STATISTICS IN MEDICINE, VOL. 11, 983–1007 (1992)

EMPIRICAL BAYES VERSUS FULLY BAYESIAN ANALYSIS OF GEOGRAPHICAL VARIATION IN DISEASE RISK

LUISA BERNARDINELLI

Istituto di Igiene e Medicina Preventiva, Università di Sassari, Via Padre Manzella 4, 07100 Sassari, Italy

AND

CRISTINA MONTOMOLI

Istituto di Scienze Sanitarie Applicate, Università di Pavia, Via Bassi 21, 27100 Pavia, Italy

SUMMARY

This paper reviews methods for mapping geographical variation in disease incidence and mortality. Recent results in Bayesian hierarchical modelling of relative risk are discussed. Two approaches to relative risk estimation, along with the related computational procedures, are described and compared. The first is an empirical Bayes approach that uses a technique of penalized log-likelihood maximization; the second approach is fully Bayesian, and uses an innovative stochastic simulation technique called the Gibbs sampler. We chose to map geographical variation in breast cancer and Hodgkin's disease mortality as observed in all the health care districts of Sardinia, to illustrate relevant problems, methods and techniques.

1. INTRODUCTION

The mapping of geographical variation in disease occurrence plays an important role in the formulation and validation of aetiological hypotheses about diseases. The construction and smoothing of disease maps has been the object of important recent methodological developments.

The classical approach consists of mapping area-specific *standardized mortality ratios*[1] (SMRs), which are maximum likelihood (ML) estimates of *relative risk* (RR) under a Poisson model of the death counts. This approach is described in Section 2 of this paper.

The SMR-based approach has been criticized by several authors, notably Clayton and Kaldor,[2] who point out that maps of SMRs are likely to be highly affected by the 'noise' that is due to the variability of the estimates. This is particularly evident when rare diseases are investigated in small areas. Areas of low population, in which both observed and expected counts are low, tend to present extreme relative risk estimates that dominate the map. Therefore the most salient features in the map are often contributed by the least reliable data and hence do not yield useful epidemiological interpretation. Mapping area-specific p-values for the significance of the corresponding SMRs addresses this difficulty. However, maps of significance are uninformative, since extreme values are likely to represent only modest increases of risk in areas with high population.

A Bayesian approach to relative risk modelling, which was first proposed by Clayton and Kaldor,[2] avoids the instability of mapping crude SMRs and the uninformativeness of mapping p-values. Clayton and Kaldor's idea was to impose a plausible structure of spatial relatedness on

0277–6715/92/080983–25$12.50

Received June 1990
Revised November 1991

the unknown relative risks by modelling them collectively as a spatial stochastic process. In Bayesian modelling, this means regarding area-specific relative risks as drawn from a *prior* multivariate distribution, whose parameters determine such aspects as the overall level of risk and geographical interdependence between those risks. After the estimation of these parameters, such a prior distribution summarizes information from all areas of the map. Consequently, each area will receive an estimate of RR which is a compromise between its SMR and inferences from information given by all the other areas. Fluctuations in RR estimates are thus reduced and a 'smoothed' map which has a better epidemiological interpretation is produced.

The Bayesian model of the RRs proposed by Clayton and Kaldor[2] combines two sub-models: one is the Poisson model of death counts, which also justifies SMRs; the other is the multivariate prior distribution of the relative risks. These two sub-models are usually viewed as first and second 'levels' of a *Bayesian hierarchical model* of relative risk. This model is described in Section 3. To convey the 'qualitative' structure of the model, that is, the conditional independence relationships between the variables, we use a *graphical* representation in which variables in the problem are represented as nodes of a graph, and conditional independence relationships are represented by missing links. The value of graphical representations of this kind in statistical modelling has been emphasized by Darroch *et al.*,[3] Edwards and Kreiner,[4] Wermuth and Lauritzen,[5] and Whittaker.[6]

A comparison between two broad classes of prior distributions for the relative risk is the most important issue discussed in this paper. The simplest prior models are models of *exchangeability*.[2] They determine a mean value towards which individual area-specific estimates are more or less displaced; this displacement depends on the intrinsic stability of the estimates and not on area location on the map. More complex prior models[2,7] incorporate the geographical structure of the map. These models impose a conditional independence structure on the set of relative risks, whereby each relative risk is conditionally independent of all other relative risks, given a small set of geographical 'neighbours'. In other words, the relative risk estimate in a given area is strongly influenced by the estimates of geographically adjacent areas, and only indirectly influenced by the estimates of all the other areas of the map. As a result, the individual estimates are more displaced towards a *local* than towards a *global* mean value.

In Section 4 we review and compare two basic approaches to relative risk estimation by using the hierarchical Bayesian model: the *empirical Bayes*[2,8-14] and the *full Bayes*[7,15] approach. The former was expressly developed to circumvent the seeming impossibility of performing the numerical integrations required by the latter. It leads to satisfactory point estimates of the relative risks, but it does not allow precise assessment of their uncertainty. As a computational approach to the empirical Bayes estimation of relative risks, Clayton[16,17] proposed a technique that is based on penalized log-likelihood maximization (PLM). This technique is reviewed in Section 4.1.

Application of the full Bayes approach has been hampered by apparent computational problems, but recent developments in Bayesian analysis provide a means of overcoming these difficulties. Such developments revolve around a Monte Carlo technique called the *Gibbs sampler*.[18] The idea is one of sampling the joint posterior distribution of the relative risks. On the basis of information contained in the simulated sample, it is possible to investigate the full posterior distribution of the relative risks. Access to this posterior distribution enables straightforward computation of point and interval estimates, and therefore to assess uncertainty in the map surface. Application of the Gibbs sampler in disease mapping was first proposed by Clayton[15] and by Besag and Jork.[7] Use of the Gibbs sampler in our application is described in Section 4.3.

Section 5 describes an application of the methods and techniques to the study of geographical variation in breast cancer and Hodgkin's disease mortality, as observed in all the health care districts of Sardinia.

This paper does not introduce new modelling or computational approaches. Its main contributions are:

(a) to provide a common framework for the many and disjointed models and computational methods described in the literature;
(b) to exploit the clarifying power of the graphical representation of a statistical model;
(c) to provide an 'anatomy' of the methods of interest, discuss their relative merits and compare them by application to real data;
(d) to provide guidelines for selecting the appropriate prior distribution, according to the epidemiological context;
(e) to discuss results obtained when applying the very powerful and promising, but still unexplored, estimation approach that is based on Gibbs sampling.

2. CLASSICAL MODEL OF RELATIVE RISKS

We shall henceforth denote probability densities by square brackets, so that joint, conditional and marginal densities respectively appear, for example, as $[X, Y]$, $[X|Y]$ and $[X]$. Multiplication of densities will be denoted by '$\times$', so that, for example, $[X, Y] = [X|Y] \times [Y]$.

In the following, ξ_i denotes the unknown relative risk for the generic ith area, $i = 1, \ldots, N$. Available data include the observed $\mathbf{Y} = \{Y_1, \ldots, Y_N\}$ and expected death counts $\mathbf{E} = \{E_1, \ldots, E_N\}$. The expected deaths are obtained by applying age/sex specific reference death rates to the area-specific population subdivided by age and sex.

The conventional approach to disease risk mapping is based on the assumption that, conditional on the E's being known, the ξ's are mutually independent. Moreover, each Y_i is usually taken to follow a Poisson distribution with mean $E_i\xi_i$, which we shall write as:

$$[Y_i|E_i, \xi_i] \sim \text{Po}(Y_i; E_i\xi_i). \tag{1}$$

Under these assumptions, the maximum likelihood estimate (MLE) of ξ_i, denoted by $\hat{\xi}_i$, coincides with the standardized mortality ratio:

$$\hat{\xi}_i = \text{SMR}_i = Y_i/E_i. \tag{2}$$

However, as we observed in the introduction, this estimator suffers from drawbacks that determine interest in a Bayesian model for relative risks, which we outline in the following section.

3. A BAYESIAN HIERARCHICAL MODEL FOR RELATIVE RISKS

3.1. General structure of the model

The model we describe is very similar to that proposed by Clayton and Kaldor.[2] To start with, let η_i denote the unknown log-relative risk for the ith area

$$\eta_i = \log(\xi_i). \tag{3}$$

η_i can be modelled as the sum of a *global mean*, denoted by μ, that expresses the overall level of the log-relative risk throughout the map, and an *area-specific effect*, denoted by ϕ_i, that represents the difference between the log-relative risk for area i and the global mean. Therefore, for the generic ith area we may write:

$$\log(\xi_i) = \eta_i = \mu + \phi_i \tag{4}$$

and re-express (1) as:

$$[Y_i|E_i, \mu, \phi_i] \sim \text{Po}(Y_i; E_i \exp(\mu + \phi_i)). \tag{5}$$

The quantity $\exp(\mu)$ actually represents the geometric mean of the relative risks over the map. The effects ϕ_i may be viewed as a surrogate for those *unknown*, or *unobserved* variables that affect risk. If it were possible to observe these variables, some would display no geographical structure, while others would display highly geographically structured patterns of variation. If the former dominated, the true ϕ_i's would exhibit *unstructured* heterogeneity, whereas if the latter dominated, the true ϕ_i's would exhibit geographically *structured* heterogeneity, in that respective risk values for a pair of neighbouring areas will be generally more alike than for a pair of arbitrary areas.

The Bayesian approach requires that we specify a *prior* multivariate distribution for the ϕ_i's. This will mirror prior belief concerning the type of geographical heterogeneity of the ϕ_i's. Commitment to a hypothesis of unstructured heterogeneity will lead to a prior assumption that the ϕ_i's are independent and identically distributed. This assumption of itself implies an *exchangeable* prior model. On the other hand, commitment to a hypothesis of structured heterogeneity will lead to the incorporation into the prior model of a spatial correlation for the ϕ_i's. We call the above prior models *simple*, since they model either structured or unstructured heterogeneity.

When we lack a prior idea as to the type of heterogeneity of the area effects, then a *mixed* rather than simple prior model of their variation may be adopted. In the mixed model, described in Clayton and Bernardinelli,[19] an extra parameter, which is estimated from the data, chooses between the structured and unstructured types of variation. An alternative model, recently suggested by Besag and Jork,[7] represents each area effect as a sum of two random components, one following an unstructured, the other a structured, pattern of heterogeneity. Applications of the latter model to real data are described in Besag and Mollié[20] and in Clayton and Bernardinelli.[19]

Issues connected with the selection and interpretation of the various prior models have yet to be fully explored. For simplicity, in this paper we consider an approach in which the researcher only uses *simple* prior models. Prior epidemiological knowledge provides guidelines for choosing the appropriate simple prior model. For example, if we believe that the determinants of risk vary on a scale that is small by comparison with the average size of the areas, we should assume the ϕ_i's to be uncorrelated, and a model of exchangeability will be appropriate. If, however, we believe these factors to vary on a larger scale, a model of spatial correlations will be appropriate. In the absence of a strong commitment to either prior model, simple models representing extreme opinions may be applied to the same data. The results thus obtained can yield instructive comparison.

In this paper we shall deal with simple prior models in which an unknown parameter, denoted by λ (formally a scale parameter of the joint density of the area effects), represents *geographical variability*. This parameter controls the amount of variation in risk distribution throughout the map.

3.2. Graphical representation of the model

The qualitative structure of the relationships between the variables in our problem is 'crystallized' in the *directed acyclic graph* (DAG) shown in Figure 1. The nodes of this graph represent variables, or group of variables, in the model. Elliptical nodes represent random quantities, while rectangular nodes represent fixed quantities. Nodes Y_i and E_i are the *known* quantities in our model, while nodes μ, λ and $\boldsymbol{\phi}_i$ are the *unknown* quantities to be estimated.

Let us introduce some terminology concerning a DAG. If a node n_A sends an arrow to a node n_B, n_A is said to be a *parent* of n_B, and n_B is said to be a *child* of n_A. The set of parents of n_B is denoted by $pa(n_B)$. The set of nodes from which a path leads to n_B are the *predecessors* of n_B.

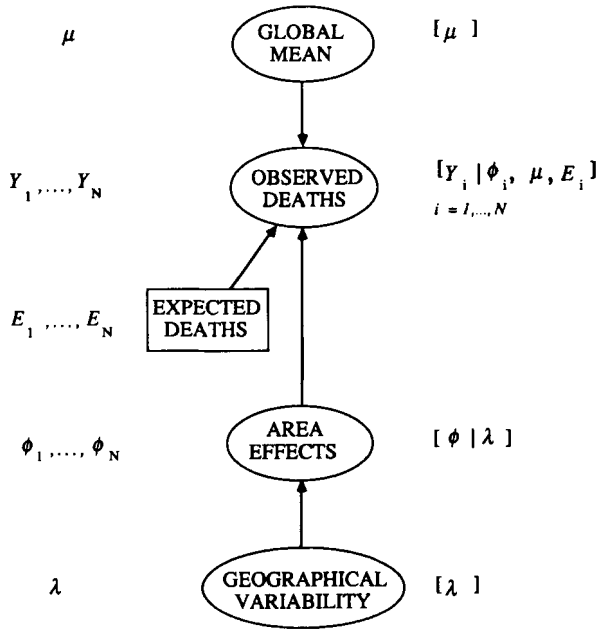


Figure 1. Directed acyclic graph (DAG) representation of the Bayesian model for relative risks

The DAG represents the assumption that each random node (variable), conditional on its parents, is probabilistically independent of all its predecessors in the graph. Thus, for example, according to Figure 1 the observed death counts **Y** are independent of the parameter λ, once we know the area-specific risk effects $\boldsymbol{\phi}$.

Under the above assumption, the joint distribution of all the random variables in the graph can be factorized into a product of terms, each term representing the conditional distribution for a random node given its parents. The DAG in Figure 1 thus represents the assumption:

$$[\mathbf{Y}, \mathbf{E}, \mu, \boldsymbol{\phi}, \lambda] = [\mathbf{Y}|\mathbf{E}, \mu, \boldsymbol{\phi}] \times [\boldsymbol{\phi}|\lambda] \times [\lambda] \times [\mu]. \tag{6}$$

From (6) it is clear that to provide a complete and consistent 'quantification' of the graphical model in Figure 1, we need to specify each of the four conditional distributions on the right hand side of (6). Let us now consider these four specifications node by node.

Node Y (observed death counts).

The generic observed death count Y_i is taken to be a Poisson variable with expectation $E_i \exp(\mu + \phi_i)$, as stated in (5). Conditional upon $\boldsymbol{\phi}$ and **E**, the Y's are mutually independent. Therefore the conditional distribution $[\mathbf{Y}|\mathbf{E}, \mu, \boldsymbol{\phi}]$, which represents the *likelihood of the data*, is given by the product:

$$[\mathbf{Y}|\mathbf{E}, \mu, \boldsymbol{\phi}] = \prod_{i=1}^{N} [Y_i|E_i, \mu, \phi_i] = \prod_{i=1}^{N} \mathrm{Po}(Y_i; E_i \exp(\mu + \phi_i)). \tag{7}$$

Node ϕ (unknown area-specific effects).

The conditional density specification $[\phi|\lambda]$ that is associated with node ϕ expresses our prior beliefs concerning the collection of the unknown area effects, viewed as a spatial process. As a convenient specification for $[\phi|\lambda]$, we choose a density from the family:

$$[\phi|\lambda] \propto \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{N} w_{i+}\phi_i\left(\phi_i - \frac{1}{w_{i+}} \sum_{j=1}^{N} w_{ij}\phi_j\right)\right] \tag{8a}$$

with

$$w_{ii} = 0, \qquad i = 1, \ldots, N \tag{8b}$$

where λ acts as a *hyperparameter*, the w_{ij}'s are prescribed non-negative *weights*, and the subscript '$i+$' denotes summation over j. Note that (8*a*) depends on the respective differences between each ϕ_i and a weighted mean of all other ϕ's. A possible choice for the 'off-diagonal' weights (w_{ij}) is to specify:

$$w_{ij} = 1 \qquad i = 1, \ldots, N, j = 1, \ldots, N, i \neq j. \tag{8c}$$

This choice induces *exchangeability* in the relative risks of different geographical areas, in the sense that given (8*c*) the right-hand side of (8*a*) remains unchanged under any permutation of $\phi_1, \phi_2, \ldots, \phi_N$. Therefore, under (8*c*), (8*a*, *b*) defines an *exchangeable prior* (EX prior), and simplifies to:

$$[\phi|\lambda]_{\mathrm{EX}} \propto \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{N} (N-1)\phi_i(\phi_i - \bar{\phi}_{-i})\right] \tag{9}$$

where $\bar{\phi}_{-i} = \dfrac{1}{(N-1)} \sum_{j \neq i} \phi_j$ represents the arithmetic mean of all the log-area effects *excluding* ϕ_i. The EX prior model (9) implies that the conditional prior distribution of the generic individual ϕ_i, given all remaining ϕ's, is Normal with mean $\bar{\phi}_{-i}$ and variance $((N-1)\lambda)^{-1}$, that is

$$[\phi_i|\phi_j, j \neq i, \lambda]_{\mathrm{EX}} \sim \mathrm{N}\left(\bar{\phi}_{-i}, \frac{1}{(N-1)\lambda}\right). \tag{10}$$

From (10) it is evident that the EX prior model tends to displace individual area effect estimates towards a *global* mean. Also, it is clear from (10) that the conditional probability of each log-area effect given all others is invariant with respect to adding a constant to all log-area effects. It follows that the EX prior density (9) is uniform with respect to the sum of the ϕ's. In Bayesian terminology, (9) is called an *improper* distribution. Later on in the text we shall discuss problems raised by this feature, and their solution by imposing appropriate constraints on the set of log-area effects.

Strict exchangeability should be relaxed when there are prior beliefs of spatial correlations in the relative risks. Spatial structure can be incorporated in (8) by choosing higher values for w_{ij} when areas i and j are geographically close (where 'close' can be defined in various ways). An extreme choice in this direction has $w_{ij} = 1$, if i and j are *contiguous* areas, and $w_{ij} = 0$, otherwise. In this case (8) yields the *conditional autoregressive prior* (CAR prior):[2, 7]

$$[\phi|\lambda]_{\mathrm{CAR}} \propto \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{N} n_i\phi_i(\phi_i - \bar{\phi}_i)\right] \tag{11}$$

where n_i represents the number of areas contiguous to the ith area, and $\bar{\phi}_i = (1/n_i) \sum_{j=1}^{N} w_{ij}\phi_j$ is the

arithmetic mean of the log-area effects for the areas which are contiguous to the ith area. The CAR prior model (11) implies that the conditional distribution of the generic individual effect ϕ_i, given all remaining ϕ's, is Normal with mean $\bar{\phi}_i$ and variance $1/\lambda n_i$, that is:

$$[\phi_i | \phi_j, j \neq i, \lambda]_{\text{CAR}} \sim \text{N}\left(\bar{\phi}_i, \frac{1}{\lambda n_i}\right). \tag{12}$$

From (12) it appears that the CAR prior model tends to displace each individual area effect estimate towards a *local* 'mean effect' $\bar{\phi}_i$. For the same reasons given in relation to the EX prior model, also the CAR prior (11) is improper.

Node λ (unknown geographical variability).

From (10) and (12) it appears that λ is inversely related to prior variance for the generic log-area effect ϕ_i. For a complete quantification of the graphical model in Figure 1, we need to specify a prior for λ, denoted by $[\lambda]$. We choose the chi-square distribution, since it preserves conjugacy with regard to the Normal conditional distribution $[\boldsymbol{\phi}|\lambda]$. More precisely, for a given positive number λ^*, we assume that the quantity $\nu\lambda/\lambda^*$ is distributed as χ^2 with ν degrees of freedom, so that

$$[\lambda] \sim \frac{\chi^2_\nu \lambda^*}{\nu}. \tag{13}$$

Since the expected value for the χ^2_ν distribution is ν, according to (13) $[\lambda]$ has the expected value λ^*. Thus, λ^* represents a prior guess for λ. Moreover, from (13) it follows that $\text{var}(\lambda) = 2\lambda^{*2}/\nu$, from which it appears that increasing ν leads to a stronger prior around λ^*. In other words, the parameter ν represents the degree of confidence in the guess λ^*.

The guess λ^* depends on whether we choose an EX or a CAR prior model. First we consider the EX model where it follows from (10) that $((N-1)\lambda)^{-1}$ is the prior variance of the (normal) distribution of each ϕ_i, and hence of each log-relative risk η_i. In the absence of prior beliefs that risks are higher or lower than the reference, it is reasonable to assume that the prior density of the ith log-relative risk, $[\eta_i]_{\text{EX}}$, is symmetric around 0. Let the interval $(-a, +a)$ be defined as containing 95 per cent of the area under $[\eta_i]_{\text{EX}}$. The idea is to guess a in such a way that the interval (e^{-a}, e^a) is reasonably consistent with our prior perception of a 95 per cent plausibility interval for the generic relative risk $\xi_i = \exp\{\eta_i\}$. In these circumstances, $(0{\cdot}5a)^2$ is a reasonable estimate of the prior variance of $[\eta_i]_{\text{EX}}$, and $\lambda^*_{\text{EX}} = \{(N-1)(0{\cdot}5a)^2\}^{-1}$ is a reasonable guess for λ under the EX prior model.

The expression (12) suggests that under the CAR model a reasonable guess λ^*_{CAR} for λ is given by λ^*_{EX} divided by an estimate of the average value for n_i. For example, if we assume that, on average, an area has 5 adjacent areas, the above rule yields $\lambda^*_{\text{CAR}} = \lambda^*_{\text{EX}}/5$.

Node μ (unknown global mean log-relative risk).

Because it is reasonable to assume that our ignorance of μ is total, we assign μ a vague *improper* prior, namely a uniform density:

$$[\mu] \sim U(-\infty, +\infty). \tag{14}$$

We conclude our description of the Bayesian model by pointing out that it can be extended easily to include the effect upon the generic ith risk of an *observed* covariate vector z_i. This effect may be incorporated by changing (4) to:

$$\eta_i = \mu + \phi_i + \beta z_i \tag{15}$$

and by providing a suitable prior for the unknown coefficients β. However, nothing essential is lost in the subsequent discussion if we ignore the term βz_i.

4. INFERENCE ABOUT THE RELATIVE RISKS: THE EMPIRICAL BAYES APPROACH AND THE FULL BAYES APPROACH

4.1. Empirical versus fully Bayesian analysis

In this section we discuss inference about the unknown risks, through the application to observed data of the previous model. Essentially, we wish to combine our 'prior' belief about the risks, which is embodied in the priors $[\boldsymbol{\phi}|\lambda]$, $[\lambda]$ and $[\mu]$, with the 'new' information contained in the data $\{\mathbf{Y}, \mathbf{E}\}$. This is a familiar operation within a Bayesian statistical framework: the rules of Bayesian analysis allow us to compute a *posterior* distribution $[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}]$ from the above priors and data. This distribution describes our prior belief about the risks as updated in the light of the data. On the basis of this posterior, it is possible to perform any desired inference about the unknown log-relative risks $\eta_i = \phi_i + \mu$.

The conditionals (7), (8), (13), (14), 'stored' in the nodes of the graphical model in Figure 1, specify the full joint posterior $[\boldsymbol{\phi}, \lambda, \mu|\mathbf{Y}, \mathbf{E}]$ of the unknown quantities in the model. More precisely, on the basis of the conditional independence structure implicit in Figure 1, this distribution may be factorized into the product:

$$[\boldsymbol{\phi}, \lambda, \mu|\mathbf{Y}, \mathbf{E}] = [\mathbf{Y}|\mathbf{E}, \boldsymbol{\phi}, \mu] \times [\boldsymbol{\phi}|\lambda] \times [\lambda] \times [\mu]. \tag{16}$$

In a *full Bayes* (FB) analysis, the computation of the desired marginal posterior $[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}]$ involves integrating (16) over λ, which is thus regarded as a nuisance parameter. Unfortunately, this integration is analytically intractable.

The seeming impossibility of performing this integration has stimulated interest in an *empirical Bayes* (EB) approach to the same estimation problem. EB inference about the unknown log-relative risks $\eta_i = \phi_i + \mu$, is not based on the true posterior $[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}]$, but on its approximation $[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}, \hat{\lambda}]$, which is conditional on a 'suitable' estimate $\hat{\lambda}$ of the parameter λ. According to the conditional independence structure implicit in Figure 1, $[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}, \hat{\lambda}]$ factorizes into a product of known terms: a normalization constant c, the likelihood of the data $[\mathbf{Y}/\mathbf{E}, \boldsymbol{\phi}, \mu]$, the prior of $\boldsymbol{\phi}$ given $\lambda = \hat{\lambda}$, and the prior of μ. By virtue of the assumption (14), this latter can be safely omitted from the product, so that we may write:

$$[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}, \hat{\lambda}] = c \times [\mathbf{Y}|\mathbf{E}, \boldsymbol{\phi}, \mu] \times [\boldsymbol{\phi}|\hat{\lambda}]. \tag{17}$$

The assumption behind the EB approach is that the location of $[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}, \hat{\lambda}]$ is a good approximation to the location of $[\boldsymbol{\phi}, \mu|\mathbf{Y}, \mathbf{E}]$. In practice, it is convenient to select the *mode* of (17) as a location estimate for $\boldsymbol{\phi}$ and μ, on the grounds that this mode is very close to the posterior mean, but easier to calculate. Hence, an EB analysis involves: (a) selecting a suitable value of $\hat{\lambda}$ to place in (17), and (b) computing the values $\hat{\boldsymbol{\phi}}$ and $\hat{\mu}$ that maximize (17). From these values point estimates $\hat{\eta}_i = \hat{\phi}_i + \hat{\mu}$, $i = 1, \ldots, N$, are derived for the log-relative risks. As we argue in the next section, this can be performed via *penalized likelihood* maximization.

The EB approach has a serious weakness. Since it conditions on a specific value of the hyperparameter, the uncertainty affecting the latter is not carried over into the estimation of the ϕ's and of μ. Hence, the obtained confidence intervals for the relative risks are too narrow. This prevents a realistic assessment of the uncertainty in the map surface. Laird and Louis[21] suggested that confidence intervals can be constructed within an EB framework by using bootstrap

methods. However, the ability of these methods to reflect fully the relevant uncertainties is in doubt.[15]

The pitfalls in the EB approach should prompt reconsideration of a full Bayes approach and an attempt to overcome the technical problems connected with the integration of the joint posterior (16). Perhaps the most attractive solution to such problems is supplied by recent innovative ideas based on sampling and resampling strategies. Clayton[15] proposed use of a Monte Carlo technique, called the Gibbs sampler, to generate samples from the posterior $[\boldsymbol{\phi}, \lambda, \mu | \mathbf{Y}, \mathbf{E}]$. From the obtained samples, summaries of the posterior distribution are computed for any desired function of $\{\boldsymbol{\phi}, \lambda, \mu\}$. For example, we may compute the posterior mean of ξ_i, where $\xi_i = \log(\phi_i + \mu)$, as a point estimate of the RR in area i, for $i = 1, \ldots, N$. More important, the Bayesian credible interval for ξ_i may also be obtained. This interval provides an assessment of the uncertainty of the estimate of ξ_i, which cannot be obtained simply by using PL maximization. The computational aspects of penalized likelihood maximization and of the Gibbs sampler are separately discussed in the next two sections.

4.2. Empirical Bayes inference via penalized likelihood maximization

As the previous section points out, in an empirical Bayes analysis of our model the estimates of $\boldsymbol{\phi}$ and μ are obtained by maximizing the posterior distribution $[\boldsymbol{\phi}, \mu | \mathbf{Y}, \mathbf{E}, \hat{\lambda}]$ with respect to these variables, where $\hat{\lambda}$ is a 'suitable' estimate of λ. Maximizing this posterior distribution is equivalent to maximizing its logarithm, which in turn, and by virtue of (17), is equivalent to maximizing the function:

$$l^*(\boldsymbol{\phi}, \mu) = \log[\mathbf{Y} | \mathbf{E}, \boldsymbol{\phi}, \mu] + \log[\boldsymbol{\phi} | \hat{\lambda}]. \tag{18}$$

Now, the first term on the right hand side of (18) for fixed $\mathbf{Y}$ and $\mathbf{E}$ is the log-likelihood of the data, hereafter abbreviated by $l(\boldsymbol{\phi}, \mu)$, while the second term may be interpreted, as we argue later, as a 'penalty function' that penalizes departures of $\boldsymbol{\phi}$ from the prior model. Therefore, as Clayton[16] points out, the quantity described in (18) may be interpreted as a *penalized log-likelihood.*[22]

The form of the penalty function $\log[\boldsymbol{\phi} | \lambda]$ depends on the prior model. If we adopt the EX prior model described by (9), then the penalized log-likelihood (18) becomes:

$$l^*(\boldsymbol{\phi}, \mu) = l(\boldsymbol{\phi}, \mu) - \frac{\hat{\lambda}(N-1)}{2} \sum_{i=1}^{N} \phi_i(\phi_i - \bar{\phi}_{-i}). \tag{19}$$

Expression (19) penalizes highly any log-area effect that differs greatly from the average of all others. There is a difficulty connected with the fact that $l^*(\boldsymbol{\phi}, \mu)$ is invariant with respect to adding an identical constant to all ϕ's and subtracting it to μ. Indeed, the penalized likelihood is flat along a linear subspace of the $(\boldsymbol{\phi}, \mu)$-space. This generates problems in the maximization of (19) with respect to $\boldsymbol{\phi}$ and μ, unless appropriate constraints on the solution are adopted. In practice, in order to avoid the mentioned ill-conditioning we adopt the constraint $\bar{\phi} = 0$, where $\bar{\phi}$ denotes arithmetic mean of all ϕ's.

If we adopt the CAR prior model described by (11), (18) becomes:

$$l^*(\boldsymbol{\phi}, \mu) = l(\boldsymbol{\phi}, \mu) - \frac{\hat{\lambda} n_i}{2} \sum_{i=1}^{N} \phi_i(\phi_i - \bar{\phi}_i). \tag{20}$$

According to (20), those log-area effects that differ greatly from their respective neighbours receive maximum penalty. In summary, under the EX prior the penalty term causes the estimates of the relative risks to 'shrink' towards the *overall* mean, while under the CAR prior the penalty causes

these estimates to shrink towards a *local* mean. In both cases, a 'smoothing' effect is obtained. Ill-conditioning of the maximization problem arises for the same reasons of the EX prior case. Again, imposing the constraint $\bar{\phi} = 0$ upon the solution will obviate the problem.

In (19) and (20), $\hat{\lambda}$ acts as a *smoothing parameter*. When it is large, the penalty function is given relatively more weight, leading to 'smoother' estimates of the effects $\{\phi_i\}$; when this parameter is equal to 0, the penalty is given no weight, which leads to the ML estimates of the effects.

The problems that remain to be described are the maximization of the penalized likelihood and the selection of a suitable value $\hat{\lambda}$ for λ. The former is only a problem in non-linear function maximization, and may be solved with various maximization methods.[16] The latter can be solved by maximizing the marginal likelihood for λ, for which an approximate expression is available (see Clayton[16] and Ogata and Katsura[23] for details).

4.3. Full Bayes inference using Gibbs sampling on the graph

In this section we show that a fully Bayesian analysis of the hierarchical model discussed in this paper can be implemented straightforwardly using an iterative Monte Carlo algorithm, called the *Gibbs sampler*, or briefly *Gibbser*. This provides a complete alternative to the empirical Bayes approach outlined in the previous section. A key reference for Gibbs sampling is Geman and Geman,[18] who discuss application of this technique to image analysis. A thorough review of the method is given by Gelfand and Smith.[24]

The Gibbs sampler uses the graphical representation of the model as the computational framework for the necessary calculations. Before we illustrate the algorithm, the graphical representation of the model in Figure 1 needs to be articulated in more detail as well as enriched by *undirected* links.

Figure 2 represents graphically the conditional independence structure of the variables of our problem under the assumption of exchangeability. This graph derives from that in Figure 1. The node 'AREA EFFECTS' has been expanded into a collection of nodes $\{\phi_i\}$, each of which represents an individual area-specific effect. The nodes that respectively represent observed and expected deaths are treated with an analogous division procedure in order that they become area-specific nodes. For simplicity, in Figure 2 we assume $N = 3$. Direct dependencies between the ϕ's are here represented as undirected links, each ϕ_i being, in this particular case, directly connected to each other $\phi_j, j \neq i$. This is consistent with the EX prior model (10), according to which each area effect depends directly on all others. Under the CAR prior model the dependencies between the ϕ's are more sparse, which reduces the degree of connectivity in the undirected portion of the graph.

Any couple of nodes at the extremes of an undirected link are said to be *brothers*. For any given node α in the graph, the 'parents', the 'children', the parents of the children and the brothers of α jointly form the so called *Markov boundary* of α. It can be demonstrated that the variables in this set render α independent from the rest of the graph.

Nodes $\{E_i\}$ and $\{Y_i\}$ represent *known* quantities. We call these *fixed* nodes. The remaining quantities, namely the effects $\{\phi_i\}$, μ and λ, are unknown, and are called *free* nodes.

Each of the nodes $\{Y_i\}$, $\{\phi_i\}$, λ and μ may be viewed as a 'submodel' that expresses the conditional distribution of the variable it represents given its parent nodes and brother nodes, if any. These conditional distributions, called *directional conditional distributions*, have already been expressed by (5), (10), (13) and (14). Moreover, for each free node α, $\alpha \in \{\phi_1, \ldots, \phi_N, \mu, \lambda\}$ the *full conditional distribution*, denoted as $[\alpha|.]$, is defined as the conditional distribution of α given *all* the remaining nodes in the graph. It may be demonstrated that the following proportionality

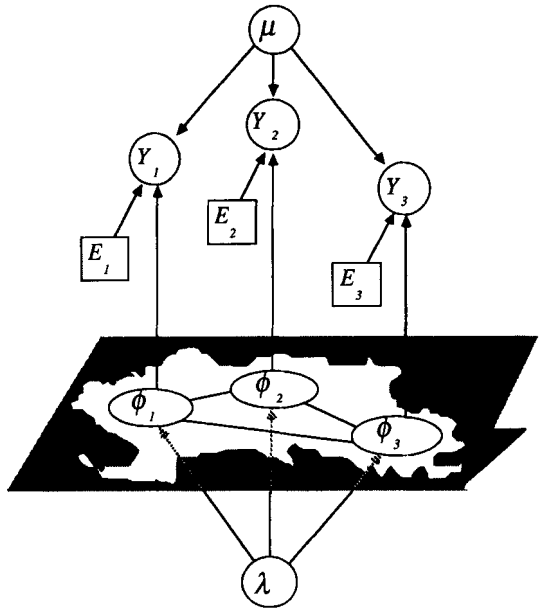


Figure 2. Graphical representation of the Bayesian model for relative risks assuming an exchangeable prior (EX) model

holds:

$$[\alpha|.] \propto [\alpha|\text{parents \& brothers of } \alpha] \times [\text{children of } \alpha|\text{parents of the children of } \alpha] \quad (21)$$

from which it appears that the full conditional distribution of α involves the rest of the graph only through the set of nodes included in the Markov boundary of α. Application of the above rule to the graphical model in Figure 2 leads to the following full conditionals for the free nodes (assuming an EX prior model):

$$[\lambda|.] \propto [\lambda] \times [\boldsymbol{\phi}|\lambda]_{EX} \quad (22)$$

$$[\phi_i|.] \propto [\phi_i|\phi_j, j \neq i, \lambda]_{EX} \times [Y_i|E_i, \mu, \phi_i] \quad i = 1, \ldots, N \quad (23)$$

$$[\mu|.] \propto [\mu] \times \prod_{i=1}^{N} [Y_i|E_i, \mu, \phi_i]. \quad (24)$$

The directional conditionals $[\boldsymbol{\phi}|\lambda]_{EX}$, $[Y_i|E_i, \mu, \phi_i]$, $[\phi_i|\phi_j, j \neq i, \lambda]_{EX}$, $[\lambda]$ and $[\mu]$, which are necessary to compute (22)–(24), are given by (5), (9), (10), (13) and (14).

The Gibbser enables us, by actually sampling the univariate full conditional distributions of the free nodes, to obtain samples from their multivariate joint posterior distribution. Thus, in our application, sampling the full conditional distributions (22)–(24), produces samples from the multivariate joint posterior distribution:

$$[\phi_1, \ldots, \phi_N, \mu, \lambda | E_1, \ldots, E_N, Y_1, \ldots, Y_N]. \quad (25)$$

To perform the Gibbs sampling, initial values are assigned to each free node. Then, node-by-node, the current value for each free node α is replaced by a value drawn from the corresponding

full conditional distribution. When all the free nodes have been 'visited', one cycle of the Gibbser is completed, and a new one begins. Many cycles are carried out. Geman and Geman[18] show that, under certain regularity conditions, this resampling process is a Markov chain that converges to an equilibrium distribution given by the joint posterior distribution of the free nodes.

In our case, a single Gibbser cycle consists of sampling the densities (22)–(24) (the missing proportionality constants do not affect the sampling). The resampling process asymptotically generates (correlated) samples from (25). From these it is straightforward to calculate any posterior summaries for both (25) and for this distribution's marginal components.

Let us now detail the generic iteration of the sampling on a node-by-node basis. As node λ is visited, the distribution (22), conditional on the current value of $\boldsymbol{\phi}$, is sampled. Our previous distributional choices (10) and (13) induce conjugacy between the product terms of (22). Hence (22) remains within the family of distributions to whicn $[\lambda]$ also belongs, namely the gamma distribution. A standard routine can be used to sample from this distribution.

The sampling distribution at the generic ϕ_i node (under the EX prior model) is obtained by substituting (5) and (10) into (23). We obtain:

$$[\phi_i|.] \propto \mathrm{N}\left(\bar{\phi}_{-i}, \frac{1}{(N-1)}\right) \times \mathrm{Po}(Y_i; E_i \exp(\mu + \phi_i)). \tag{26}$$

Every time the Gibbser visits ϕ_i, a new prior conditional mean $\bar{\phi}_{-i}$ is computed by taking the average of the current values for $\{\phi_j, j \neq i\}$. Such a prior mean, as well as the current values of μ, and λ, are replaced in (26) to obtain (up to a constant) the density from which the new value for ϕ_i is to be sampled. This density is not standard, since conjugacy does not hold between the terms in the right-hand-side of (26). To sample it, we used a technique called the *weighted bootstrap*, which was proposed by Smith and Gelfand.[25] However, (26) is clearly log-concave with respect to ϕ_i. In such a case, the adaptive rejection sampling algorithm proposed by Gilks and Wild[26] should offer a better and more efficient sampling procedure.

We have used the weighted bootstrap algorithm also for sampling the full conditional for the μ node.

It should be pointed out that the presence of linear dependencies within the $(\mu, \boldsymbol{\phi})$ vector raises difficulties that are rather hard to deal with in a fully Bayesian framework. As noted by Clayton,[27] there are linear transformations of $(\mu, \boldsymbol{\phi})$ that do not affect the posterior probability for these parameters, both under an EX and under a CAR prior. For example, under the EX model, the prior density $[\mu, \boldsymbol{\phi}|\lambda]_{\mathrm{EX}} = [\boldsymbol{\phi}|\lambda]_{\mathrm{EX}}[\mu]$ is invariant with respect to a transformation that reduces each ϕ_i and increments μ by the same constant. Nor does this transformation affect the likelihood $[\mathbf{Y}|\mathbf{E}, \mu, \boldsymbol{\phi}]$, for it leaves the log-relative risks $\eta_i = \mu + \phi_i$ unchanged. Thus the vector $(\mu, \boldsymbol{\phi})$ is free to vary within a subspace without any consequent change in its posterior distribution, from which indeterminacy of the solution follows.

The natural cure for this problem is to apply a constraint to the vector $(\mu, \boldsymbol{\phi})$. At the end of each Gibbser cycle, it is simply necessary to adjust the sampled values of $(\mu, \boldsymbol{\phi})$ in such a way that the η's remain unchanged, but the constraint is satisfied. Under the EX prior model one possibility is to adopt the linear constraint $\Sigma\phi_i = 0$. This constraint may be enforced, at the end of each Gibbser cycle, by reducing each of the ϕ's by their overall mean and by incrementing μ by the same amount. An identical constraint is appropriate under the CAR-prior model.

Following the resampling process, there is a 'summarization' phase, in which calculations are performed upon the generated samples to obtain the desired estimates. We estimate the posterior mean $\hat{\eta}_i$ of the ith log-relative risk by taking the average over the repeated samples $\{\eta_i^{(0)}, \ldots, \eta_i^{(L)}\}$ generated by the Gibbser, where L denotes the number of cycles performed, and $\eta_i^{(j)}$ is calculated

from $\eta_i^{(j)} = \mu^{(j)} + \phi_i^{(j)}$. Other statistical summaries of the posterior distribution for η_i, such as its standard deviation sd_i, may be computed from the series $\{\eta_i^{(0)}, \ldots, \eta_i^{(L)}\}$. For example, a 95 per cent Bayesian credible interval for the true η_i is delimited by the 2·5th and 97·5th percentiles of the empirical distribution defined by $\{\eta_i^{(0)}, \ldots, \eta_i^{(L)}\}$.

Moreover, for each i, $i = 1, \ldots, N$, we compute the proportion of Gibbser cycles $j = 1, \ldots, L$ in which $\eta_i^{(j)}$ is greater than the global posterior mean given by:

$$\frac{1}{N \times L} \sum_{i=1}^{N} \sum_{j=1}^{L} \eta_i^{(j)}. \tag{27}$$

This proportion, may be considered as an analogue of the significance test on the evidence that the ith risk is greater or less than the mean relative risk.[28]

Finally, at each jth cycle of the Gibbser, we rank the $\{\eta_1^{(j)}, \ldots, \eta_N^{(j)}\}$ so as to obtain cycle-specific ranks $\{r_1^{(j)}, \ldots, r_N^{(j)}\}$ for the N log-relative risks. Then, for each ith area we compute area-specific percentiles $Q_{i,5}, \ldots, Q_{i,95}$ from the series $\{r_i^{(0)}, \ldots, r_i^{(L)}\}$. These percentiles may be thought of as defining a 90 per cent Bayesian credible interval for the rank of the ith area.

5. BREAST CANCER AND HODGKIN'S DISEASE

To illustrate the methods described in this paper, we applied them to, and compared the results obtained from, a study of geographical variation in cancer mortality as observed in all the health care districts (USLs = Unita' Sanitarie Locali) of Sardinia. Breast cancer (BC) and Hodgkin's disease (HD) were analysed. Before considering the specific data available, we give, in the following section, a brief epidemiological background for the two diseases involved.

5.1. Brief epidemiological background

It has been suggested that the risk of BC is modulated by two main factors: dietary practice and reproductive history.[29,30] Other factors, such as age at first marriage for women, living in an urban area, family income, and degree of education, also seem to be involved.[31] Since these factors are likely to vary on a scale which is larger than that represented by the average USL, we expected mortality from BC in Sardinia to form geographical clusters involving adjacent USLs.

Our expectations concerning the pattern of geographical variation in HD, however, were markedly different. The presence of HD clusters on a very small, typically family, scale[32,33] had already been reported in the literature. Moreover, it had been hypothesized that HD is transmissable.[34] These findings suggested the presence of risk variation on a considerably smaller scale than that represented by the average USL. The literature offered no evidence, or even suggestion, of variation in HD incidence on a larger scale. It was therefore predictable that we would not find a marked correlation between HD death rates in adjacent USLs.

The above considerations suggest the CAR model as the more appropriate prior for analysing BC data, and the EX model as the more appropriate for analysing HD data. Nevertheless, we carried out multiple analyses, so that each set of data was analysed under both prior models. In this way, we were able to evaluate agreement between the results obtained under the two prior models.

5.2. The data

Sardinia is divided into 22 USLs (Unità Sanitarie Locali), or health care districts. The data analysed include yearly deaths from (a) breast cancer and (b) Hodgkin's disease, as observed in

each of these districts during the 1983–1985 period. The individual mortality records and the 1981 Sardinia residence population, the latter subdivided according to age, sex and geographic area, were provided by the Italian Central Institute of Statistics (ISTAT). The causes of death were coded according to the IXth International Classification of Disease (ICD IX).

Italian age/sex specific mortality rates (calculated on the basis of five year age groups) for 1981 were applied to area-specific populations in the various age/sex categories to compute the expected number of yearly deaths (E_i) in each area i. Age was truncated at 75 years.

5.3. Results

Breast cancer

The data and the results for the breast cancer analysis are summarized for each USL in Table I. This table has four main sections. The first reports the raw data for each USL, including the observed and the expected deaths during the period 1983–1985. The remaining main sections report relative risk (RR) estimates and summary statistics obtained by the three estimation approaches considered in this paper, the ML approach, the empirical Bayes (EB) approach and the full Bayes (FB) approach. For each of the Bayesian approaches, the table reports both the results obtained under the exchangeable prior (EX) model and those obtained under the conditional autoregressive (CAR) model. For each relative risk estimate, we report, wherever obtainable, the 95 per cent credible interval (denoted by 95 per cent C.I.), the probability (denoted by $P_{>\text{mean}}$) that the relative risk is greater than its global mean value, and the 5th, 50th and 95th percentiles (respectively, denoted by Q_5, Q_{50} and Q_{95}) of the relative risk rank distribution. The FB estimates and tests were obtained by running the Gibbser until convergence (for about 500 cycles), and then by collecting information from a further 5000 cycles, of which we stored every sample for the subsequent computation of the point estimates and summary statistics.

The expected number of deaths varies widely according to area, which reflects high variation in the area population sizes. SMRs range from 0·63 to 1·45, but much of this variation is Poisson. The Bayesian estimates of the relative risks correct this dispersion, since the pooled EB and FB estimates range from 0·80 to 1·21.

Figure 3 (*a*)–(*e*) provide further insight into the differences between the estimates obtained from breast cancer data by the various methods. The maps reported in this figure were constructed as follows: the five series of estimates (corresponding to the different estimation methods) were pooled into a single series. The obvious outliers were removed from this series and the range of the remaining RR values was divided into seven classes of equal size. The class limits were so adjusted that clear values would obtain for the cut-off points, and a grey level was associated with each class. These map construction measures provide the five maps with a common set of cut-off points, thus making the maps more comparable.

Figure 3(*a*) displays the SMRs; Figures 3(*b*) and (*c*) display EB estimates obtained under the EX and CAR models. Figures 3(*d*) and (*e*) display FB estimates obtained under the EX and CAR models.

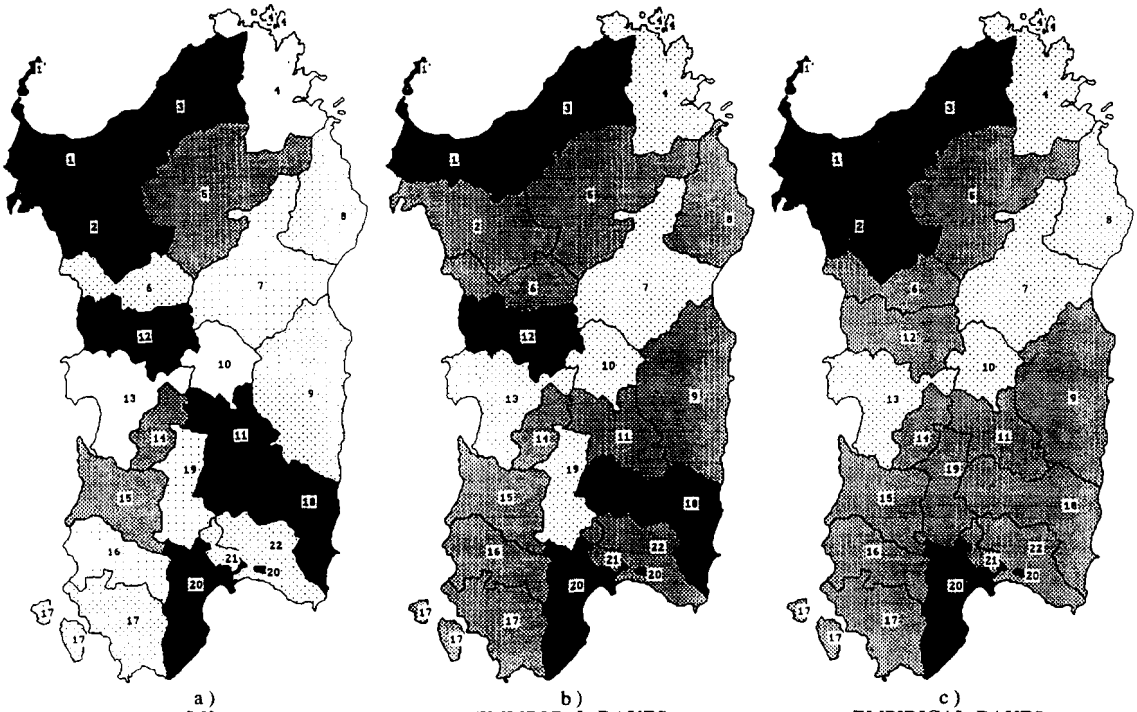
The tendency of the Bayesian estimates to be pulled towards a collective mean results in 'smoother' maps than those provided by ML estimates. Note that the Bayesian maps have no 'black' or 'white' areas. It is also interesting to note that FB are less smoothed than the EB estimates.

The maps highlight contrasts in the geographical distribution of risk; the north-western and the south-eastern parts of Sardinia appear at higher risk. This clustering pattern is fairly stable in all the maps, although, as expected, it emerges more clearly in the maps obtained under the CAR

Table I. Breast cancer mortality, Sardinia 1983–1985. Maximum likelihood (ML), empirical Bayes (EB), full Bayes (FB) estimates of relative risks

	Data		ML		Empirical Bayes		Full Bayes											
							EX model						CAR model					
USL code	Observed deaths	Expected deaths	SMR	95 per cent C.I.	EX model RR	CAR model RR	RR	95 per cent C.I.	$P_{>mean}$	Q_5	Q_{50}	Q_{95}	RR	95 per cent C.I.	$P_{>mean}$	Q_5	Q_{50}	Q_{95}
1	67	54·36	1·23	(0·96–1·56)	1·15	1·18	1·14	(0·93–1·41)	0·91	9	18	22	1·20	(0·97–1·50)	0·97	12	19	22
2	30	25.52	1·17	(0·79–1·68)	1·08	1·12	1·07	(0·82–1·41)	0·71	5	15	21	1·13	(0·83–1·53)	0·82	7	17	22
3	22	15·51	1·42	(0·89–2·14)	1·15	1·15	1.16	(0·84–1·59)	0·83	7	18	22	1·21	(0·85–1·74)	0·88	8	19	22
4	13	20·58	0·63	(0·34–1·08)	0·87	0·90	0·83	(0·61–1·14)	0·13	1	4	14	0·80	(0·56–1·16)	0·15	1	5	15
5	17	15·63	1·09	(0·63–1·74)	1·03	1·05	1·03	(0·75–1·43)	0·59	3	13	22	1·04	(0·77–1·41)	0·67	5	14	21
6	12	13·04	0·92	(0·47–1·61)	0·98	1·00	0·97	(0·69–1·37)	0·45	2	10	21	0·97	(0·66–1·42)	0·50	2	11	21
7	21	26·66	0·79	(0·49–1·20)	0·91	0·93	0·89	(0·66–1·20)	0·24	1	7	17	0·85	(0·64–1·14)	0·18	2	6	15
8	8	9·34	0·86	(0·37–1·68)	0·98	0·90	0·95	(0·67–1·36)	0·42	1	10	20	0·84	(0·50–1·40)	0·28	1	6	20
9	15	17·51	0·86	(0·48–1·41)	0·96	0·96	0·85	(0·61–1·19)	0·19	1	5	16	0·81	(0·56–1·16)	0·15	1	5	15
10	5	7·66	0·65	(0·21–1·52)	0·94	0·93	0·92	(0·63–1·33)	0·34	1	8	20	0·84	(0·57–1·24)	0·21	1	6	17
11	12	8·58	1·40	(0·72–2·45)	1·09	1·03	1·09	(0·78–1·54)	0·71	4	15	22	1·05	(0·72–1·54)	0·66	4	14	22
12	19	14·26	1·33	(0·80–2·08)	1·11	1·02	1·12	(0·82–1·53)	0·77	5	17	22	1·06	(0·74–1·51)	0·67	4	14	22
13	20	28·66	0·70	(0·43–1·07)	0·87	0·88	0·85	(0·62–1·16)	0·17	1	5	15	0·81	(0·58–1·13)	0·12	1	5	15
14	5	5·28	0·95	(0·31–2·21)	1·00	0·97	0·99	(0·67–1·44)	0·48	2	11	21	0·93	(0·61–1·42)	0·41	1	10	21
15	19	17·45	1·09	(0·65–1·70)	1·04	0·98	1·03	(0·76–1·39)	0·59	3	13	21	0·99	(0·70–1·39)	0·55	3	12	20
16	13	15·55	0·84	(0·44–1·43)	0·95	0·97	0·93	(0·66–1·30)	0·37	1	8	19	0·92	(0·63–1·35)	0·40	1	9	20
17	24	26·18	0·92	(0·59–1·36)	0·97	0·98	0·95	(0·71–1·27)	0·39	2	9	19	0·95	(0·68–1·33)	0·44	2	10	20
18	20	13·80	1·45	(0·88–2·23)	1·15	1·09	1·16	(0·84–1·60)	0·82	7	18	22	1·16	(0·81–1·65)	0·84	7	18	22
19	14	18·04	0·78	(0·42–1·30)	0·93	0·99	0·92	(0·66–1·29)	0·33	1	8	19	0·95	(0·69–1·32)	0·46	3	10	19
20	102	84·63	1·21	(1·00–1·47)	1·15	1·12	1·16	(0·97–1·39)	0·96	11	18	22	1·16	(0·97–1·40)	0·97	12	18	22
21	8	8·87	0·90	(0·39–1·78)	0·99	1·01	0·95	(0·66–1·37)	0·42	1	9	21	0·98	(0·63–1·52)	0·52	2	12	22
22	17	18·12	0·94	(0·55–1·50)	0·98	1·00	0·97	(0·71–1·32)	0·46	2	10	20	0·97	(0·66–1·42)	0·50	2	11	21

Exchangeable prior model (EX), conditional autoregressive model (CAR), standardized mortality ratio (SMR), relative risk (RR), probability that relative risk greater than its global mean value ($P_{>mean}$), 5 per cent (Q_5), 50 per cent (Q_{50}) and 95 per cent (Q_{95}) percentiles of the relative risk distribution



a)
ML

b)
EMPIRICAL BAYES
EX MODEL

c)
EMPIRICAL BAYES
CAR MODEL

Figure 3. (*a*)–(*c*)

model. Agreement between the maps increases our confidence that this cluster is 'real', rather than simply a product of random variation. Moreover, the presence of clusters of risk is consistent with the expectations in Section 5.1.

Let us now consider the summary statistics associated with the individual RR estimates, as shown in Table I. These statistics, obtained as a product of Gibbs sampling, allow us to assess the uncertainty in the estimates and to perform significance tests.

An inspection of the C.I.s reveals that none of the full Bayes RR estimates differs significantly from 1, that is, no significant deviations from the Italian population risk were detected. Nevertheless, the C.I.s of some north-western and south-eastern areas (1, 2, 3, 18 and 20) shift markedly towards values that are greater than 1. This phenomenon is more evident under the CAR than under the EX model.

Moreover, summary statistics allow us to test whether a given area-specific RR is greater or less than a specific reference value, for example the mean regional value (MRV). Consider, for example, the map of full Bayes CAR estimates displayed in Figure 4(*a*). This map differs from Figure 3(*e*) only in that new cut-off values specifically computed from this series of estimates have been used. We complemented this map with a map of the corresponding $P_{>\text{mean}}$ values (Figure 4(*b*)), which displays the posterior probability that each area's RR will exceed the MRV.

As a key to the interpretation of the $P_{>\text{mean}}$ map of Figure 4(*b*), we may consider that there is a 'high posterior probability' that the RR of each black area is higher than the MRV; there is an 'indication' that the RR of each slightly lighter area is higher than the MRV; there is 'equivocal

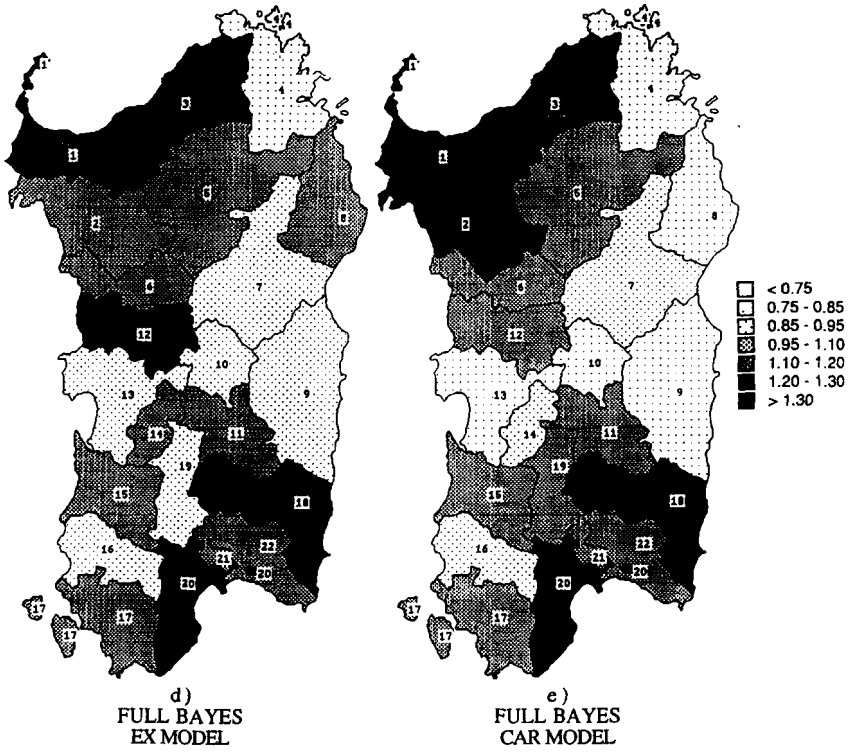


Figure 3. Breast cancer mortality, Sardinia 1983–1985
(*a*) Maximum likelihood (ML), (*b*) Empirical Bayes (EB) EX model (*c*) Empirical Bayes CAR model (*d*) Full Bayes EX model (*e*) Full Bayes CAR model

evidence' that the RR of each medium tone area is higher than the MRV; there is an 'indication' that the RRs of the lightly shaded areas are lower than the MRV and, finally, there is a 'high posterior probability' that the RR of each white area is lower than the MRV. According to these rules, the $P_{>\text{mean}}$ map shows areas in both the north-western and the south-eastern parts of Sardinia in which there is both a high probability and an indication that the RRs are higher than the MRV; these two parts are separated by a portion of land that extends from area 4 to area 13, in which there is an indication that the RRs are lower than the MRV.

The examination of Q_5, Q_{50} and Q_{95} in Table I corroborates this interpretation: the areas 1, 2, 3, 18 and 20 have the highest median rank and the highest lower and upper rank Bayesian credible limits.

Hodgkin's disease

This section presents the results of an analysis of geographical variation in Hodgkin's disease (HD) mortality. The data, estimates and summary statistics are presented in Table II, the structure of which replicates that of Table I.

Since HD is relatively rare, area-specific numbers of observed deaths are low, and four areas show no deaths at all. This causes a high Poisson variation in the SMRs, which range between 0

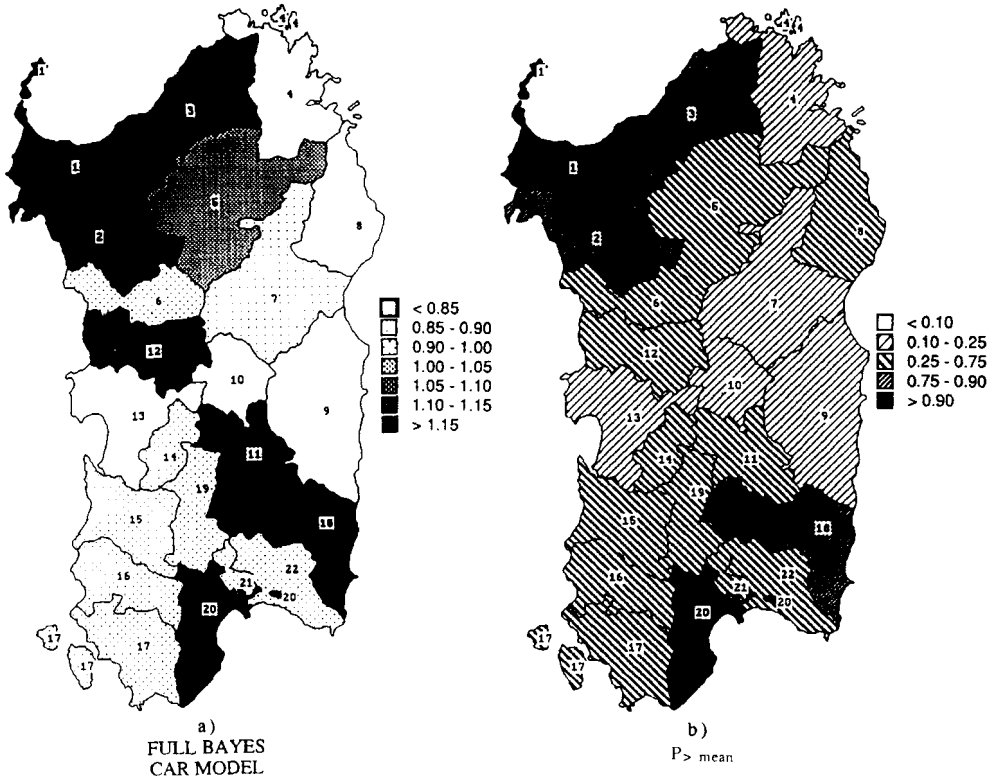


Figure 4. Breast cancer mortality, Sardinia 1983–1985
(*a*) Map of Full Bayes estimates for relative risks (*b*) map of the corresponding $P_{>\text{mean}}$ values

and 8·89; the most extreme values tend to occur in the areas of lowest population. Bayesian RR estimates show less dispersion than the SMRs; the former vary from 0·99 to 4·06. In conclusion, the spatial variation shown in the SMR map (mostly an effect of disease rareness) poorly reflects genuine geographical heterogeneity.

Meaningful differences emerge when we compare estimates obtained by the two Bayesian approaches. FB estimates tend more to shrink towards the overall mean than do the corresponding EB estimates. Under the EX model, the former range from 1·18 to 2·40, while the latter range from 0·99 to 4·06; under the CAR model, the former range from 1·03 to 3·09, while the latter range from 1·04 to 3·51.

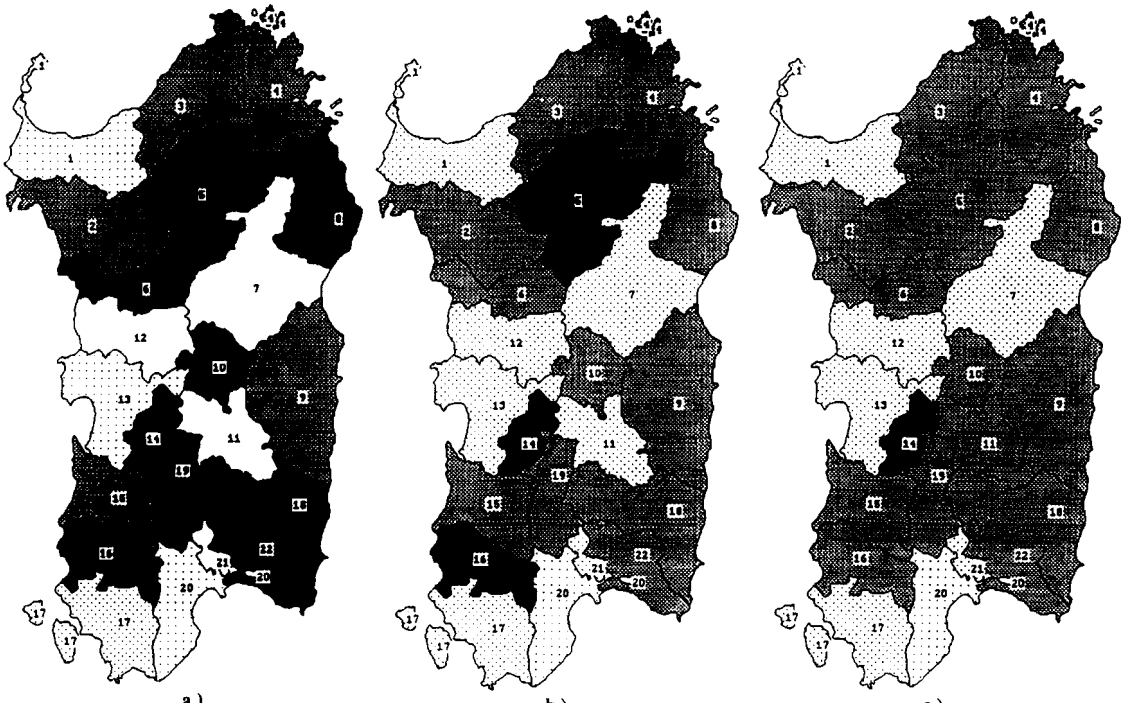
Let us now consider the maps, which are reported in Figure 5. Figure 5(*a*) displays SMRs; Figures 5(*b*) and (*c*) display EB estimates under the EX and the CAR models. Figures 5(*d*) and (*e*) display FB estimates under the EX and the CAR models. The choice of cut-off values for map visualization used the same criteria as the breast cancer maps.

At first sight, the five maps look quite different. The ML map is entirely dominated by the SMRs of the areas with very low population, which visually form spatial clusters. However, in the

Table II. Hodgkin's disease mortality, Sardinia 1983–1985. Maximum likelihood (ML), empirical Bayes (EB), full Bayes (FB) estimates of relative risks

	Data		ML		Empirical Bayes		Full Bayes											
							EX model						CAR model					
USL code	Observed deaths	Expected deaths	SMR	95 per cent C.I.	EX model RR	CAR model RR	RR	95 per cent C.I.	$P_{>mean}$	Q_5	Q_{50}	Q_{95}	RR	95 per cent C.I.	$P_{>mean}$	Q_5	Q_{50}	Q_{95}
1	5	4·79	1·04	(0·34–2·43)	1·30	1·34	1·42	(0·83–2·44)	0·21	1	7	17	1·32	(0·69–2·50)	0·18	1	6	16
2	3	1.47	2·04	(0·42–5·96)	2·01	1·95	1·79	(0·98–3·26)	0·50	3	12	21	1·84	(0·80–4·23)	0·53	3	12	21
3	2	0·90	2·22	(0·27–8·02)	2·07	1·95	1.78	(0·94–3·38)	0·54	2	12	21	1·81	(0·75–4·37)	0·53	2	12	21
4	3	1·51	1·99	(0·41–5·80)	1·98	2·04	1·77	(0·97–3·22)	0·52	2	12	21	1·87	(0·87–4·02)	0·56	3	12	21
5	3	0·88	3·41	(0·70–9·96)	2·54	2·20	1·96	(0·98–3·90)	0·63	3	15	22	2·06	(1·00–4·25)	0·66	5	14	21
6	2	0·53	3·77	(0·46–13·62)	2·48	2·26	1·91	(0·94–3·87)	0·60	2	14	22	2·02	(0·84–4·87)	0·60	3	14	22
7	0	0·45	0·00	(0·00–8·15)	1·56	1·86	1·60	(0·79–3·24)	0·42	1	9	20	1·64	(0·70–3·84)	0·44	2	10	19
8	1	0·27	3·70	(0·09–20·63)	2·25	2·30	1·83	(0·91–3·69)	0·55	2	13	21	2·10	(0·59–7·51)	0·61	1	15	22
9	1	0·46	2·17	(0·05–12·11)	2·02	2·13	1·75	(0·88–3·48)	0·49	2	12	21	1·88	(0·70–5·04)	0·55	2	13	21
10	1	0·19	5·26	(0·13–29·31)	2·37	2·08	1·81	(0·86–3·84)	0·52	2	12	22	1·89	(0·73–4·93)	0·57	2	13	21
11	0	0·14	0·00	(0·00–26·21)	1·81	2·26	1·66	(0·81–3·43)	0·45	1	11	21	2·00	(0·78–5·13)	0·58	3	14	21
12	0	0·56	0·00	(0·00–6·55)	1·49	1·59	1·58	(0·79–3·14)	0·40	1	9	20	1·48	(0·55–4·01)	0·35	1	9	20
13	3	2·80	1·07	(0·22–3·13)	1·42	1·61	1·50	(0·85–2·66)	0·30	1	8	19	1·51	(0·72–3·17)	0·35	2	9	18
14	4	0·45	8·89	(2·42–22·76)	4·06	3·51	2·40	(1·06–5·40)	0·77	5	19	22	3·09	(1·14–8·38)	0·87	8	20	22
15	2	0·98	2·04	(0·25–7·37)	2·00	2·22	1·79	(0·93–3·42)	0·53	2	12	21	1·98	(0·90–4·32)	0·61	3	14	21
16	3	0·70	4·28	(0·88–12·51)	2·79	2·28	2·01	(1·00–3·93)	0·67	3	15	22	2·06	(0·78–5·45)	0·60	3	14	22
17	1	0·74	1·35	(0·03–7·52)	1·78	1·47	1·66	(0·84–3·27)	0·45	2	10	21	1·40	(0·47–4·18)	0·33	1	7	20
18	2	0·75	2·67	(0·32–9·63)	2·21	2·29	1·86	(0·97–3·56)	0·57	2	13	22	2·03	(0·85–4·82)	0·61	3	14	21
19	2	0·59	3·39	(0·72–12·23)	2·40	2·18	1·90	(0·98–3·68)	0·59	3	14	21	2·00	(0·89–4·51)	0·61	4	13	21
20	6	8·14	0·74	(0·27–1·60)	0·99	1·04	1·18	(0·70–1·99)	0·07	1	3	13	1·03	(0·55–1·93)	0·04	1	3	11
21	0	0·46	0·00	(0·00–7·98)	1·55	1·38	1·58	(0·79–3·17)	0·41	1	9	20	1·26	(0·42–3·74)	0·25	1	6	19
22	1	0·34	2·94	(0·07–16·38)	2·15	2·04	1·83	(0·90–3·74)	0·55	2	13	22	1·85	(0·53–6·55)	0·53	1	12	22

Exchangeable prior model (EX), conditional autoregressive model (CAR), standardized mortality ratio (SMR), relative risk (RR), probability that relative risk greater than its global mean value ($P_{>mean}$), 5 per cent (Q_5), 50 per cent (Q_{50}) and 95 per cent (Q_{95}) percentiles of the relative risk distribution



a)
ML

b)
EMPIRICAL BAYES
EX MODEL

c)
EMPIRICAL BAYES
CAR MODEL

Figure 5. (*a*)–(*c*)

Bayesian maps these clusters disappear, which suggests that clusters are simply due to random variation. This is consistent with the expectations in Section 5.1.

As found in the analysis of breast cancer, the FB maps appear less smoothed than the corresponding EB maps. Notwithstanding the differences, there are a few distinctive features that are stable throughout the five maps: the very high rate in area 14, and the low rates in areas 1 and 20. We note that area 14 is the only one for which the test of comparison of the SMR to one is statistically significant; also significant are the tests of comparison to one, performed by using credible intervals, of the FB RR estimates of area 14.

As for breast cancer, we can test whether the area-specific RR is greater than the MRV by considering the map of $P_{>\text{mean}}$ values. Consider Figure 6, which reports the map of Figure 5(*d*) as modified by new specific cut-off values and as complemented by a map of the corresponding $P_{>\text{mean}}$ values (Figure 6(*b*)).

If we interpret this latter map according to the same criteria as used for Figure 4(*b*), it appears that in area 14 there is a high probability that the RR value is higher than the mean regional value (MRV); in area 1 there is an indication that the RR is lower than the MRV; in area 20 there is a high probability that the RR is lower than the MRV. In the remaining areas there is equivocal evidence that the RRs are higher than the MRV. The foregoing is also confirmed by the Q's, since these show the highest values for area 14 and the lowest for areas 1 and 20.

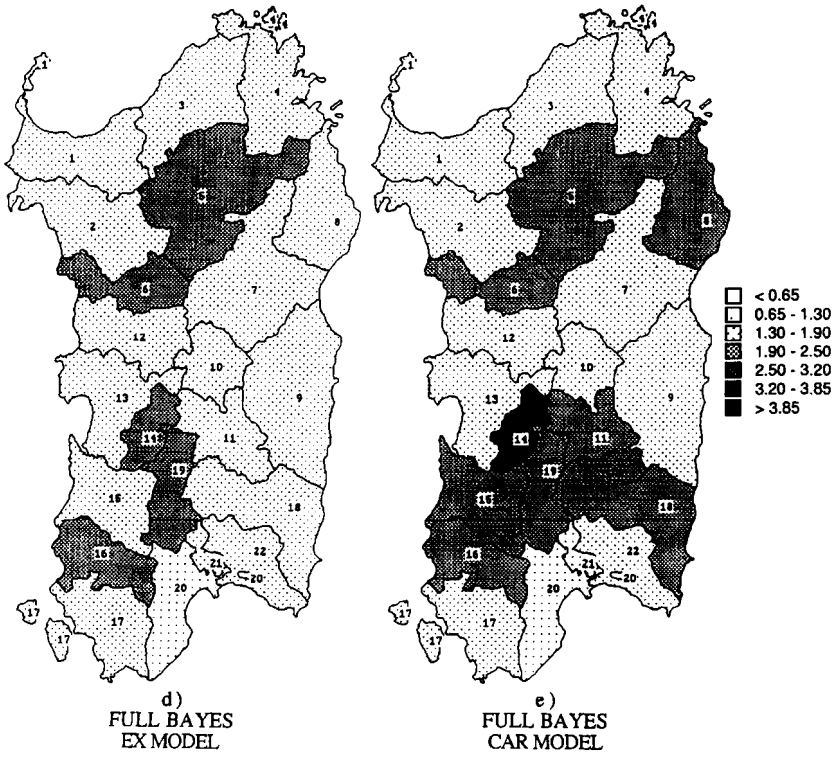


Figure 5. Hodgkin's disease mortality, Sardinia 1983–1985
(*a*) Maximum likelihood (ML), (*b*) Empirical Bayes (EB) EX model (*c*) Empirical Bayes CAR model (*d*) Full Bayes EX model (*e*) Full Bayes CAR model

6. DISCUSSION

In this section we consider a number of general questions on the statistical technology reviewed in this paper.

The first question we should raise is what are the advantages of the Bayesian approach as compared to the SMR-based maximum likelihood approach to mapping disease risk? The answer should be evident from the examples in the previous section. Because of overdispersion, the variability of the SMRs throughout the map only partly reflects genuine geographical variation, the remaining variance being due to Poisson variability. Working on a fine geographical grid and/or on a rare disease exacerbates this problem, as was evident in our application to Hodgkin's disease: high SMRs occurred in the areas with the smallest population, yielding a map which was dominated by the least reliable information.

The Bayesian approach corrects the above pitfall. For each area, this approach combines the information embodied in the area's data with the information from the rest of the map, as embodied in the prior model. The resulting RR estimate is a compromise between the SMR and the mean of the map: the poorer the data, the more this estimate will be displaced towards the global mean (for the EX model) or towards a local mean (for the CAR model). This mechanism tends to 'smooth' those portions of the map where the information is unreliable, but will preserve

HODGKIN'S DISEASE MORTALITY

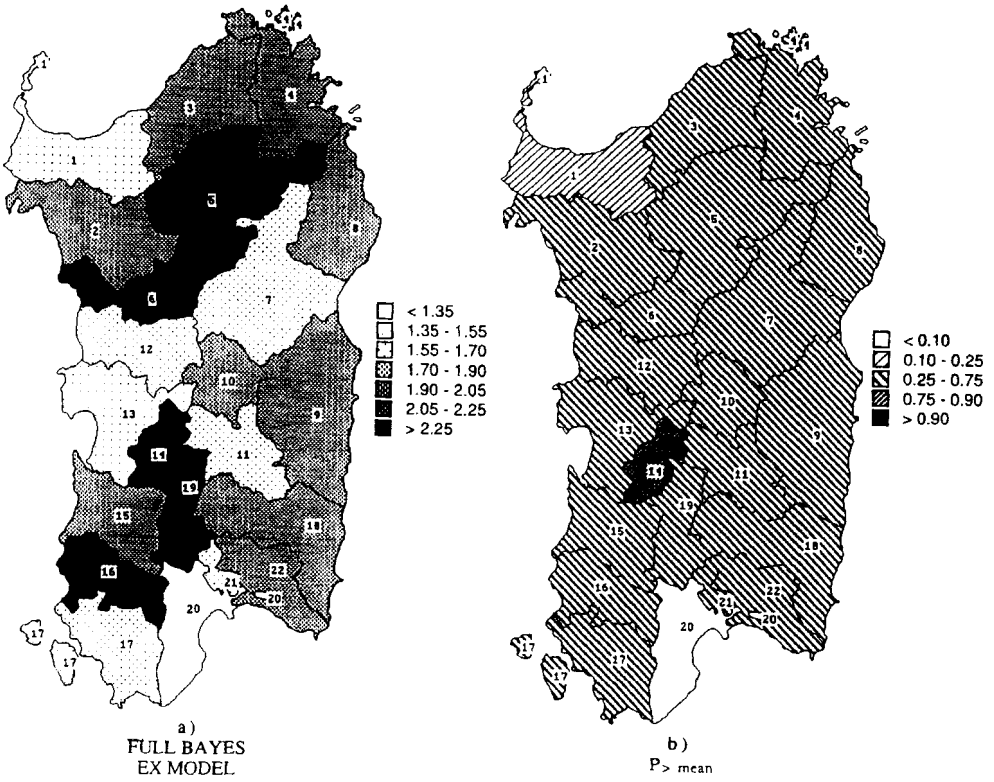


Figure 6. Hodgkin's disease mortality, Sardinia 1983–1985
(*a*) Map of Full Bayes estimates for relative risks (*b*) map of the corresponding $P_{>\text{mean}}$ values

those patterns which have high statistical significance. For example, in all Hodgkin's disease maps, the area 14, where large numbers of deaths were recorded, stands out clearly as a very high-risk area. In conclusion, the 'selective smoothing' effect yields results that are visually more pleasing and more interpretable epidemiologically.

Thanks to the 'built-in' smoother, the Bayesian approach allows us to analyse the data on a very fine geographical scale. This facilitates the identification of highly localized risk sources, for example point-source pollutants, or the detection of cluster shapes that intersect strangely with the given areas. The other advantages of the Bayesian approach are that it enables us: (a) to perform hypothesis testing upon individual RR estimates without having to adjust the *p*-values for multiplicity, and (b) to design prior models in such a way that they reflect specific prior epidemiological knowledge.

However, Bayesian modelling of spatially structured data leaves room for further research, especially as regards methods for suggesting and criticizing prior models.

The second question we should ask is which of the two Bayesian approaches (empirical Bayes/penalized likelihood and full Bayes/Gibbs sampling) is more powerful? Without doubt, the

full Bayes/Gibbs sampling approach is more powerful. True, the Gibbs sampler only yields approximate estimates, but their degree of precision is limited only by the allocated computing time. Conversely, the EB approach gives estimates which are conditional on a point estimate for the smoothing parameter, and hence they are not truly exact!

Moreover, as we remarked in Section 4.1, the access afforded by the Gibbs sampling to the *full* posterior distributions of the estimates enables us to describe the uncertainty affecting these estimates, and to perform a variety of hypothesis tests by calculating appropriate statistics. We have seen that a useful area-specific statistic is the $P_{>\text{mean}}$, the posterior probability that the area's mortality exceeds average mortality throughout the map. Further statistics may be synthesized according to specific interests, for example, the posterior probability that each area's mortality exceeds the *national* average. Moreover, as we have seen, the Gibbs sampler has the power to compute summary statistics for the distribution of ranks, which would be extremely hard to obtain with a different computational approach.

Both PL maximization and the Gibbs sampler estimation can be very easily extended to allow for *covariate* effects. An improper uniform prior for the vector of the covariate coefficients is easy to integrate into the penalized log-likelihood format. When adopting a Gibbs sampling approach it is sufficient to represent covariates and their respective coefficients as additional nodes in the graphical model on which the Gibbs sampling computations are performed. A full Bayes/Gibbs sampling approach enables us to compute posterior credible intervals for the unknown coefficients, so as to facilitate covariate selection. These approaches can be extended to deal with time. A full Bayes/Gibbs Sampling approach similar to ours is currently used in the context of temporal data analysis.[35] This sets the stage for simultaneous modelling of spatial and temporal evolution of risk.

The third question is what are the relative computational merits of the PL maximization approach and of the Gibbs sampler? The major computational limitation of the PL maximization approach is that it requires the inversion of an N-dimensional matrix, which severely limits the number of areas that can be analysed. The Gibbs sampler avoids this problem, since it proceeds by simple computational steps that are 'local' to the nodes of a graph, so that its complexity increases only linearly with the number of the areas. However, the Gibbs sampler is computationally more demanding than PL maximization. Constant control is required when the algorithm is running, so as to monitor the algorithm convergence, to choose a policy for the selection of the samples to be processed, to restart the sampler whenever appropriate, and so on. A major question concerns the convergence properties of the sampler, and potential doubts as to when the sampling process has reached an equilibrium. These issues require further research work.

To summarize, we believe that the full Bayes/Gibbs sampling approach is preferable to the empirical Bayes/PL approach in that the former offers greater flexibility and convenience in the statistical analysis of geographical variation in disease risk. More generally, the Gibbs sampler represents an innovative solution to the computational problems involved in all areas of Bayesian modelling. The advantages of the Gibbs sampler outweigh the computational demands it makes, especially if we consider the progress that we may confidently expect in computer hardware (namely parallel computers) and in the knowledge of the behaviour of the algorithm in an array of diverse analysis situations.

ACKNOWLEDGEMENTS

This research is supported by National Research Council (CNR) grant 91.00084.PF41, within the project 'Prevention and Control of Disease Factors', subproject 'Community Medicine'. It was also supported in part by the British Council and by the International Union Against Cancer

(UICC). The authors are indebted to David Clayton for the access they enjoyed to his previous related work. Thanks are also due to Wally Gilks and Carlo Berzuini for their useful advice. The valuable comments of the referees are also acknowledged.

REFERENCES

1. Breslow, N. E. and Day, N. E. *The Standardized Mortality Ratio*, Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences, Elsevier Science Publishers V. B., North Holland, 1985.
2. Clayton, D. and Kaldor, J. 'Empirical Bayes estimates of age-standardized relative risks for use in disease mapping', *Biometrics*, **43**, 671–681 (1987).
3. Darroch, J. N., Lauritzen, S. L. and Speed, T. P. 'Markov fields and log-linear models for contingency tables', *The Annals of Statistics*, **8**, 522–552 (1980).
4. Edwards, D. and Kreiner, S. 'The analysis of contingency table by graphical models', *Biometrika*, **70**, 553–562 (1983).
5. Wermuth, N. and Lauritzen, S. L. 'Graphical and recursive models for contingency tables', *Biometrika*, **70**, 537–552 (1983).
6. Whittaker, J. *Graphical Models in Applied Statistics*, Wiley Series in Probability and Mathematical Statistics, 1990.
7. Besag, J., York, J. and Mollie, A. 'Bayesian image restoration, with applications in spatial statistics (with discussion)', *Annals of the Institute of Statistical Mathematics*, **43**, 1–59 (1991).
8. Morris, C. N. 'Parametric empirical Bayes inference: theory and applications (with discussion)', *Journal of the American Statistical Association*, **78**, 47–65 (1983).
9. Maritz, J. S. and Lwin, T. *Empirical Bayes Methods*, Monographs on Statistics and Applied Probability 35, 2nd edn, Chapman and Hall, 1989.
10. Tsutakawa, R. K., Shoop, G. L. and Marienfeld, C. J. 'Empirical Bayes estimation of cancer mortality rates', *Statistics in Medicine*, **4**, 201–212 (1985).
11. Manton, K. G., Stallard, E., Woodbury, M. A., Riggan, W. B., Creason, J. P. and Mason, T. J. 'Statistically adjusted estimates of geographic mortality profiles', *Journal of the National Cancer Institute*, **78**, 805–815 (1987).
12. Tsutakawa, R. K. 'Mixed model for analyzing geographic variability in mortality rates', *Journal of the American Statistical Association*, **83**, 37–42 (1988).
13. Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason, J. P. and Pellom, A. C. 'Empirical Bayes procedures for stabilizing maps of U.S. cancer mortality rates', *Journal of the American Statistical Association*, **84**, 637–650 (1989).
14. Richardson, S. and Mollié, A. 'Empirical Bayes estimates of cancer mortality rates using spatial models', *Statistics in Medicine*, **10**, 95–112 (1991).
15. Clayton, D. 'Hierarchical Bayesian models in descriptive epidemiology', *Proceedings of the XIVth International Biometrics Conference*, 201–213 (1989).
16. Clayton, D. 'Penalized likelihood methods for mapping disease incidence', Technical Report, University of Leicester, 1990.
17. Cantwright, R., Alexander, F., MacKinley, P., Ricketts, T., Hegho, S. and Clayton, D. *Leukemias and Lymphomas, an Atlas of Distribution within Areas of England and Wales, 1984–1988*, Leukemia Research Fund, 1990.
18. Geman, S. and Geman, D. 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741 (1984).
19. Clayton, D. and Bernardinelli, L. 'Bayesian methods for mapping disease risk', in Cuzick, J. and Elliott, P. (eds), *Small Area Studies in Geographical and Environmental Epidemiology*, Oxford University Press, (to appear).
20. Besag, J. and Mollié, A. 'Bayesian mapping of mortality rates', *Bulletin of the International Statistical Institute*, 47th session, **1**, 127–128 (1989).
21. Laird, N. M. and Louis, T. A. 'Empirical Bayes confidence intervals based on bootstrap samples (with discussion)', *Journal of the American Statistical Association*, **82**, 739–757 (1987).
22. Green, P. J. 'Penalized likelihood for general semi-parametric regression model', *International Statistical Review*, **55**, 245–259 (1987).
23. Ogata, Y. and Katsura, K. 'Likelihood analysis of spatial inhomogeneity for marked point patterns', *Annals of the Institute of Statistical Mathematics*, **40**, 29–39 (1988).

24. Gelfand, A. E. and Smith, A. F. M. 'Sampling based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409 (1990).
25. Smith, A. F. M. and Gelfand, A. E. 'Bayesian statistic without tears: a sampling–resampling perspective', Technical Report, Department of Mathematics, Imperial College, London, 1990.
26. Gilks, W. R. and Wild, P. 'Adaptive rejection sampling for Gibbs sampling', *Applied Statistics*, **41**(2), 337–348 (1992).
27. Clayton, D. 'A 'C' library for Gibbs sampling in generalised linear mixed models', (unpublished manuscript).
28. Meng, C. Y. K. and Dempster, A. P. 'A Bayesian approach to the multiplicity problem for significance testing with binomial data', *Biometrics*, **43**, 301–311 (1987).
29. Verreault, R., Brisson, J., Deschenes, L. Naud, F., Meyer, F. and Bélanger, L. 'Dietary fat in relation to prognostic indicators in breast cancer', *Journal of the National Cancer Institute*, **80**, 819–825 (1988).
30. Gaskill, S. P., McGuire, W. L., Osborne, C. K. and Stern, M. P. 'Breast cancer mortality and diet in the United States', *Cancer Research*, **39**, 3628–3637 (1979).
31. Lowe, C. R. and MacMahon, B. 'Breast cancer and reproductive history of women in South Wales', *Lancet*, **i**, 153–157 (1970).
32. Robertson, S. J., Lowman, J. T., Grufferman, S., Kotsyu, D., Van der Horst, L., Matthew, T. J., Borowitz, M. J. and Bigner, S. H. 'Familial Hodgkin's disease: a clinical and laboratory investigation', *Cancer*, **59**, 1314–1319 (1987).
33. Vianna, N. J. and Polan, A. K. 'Epidemiologic evidence for transmission of Hodgkin's disease', *New England Journal of Medicine*, **289**, 499–502 (1973).
34. Grufferman, S. 'Clustering and aggregation of exposure in Hodgkin's disease', *Cancer*, **39**, 1829–1833 (1977).
35. Berzuini, C. and Clayton, D. 'Bayesian inference on the Lexis diagram', Technical Report, Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy, 1992.