1

Disease Mapping Basics

The representation and analysis of maps of disease incidence data is now established as a basic tool in the analysis of regional public health. One of the earliest examples of disease mapping is the map of the addresses of cholera victims related to the locations of water supplies given by Snow (1854). In that case, the street addresses of victims were recorded and their proximity to putative pollution sources (water supply pumps) was assessed.

The subject area of disease mapping has developed considerably in recent years. This growth in interest has led to a greater use of geographical or spatial statistical tools in the analysis of data both routinely collected for public health purposes and in the analysis of data found within ecological studies of disease relating to explanatory variables. The study of the geographical distribution of disease can have a variety of uses. The main areas of application can be conveniently broken down into the following classes: (1) disease mapping, (2) disease clustering, and (3) ecological analysis. In the first class, usually the object of the analysis is to provide (estimate) the true *relative risk* of a disease of interest across a geographical study area (map): a focus similar to the processing of pixel images to remove noise. Applications for such methods lie in health services resource allocation, and in disease atlas construction (see, for example, Pickle *et al*., 1999). The second class, that of disease clustering, has particular importance in public health surveillance, where it may be important to be able to assess whether a disease map is clustered and where the clusters are located. This may lead to examination of potential environmental hazards. A particular special case arises when a known location is thought to be a potential pollution hazard. The analysis of disease incidence around a putative source of hazard is a special case of cluster detection called focused clustering. The third class, that of ecological analysis, is of great relevance within epidemiological research, as its focus is the analysis of the geographical distribution of disease in relation to explanatory covariates, usually at an aggregated spatial level. Many issues relating to disease mapping are also found in this area, in addition to issues relating specifically to the incorporation of covariates.

Disease Mapping with WinBUGS and MLwiN A. Lawson, W. Browne and C. Vidal Rodeiro
 ISBN: 0-470-85604-1 (HB)

In this volume, we focus on the issues of modelling. While the focus here is on *statistical* methods and issues in disease mapping, it should be noted that the results of such statistical procedures are often represented visually in mapped form. Hence, some consideration must be given to the purely cartographic issues that affect the representation of geographical information. The method chosen to represent disease intensity on the map, be it colour scheme or symbolic representation, can dramatically affect the resulting interpretation of disease distribution. It is not the purpose of this review to detail such cognitive aspects of disease mapping, but the reader is directed to some recent discussions of these issues (MacEachren, 1995; Monmonier, 1996; Pickle and Hermann, 1995; Walter, 1993).

1.1 DISEASE MAPPING AND MAP RECONSTRUCTION

To begin, we consider two situations which commonly arise in studies of the geographic distribution of disease. These situations are defined by the form of the mapped data which arises in such studies. First a study area or window is defined and within this area for a fixed period of time the locations of cases of a specified disease are recorded. These locations are usually residential addresses (street address or, at a higher spatial scale, zip code (USA) or post code unit (UK)). When such addresses are known it is possible to proceed by direct analysis of the case locations. This is termed *case-event* analysis. Often this analysis requires the use of point process models and associated methodology. This form of analysis is reviewed in Lawson (2001, Chapters 4 and 5) and elsewhere (see, for example, Elliott *et al.* (2000, Chapter 6)). Due to the requirements of medical confidentiality, it is often not possible to obtain data at this level of resolution and so resort must be made to the analysis of *counts* of cases within small areas within the study window. These small areas are arbitrary regions usually defined for administrative purposes, such as census tracts, counties, municipalities, electoral wards or health district regions. Data of this type consist of counts of cases within tracts and the analysis of this data is termed *tract count* analysis. In this volume we focus exclusively on tract count analysis. An example of the analysis of case-event data with a Bernouilli model using WinBUGS is given in Congdon (2003, Chapter 7).

Essentially the count is an aggregation of all the cases within the tract. By aggregation, the individual case spatial references (locations) are lost and therefore any georeference of the count is related to the tract 'location'. Often this is represented by the tract centroid. In a chosen study window there is found to be m tracts. Denote the counts of disease within the m tracts as $\{y_i\},\ i = 1, \ldots, m$. Figure 1.1 displays a tract count example.

This example is of the 46 counties of South Carolina in which were collected the congenital abnormality death counts for the year 1990.

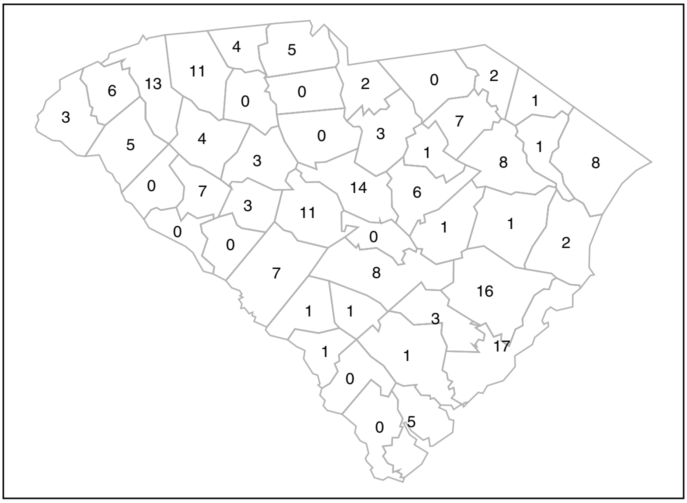

Figure 1.1 South Carolina congenital abnormality deaths 1990 by counties.

1.2 DISEASE MAP RESTORATION

1.2.1 Simple statistical representations

The representation of disease-incidence data can vary from pictorial representation of counts within tracts, to the mapping of estimates from complex models purporting to describe the structure of the disease events. In this section, we describe the range of mapping methods from simple representations to model-based forms. The geographical incidence of disease has as its fundamental unit of observation, the address location of cases of disease. The residential address (or possibly the employment address) of cases of disease contains important information relating to the type of exposure to environmental risks. Often, however, the exact address locations of cases are not directly available, and one must use instead counts of disease in arbitrary administrative regions, such as census tracts or postal districts.

1.2.1.1 Crude representation of disease distribution

The simplest possible mapping form is the depiction of disease rates at specific sets of locations. For counts within tracts, this is a pictorial representation of the number of events in the tracts plotted at a suitable set of locations (e.g., tract centroids). The locations of case-events within a spatially heterogeneous population can display a small amount of information concerning the overall pattern of disease events within a window. However, any interpretation of the structure of these events is severely limited by the lack of information concerning the spatial distribution of the background population which might be 'at risk' from the disease of concern and which gave rise to the cases of disease. This population also has a spatial distribution and failure to take account of this spatial variation severely limits the ability to interpret the resulting case-event map. In essence, areas of high density of 'at risk' population would tend to yield high incidence of case-events and so, without taking account of this distribution, areas of high disease intensity could be spuriously attributed to excess disease risk.

In the case of counts of cases of disease within tracts, similar considerations apply when crude count maps are constructed. Here, variation in population density also affects the spatial incidence of disease. It is also important to consider how a count of cases could be depicted in a mapped representation. Counts within tracts are totals of events from the whole tract region. If tracts are irregular, then a decision must be made to either 'locate' the count at some tract location (e.g. tract centroid, however defined) with suitable symbolization, or to represent the count as a fill colour or shade over the whole tract (choropleth thematic map). In the former case, the choice of location will affect interpretation. In the latter case, symbolization choice (shade and/or colour) could distort interpretation also, although an attempt to represent the whole tract may be attractive.

In general, methods that attempt to incorporate the effect of background 'at risk' population (termed: *at risk background*) are to be preferred. These are discussed in the next section.

1.2.1.2 Standardized mortality/morbidity ratios and standardization

To assess the status of an area with respect to disease incidence, it is convenient to attempt to first assess what disease incidence should be locally 'expected' in the tract area and then to compare the observed incidence with the 'expected' incidence. This approach has been traditionally used for the analysis of counts within tracts. Traditionally, the ratio of observed to expected counts within tracts is called a Standardized Mortality/Morbidity Ratio (SMR) and this ratio is an estimate of *relative risk* within each tract (i.e., the ratio describes the odds of being in the disease group rather than the background group). The justification for the use of SMRs can be supported by the analysis of likelihood models with multiplicative expected risk (see, for example, Breslow and Day, 1987). In Section 1.2.3.1, we explore further the connection between likelihood models and tract-based estimators of risk. Figure 1.2 displays the SMR thematic map for congenital abnormality deaths within South Carolina, USA, for the year 1990 based on expected rates calculated from the South Carolina 1990–1998 statewide rate per 1000 births.

Define y_i as the observed count of the case disease in the ith tract, and e_i as the expected count within the same tract. Then the SMR is defined as:

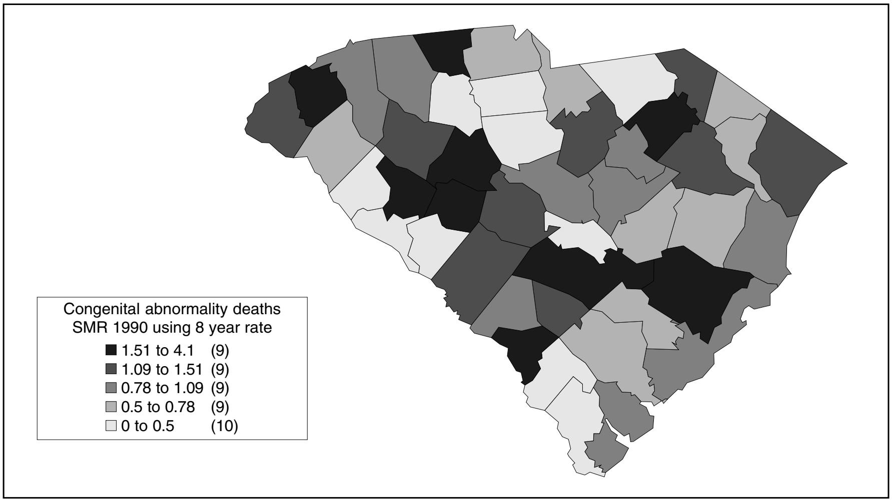

Figure 1.2 South Carolina congenital abnormality deaths 1990: SMRs.

$$\widehat{\theta}_i = \frac{y_i}{e_i}. \tag{1.1}$$

In this case it must be decided whether to express the $\widehat{\theta}_i$ as fill patterns in each region, or to locate the result at some specified tract location, such as the centroid. If it is decided that these measures should be regarded as continuous across regions then some further interpolation of $\widehat{\theta}_i$ must be made (see, for example, Breslow and Day, 1987, pp. 198–9).

SMRs are commonly used in disease map presentation, but have many drawbacks. First, they are based on ratio estimators and hence can yield large changes in estimate with relatively small changes in expected value. In the extreme, when a (close to) zero expectation is found the SMR will be very large for any positive count. Also the zero SMRs do not distinguish variation in expected counts, and the SMR variance is proportional to $1/e_i$. The SMR is essentially a saturated estimate of relative risk and hence is not parsimonious.

1.2.2 Informal methods

To circumvent the problems associated with SMRs a variety of methods have been proposed. Some of these are relatively informal or nonparametric and others highly parametric. In the rest of this volume we will concentrate on the model-based relative risk estimation methods. However, it is useful here to present briefly some notes on alternative methods.

One approach to the improvement of relative risk estimation is to employ smoothing tools on SMRs to reduce the noise. These tools could be based on interpolation methods, or more commonly on nonparametric smoothers such as kernel regression (Nadaraya-Watson, local linear) (Bowman and Azzalini, 1997), and partition methods (Ferreira *et al*., 2002). A variety of exploratory data analysis (EDA) methods have also been advocated (see, for example, Cressie, 1993). These methods usually require the estimation of a smoothing constant which describes the overall behaviour of the relative risk surface. Some local methods are also available. Generalized additive models have also been proposed and these have the advantage of allowing the incorporation of covariates (see, for example, Kelsall and Diggle, 1998).

1.2.3 Basic models

When more substantive hypotheses and/or greater amounts of prior information are available concerning the problem, then it may be advantageous to consider a model-based approach to disease map construction. Model-based approaches can also be used in an exploratory setting, and if sufficiently general models are employed then this can lead to better focusing of subsequent hypothesis generation. In what follows, we consider first likelihood models for case event data and then discuss the inclusion of extra information in the form of random effects.

1.2.3.1 Likelihood models

Usually the basic model for case-event data is derived from the following assumptions:

(1) Individuals within the study population behave independently with respect to disease propensity, after allowance is made for observed or unobserved confounding variables.
(2) The underlying at risk background intensity has a continuous spatial distribution, within a specified boundary.
(3) The case-events are unique, in that they occur as single spatially separate events.

Assumption (1) above allows the events to be modelled via a likelihood approach, which is valid conditional on the outcomes of confounder variables. Further, assumption (2), if valid, allows the likelihood to be constructed with a background continuous modulating intensity function representing the ‘at risk’ background. The uniqueness of case-event locations is a requirement of point process theory (the property called orderliness: see, for example, Daley and Vere-Jones, 1988), which allows the application of Poisson-process models in this analysis. Assumption (1) is generally valid for non-infectious diseases. It may also be valid for infectious diseases if the information about current infectives were known at given time points. Assumption (2) will be valid at appropriate scales of analysis. It may not hold when large areas of a study window include zones of zero population (e.g. harbours/industrial zones). Often models can be restricted to exclude these areas however. Assumption (3) will usually hold for relatively rare diseases but may be violated when households have multiple cases and these occur at coincident locations. This may not be important at more aggregate scales, but could be important at a fine spatial scale. Remedies for such non-orderliness are the use of de-clustering algorithms (which perturb the locations by small amounts), or analysis at a higher aggregation level. Note that it is also possible to use a conventional case-control approach to this problem (Diggle *et al.*, 2000).

In the case of observed counts of disease within tracts, the Poisson-process assumptions given above mean that the counts are Poisson distributed with, for each tract, a different expectation. Often at this point a simplifying assumption is made where the ith tract count expectation is regarded as being a function of a parameter within a model hierarchy, without considering the spatial continuity of the intensity. This assumption leads to considerable simplifications and the distribution of the tract counts is often assumed to be

$$y_i \sim Poisson(e_i \theta_i),$$

where θ_i is assumed to be a constant relative risk parameter. In this definition the expected value of the count is a multiplicative function of the expected count/rate (e_i) and a relative risk. This is the classic model assumed in many disease mapping studies. The log-likelihood associated with this model is, bar a constant, given by:

$$l = \sum_{i=1}^{m} y_i \ln(e_i \theta_i) - \sum_{i=1}^{m} e_i \theta_i.$$

Note that by differentiation the saturated maximum likelihood estimator of θ_i is just y_i/e_i, the SMR.

This model makes a number of assumptions. First it is assumed that any excess risk in a tract will be expressed beyond that described by e_i. For example, the expected rate (e_i) can be estimated in a variety of ways. Often external standardization is used, where known supra-regional rates for different age $\times$ sex groups are applied to the local population in each tract. The use of external standardization alone to estimate the expected counts/rates within tracts may provide a different map from that provided by a combination of external standardization and measures of tract-specific deprivation (e.g. deprivation indices (Carstairs, 1981)). If any confounding variables are available and can be included within the estimate of the at risk background, then these should be considered for inclusion. Examples of confounding variables could be found from national census data, particularly relating to socioeconomic measures. These measures are often defined as 'deprivation' indicators, or could relate to lifestyle choices. For example, the local rate of car ownership or percentage unemployed within a census tract or other small area, could provide a surrogate measure for increased risk, due to correlations between these variables and poor housing, smoking lifestyles, and ill-health. Hence, if it is possible to include such variables, then any resulting map will display a close representation of the 'true' underlying risk surface. When it is not possible to include such variables, it is sometimes possible to adapt a mapping method to include covariates of this type within regression setting.

1.2.3.2 Fixed effects

Usually the focus of attention when more sophisticated models are applied in disease mapping is the relative risk. Hence, all the models we will examine in this volume will be models for the $\{\theta_i\}$. One simple model for the relative risks would be to suppose that there could be a spatial trend or long-range variation over the study area. To do this we can construct a model which is a function of the spatial coordinates of the tract centroids: $\{x_{1i}, x_{2i}\}$ representing eastings and northings, say. Simple forms of spatial trend can be modelled by using the centroid coordinates or functions of the coordinates as covariates and assuming a regression-type model. As the relative risks must be positive it is usual to model the logarithm of the relative risk as a linear function. Hence, in this case we could have:

$$\theta_i = \exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\}. \tag{1.2}$$

This model includes a constant rate ($\exp\{\beta_0\}$) which captures the overall rate across the whole study region, and two linear parameters: β_1, β_2. This model describes a planar trend across the study region, and can be easily extended to include higher-order trend surfaces by adding power functions of the coordinates. Here we have used centroid locations as covariates, and indeed this model can be generalized simply when you observe other covariates measured within the tracts. For example it may be possible to include deprivation scores for each tract or census variables such as percentage unemployed or percentage car ownership. In general, assume that the intercept (constant rate) term is defined for a variable x_{0i} which is 1 for each tract. Hence we can specify the model compactly as

$$\theta_i = \exp\{\mathbf{x}_i\boldsymbol{\beta}\},$$

where $\mathbf{x}$ is a $m \times p$ matrix consisting of $p - 1$ covariates, $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector and $\mathbf{x}_i$ denotes the ith observation row of $\mathbf{x}$.

This type of fixed effect model can be fitted in conventional statistical packages which allow Poisson regression or log-linear modelling. The *glm* function in R or S-Plus with a log link and log offset of the $\{e_i\}$ can be used, for example.

1.2.3.3 Random effects

In the sections above some simple approaches to mapping counts within tracts have been described. These methods assume that once all known and observable confounding variables are included then the resulting map will be clean of all artefacts and hence depicts the true excess risk surface. However, it is often the case that unobserved effects could be thought to exist within the observed data and that these effects should also be included within the analysis. These effects are often termed *random* effects, and their analysis has provided a large literature both in statistical methodology and in epidemiological applications (see, for example, Manton *et al*., 1981; Tsutakawa, 1988; Breslow and Clayton, 1993; Clayton, 1991; Best and Wakefield, 1999; Lawson, 2001; Richardson, 2003). Within the literature on disease mapping, there has been a considerable growth in recent years in modelling random effects of various kinds. In the mapping context, a random effect could take a variety of forms. In its simplest form, a random effect is an extra quantity of variation (or variance component) which is estimable within the map and which can be ascribed a defined probabilistic structure. This component can affect individuals or can be associated with tracts or covariates. For example, individuals vary in susceptibility to disease and hence individuals who become cases could have a random component relating to different susceptibility. This is sometimes known as *frailty*. Another example is the interpolation of a spatial covariable to the locations of case events or tract centroids. In that case, some error will be included in the interpolation process, and could be included within the resulting analysis of case or count events. Also, the locations of case-events might not be precisely known or subject to some random shift, which may be related to uncertain residential exposure. Finally, within any predefined spatial unit, such as tracts or regions, it may be expected that there could be components of variation attributable to these different spatial units. These components could have different forms depending on the degree of prior knowledge concerning the nature of this extra variation. For example, when observed counts, thought to be governed by a Poisson distribution, display greater variation than expected (i.e. variance $>$ mean), it is sometimes described as overdispersion. This overdispersion can occur for various reasons. Often it arises when clustering occurs in the counts at a particular scale. It can also occur when considerable numbers of cells have zero counts (sparseness), which can arise when rare diseases are mapped. In spatial applications, it is important furthermore to distinguish two basic forms of extra variation. First, as in the aspatial case, a form of independent and spatially uncorrelated extra variation can be assumed. This is often called *uncorrelated heterogeneity* (see, for example, Besag *et al.*, 1991). Another form of random effect is that which arises from a model where it is thought that the spatial unit (such as case-events, tracts or regions) is correlated with neighbouring spatial units. This is often termed *correlated heterogeneity*. Essentially, this form of extra variation implies that there exists spatial autocorrelation between spatial units (see, for example, Cliff and Ord (1981) for an accessible introduction to spatial autocorrelation). This autocorrelation could arise for a variety of reasons. First, the disease of concern could be naturally clustered in its spatial distribution at the scale of observation. Many infectious diseases display such spatial clustering, and a number of apparently non-infectious diseases also cluster (see, for example, Cuzick and Hills, 1991; Glick, 1979). Second, autocorrelation can be induced in spatial disease patterns by the existence of unobserved environmental or frailty effects. Hence the extra variation observed in any application could arise from confounding variables that have not been included in the analysis. In disease mapping examples, this could easily arise when simple mapping methods are used on SMRs with just basic age–sex standardization.

In the discussion above on heterogeneity, it is assumed that a global measure of heterogeneity applies to a mapped pattern. That is, any extra variation in the pattern can be captured by including a general heterogeneity term in the mapping model. However, often spatially-specific heterogeneity may arise where it is important to consider local effects as well as, or instead of, general heterogeneity. To differentiate these two approaches, we use the term *specific* and *nonspecific* heterogeneity. Specific heterogeneity implies that spatial locations are to be modelled locally; for example, clusters of disease are to be detected on the map. In contrast, 'nonspecific' describes a global approach to such modelling, which does not address the question of the location of effects. In this definition, it is tacitly assumed that the locations of clusters of disease can be regarded as random effects themselves. Hence, there are strong parallels between image processing tasks and the tasks of disease mapping.

Random effects can take a variety of forms and suitable methods must be employed to provide correctly estimated maps under models including these effects. In this section, we discuss simple approaches to this problem from a frequentist, multilevel and Bayesian viewpoint.

A frequentist approach. In what follows, we use the term 'frequentist' to describe methods that seek to estimate parameters within a hierarchical model structure. The methods *do* assume that the random effects have mixing (or prior) distributions. For example a common assumption made when examining tract counts is that $y_i \sim Poisson(e_i\theta_i)$ independently, and that $\theta_i \sim Gamma(\alpha, \beta)$. This latter distribution is often assumed for the Poisson relative risk parameter and provides for a measure of overdispersion relative to the Poisson distribution itself, depending on the α, β values used. The joint distribution is now given by the product of a Poisson likelihood and a gamma distribution. At this stage a choice must be made concerning how the random intensities are to be estimated or otherwise handled. One approach to this problem is to average over the values of θ_i to yield what is often called the *marginal* likelihood. Having averaged over this density, it is then possible to apply standard methods such as maximum likelihood. This is usually known as marginal maximum likelihood (see, for example, Bock and Aitkin, 1981; Aitkin, 1996b). In this approach, the parameters of the gamma distribution are estimated from the integrated likelihood. A further development of this approach is to replace the gamma density with a finite mixture. This approach is essentially nonparametric and does not require the complete specification of the parameter distribution (see, for example, Aitkin, 1996a).

Although the example specified here concerns tract counts, the method described above can equally be applied to case-event data, by inclusion of a random component in the intensity specification.

A Bayesian approach. It is natural to consider modelling random effects within a Bayesian framework. First, random effects naturally have prior distributions and the joint density discussed above is proportional to the posterior distribution for the parameters of interest. Hence, the development of full Bayes and empirical Bayes (posterior approximation) methods has progressed naturally in the field of disease mapping. The prior distribution(s) for the ($\boldsymbol{\theta}$, say) parameters in the intensity specification $e_i\theta_i$, have hyperparameters (in the Poisson–gamma example above, these were α, β). These hyperparameters can also have hyperprior distributions. The distributions chosen for these parameters depend on the application. In the full Bayesian approach, inference is based on the posterior distribution of $\boldsymbol{\theta}$ given the data. However, as in the frequentist approach above, it is possible to adopt an intermediate approach where the posterior distribution is approximated in some way, and subsequent inference may be made via 'frequentist-style' estimation of parameters or by computing the approximated posterior distribution. In the tract-count example, approximation via intermediate prior-parameter estimation would involve the estimation of α and β, followed by inference on the estimated posterior distribution (see, for example, Carlin and Louis, 1996, pp. 67–8).

For count data, a number of examples exist where independent Poisson distributed counts (with constant within-tract rate) are associated with prior distributions of a variety of complexity. The earliest examples of such a Bayesian mapping approach can be found in Manton *et al.* (1981) and Tsutakawa (1988). Also, Clayton and Kaldor (1987) developed a Bayesian analysis of a Poisson likelihood model where y_i has expectation $e_i\theta_i$, and found that with a prior distribution given by $\theta_i \sim Gamma(\alpha, \beta)$, the Bayes estimate of θ_i is the posterior expectation:

$$\frac{y_i + \alpha}{e_i + \beta}. \tag{1.3}$$

Hence one could map directly these Bayes estimates. Now, the distribution of θ_i conditional on y_i is $Gamma(y_i + \alpha,\ e_i + \beta)$ and a Bayesian approach would

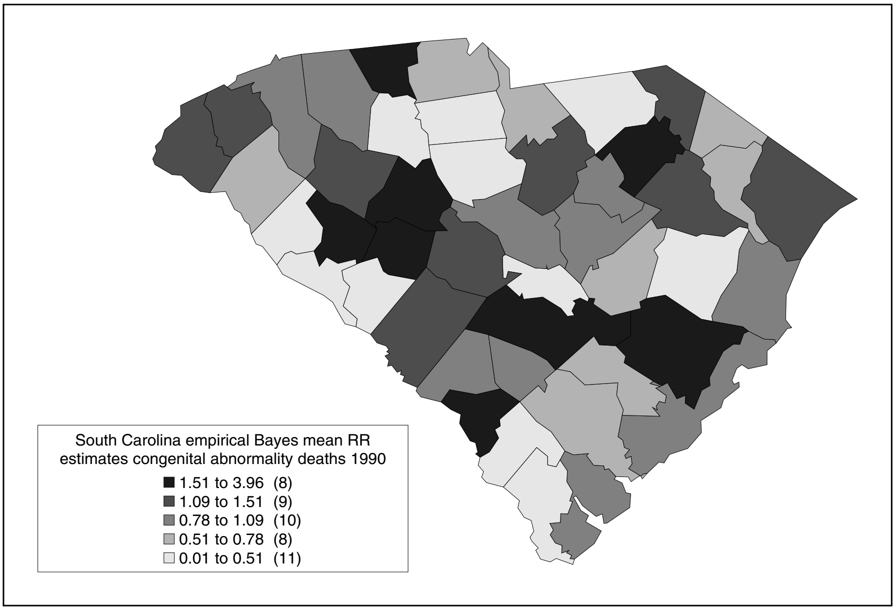

Figure 1.3 Empirical Bayes mean relative risk (RR) estimates.

require summarization of θ_i from this posterior distribution. In practice, this is often obtained by generation of realizations from this posterior and then the summarizations are empirical (e.g. Markov Chain Monte Carlo (MCMC) methods). Figure 1.3 displays the empirical Bayes estimates under the Poisson–gamma model with α and β estimated as in Clayton and Kaldor. Note that in contrast to the SMR map (Figure 1.2), Figure 1.3 presents a smoother relative risk surface.

Other approaches and variants in the analysis of simple mapping models have been proposed by Tsutakawa (1988), Marshall (1991) and Devine and Louis (1994). In the next section, more sophisticated models for the prior structure of the parameters of the map are discussed.

1.2.4 Advanced Bayesian models

Many of the models discussed above can be extended to include the specification of prior distributions for parameters and hence can be examined via Bayesian methods. In general, we distinguish here between empirical Bayes methods and full Bayes methods, on the basis that any method which seeks to approximate the posterior distribution is regarded as empirical Bayes (Bernardo and Smith, 1994). All other methods are regarded as full Bayes. This latter category includes maximum a posteriori estimation, estimation of posterior functionals, as well as posterior sampling.

1.2.4.1 Empirical Bayes methods

The methods encompassed under the definition above are wide-ranging, and here we will only discuss a subset of relevant methods. The first method considered by the earliest workers was the evaluation of simplified (constrained) posterior distributions. Manton *et al.* (1981) used a direct maximization of a constrained posterior distribution, Tsutakawa (1988) used integral approximations for posterior expectations, while Marshall (1991) used a method of moments estimator to derive shrinkage estimates. Devine and Louis (1994) further extended this method by constraining the mean and variance of the collection of estimates to equal the posterior first and second moments.

The second type of method which has been considered in the context of disease mapping is the use of likelihood approximations. Clayton and Kaldor (1987) first suggested employing a quadratic normal approximation to a Poisson likelihood, with gamma prior distribution for the intensity parameter of the Poisson distribution and a spatial correlation prior. Extensions to this approach lead to simple generalized least squares (GLS) estimators for a range of likelihoods (Lawson, 1994; 1997).

A third type is the Laplace asymptotic integral approximation, which has been applied by Breslow and Clayton (1993) to a generalized linear modelling framework in a disease mapping example. This integral approximation method allows the estimation of posterior moments and normalizing integrals (see, for example, Bernardo and Smith, 1994, pp. 340–4). A further, but different, integral approximation method is where the posterior distribution is integrated across the parameter space: that is, the nuisance parameters are 'integrated out' of the model. In that case the method of nonparametric maximum likelihood (NPML) can be employed (Bock and Aitkin, 1981; Aitkin, 1996b; Clayton and Kaldor, 1987). Another possibility is to employ Linear Bayes methods (Marshall, 1991).

1.2.4.2 Full Bayes methods

Full posterior inference for Bayesian models has now become available, largely because of the increased use of MCMC methods of posterior sampling. The first full sampler reported for a disease mapping example was a Gibbs sampler applied to a general model for intrinsic autoregression and uncorrelated heterogeneity by Besag *et al.* (1991). Subsequently, Clayton and Bernardinelli (1992), Breslow and Clayton (1993) and Bernardinelli *et al.* (1995) have adapted this approach to mapping, ecological analysis and space–time problems.

This has been facilitated by the availability of general Gibbs sampling packages such as BEAM and BUGS (GeoBUGS and WinBUGS). Such Gibbs sampling methods can be applied to putative source problems as well as mapping/ecological studies. Alternative, and more general, posterior sampling methods, such as the Metropolis–Hastings algorithm, are currently not separately available in a packaged form, although these methods can accommodate considerable variation in model specification. WinBUGS does provide such estimators when non-convex posterior distributions are encountered. Metropolis–Hastings algorithms have been applied in comparison to approximate maximum a posteriori (MAP) estimation by Lawson *et al.* (1996) and Diggle *et al.* (1998); hybrid Gibbs–Metropolis samplers have been applied to space–time problems by Waller *et al.* (1997). In addition, diagnostic methods for Bayesian MCMC sample output have been discussed for disease mapping examples by Zia *et al.* (1997). Developments in this area have been reviewed recently (Lawson *et al.*, 1999; Elliott *et al.*, 2000; Lawson, 2001).

1.2.5 Multilevel modelling approaches

An alternative to the above specification can be considered where a log-linear form is specified:

$$\theta_i = \exp\{\beta_0 + v_i\},$$

where the random term has a zero mean Gaussian distribution, i.e. $v_i \sim N(0, \sigma_v^2)$ and σ_v^2 is the variance of the random effects v.

This model may be rewritten (in terms of counts rather than rates) as

$$y_i \sim Poisson(\mu_i), \;\; log(\mu_i) = log(e_i) + \beta_0 + v_i.$$

Here $\mu_i = e_i\theta_i$ and the $log(e_i)$ are treated as known 'offset' terms.

Generally, multilevel models (see, for example, Goldstein, 1995) are fitted to data that possess levels of clustering in their structure. In disease mapping and geographical applications in general such levels would be different levels of geographical aggregation, for example, census tracts nested within counties nested within countries. For each level of geography we could then fit normally distributed random effects so for example if we had data on census tracts nested within counties we could fit

$$y_{ij} \sim Poisson(\mu_{ij}), \;\; log(\mu_{ij}) = log(e_{ij}) + \beta_0 + v_j + u_{ij},$$

where both the county and tract random effects have Gaussian distributions, i.e. $v_j \sim N(0, \sigma_v^2)$ and $u_{ij} \sim N(0, \sigma_u^2)$.

Poisson response multilevel models can be fitted using either frequentist or Bayesian approaches. Frequentist approaches generally involve some approximations, for example the software package MLwiN (Rasbash *et al.*, 2000) uses quasi-likelihood methods that involve Taylor series approximations (Goldstein, 1991; Goldstein and Rasbash, 1996) to transform the problem so that it can be fit using the iterative general least squares algorithm (IGLS; Goldstein (1986)). Other common frequentist approaches include Laplace approximations (Raudenbush *et al.*, 2000) and Gaussian quadrature (e.g. Rabe-Hesketh *et al.*, 2001). Bayesian estimation involves extending the models to include prior distributions and fitting using MCMC estimation.

Fitting higher-level random effects may account for some spatial autocorrelation in the data but multilevel models can also more directly account for spatial correlations via their extension to multiple-membership models (Hill and Goldstein, 1998). Although originally used to account for missing unit identifiers, Langford *et al.* (1999) showed how to use such models to fit spatial data. The model can be described as

$$y_i \sim Poisson(\mu_i), \; log(\mu_i) = log(e_i) + \beta_0 + v_i + \sum_{j \in neigh(i)} w_{ij} u_j,$$

where $u_i \sim N(0, \; \sigma_u^2)$ and $v_i \sim N(0, \; \sigma_v^2)$. Here we have, for each observation, an unstructured random effect v_i and a group of (weighted) 'neighbour' random effects u_j. Browne *et al.* (2001) consider Bayesian extensions of this model as a member of the family of models that they call multiple-membership multiple-classification (MMMC) models. Langford *et al.* (1999) also introduce a correlation between the two sets of random effects giving u and v a multivariate Gaussian distribution. Multilevel approaches will be discussed in greater detail in Chapter 3.