversion of a $p \times p$ covariance matrix. This of course allows for more detailed covariance modeling.

For the ICAR convolution model the full hierarchical specification, for a Poisson data model, would be:

$$\begin{aligned} y_i|e_i, \theta_i &\sim \text{Poisson}\left(e_i\theta_i\right) \\ \log\theta_i &= \beta_0 + v_i + u_i, \\ \beta_0 &\sim N\left(0, \tau_0^{-1}\right) \\ v_i &\sim N\left(0, \tau_v^{-1}\right) \\ u_i|\left\{u_j\right\}_{-i} &\sim N\left(\overline{u}_{\delta_i}, \tau_u^{-1}/n_{\delta_i}\right) \\ \tau_* &\sim Ga\left(a, b\right). \end{aligned} \tag{10}$$

In a full Bayesian analysis, all parameters $(\beta, \mathbf{u}, \mathbf{v}, \tau_* \ldots)$ would be assigned prior distributions, and usually posterior sampling of these parameters via Markov chain Monte Carlo (McMC) algorithms would be chosen. A fast alternative would be to use Integrated Nested Laplace Approximation software (INLA:www.R-inla.org) which provides numerical approximations to posterior distributions and has BYM convolution models as a built-in feature (see Section on Computation). The BYM/convolution model has been evaluated via extensive simulation and found to be relatively robust in risk estimation and cluster estimation performance even under mis-specified models.[19–21]

For case-event data, point process models must be considered initially. A heterogeneous Poisson process model could be considered for p case events $\{\mathbf{s}_i\}$ $i = 1,..., p$. It is possible to extend such a model to deal with random effects also. However, when a control disease is also available, then it is possible to consider a simpler conditional logistic analysis. Define the joint realization of p cases and q controls as $i = 1,..., p$ for the cases and $i = p+1, \ldots, p+q$ for the controls. Assume that the first-order intensity of the cases is $\lambda(s,\theta) = \rho\lambda_0(s,\theta)\lambda_1(s,\theta)$ and of the controls $\lambda_0(s,\theta)$. Define the binary indicator variable y_i as follows:

$$y_i = \begin{cases} 1 \text{ if } \mathbf{s}_i \text{ is a case} \\ 0 \text{ otherwise} \end{cases}$$

then the conditional probability of a case at s_i is just

$$\frac{\rho\lambda_1\left(\mathbf{s}_i, \theta\right)}{1 + \rho\lambda_1\left(\mathbf{s}_i, \theta\right)}.$$

Hence, the likelihood of the realization is a logistic likelihood[22] specified by

$$L\left(\theta|\left\{\mathbf{s}_i\right\}\right) = \prod_{i=1}^{p+q} \frac{\left[\rho\lambda_1\left(\mathbf{s}_i, \theta\right)\right]^{y_i}}{1 + \rho\lambda_1\left(\mathbf{s}_i, \theta\right)}. \tag{11}$$

A suitable specification for the relative risk $\lambda_1(\mathbf{s}_i, \theta)$ could be $\log\lambda_1\left(\mathbf{s}_i, \theta\right) = \mathbf{x}_i^{\mathrm{T}}\beta + v_i + u_i$, where any covariates would have to be available at all case and control locations. Note that a model without covariates only requires random effect estimates at locations. Specifying suitable prior distributions for such a model is not difficult and, for example, first-order neighborhoods of points can be obtained from tesselation information,[23,18,24] and so MRF prior distributions can be specified. Alternative semi-parametric models have been suggested by Kelsall and Diggle.[25]

Disease Clustering

In this area, the focus is not on reduction of noise, *per se*, but the assessment of the clustering tendency of the map and in particular the assessment of which areas

of a map display clustering. Here, clustering could be around a known putative source of hazard (*focused* clustering) or have no known locations of clustering (*non-focused* clustering).

It is also possible to consider hierarchical modeling of clusters. In general, the model formulation may not differ greatly from that of relative risk estimation, depending largely on the definition of clusters and clustering. Here we only consider focused clustering as it relates closely to standard mixed regression models. Non-focused clustering is discussed in detail in Ref 14, ch. 6.

Focused Clustering

Focused clustering usually assumes that some form of distance decrease in risk happens around a fixed known point or points. For example, the count data model can be defined as

$$y_i \sim \text{Poisson}\left(e_i\theta_i\right)$$
$$\log\theta_i = \log\left(1+\exp\left\{-\alpha d_i\right\}\right) + x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}\gamma,$$

where d_i is a distance measured at the small area from the focus point (such as, a chimney, mobile phone mast, or waste dump site). Here the extra covariates appear in x_i^{T}, while the z_i^{T} is the ith row of a matrix of random effects and γ is a 0/1 vector. In this case focus is on inference concerning α, as this defined the distance relation. Within x_i^{T} there could also be directional terms such as $\cos(\phi)$ and $sin(\phi)$, where ϕ is the angle between the area (centroid) and the focus point. This can be used to detect any directional concentration of risk (which could be important particularly if an air pollution risk is possible).

For case events, the case-event locations are often assumed to follow a heterogeneous Poisson process with first-order intensity $\lambda(\mathbf{s})$. Denote this as $\{\mathbf{s}_i\} \sim PP(\lambda(\mathbf{s}))$. If a control disease is available and the conditional logistic likelihood (11) is assumed then the intensity can be parameterized as:

$$\log\lambda_1\left(\mathbf{s}_i,\theta\right) = \log\left(1+\exp\left\{-\alpha d_i\right\}\right) + x_i^{T}\beta + z_i^{T}\gamma,$$

where d_i is the distance from any case or control event to the focus point. Directional effects can be included here also as for count data. When fixed effects are included only with no covariates, then a frequentist approach would allow the estimation of α via maximum likelihood. Equally, if a Bayesian approach is assumed then all parameters would have prior distributions and the resulting posterior distribution would usually be sampled. A recent summary of this area of Bayesian modeling appears in Ref 14, ch. 8.

Computation

Full Bayesian analysis of spatial hierarchical models such as (7), (8), and (10) is implemented on WinBUGS (available free at http://www.mrc-bsu.cam.ac.uk/bugs), OpenBUGS (www.openbugs.net), and INLA (www.R-INLA.org). Many WinBUGS or OpenBUGS programs are available to use at: http://academic departments.musc.edu/phs/research/lawson/data.htm.

Also the GeoBUGS manual[26] contains a wide variety of examples with different prior distributional assumptions. Appendix D of Lawson[14] also has examples of INLA code for Bayesian spatial models. Discussion of the use of INLA for disease mapping can be found in Refs 27–30.

Ecological Analysis

In this area, the relation between disease incidence and explanatory variables is the focus, and this is usually carried out at an aggregate level, such as with counts in small areas. Many issues of bias and misclassification error can arise with ecological data and the interested reader is referred to Wakefield[6,9,31,32] for further insights.

Two important areas of concern are related to scale aggregation issues, modifiable areal unit problem (MAUP) and misaligned data problem (MIDP). The MAUP concerns the scalability of models and that whether, at different spatial scales, a model is valid. In general, this is unlikely to be the case as far as covariance structure is concerned as this would lead to fractal covariances which are not found commonly. However, the labeling of scales of relevance of models is important and the extent to which a model can be scaled is relevant in many applications. A related but different issue is how to use different scales of data within one analysis, i.e., should individual level data be used in preference to aggregated data or can they be combined. This is the focus of current research.

The MIDP is related to the last issue, but specifically addresses the issue of combining data from different spatial scales to provide analysis at one level. For example, health outcomes (disease incidence etc.) may be observed within census tracts and we have available pollution measurements at monitoring sites around the study area. To make inferences about the health outcomes we want to use the pollution data relevant to the census tracts. One simple solution would be to block Krige the pollution data to provide block estimates for each of the tracts (see, e.g., Refs 33 and 13, ch. 6). This would ignore the error in the interpolation of the pollution data of course and a better approach is to consider a model where the true exposure is modeled within the health model but the pollution model is jointly estimated.

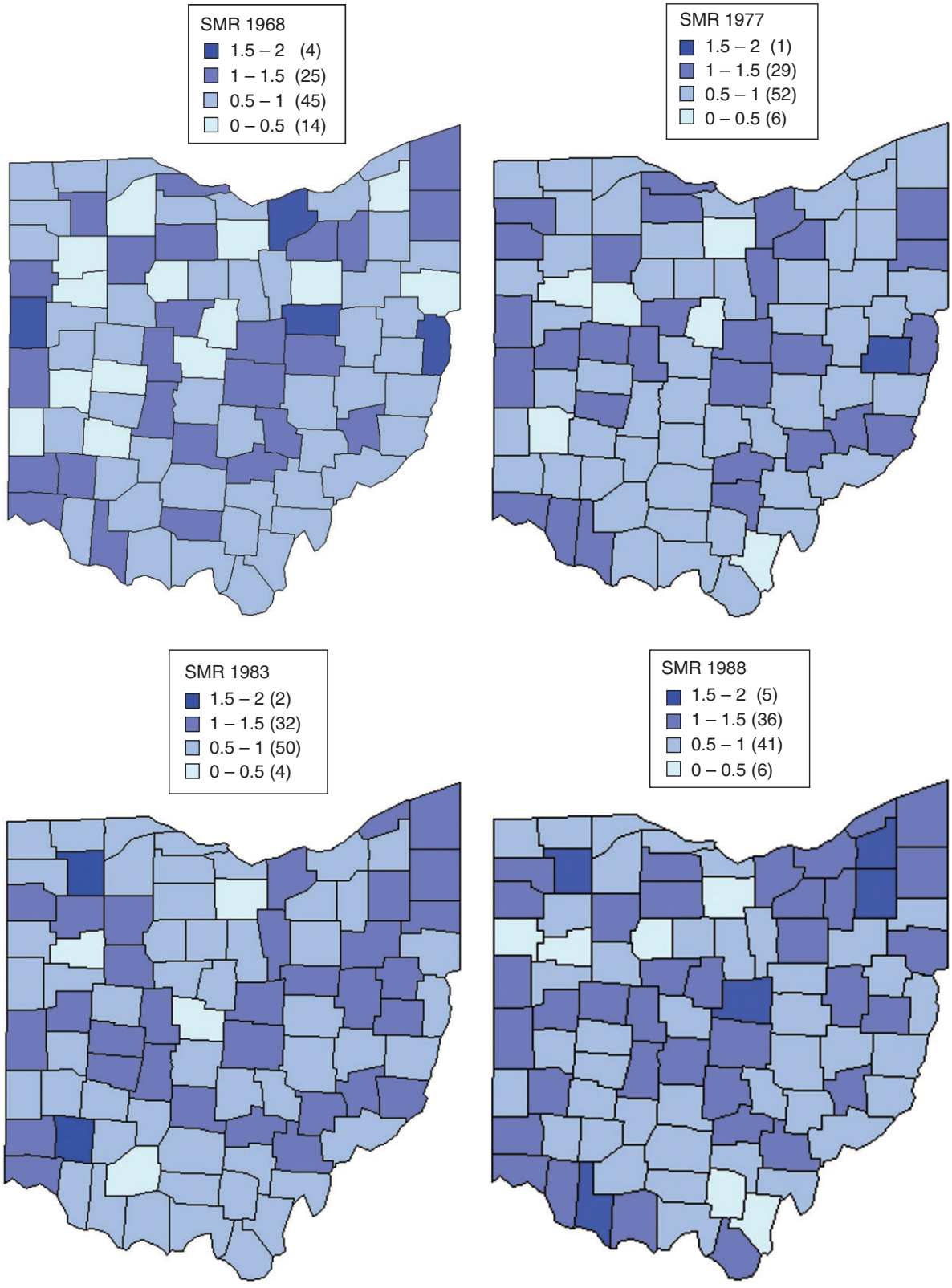


FIGURE 3 | Ohio county map: respiratory cancer standardized mortality ratios (SMRs) for four selected years: 1968, 1977, 1983, and 1988.

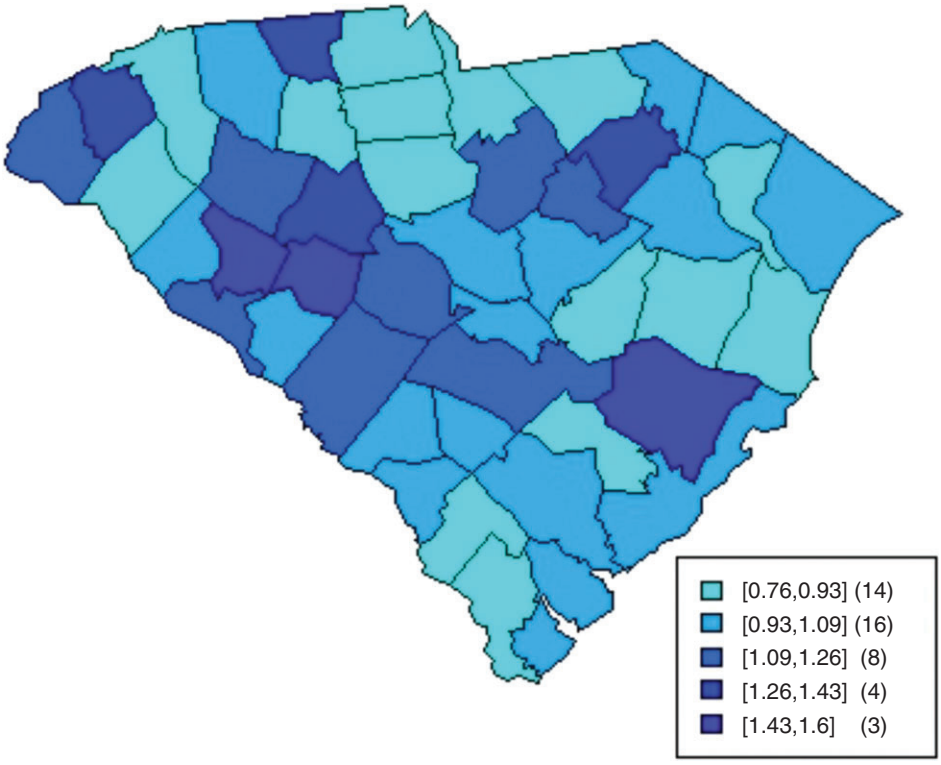


FIGURE 4 | Relative risk posterior mean estimates from the BYM convolution model with a poverty covariate included.

The model often assumed for count data is of the form

$$y_i \sim \text{Poisson}\left(e_i\theta_i\right)$$
$$\log\theta_i = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}\gamma.$$

Assume we observe count data (y_i) and also observe measurements $\{x(\mathbf{s}_j)\}$ made at q sites. Assume that the measures have mean $E(x(\mathbf{s}_j)) = \mu(\mathbf{s}_j)$, and they are multivariate normal with covariance $\text{cov}(x(\mathbf{s}_i), x(\mathbf{s}_j)) = \sigma_{xij}$. Also Σ is the covariance matrix with ijth element σ_{xij}. For a block, the mean is defined as $\mu_{B_i} = |B_i|_i^{-1}\mu\left(\mathbf{s}_j\right)d\mathbf{u}$, where B_i denotes the ith area, and an estimate is $\widehat{\mu}_{B_i} = |B_i|_i^{-1}\widehat{\mu}\left(\mathbf{s}_j\right)d\mathbf{u}$. We could assume in this case:

$$\left\{y_i\right\} i\text{thPoisson}\left(e_i\theta_i\right)$$
$$\log\theta_i = \beta\mu_{B_i} + z_i^{\mathrm{T}}\gamma,$$

and jointly with

$$\left\{x\left(\mathbf{s}_j\right)\right\} \sim N_q\left(\mu\left(\mathbf{s}_j\right), \Sigma\right),$$

μ_{B_i} can be estimated and the associated error can be accounted for. Similar considerations can apply to case-event data.

Space-Time Modeling

The extension of mapping models to space time is straightforward in the case of counts within areas within time periods. Figure 3 displays sequences of maps of respiratory cancer for 4 year periods in the counties of the US State of Ohio. Space-time variation in risk is apparent from the variation from year to year for given counties.

For example, yearly counts of disease within small areas can be handled relatively straightforwardly. In this area, the focus is the construction of methods which, usually, examine the spatio-temporal variation of disease. A typical count data model (for counts y_{ij} in the ith region and jth time period) might be

$$y_{ij} \sim \text{Poisson}\left(e_{ij}\theta_{ij}\right)$$
$$\log\theta_{ij} = \alpha + (\text{covariates}) + u_i^{\mathrm{T}}\gamma + w_j^{\mathrm{T}}\xi + z_{ij}$$

where $u_i^{\mathrm{T}}\gamma$ is a sum of spatial random components (γ is a unit vector), $w_j^{\mathrm{T}}\xi$ is a sum of temporal effects (ξ is also a unit vector), and z_{ij} is a space–time interaction effect. This formulation can lead to a rich variety of models depending on the definition of the structure of the components. Knorr-Held[34] discusses various possibilities in the Bayesian context.

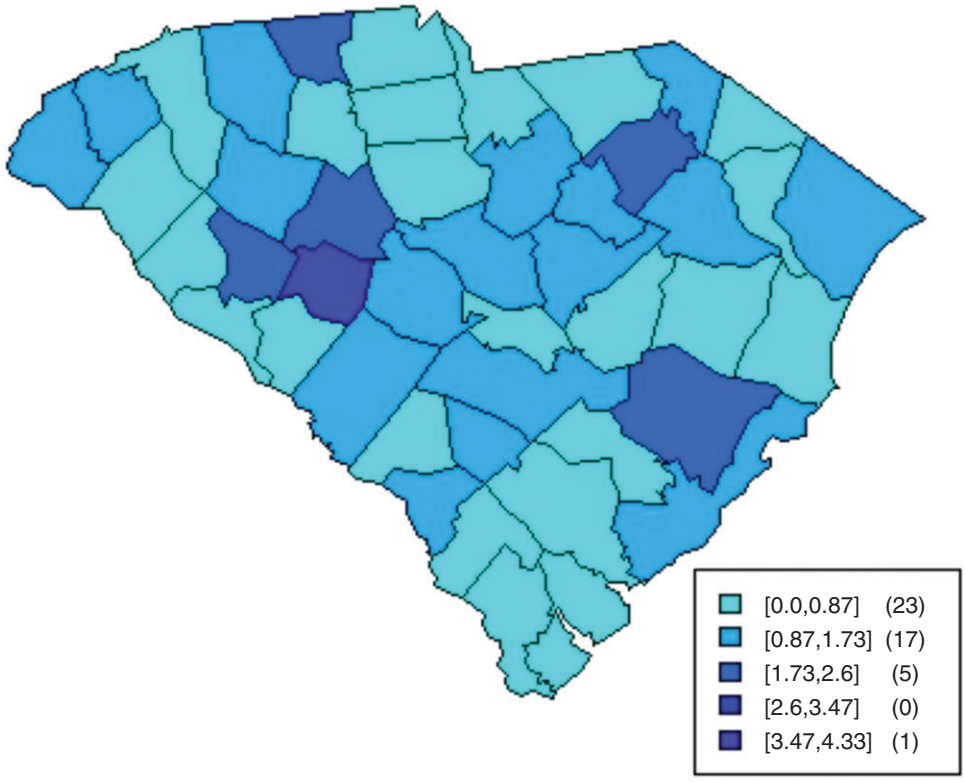


FIGURE 5 | Relative risk estimates from the uncorrelated heterogeneity (UH) model with no poverty covariate.

CASE STUDY EXAMPLE

The county level congenital death data for 1990 for South Carolina will be the focus of this exercise in application of hierarchical Bayesian models to small-area health data (see Figure 2). In this example, we have the count of deaths and expected rates (computed from the state wide rate for 1990). In addition we have a potential explanatory covariate based on the US census: percentage population in the county below the official poverty income level. We have considered four different models for these data, involving combinations of random effects and fixed effects (intercept and poverty covariate). The models are

$$\log\left(\theta_i\right) = \beta_0 + v_i \tag{12}$$

$$\log\left(\theta_i\right) = \beta_0 + \beta_1 x_i + v_i \tag{13}$$

$$\log\left(\theta_i\right) = \beta_0 + v_i + u_i \tag{14}$$

$$\log\left(\theta_i\right) = \beta_0 + \beta_1 x_i + v_i + u_i \tag{15}$$

where v_i has a zero mean Gaussian prior distribution and u_i has an ICAR prior distribution, and the regression parameters also have zero mean Gaussian prior distributions. The precisions have non-informative SD-uniform prior distributions.[35]

Fitting of models 1–4 using posterior sampling on WinBUGS led to the following results. We examined the deviance information criterion (DIC[36]) to assist in the evaluation of the merits of each model. For these data the DICs are not greatly different and certainly cannot lead to a clear choice. The range of DIC is 171.1 (model 1) to 172.1 (model 2), with the other models on 171.9. The lowest DIC is for the random intercept model (model 1), but the DIC differences are small. The pD (effective number of parameters) can be used as a secondary criterion for parsimony. Based on this model 1 is most parsimonious. This suggests that there is little residual spatial structure in these data after removal of random noise, and that a UH model is adequate to describe the variation. To demonstrate the resulting posterior estimates of risk Figures 4 and 5 display the θ_i estimates across the 46 counties of SC under the BYM model with poverty covariate (Figure 4) and the simpler UH model (Figure 5). There is little difference between these estimated risk maps, except that the UH model shows more variability in risk.

CONCLUSIONS

In this review, we have attempted to outline the main features of hierarchical modeling in the context

of spatial epidemiology. This has inevitably been a limited review as there exists a wide ranging of models now developed for specialist area with the subject. However, it is hoped that this overview will provide enough insight into the basic ideas and methods in the area to allow for easy application of these methods for those unfamiliar with the area.

REFERENCES

1. Bernardinelli L, Clayton DG, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Stat Med* 1995, 14:2433–2443.
2. Lawson AB. *Statistical Methods in Spatial Epidemiology*. 2nd ed. New York: John Wiley & Sons; 2006.
3. Fleiss JL, Levin B, Paik MC. *The Standardization of Rates, in Statistical Methods for Rates and Proportions*. 3rd ed. New York: John Wiley & Sons; 2004.
4. Elliott P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. *Stat Methods Med Res* 1995, 4:137–159.
5. Lawson AB. *Statistical Methods in Spatial Epidemiology*. New York: John Wiley & Sons; 2001.
6. Wakefield J. A critique of statistical aspects of ecological studies in spatial epidemiology. *Environ Ecol Stat* 2004, 11:31–54.
7. Salway R, Wakefield J. Sources of bias in ecological studies of non-rare events. *Environ Ecol Stat* 2005, 12:321–347.
8. Haneuse S, Wakefield J. Ecological inference incorporating spatial dependence. In: King G, Rosen O, Tanner M, eds. *Ecological Inference: New Methodological Strategies*, Chapter 12. Cambridge University Press; 2004, 266–301.
9. Wakefield J, Haneuse S, Dobra A, Teeple E. Bayes computation for ecological inference. *Stat Med* 2011, 30:1381–1396.
10. Carstairs V. Small area analysis and health service research. *Community Med* 1981, 3:131–139.
11. Elliott P, Wakefield JC, Best NG, Briggs DJ, eds. *Spatial Epidemiology: Methods and Applications*. London: Oxford University Press; 2000.
12. Lawson AB, Böhning D, Lessafre E, Biggeri A, Viel JF, Bertollini R, eds. *Disease Mapping and Risk Assessment for Public Health*. Chichester: John Wiley & Sons; 1999.
13. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. London: Chapman and Hall/CRC Press; 2004.
14. Lawson AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. 2nd ed. New York: CRC Press; 2013.
15. Besag J, York J, Mollié A. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math* 1991, 43:1–59.
16. Kunsch H. Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika* 1987, 74:517–524.
17. Rue H, Held L. *Gaussian Markov Random Fields: Theory and Applications*. New York: Chapman&Hall/CRC; 2005.
18. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent gaussian models by using integrated nested laplace approximations. *J R Stat Soc B* 2009, 71:319–392.
19. Lawson AB, Biggeri A, Boehning D, Lesaffre E, Viel J-F, Clark A, Schlattmann P, Divino F. Disease mapping models: an empirical evaluation. *Stat Med* 2000, 19:2217–2242. Special issue: Disease Mapping with emphasis on evaluation of methods.
20. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* 2005, 14:35–59.
21. Rotejanaprasert C. Evaluation of cluster recovery for small area relative risk models. *Stat Methods Med Res* to appear. doi: 10.1177/09622802145-27382.
22. Diggle P. Point process modellng in environmental epidemiology. In: Barnett V, Turkman K, eds. *Statistics in the Environment SPRUCE I*. New York: John Wiley & Sons; 1993.
23. Baddeley A, Turner R. Spatstat: an r package for anayzing spatial point patterns. *J Stat Softw* 2003, 12:1–42.
24. Lawson AB. Bayesian point event modeling in spatial and environmental epidemiology: a review. *Stat Methods Med Res* 2012, 21:509–529.
25. Kelsall J, Diggle P. Spatial variation in risk of disease: a nonparametric binary regression approach. *Appl Stat* 1998, 47:559–573.
26. Thomas, A, Best N, Lunn D, Arnold R Spiegelhalter D. Geobugs User Manual v 1.2. 2004. Available at: www.mrc-bsu.cam.ac.uk/bugs. (Accessed 2 January, 2014).
27. Schrodle B, Held L, Rieber A, Danuser J. Using integrated nested laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. *Appl Stat* 2011, 60:261–279.
28. Schrodle B, Held L. Spatio-temporal disease mapping using inla. *Environmetrics* 2011b, 22:725–734.
29. Schrodle B, Held L. A primer on disease mapping and ecological regression using inla. *Comput Stat* 2011a, 26:241–258.

30. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with r-inla. *Spat Spat-Temp Epidemiol* 2013, 4:33–49.
31. Wakefield J, Shaddick G. Health-exposure modeling and the ecological fallacy. *Biostatistics* 2006, 7:438–455.
32. Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology*. London: Chapman& Hall; 2004.
33. Schabenberger O, Gotway C. *Statistical Methods for Spatial Data Analysis*. London: Chapman&Hall; 2004.
34. Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Stat Med* 2000, 19:2555–2567.
35. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 2006, 1:515–533.
36. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian deviance, the effective number of parameters and the comparison of arbitrarily complex models. *J R Stat Soc B* 2002, 64:583–640.

