STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:883–895
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2410

Impact of prior choice on local Bayes factors for cluster detection

Ronald E. Gangnon*,†

Department of Biostatistics and Medical Informatics and Department of Population Health Sciences, University of Wisconsin—Madison, 610 N. Walnut Street, Madison, Wisconsin 53726, U.S.A.

SUMMARY

In this paper, we evaluate the usefulness of local Bayes factors as a tool for spatial cluster detection. In particular, we consider whether local Bayes factors from models with a fixed, but overly large number of clusters can consistently identify the evidence for clustering for a variety of prior specifications for the cluster locations. We also investigate the robustness of the local Bayes factor to the number of clusters included in the model. We explore the impacts of prior choice for cluster location and the number of clusters on posterior inference for disease rates. We conduct the comparison by analysing data on 1990 breast cancer incidence in Wisconsin. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: Bayes factor; cluster detection; Markov chain Monte Carlo; sensitivity to priors

1. INTRODUCTION

The detection of spatial clusters is an important problem in spatial epidemiology. Spatial cluster detection should be distinguished from two other related problems: global clustering and focused clustering. Global clustering is a tendency for cases to occur near other cases throughout the entire study region. Focused clustering is the assessment of the pattern of disease risk around one or more pre-specified locations. Spatial cluster detection, on the other hand, is the assessment of the evidence for one or more clusters, small areas of increased or decreased disease incidence. Here, the locations of the clusters are unknown, and the identification of cluster locations is a major goal of the analysis. See References [1, 2] for additional discussion of these distinctions.

Spatial cluster detection has typically been approached in a hypothesis testing framework. Initially, Openshaw *et al.* [3] proposed the geographical analysis machine (GAM) as a tool for exploratory cluster detection. The GAM is a large-scale search for nominally significant circular clusters across the entire study region. Turnbull *et al.* [4] and Besag and Newell [1]

*Correspondence to: Ronald E. Gangnon, 603 WARF Office Building, Department of Population Health Sciences, University of Wisconsin—Madison, 610 N. Walnut Street, Madison, Wisconsin 53726, U.S.A.
†E-mail: ronald@biostat.wisc.edu

Contract/grant sponsor: National Cancer Institute; contact/grant numer: CA82004

Received 31 August 2005
Accepted 1 September 2005

proposed more rigorous alternatives to the GAM based on circles of fixed population radius and circles of fixed case radius, respectively. The spatial scan statistic [5, 6] provides a statistically valid method for evaluating an arbitrary collection of potential clusters. Other tests for spatial cluster detection are the weighted average likelihood ratio (WALR) statistic [7] and the weighted average likelihood ratio scan (WALRS) statistic [8].

A less common, but potentially very attractive, approach to spatial cluster detection uses a Bayesian framework for inference about explicit cluster models. Lawson and Clark [9] and Lawson [10] proposed a Cox cluster process model for the identification of cluster locations when exact case and control locations are known. Lawson and Clark [11] apply the point process model to case-count data using a data augmentation algorithm to impute exact locations for the cases and controls.

Gangnon and Clayton [12] proposed a very flexible clustering model in which the study region is divided into a large background area and a small number of clusters. The size and shape of the clusters are flexible, but controlled by a user-specified prior. Inferences were obtained using a randomized search algorithm. Gangnon and Clayton [13, 14] considered a similar model, but restricted attention to circular clusters as in the GAM. By restricting the set of potential clusters, one can more easily define the prior for the clusters and incorporate covariate effects and extra-Poisson variation as well as obtain inferences using Markov chain Monte Carlo (MCMC) techniques.

Gangnon and Clayton [13] treated the number of clusters as a parameter to be estimated. A discrete uniform prior was specified, and posterior samples were obtained using a reversible jump MCMC (RJMCMC) algorithm [15]. Gangnon and Clayton [14] proposed using a fixed, but overly large, number of clusters. They advocated the use of local Bayes factors, the ratio of the posterior odds for cluster membership to the prior odds for cluster membership, at a given location as a tool for minimizing the impact of the choice of the number of clusters in the model on inferences about the clusters. Using data sets on leukaemia incidence in upstate New York and breast cancer incidence in Wisconsin, they demonstrated both the robustness of the local Bayes factors to the number of clusters included in the model and close agreement with formal inference about the number of clusters using RJMCMC algorithms. Here, we consider the robustness of the local Bayes factors to different choices for the prior distribution on the clusters. To do so, we revisit the analysis of the data on breast cancer incidence in Wisconsin. We also consider the effects of both the prior on the clusters and the number of clusters on inferences about the disease risks.

In Section 2, we present the clustering model proposed by Gangnon and Clayton [13, 14] and give three different specifications for the prior distribution for the clusters. In Section 3, we describe MCMC techniques for obtaining posterior inferences and discuss the local Bayes factor as a tool for inference about cluster locations. In Section 4, we present an analysis of breast cancer incidence in Wisconsin to investigate the robustness of the inferences about cluster membership and disease risk to these choices of prior distribution on clusters. In Section 5, we present some concluding remarks.

2. STATISTICAL MODEL

The available data consist of $(y_i, E_i, \mathbf{x}_i)_{i=1}^N$, where y_i is the number of cases of disease in region i, E_i is the expected number of cases of disease in region i (calculated using internal

or external standardization) and $\mathbf{x}_i = (x_{1i}, x_{2i})$ is the vector of co-ordinates of the geographic centroid of region i. We assume that y_i each follow a Poisson distribution with mean $\rho_i E_i$ and that y_i are conditionally independent given the Poisson mean, where ρ_i is the standardized incidence ratio (SIR) for region i. We adopt a log-linear model for ρ_i, e.g. $\log(\rho_i) = \alpha + \phi_i + \varepsilon_i$, where α is an intercept, ϕ_i is a spatial clustering effect and ε_i is a spatially uncorrelated random effect. The basic framework of the model is similar to the models adopted by Clayton and Kaldor [16] and Besag *et al.* [17], but our specification of the spatial effect ϕ_i is different to reflect our interest in cluster detection rather than spatial smoothing.

The spatial clustering effect ϕ_i is given by $\phi_i = \sum_{j=1}^{k} \theta_j \delta_{(\mathbf{c}_j, r_j)}(\mathbf{x}_i)$, where k is the number of clusters, $\delta_{(\mathbf{c}_j, r_j)}(\mathbf{x}_i)$ is an indicator variable of membership in a circular cluster of radius r_j centred at $\mathbf{c}_j$, and θ_j is the log relative risk associated with cluster j. Specifically, $\delta_{(\mathbf{c}_j, r_j)}(\mathbf{x}_i) = 1$ if $d(\mathbf{x}_i, \mathbf{c}_j) \leqslant r_j$ and $\delta_{(\mathbf{c}_j, r_j)}(\mathbf{x}_i) = 0$ otherwise, where d is the Euclidean metric. To eliminate the possibility of empty clusters, we will select the cluster centres $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k$ from the cell centroids, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$. The cluster radii $r_1, r_2, \ldots, r_k$ are allowed to range from 0 up to a fixed maximum radius $r_{\max}$. To identify the m_s unique clusters centred at $\mathbf{x}_s$ for $s = 1, 2, \ldots, N$, we let $0 = r_{s,1} < r_{s,2} < \cdots < r_{s,m_s} \leqslant r_{\max}$ be the ordered distances from the centroid of cell s to the centroids of all cells, truncated at $r_{\max}$. (If two or more centroids are equidistant from the centroid of cell s, the common distance is only listed once.) So, the clusters are selected from the set $\{(\mathbf{x}_s, r_{st}); t = 1, 2, \ldots, m_s, s = 1, 2, \ldots, N\}$. Although we illustrate the methodology using this specific set of clusters, we note that the approach to inference described here is quite general and could be applied to any discrete set of clusters.

As we will adopt a Bayesian approach to inference, we need to specify prior distributions for each of the parameters in the model. The intercept α is given a flat prior. The spatially uncorrelated random effects $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ are independent and identically distributed (iid) normal with mean 0 and variance $1/\tau$. The precision of the random effects τ is given a conjugate gamma prior. Typically, we will use a gamma prior with mean 100 and standard deviation 100 so that, with 95 per cent probability, the variance $1/\tau$ falls between 0.003 and 0.40. A variance of 0.40 will imply a relative risk of roughly 12 between regions at the 2.5th and 97.5th percentiles; a variance of 0.003 will imply that the same relative risk is just 1.2.

For the spatial clustering effect, we adopt an iid prior specification for the k clusters; we also assume *a priori* independence of the cluster location and its associated log relative risk. For the cluster log relative risk, we use a normal prior, e.g. $\theta_1 \sim \mathrm{N}(0, \sigma_\theta^2)$. Typically, we take σ_θ^2 to be 0.355 so that, *a priori*, $P(1/4 < e^{\theta_1} < 4) = 0.99$. For the cluster (centre and radius), we will need to specify a probability measure on the discrete set $\{(\mathbf{x}_s, r_{st});\ t = 1, 2, \ldots, m_s, s = 1, 2, \ldots, N\}$, e.g. $P(\mathbf{x}_s, r_{st}) = p_{st}$.

Finally, we consider the number of clusters, k. We can proceed in one of two ways: k is a parameter to be estimated or k is a fixed constant. In the former case, inference is performed using a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm [15] as described previously [13]. In the latter case, we follow Gangnon and Clayton [14] and select k as an upper bound on the true number of clusters. If the true number of clusters, say k_0, is no greater than k, the underlying model is, in fact, correct, albeit possibly overparameterized. That is, if $\phi_i = \sum_{j=1}^{k_0} \theta_j \delta_{(\mathbf{c}_j, r_j)}(\mathbf{x}_i)$, then $\phi_i = \sum_{j=1}^{k} \theta_j \delta_{(\mathbf{c}_j, r_j)}(\mathbf{x}_i)$, where $\theta_{k_0+1} = \theta_{k+2} = \cdots = \theta_k \equiv 0$ and $(\mathbf{c}_{k_0+1}, r_{k_0+1}), (\mathbf{c}_{k_0+2}, r_{k_0+2}), \ldots, (\mathbf{c}_k, r_k)$ are arbitrary. Thus, one would expect similar behaviour in the posterior, e.g. a concentration of mass on the k_0 true clusters along with essentially arbitrary placement of the $k - k_0$ excess clusters with cluster risks near 0.

2.1. Priors for circular clusters

We now consider three specifications of the prior probability p_{st} of circular clusters $(\mathbf{x}_s, r_{st})$ for $t = 1, 2, \ldots, m_s, s = 1, 2, \ldots, N$. The prior probability of cluster membership for each ZIP code area for each prior with $k = 10$ is displayed in the first column of Figure 1.

Prior #1. The dartboard prior: The first distribution considered is the dartboard prior [7]. The dartboard prior is a discrete approximation to the uniform selection of a circle with a radius no greater than $r_{\max}$ within the study region. First, one of the potential cluster centres $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is selected by 'throwing a dart' at the study region, e.g. the probability of selecting $\mathbf{x}_s$ as the cluster centre is a_s/A, where a_s is the area of region s and $A = \sum_{s=1}^{N} a_s$ is the area of the entire study region. The radius of the circle is then selected from a uniform distribution on $[0, r_{\max}]$. Thus, the probability of selecting $(\mathbf{x}_s, r_{st}), t = 1, 2, \ldots, m_s$, $s = 1, 2, \ldots, N$ is

$$p_{st} = \frac{a_s}{A} \frac{r_{s,t+1} - r_{st}}{r_{\max}}$$

where $r_{s,m_s+1} = r_{\max}$. The probability that region i belongs to a single cluster, $p_i = \sum_{s,t} p_{st}\ \delta_{(\mathbf{x}_s, r_{st})}(\mathbf{x}_i)$, is roughly constant for this prior.

Prior #2. The naive uniform prior: A simpler alternative to the dartboard prior is a naive, seemingly uniform prior. This distribution assigns equal mass to each of the $m = \sum_{s=1}^{n} m_s$ possible clusters. Thus, the probability of selecting $(\mathbf{x}_s, r_{st}), t = 1, 2, \ldots, m_s, s = 1, 2, \ldots, N$ is

$$p_{st} = \frac{1}{m}$$

For this prior, the probability that region i belongs to a single cluster from this prior is proportional to the number of clusters overlapping region i. For many commonly used small regions, e.g. census tracts or zip codes, urban areas tend to have a relatively large number of geographically small regions, while rural areas tend to have a relatively small number of geographically large regions. Thus, small regions in urban areas belong to more potential clusters and thus have higher prior probabilities of cluster memberships, while small regions in rural areas belong to fewer potential clusters and have lower prior probabilities of cluster membership.

Prior #3. Variant of dartboard prior: The third prior distribution is a variant of the dartboard prior, which can incorporate prior information about likely cluster locations. Here, the cluster centres, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, are partitioned into $G \geqslant 2$ subsets, denoted $S_1, S_2, \ldots, S_G$. One of these subsets is selected at random with probability $P(S_1), P(S_2), \ldots, P(S_G)$. One of the circles centred inside this subset of the cluster centres is then selected from the corresponding restricted dartboard prior. Thus, the probability of selecting $(\mathbf{x}_s, r_{st}), t = 1, 2, \ldots, m_s, s = 1, 2, \ldots, N$ is

$$p_{st} = \sum_{g=1}^{G} P(S_g) \frac{a_s}{A(S_g)} \frac{r_{s,t+1} - r_{st}}{r_{\max}} \mathbf{1}\{\mathbf{x}_s \in S_g\}$$

where $A(S_g) = \sum_{s:\mathbf{x}_s \in S_g} a_s$, $g = 1, 2, \ldots, G$, and the indicator function $\mathbf{1}\{A\}$ takes the value 1 if A is true and 0 otherwise.

This family of distributions is flexible enough that we can assign relatively high prior probabilities of cluster membership to any collection of cells. For the analyses of the Wisconsin breast cancer data in Section 4, we use this prior distribution with $G = 2$. The two subsets of

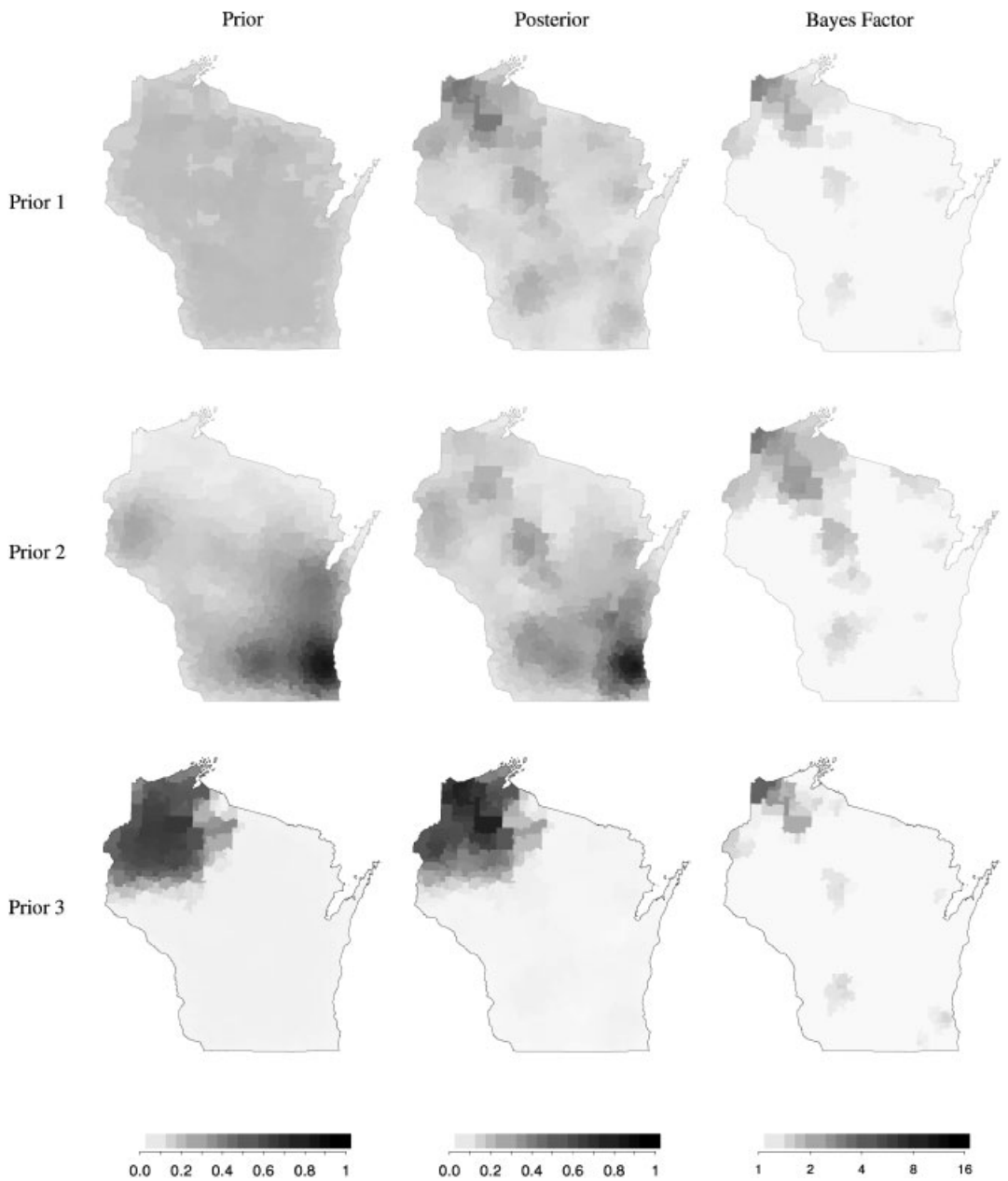

Figure 1. Inferences about cluster locations in the Wisconsin breast cancer data using models with three different priors for circular clusters and fixed $k = 10$: prior probability of cluster membership $P(\phi_i \neq 0)$, posterior probability of cluster membership $P(\phi_i \neq 0 \mid y_1, y_2, \ldots, y_n)$, and Bayes factor for cluster membership.

the potential cluster centres are S_1, the centroids of zip code areas in an eight-county region in the northwest corner of the state and S_2, the centroids of the remaining zip codes. We assign $P(S_1) = 0.9$ and $P(S_2) = 0.1$, creating a high probability of clusters in the northwest corner of Wisconsin and a low probability of clusters elsewhere.

3. POSTERIOR CALCULATION

Conditional on the k clusters $(\mathbf{c}_1, r_1), (\mathbf{c}_2, r_2), \ldots, (\mathbf{c}_k, r_k)$, the above model is a hierarchical Poisson generalized linear model with parameters α, $\theta_1, \theta_2, \ldots, \theta_k$ and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Techniques for sampling from the posterior distribution for general models of this type are described in Reference [18] and for this specific model in Reference [13]. Specifically, the quadratic approximation to the likelihood, which is conjugate to the normal priors for α, $\theta_1, \theta_2, \ldots, \theta_k$ and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, is used to develop proposal distributions for a Metropolis–Hastings algorithm [19]. Posterior samples for τ are obtained from its conjugate full conditional distribution.

Only the updates for the k cluster locations $(\mathbf{c}_1, r_1), (\mathbf{c}_2, r_2), \ldots, (\mathbf{c}_k, r_k)$ remain to be specified. We update each cluster location in turn using its full conditional distribution, e.g. the probability that cluster $(\mathbf{x}_s, r_{st})$, $t = 1, 2, \ldots, m_s$; $s = 1, 2, \ldots, N$ is selected as the update for the current cluster $(\mathbf{c}_1, r_1)$ is given by

$$q(\mathbf{x}_s, r_{st}) = \frac{p_{st} \exp\{\theta_1 y_{st} - e^{\theta_1} E_{st}\}}{\sum_{s,t} p_{st} \exp\{\theta_1 y_{st} - e^{\theta_1} E_{st}\}}$$

where $y_{st} = \sum_{i=1}^{n} y_i \delta_{(\mathbf{x}_s, r_{st})}(\mathbf{x}_i)$ is the observed number of cases inside cluster $(\mathbf{x}_s, r_{st})$ and $E_{st} = \sum_{i=1}^{n} \rho_i \exp\{\theta_1[1 - \delta_{(\mathbf{c}_1, r_1)}(\mathbf{x}_i)]\} E_i \delta_{(\mathbf{x}_s, r_{st})}(\mathbf{x}_i)$ is the expected number of cases inside cluster $(\mathbf{x}_s, r_{st})$ without cluster 1 in the model (based on the current values of the other model parameters). This proposal of a replacement cluster is accepted with probability 1. In practice, we often use a truncated version of this proposal distribution, e.g. $q'(\mathbf{x}_s, r_{st}) \propto q(\mathbf{x}_s, r_{st})$ if $q(\mathbf{x}_s, r_{st}) / \sup_{s,t} q(\mathbf{x}_s, r_{st}) \geqslant 1/W$ and $q'(\mathbf{x}_s, r_{st}) = 0$ otherwise for some constant $W > 1$. For the truncated version, the proposal is accepted with probability 1 if $q(\mathbf{c}_1, r_1) / \sup_{s,t} q(\mathbf{x}_s, r_{st}) \geqslant 1/W$ and rejected otherwise. For sufficiently large values of W, this latter condition will rarely, if ever, apply. In our applications, we will use $W = 10\,000$.

An update of the chain consists of a single update of the k cluster locations $(\mathbf{c}_1, r_1)$, $(\mathbf{c}_2, r_2), \ldots, (\mathbf{c}_k, r_k)$ followed by $m \geqslant 1$ updates of the intercept α, the cluster log relative risks $\theta_1, \theta_2, \ldots, \theta_k$, the random effects $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ and the precision of the random effects distribution τ. In our applications, we will use $m = 10$.

3.1. Parameters of interest

Our primary interest is inference about the spatial clustering effects $\phi_1, \phi_2, \ldots, \phi_n$. In particular, we are interested in the cluster membership indicator for each region, $\mathbf{1}\{\phi_i \neq 0\}$. Given the dichotomous nature of this parameter, the posterior mean, $P(\phi_i \neq 0 \mid y_1, y_2, \ldots, y_n)$, serves as a complete summary of its posterior distribution. The posterior probability of cluster membership for region i depends not only on the evidence in the data for a cluster involving region i, but also on the prior for the cluster locations $\{p_{st}, t = 1, 2, \ldots, m_s, s = 1, 2, \ldots, N\}$ and the number of clusters k. To identify the evidence in the data for clustering, we suggest using the Bayes

factor for clustering in region i, the ratio of the posterior odds in favour of $\mathbf{1}\{\phi_i \neq 0\}$ to the prior odds in favour of $\mathbf{1}\{\phi_i \neq 0\}$. The Bayes factor is given by

$$\mathrm{BF}_i = \frac{P(\phi_i \neq 0 \mid y_1, y_2, \ldots, y_n)/\{1 - P(\phi_i \neq 0 \mid y_1, y_2, \ldots, y_n)\}}{\{1 - (1 - p_i)^k\}/(1 - p_i)^k}$$

where $p_i = \sum_{s,t} p_{st}\delta_{(\mathbf{x}_s, r_{st})}(\mathbf{x}_i)$ is the probability that region i belongs to a single cluster selected from the prior. The use of Bayes factors should largely minimize, but will not completely eliminate, the influence of the prior on cluster locations and the number of clusters. Thus, we might hope that the Bayes factors in favour of clustering for each region will be reasonably comparable for different priors for the cluster locations and different numbers of clusters. The use of local Bayes factors may also tend to minimize edge effects due to the use of circular clusters, since the Bayes factor naturally takes into account the artificially depressed prior probability of cluster membership near the edge of the study region. Gangnon and Clayton [14] explore the robustness of the Bayes factors for different numbers of clusters with a single prior for the cluster locations. Here, we expand their study to investigate multiple priors for the cluster locations.

In addition to the cluster memberships, we are also interested in inferences about the standardized incidence ratios $\rho_1, \rho_2, \ldots, \rho_n$. The posterior distribution of ρ_i will be summarized in terms of the posterior mean (or median) of ρ_i and posterior standard deviation of $\log(\rho_i)$. We anticipate that the posterior means will be relatively robust to the prior for cluster locations and the number of clusters. We also expect that the posterior uncertainty about ρ_i will generally increase along with the prior probability that region i belongs to a cluster.

4. EXAMPLE: WISCONSIN BREAST CANCER DATA

To assess the robustness of the Bayes factor as a tool for assessing the evidence for clustering with respect to the prior for cluster locations and the number of clusters in the model, we revisit the Wisconsin breast cancer data set, previously analysed by Gangnon and Clayton [14]. Data are available for 716 ZIP code areas. For each ZIP code area, the count of incident breast cancer cases is available for the Wisconsin State Cancer Registry, and the age-specific female populations (in 5-year intervals) are available from the Census Bureau. For each ZIP code area, we calculated an expected number of breast cancer cases using indirect, internal age standardization.

Here, we will consider the same set of circular clusters considered previously [14]. The set of potential clusters consists of the 29 462 circular clusters centred at the zip code centroids with $r_{\max} = 50$ km. We used the three prior distributions for the cluster locations described in Section 2.1. For each cluster prior, we considered three different choices for the number of clusters in the model: $k = 5$, 10 and 20.

To identify convergence of the Markov chains, we followed the strategy suggested by Gelman and Rubin [20]. We ran five independent Markov chains. For each chain, we used a run-in of 1 million iterations; we kept every 100th sample from the next 1 million iterations as our sample for inference. Gelman–Rubin statistics were used to assess convergence for all parameters of interest. In addition, a subset of the parameters were graphically monitored across the 5 chains. Based on these assessments, there were no substantial differences in the

samples across the chains (for each of the 9 models under consideration), and we concluded that the chains had converged.

In Figure 1, we display, for $k = 10$, the prior probability of cluster membership, the posterior probability of cluster membership, and the local Bayes factor for cluster membership for each ZIP code area using the three priors for cluster locations discussed in Section 3. For Prior 1, the prior probability of cluster membership is roughly uniform across the entire state. For Prior 2, the prior probability of cluster membership is relatively high in the southeast corner of the state, corresponding to the major cities in Wisconsin (Milwaukee, Madison and Green Bay) and relatively low in sparsely populated northern Wisconsin. For Prior 3, the prior probability of cluster membership is, by design, relatively high in the northwest corner of the state and relatively low elsewhere.

The maps of the posterior probability of cluster membership, to a large extent, mirror the maps of the prior probability of cluster membership. Because of this, one cannot directly evaluate the evidence for clustering from the posterior map, but one must evaluate the posterior map in comparison to the prior map. An informal visual comparison of the two prior and posterior maps is actually quite informative. It is relatively easy to identify an increase in mass for the northwest corner of the state in all three maps. However, it is difficult to evaluate the magnitude of the increase and hence the evidence for clustering in the data.

As noted previously, the maps of the local Bayes factors for cluster membership should provide a consistent assessment of the evidence for clustering at each location, regardless of the chosen prior. The three maps of the local Bayes factor for cluster membership in Figure 1 are strikingly similar. If we interpret the Bayes factor using the scale proposed by Kass and Raftery [21], all three maps show modest, but positive evidence (a Bayes factor of 3–7) for a (low risk) cluster in the extreme northwest corner of Wisconsin, which includes the city of Superior. There is very weak evidence (a Bayes factor of 1–3) for clustering in several other regions across the state. For the most part, the same areas are consistently identified in all three maps.

There are, of course, some notable differences between the maps. For example, the potential areas of clustering in the north and west portions of the state indicated using prior 2 are quite a bit larger than the corresponding areas identified using prior 1, which are generally larger than those identified using prior 3. One might hypothesize that these differences are related to the intensity of sampling for clusters in this portion of the state. A sampler based on prior 3 heavily samples clusters in the northwest corner of the state and thus can produce a more refined assessment of the location of clusters there, while a sampler based on prior 1 lightly samples clusters in the northwest corner of the state and thus can only provide a crude assessment of the location of clusters in that portion of the state.

In Figure 2, we display the local Bayes factor for cluster membership for the three priors for cluster locations given in Section 3 and three values for k $(5, 10, 20)$. Note that the second column of Figure 2 is the same as the third column of Figure 1. Within each row, we observe relatively minor differences in the maps of the local Bayes factors for different values for k for all three priors for cluster locations. Thus, the choice of k appears to have minimal impact on the results. Within each column, we observe somewhat greater, although still relatively modest, differences in the local Bayes factors for the different priors for the cluster locations. In summary, it appears that local Bayes factors from models with a fixed, but overly large number of clusters can consistently identify the evidence for clustering for a variety of prior specifications for the cluster locations.

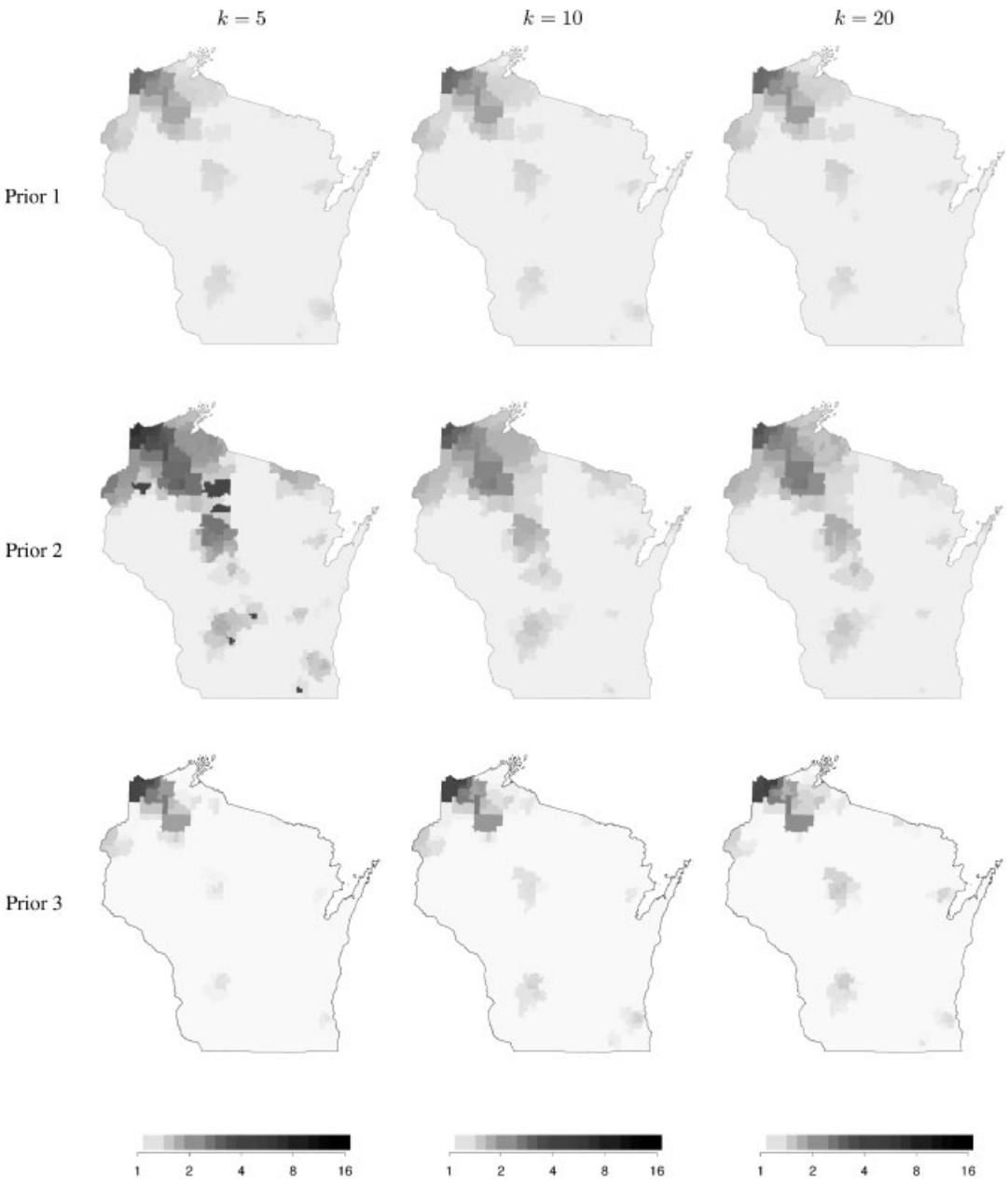

Figure 2. Local Bayes factors for ZIP code-specific cluster memberships for the Wisconsin breast cancer data using three different priors for circular clusters and fixed $k = 5, 10, 20$.

In Figures 3 and 4, we display the posterior means for the standardized incidence ratio ρ_i and the posterior standard deviations of $\log(\rho_i)$, respectively, for this same set of nine models. For priors 1 and 3, we observe similar behaviour of these posterior summary statistics. As k increases, the posterior means are more variable, e.g. less shrinkage towards an SIR of 1, and the posterior standard deviations increase. This is particularly true for ZIP code areas in the northwest corner of the state, the portion of the state in which we find the strongest evidence for clustering in the data. These observations are to be expected. The presence of clusters in a particular area will result in shrinkage towards the local mean rather than global mean and a corresponding increase in the posterior uncertainty. Increasing k will naturally increase the probability of any ZIP code area belonging to a cluster, and the effect should be most notable in areas with the strongest evidence (in the data) for clustering. The same type of effect is observed when comparing the results using prior 1 to the results using prior 3. Because prior 3 places higher prior (posterior) probability on clusters in the northwest portion of the state than prior 1, we observe less shrinkage of the posterior means towards 1 and greater posterior uncertainty in that region of the state when we use prior 3.

For prior 2, we observe very different behaviour of the posterior distributions for standardized incidence ratios. Here, as k increases, the posterior means are largely unchanged, and the posterior standard deviations actually decrease. We suspect that this unusual behaviour may be due to the conflict between the prior (which favours clusters in the southeast) and the data (which favour clusters in the northwest). This conflict between the prior and the data, which is most pronounced for $k = 5$ and gradually declines as k is increased, results in the high posterior uncertainty regarding the cluster locations and hence high posterior uncertainty in the SIRs. Even with $k = 20$ clusters in the model, the posterior probability of a cluster in the northwest corner of the state remains quite low under prior 2, and hence the posterior means remain quite stable.

5. DISCUSSION

In this paper, we revisited the spatial clustering model proposed by Gangnon and Clayton [13, 14]. In prior work, we have considered both direct inference about the number of clusters k supported by the data and indirect inference about clustering based on local Bayes factors. Previously, we demonstrated robustness of the local Bayes factor for cluster membership to the assumed value for k using a relatively uniform prior for the clusters, the 'dartboard' prior [14]. Here, we explored the impact of three different prior specifications—the dartboard prior, a naive uniform prior, and an informative variant of the dartboard prior—on the local Bayes factors for cluster membership. As noted previously, the choice of k appears to have minimal impact on the results. There are slightly greater, but still quite modest, differences in the local Bayes factors for the different priors on the clusters. Overall, the local Bayes factors are quite robust to both the prior on clusters and the number of clusters.

In contrast, inferences about the disease risks appear to depend on the choice of the prior on clusters and the number of clusters. In our analyses, this is particularly apparent in areas with the most evidence for clustering. This lack of robustness can be explained by noting that the posterior distribution for the disease risk in a given cell is a mixture of two distributions—the posterior conditional on the cell belonging to one or more clusters and the posterior conditional on the cell not belonging to a cluster. In areas with evidence favouring clustering, these two

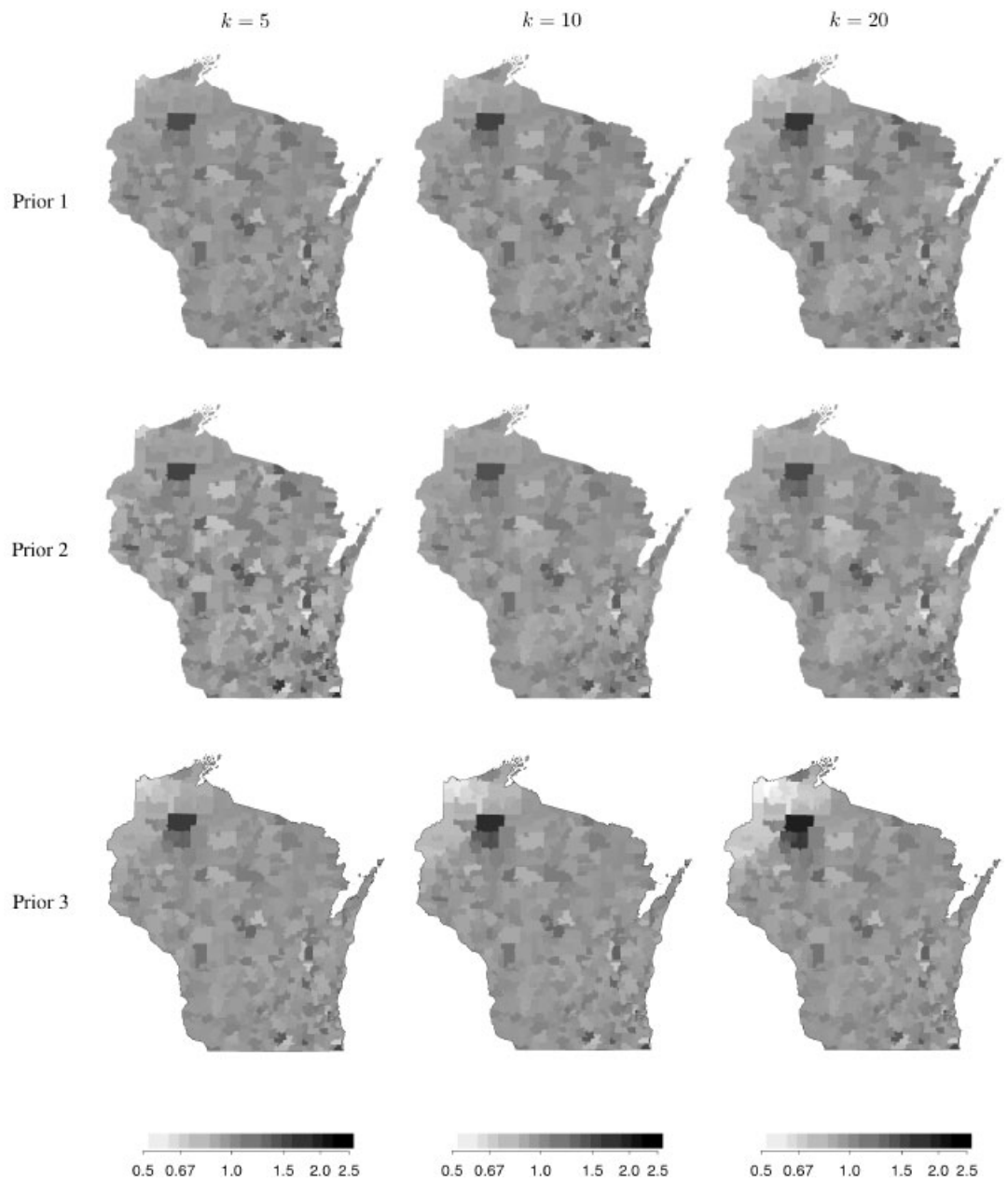

Figure 3. Posterior means for ZIP code-specific breast cancer risk for the Wisconsin breast cancer data using three different priors for circular clusters and fixed $k = 5, 10, 20$.

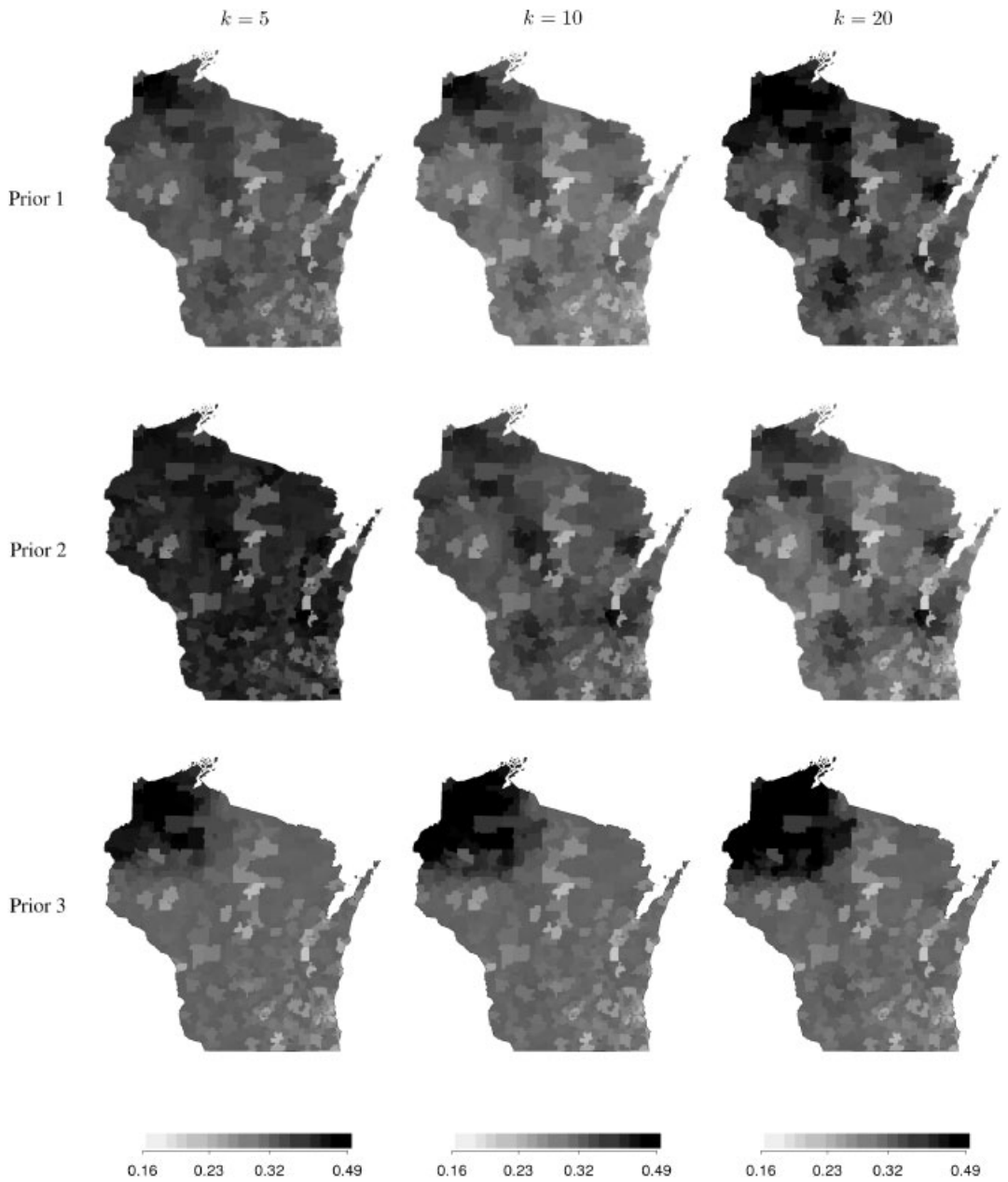

Figure 4. Posterior standard deviations for the log ZIP code-specific breast cancer risk for the Wisconsin breast cancer data using three different priors for circular clusters and fixed $k = 5, 10, 20$.

distributions will be quite different, and so the relative weighting of the two, which depends heavily on the number of clusters and the prior on clusters, will have a substantial impact on the posterior. Since the primary goal of our analysis is cluster detection, this lack of robustness is not a major concern. However, in future work, we hope to explore methods for more robust posterior inference about the disease risks in these model, perhaps by exploiting the decomposition of the posterior described above.

ACKNOWLEDGEMENTS

The author would like to thank Jane McElroy, Brett Moore, John Hampton, Patrick Remington and Amy Trentham-Dietz for their assistance with the Wisconsin breast cancer data set.

REFERENCES

1. Besag J, Newell J. The detection of clusters of rare diseases. *Journal of the Royal Statistical Society*, *Series A* 1991; **154**:143–155.
2. Kulldorff M. Statistical method for spatial epidemiology: tests for randomness. In *GIS and Health for Europe*, Gatrell A, Løytønen M (eds). Taylor & Francis: London, 49–62.
3. Openshaw S, Craft AW, Charlton M, Birch JM. Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet* 1988; **1**:272–273.
4. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. Monitoring for clusters of disease; Application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 1990; **132**:S136–S143.
5. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995; **14**:799–810.
6. Kulldorff M. A spatial scan statistic. *Communications in Statistics*: *Theory and Methods* 1997; **26**:1481–1496.
7. Gangnon RE, Clayton MK. A weighted average likelihood ratio test for spatial clustering of disease. *Statistics in Medicine* 2001; **20**:2977–2987.
8. Gangnon RE, Clayton MK. Likelihood-based tests for spatial clustering of disease. *Environmetrics* 2004; **15**:797–810.
9. Lawson A, Clark A. Small area cluster modeling via RJMCMC methods. *Journal of the National Institute of Public Health* 1999; **48**:113–120.
10. Lawson AB. Cluster modeling of disease incidence via RJMCMC methods: a comparative evaluation. *Statistics in Medicine* 2000; **19**:2361–2376.
11. Lawson AB, Clark A. Markov chain Monte Carlo methods for putative sources of hazard and general clustering. In *Disease Mapping and Risk Assessment for Public Health*, Lawson AB, Bohning D, Biggeri A, Viel J-F, Bertollini R (eds), Chapter 9. Wiley: New York, 1999.
12. Gangnon RE, Clayton MK. Bayesian detection and modeling of spatial disease clustering. *Biometrics* 2000; **56**:922–935.
13. Gangnon RE, Clayton MK. A hierarchical model for spatially clustered disease rates. *Statistics in Medicine* 2003; **22**:3213–3228.
14. Gangnon RE, Clayton MK. Cluster detection using Bayes factors from over-parameterized cluster models. *Environmental and Ecological Statistics* 2005, in press.
15. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995; **82**:711–732.
16. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; **43**:671–681.
17. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 1991; **43**:1–59.
18. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman & Hall: London, 1995.
19. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970; **57**:97–109.
20. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**:457–472.
21. Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association* 1995; **90**:773–795.

