A comparison of focused score tests and Bayesian hierarchical models for detecting spatial disease clustering

Elizabeth G. HILL, Andrew S. ALLEN and Lance A. WALLER[1)]

We compare performance of two approaches for detecting disease clustering (spatially localized increases in disease rates) in maps of incidence data. The first is a score test for detecting clustering around a prespecified source of suspected increased risk (e.g. a hazardous wastes site). This method is simple to calculate, and based on a hypothesis-testing paradigm. The second is based on estimates of region-specific standardized morbidity ratios (SMRs) from a hierarchical Bayesian model with spatial random effects. This method involves greater computational effort than the score test, but provides much more inferential output than a simple 'reject/fail to reject the null hypothesis' conclusion, namely model-based estimates of SMRs for each region. We contrast the methods' performance in a frequentist manner, comparing the power of the score test to the proportion of times the posterior median SMR exceeds an extreme quantile of the sampling distribution of this estimate in the absence of clustering. We find the Bayesian approach achieves comparable frequentist performance to the score test, and seems to provide greater discriminatory accuracy in determining the center of a true cluster.

1 Introduction

Analysis of spatial patterns of incident cases of disease plays a key role in exploratory epidemiologic investigations and disease surveillance. Typically, investigators assess whether the observed cases are significantly clustered within a heterogeneously distributed population. Numerous hypothesis tests are available to detect clustering, providing the investigator with simple measures that quantify the significance of the clustering observed in the data. That is, the tests quantify whether the observed amount of clustering is more than that expected by chance. Lawson and Waller[1)], Elliott *et al.*[2)], and Marshall[3)] review these methods.

Besag and Newell[4)] distinguish between 'focused' and 'general' tests to detect clustering. The distinction between the two is that the former concentrate on regions around one or more foci (e.g. a hazardous waste site or contaminated well) thought to be related a priori to increased disease incidence, while the latter are applied to geographic regions with no prior belief as to the putative causes of disease. Waller and Lawson[5)] compare the power of focused tests to detect clustering, and Tango[6)] compares the power of both focused and general tests.

In addition to simple tests of clustering, epidemiologists often wish to compare disease rates between regions (e.g. census enumeration districts) within a study area. Throughout this paper, we use the term 'region' to denote small administrative districts that subdivide the study area. Unfortunately, rare disease rate estimates are difficult to compare directly due to the variability in the size of the denominator. Regions with small at-risk population sizes generate rate estimates with large variances. Often these estimates appear highly discrepant from the overall disease rate. Recently, Bayesian models have gained popularity in mapping disease rates due to their characteristic 'borrowing-of-strength'. That is, regional rate estimates incorporate information from neighboring regions, resulting in smoothed posterior estimates of the rate of disease. Compared to the standard 'reject/fail to reject' conclusion from statistical hypothesis tests, these Bayesian hierarchical models provide a wealth of information about the entire study region. Rather than attempting to quantify clustering by a single measure, posterior parameter estimates provide a mappable picture of how the disease is clustered. This increase in information requires increased computational intensity,

1) contact author
lwaller@sph.emory.edu
Department of Biostatistics
Rollins School of Public Health
Emory University
1518 Clifton Road NE
Atlanta, GA 30322-5730
(404)727-1057 fax : (404)727-1370

as outlined in Section 3.3. Clayton and Bernardinelli[7] and Best *et al.*[8] summarize current Bayesian models used to analyze spatially-referenced health and exposure data.

The purpose of this paper is to compare frequentist (hypothesis-testing) and Bayesian methods in their ability to detect clustering of a rare disease. Typically such comparisons require an assessment of how often a given method detects clustering when clustering is known to exist (the power of the test). In the hypothesis-testing paradigm, if the theoretical distribution of the test statistic is known, a critical value for a one-sided upper-tail test is equal to the $(1-\alpha)$ 100% quantile of the distribution under the null hypothesis (no clustering). An exact measure of power is obtained as the probability of the test statistic exceeding the critical value under a specified alternative hypothesis (model of clustering). Often the sampling distribution of the test statistic is unknown or difficult to calculate, making exact power calculations impossible. However, if simulating the data under the null hypothesis is relatively easy (as is usually the case with disease clustering studies), simulation (Monte Carlo) studies offer a viable alternative to provide reliable estimates of power. This entails repeatedly simulating data under the null hypothesis and obtaining the value of the test statistic for each repetition, resulting in an empirical estimate of its sampling distribution under the null hypothesis. By simulating the data repeatedly under a specified alternative (clustering model), we estimate power by the proportion of data sets for which the resulting value of the test statistic exceeds the $(1-\alpha)$ $100th$ percentile of the empirical distribution obtained under the null hypothesis.

Bayesian methods are framed in a modeling rather than hypothesis-testing paradigm and do not easily offer 'detect/not detect' conclusions. To investigate performance, we simulate clustering using a well-known Scottish lip-cancer data set[9] and distribute the observed 536 cases among Scotland's 56 districts by specifying an exposure of 1.0 at an artificial focus and allowing the exposure to decrease to 0.0 with increasing distance from the focus. We propose a measure of 'power' for the Bayesian approach based on the empirical distribution of the median standardized mortality ratio (SMR) obtained from the posterior distribution of SMR in each region, simulated under a specification of no increased relative risk. In effect, we use the estimated SMR as a summary index of clustering and assess its performance in a manner similar to a hypothesis test using the estimated SMR as a test statistic. We then compare these measures with the frequentist power of a focused score test. In Section 2 we describe the data. We formulate our methods in Section 3. In Section 4 we present our findings, and conclude with a summary and discussion of various issues in Section 5.

2 Data

Clayton and Kaldor[9] first analyzed the incidence of male lip-cancer in each of the 56 districts of Scotland using an empirical Bayes method. The data contain observed and age-adjusted expected numbers of cases for the six year period from 1975 to 1980, resulting in a total of 536 cases in roughly 15 million person-years-at-risk[10] (pp. 536-537). The actual locations of the cases are unavailable. Instead, the data include the latitude and longitude of the centroid of each region (district), and therefore we assume all cases falling within a given region occur at its centroid. The regions are numbered 1 to 56, with one corresponding to the highest observed (crude) incidence rate. Figure 1 shows expected numbers of male lip-cancer cases. Higher numbers of expected cases occur in the more densely populated regions of Glasgow and Edinburgh, and surrounding districts. The rural regions to the north and south have fewer numbers of expected cases due to their smaller population sizes.

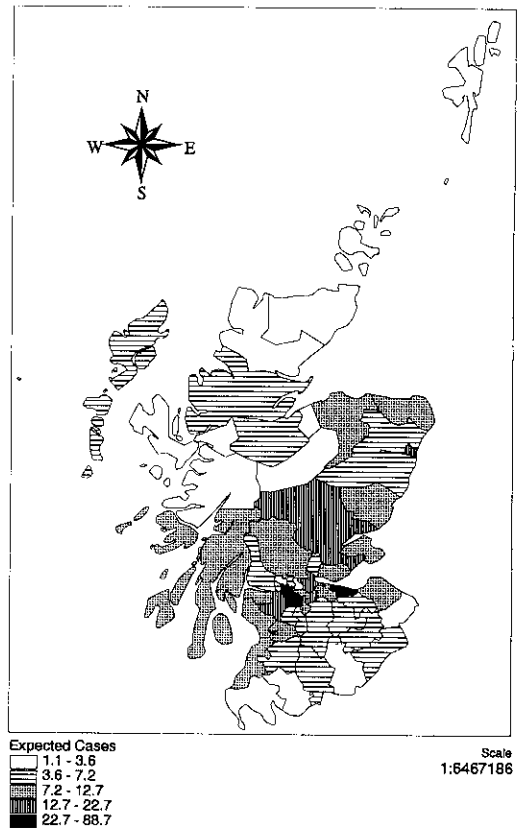


Figure 1 Regional age-adjusted expected number of lip cancer cases in Scotland from 1975 to 1980.

3 Methods

3.1 Frequentist null and alternative hypotheses

We consider a study region divided into k distinct regions, each with E_i expected cases and y_i observed cases, where y_i is the observed value of the random variable Y_i. The total numbers of expected and observed cases over the study region are denoted E_+ and y_+, respectively. We are interested in examining whether cases are significantly clustered around a focus i_0 by testing the null hypothesis

H_0: $\{Y_i\}$ are independent Poisson random variables with $E(Y_i)=E_i$, $i=1, \ldots, k$.

That is, under the null hypothesis each individual at risk is equally likely to be a case, and the expected number of cases varies only due to the heterogeneity in population density expressed in the E_i's.

The alternative hypothesis is

$$H_A: E(Y_i)=E_i \exp(\beta g_i), \tag{1}$$

where $\beta=\ln(RR_{i_0})$, RR_{i_0} is the relative risk of disease at the focus, and g_i is a measure of exposure between the ith region and the focus. True exposure information is rarely available, and Tango[6] proposes as a reasonable surrogate the function

$$a_{i_0,i}=\exp(-d_{i_0,i}/\tau), \tag{2}$$

a non-increasing function of the Euclidean distance $d_{i_0,i}$ between the centroid of the ith cell and the focus i_0, with scale parameter τ.

It is well-known that the null distribution of the $\{Y_i\}$ conditional on the observed number of cases, y_+, is multinomial. The conditional null hypothesis becomes

$$H_0: Y_1, \ldots, Y_k | y_+ \sim \text{multinomial} (y_+, E_1/E_+, \ldots, E_k/E_+) \tag{3}$$

with conditional alternative

$$H_A: Y_1, \ldots, Y_k | y_+ \sim \text{multinomial} (y_+, \pi_1, \ldots, \pi_k), \tag{4}$$

where

$$\pi_i=\frac{E_i \exp(\beta g_i)}{\sum_{i=1}^{k} E_i \exp(\beta g_i)}, \quad i=1, \ldots, k.$$

As in the unconditional alternative, a reasonable surrogate for the exposure function g_i is a decreasing function of distance, such as the exponential function $a_{i_0,i}$ previously described in (2).

3.2 Focused score test

Waller *et al.*[11] and Lawson[12] formulate a focused score test which is locally most powerful against the alternative stated in (1). The particular form of the test fits into the larger class of clustering statistics considered by Tango[6], and is given by

$$U=y+\sum_{i=1}^{k}\{g_i(r_i-p_i)\}$$

where $r_i=y_i/y_+$ and $p_i=E_i/E_+$. The statistic sums the differences in observed and expected proportions of cases, weighted by exposure to the focus. The variance of the score is the Fisher information given by

$$Var(U)=y_+\left[\sum_{i=1}^{k} g_i^2 p_i-\left(\sum_{i=1}^{k} g_i p_i\right)^2\right].$$

Following Tango[6], we standardize the score statistic giving

$$U_s=\sqrt{y_+}\left(\frac{a'(r-p)}{\sqrt{a' V_p a}}\right) \overset{a}{\sim} N(0,1),$$

where $a=(a_{i_0,1}, \ldots, a_{i_0,k})'$, $r=(y_1/y_+, \ldots, y_k/y_+)'$, $p=(E_1/E_+, \ldots, E_k/E_+)'$, and $V_p=\text{diag}(p)-pp'$. We obtain a large positive value of U_s when the observed proportions of cases are greater than expected, indicating possible clustering of cases around the focus. We reject the null hypothesis based on a one-sided upper-tail test conducted at $\alpha=0.05$.

3.3 Bayesian Hierarchical Model

Following Besag, York, and Mollié[13], we formulate a Bayesian hierarchical model of disease incidence where the $\{Y_i|\mu_i\}$ are conditionally independently distributed Poisson random variables with expected number of cases $E_i \exp(\mu_i)$, and the log relative risk μ_i is a linear combination of fixed and random effects. Specifically, $\mu_i=x_i'\beta+\theta_i+\phi_i$, where x_i represents a vector of covariate values associated with region i, β is the corresponding vector of parameters, θ_i is a residual associated with excess heterogeneity, and ϕ_i is a residual measuring spatial similarity. We assign independent, mean zero normal prior distributions to the excess heterogeneity random effects, $\{\theta_i\}$. These random effects incorporate extra Poisson variation into the model, perhaps due to omitted covariates effecting the probability of disease, but not in any clear spatial pattern. In contrast, the spatial residuals, $\{\phi_i\}$, model the effect of unmeasured covariates related to disease incidence that, if observed, would be more similar for neighboring regions than for those selected at random. We model the ϕ_i's using a Gaussian conditional autoregressive (CAR) prior[14], where the expected value of $\phi_i|\phi_{j\neq i}$ is simply the average of the residuals in contiguous regions and the variance is proportional to the number of neighboring regions.

In our application, the covariate of interest is expo-

sure to the focus, resulting in the following full specification of the model:

$$Y_i|\mu_i \overset{ind}{\sim} \text{Poisson}\,(E_i \exp(\mu_i)),$$
$$\mu_i=\beta g_i+\theta_i+\phi_i,$$
$$\theta_i \overset{ind}{\sim} N(0,\eta),$$
$$\phi_i|\phi_{i\neq j}\sim N\left(\frac{\sum_{j\sim i, j\neq i}\phi_j}{m_i}, \frac{\kappa}{m_i}\right),$$
$$\eta^{-1}\sim \text{Ga}(0.1, 0.1),$$
$$\kappa^{-1}\sim \text{Ga}(0.1, 0.1),$$
and
$$\beta\sim N(0, 10000),$$

where the notation $j\sim i$ indicates that region j is adjacent to region i, and m_i is the total number of such regions associated with region i. The hyperprior distributions for η and κ are similar to those used by Best *et al.*[8] where *Ga (a,b)* denotes a Gamma distribution with mean a/b and variance a/b^2.

We fit the hierarchical model by implementing a Markov Chain Monte Carlo (MCMC) method known as Gibbs sampling using the BUGS (Bayesian inference Using Gibbs Sampler) software package[15),16)]. We assessed convergence at several levels of relative risk using Gelman and Rubin's[17)] statistic, trace plots, and lag 1 autocorrelation. We used a burn-in of 2000, and retained the last 10,000 iterations of the chain. Finally, note that although we simulated data under various alternatives by specifying values of the exposure parameter β, we intentionally fit the Bayesian model with the exposure covariate excluded, thereby allowing the trend to be expressed as spatial correlation through the ϕ_i's. Our goal is to apply the method in a setting where the investigator does not know the true form of the cluster (as measured by the exposure function), and to see if the spatial residuals (the ϕ_i's) will detect the cluster we create in the simulated data.

3.4 Simulations

To assess performance of the testing and the modeling approaches, we simulate disease counts in the Scotland data based on a single focus in the region containing Glasgow (district 49). Region 49 has nine adjacent neighbors ranging in distance from 0.13 to 0.47 units from its centroid. Figure 2 shows the locations of Glasgow and its neighboring districts relative to the rest of Scotland. We model exposure to district 49 according to the exposure function described by Tango[6)], namely

$$g_i=\exp\left(-4\left(\frac{d_{49,i}}{L}\right)^2\right),$$

where L is the maximum distance from the focus with an appreciable increase in risk. We chose $L=0.5$ units which approximately corresponds to the maximum

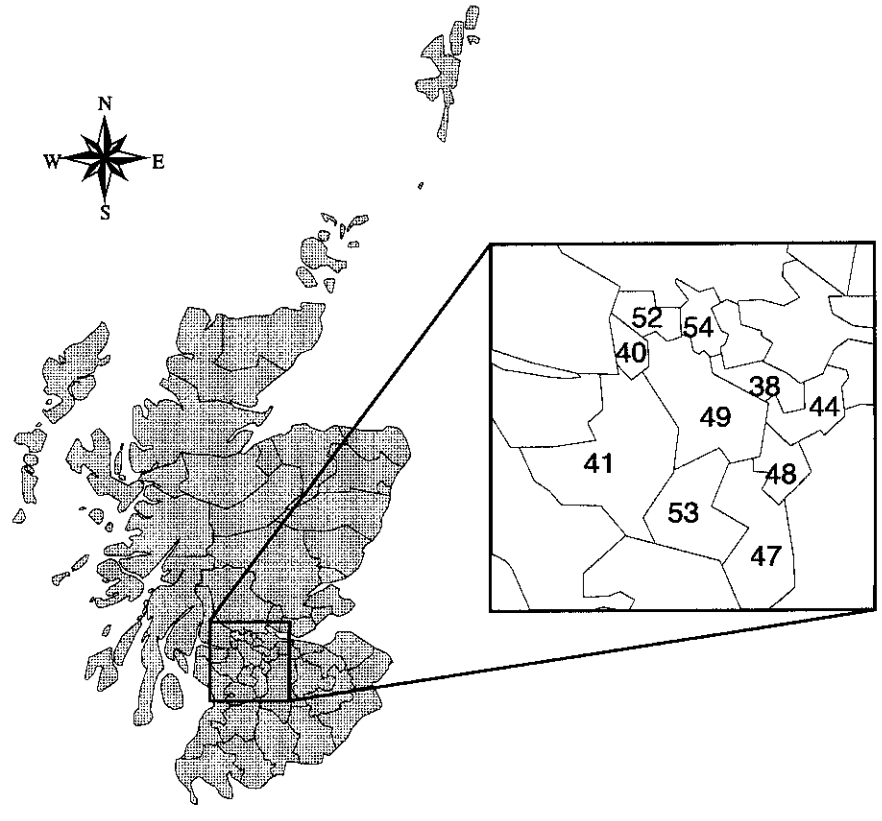


Figure 2 Glasgow (region 49) and its nine adjacent neighbors.

distance between the centroid of region 49 and the adjacent districts. For districts beyond a distance of 0.5 units from the 49th, exposure to the focus was essentially zero. We specified a level of relative risk at the focus in each simulation, which in turn established the associated value of β (recall $\beta=\ln(RR_{49})$). For the focused score test, we specified a total of 100 distinct levels of relative risk between 1.0 and 3.0 and for the Bayesian hierarchical model, we specified relative risk at the focus equal to 1.0, 1.1, ..., 1.9, 2.0, and 3.0. We then distributed the 536 cases among the 56 regions based on the conditional alternative specified in (4).

For the focused score test, we selected $\tau=0.2$ units in (2) so that distances beyond 0.5 units (the simulated exposure radius) received little weight. As with the hierarchical model, we define a score test as might be applied in the absence of detailed exposure information, rather than define the test specific to the exposure function generating the data. Using 10,000 Monte Carlo simulations at each level of increased risk, we calculated the empirical power of the focused score test to detect the alternative as the proportion of simulated data sets for which we obtained a significant result at $\alpha=0.05$.

As the Bayesian model does not result in a dichotomous 'detect clustering/fail to detect clustering' result, we use the posterior distribution of region specific SMRs obtained from the hierarchical model in a manner similar to a frequentist test statistic. We do not propose this approach as a competing *testing* approach, but rather use the approach to determine how often estimates arising from the hierarchical model would trigger clustering conclusions when in fact clustering does exist.

Let M_i be the median of the posterior distribution of SMR in region i. We simulated data according to the conditional null hypothesis stated in (3) and fit the hierarchical model. We obtained the median of the posterior distribution of SMR in region i and repeated the process 1000 times, providing an empirical estimate of the sampling distribution of M_i under the probability model based on the null hypothesis of no clustering (3). We generated data 100 times for each alternative value of relative risk according to (4), fit a hierarchical model, and obtained the collection $\{m_{i,1}, \ldots, m_{i,100}\}$, where $m_{i,j}$ is the jth realization of the random variable M_i under a specified alternative ($j=1, \ldots, 100$). Let $m^{(0)}_{i,(950)}$ be the $95th$ percentile of the estimated sampling distribution of M_i under the null hypothesis. Using this as a critical value, empirical 'power' in region i was defined in the usual manner as the proportion of $\{m_{i,1}, \ldots, m_{i,100}\}$ exceeding $m^{(0)}_{i,(950)}$. Note that due to the computational burden associated with each MCMC implementation of the Bayes model, fewer simulated data sets were used for each level of relative risk than in the implementation of the focused score test.

4 Results

Figure 3 compares empirical power of the focused score test and the 'power' of M_{49} for detecting various departures from the null model when the focus being tested is correctly specified (Glasgow). Again, we note that while the focus *location* is correctly specified, neither approach uses precisely the same cluster model which generates the data. From Figure 3, it appears the focused score test and the M_{49}-based test have comparable power for detecting clustering at a correctly specified region. We note that since the Bayesian model has far fewer estimates of 'power' in the lower tail of the plot the behavior of the Bayesian model in this region is not very well characterized and therefore it is difficult to tell with certainty which test dominates here.

In addition to the power of the methods to detect clustering about a focus, the ability to discriminate the true focus from a collection of close regions, should also be considered when evaluating different methods. Figure 4 plots the power for the focused score test centered at regions: 49 (true focus), 54, 41, and 44. As can be seen in Figure 4, the focused score test has high power for detecting clusters even when the test's focus is the centroid of a region away from the true focus of the clustering; i.e., the test does not discriminate well between the true focus and a focus located at the centroid of a neighboring region. In contrast, Figure 5, a plot of the 'power' of the M_i-based approach for regions close to the true focus, shows much better discrimination between the true focus and neighboring regions. This discrimination ability can also be see in Figure 6 which maps the 'power' of M_i for regions $i=54$, 44, *and* 41, all close to the true focus. However, it should be pointed out that the ability of the focused score test to differentiate between close regions is largely due to the weight vector used. In order to get discrimination properties similar to the M_i-based test we decreased τ to 0.03 units, a level which assigns appreciable weight only at the focus. Figure 7 plots the same power curves as are plotted in Figure 4 but with $\tau=0.03$. This will, in effect, decrease the range at which regions receive significant weighting. Notice, however, that the power of the correctly specified test decreases as well.

Another benefit of the Bayesian modeling paradigm is that one has the entire posterior distribution of quan-

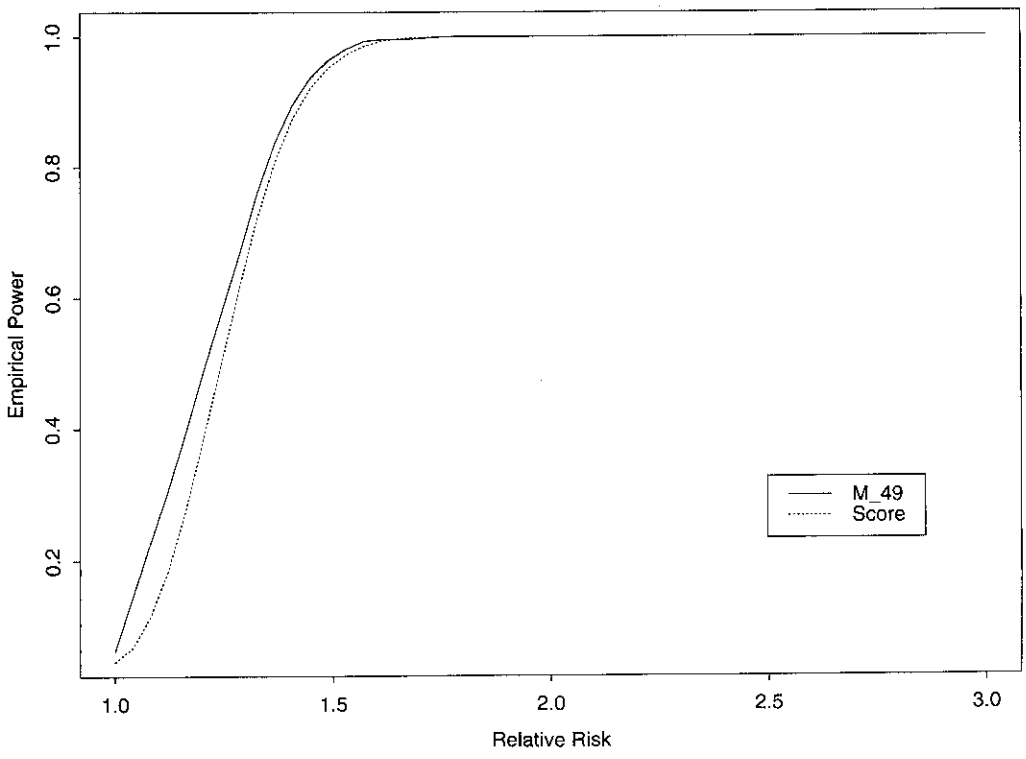


Figure 3 Power curve for the focused score test and 'power' (see text) for the index M_{49} (posterior median SMR) based on a hierarchical Bayesian model.

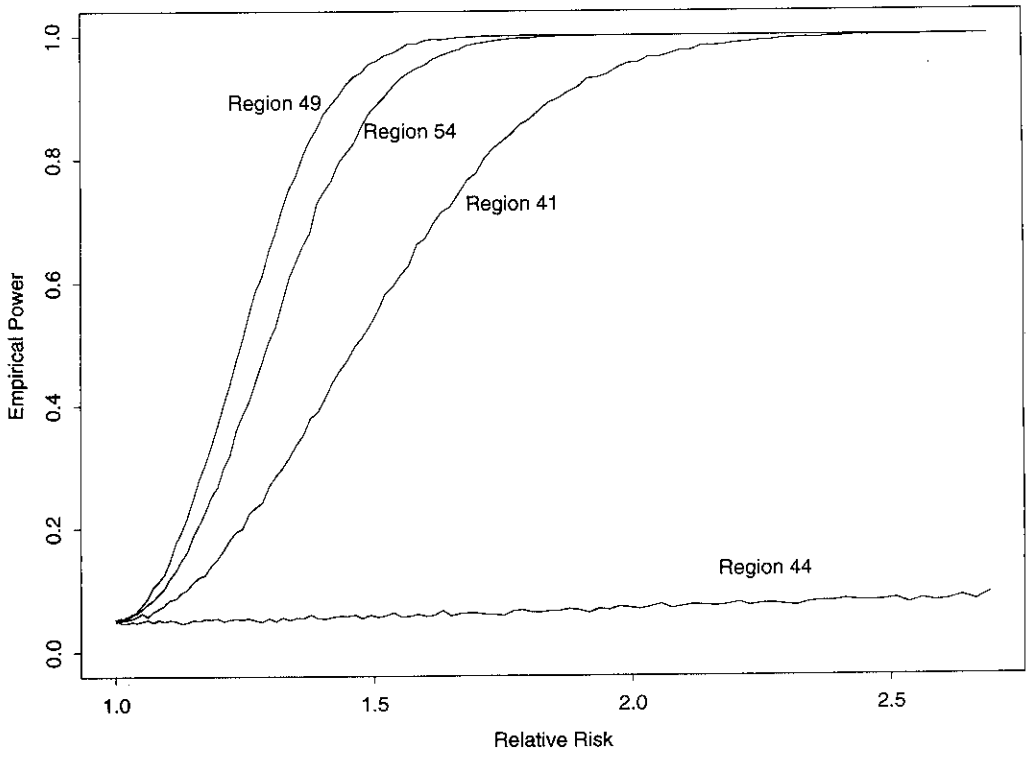


Figure 4 Empirical power of the focused score test where $\tau=0.2$ (see text). Each graph presents power where the focus of the test is specified by the named region, but we simulate data in each case with the focus centered at Glasgow (region 49).

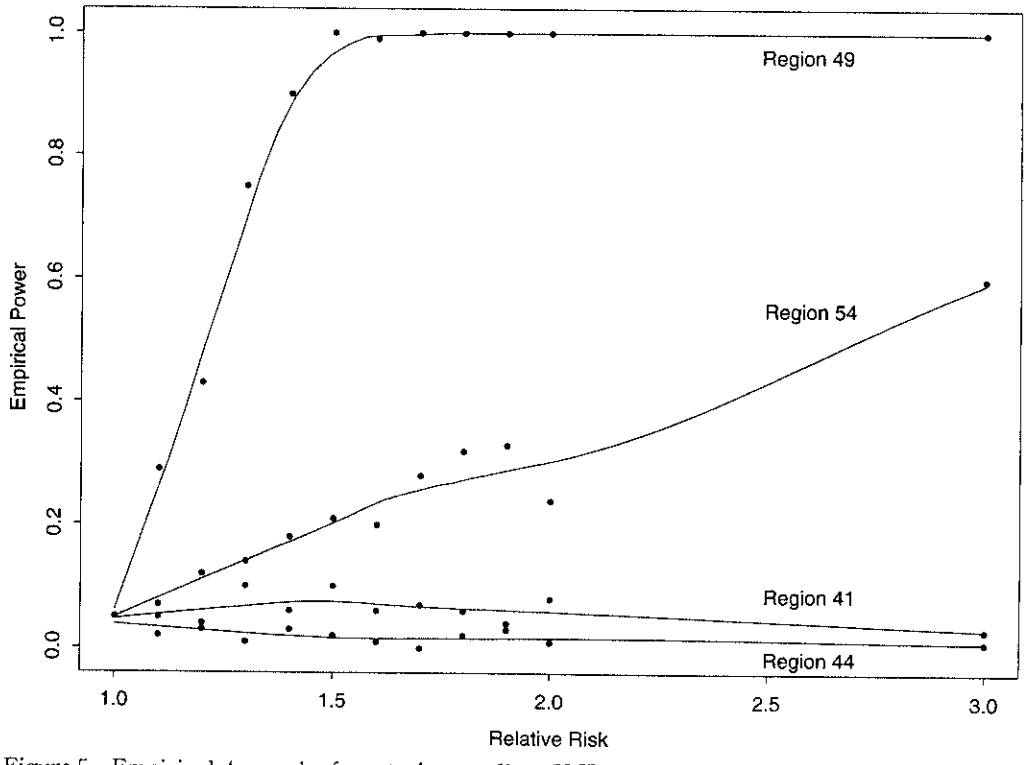


Figure 5 Empirical 'power' of posterior median SMRs M_{49}, M_{54}, M_{41}, and M_{44}. Each graph presents power where the focus of the test is specified by the named region, but we simulate data in each case with the focus centered at Glasgow (region 49).

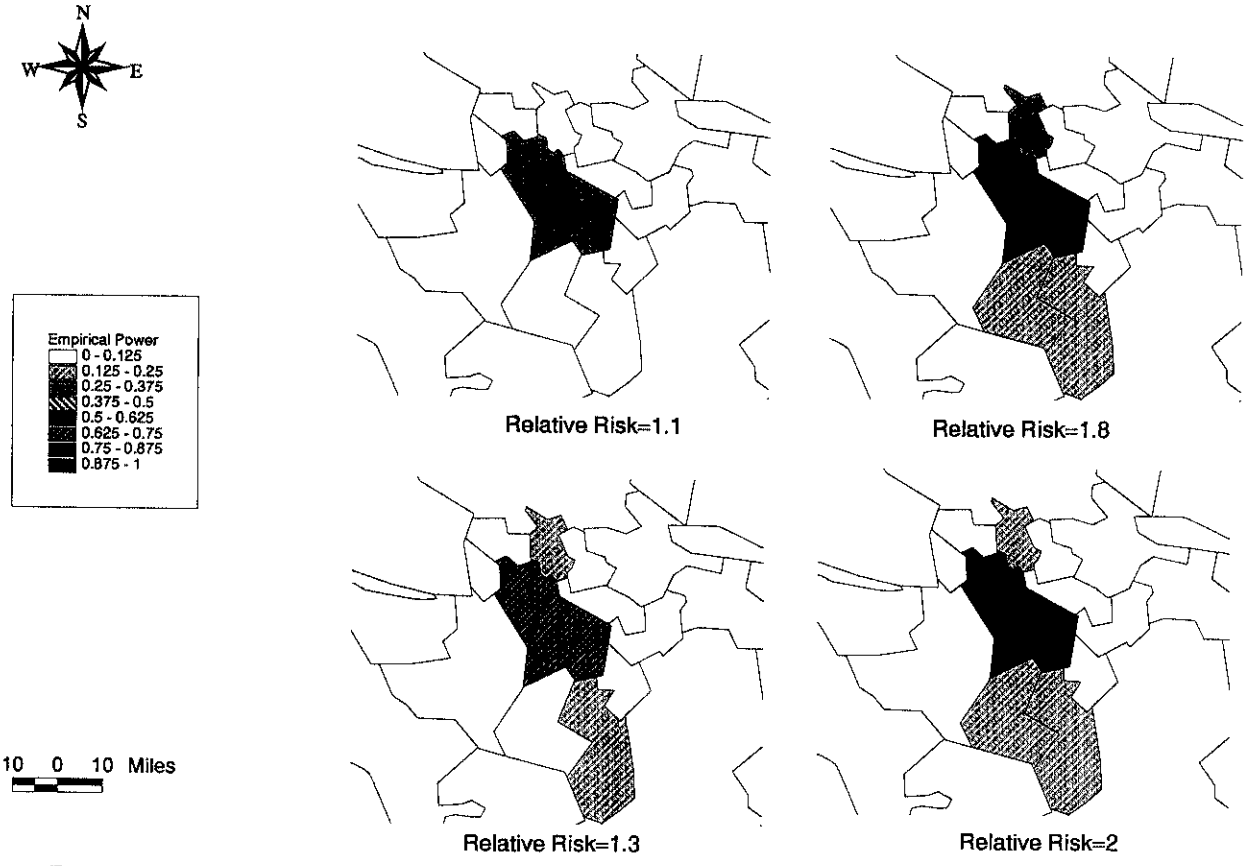


Figure 6 Map of empirical 'power' of M_i for the simulated focus (Glasgow) and its nine adjacent neighbors.

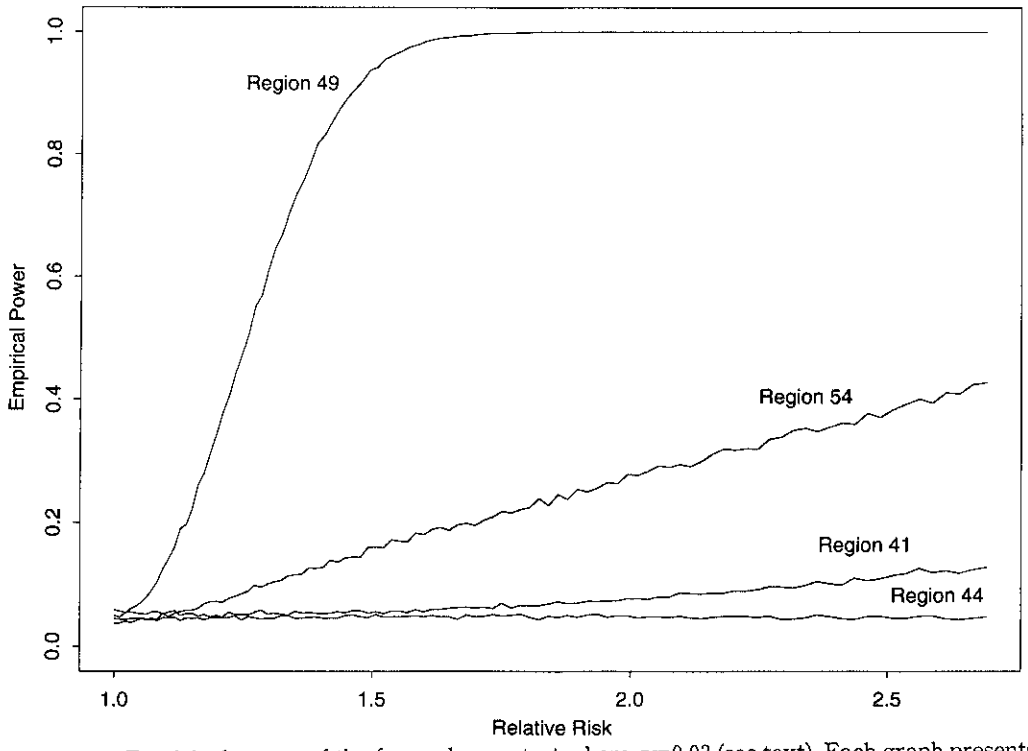


Figure 7 Empirical power of the focused score test where $\tau=0.03$ (see text). Each graph presents power where the focus of the test is specified by the named region, we simulate the data in each case with the focus centered at Glasgow (region 49).

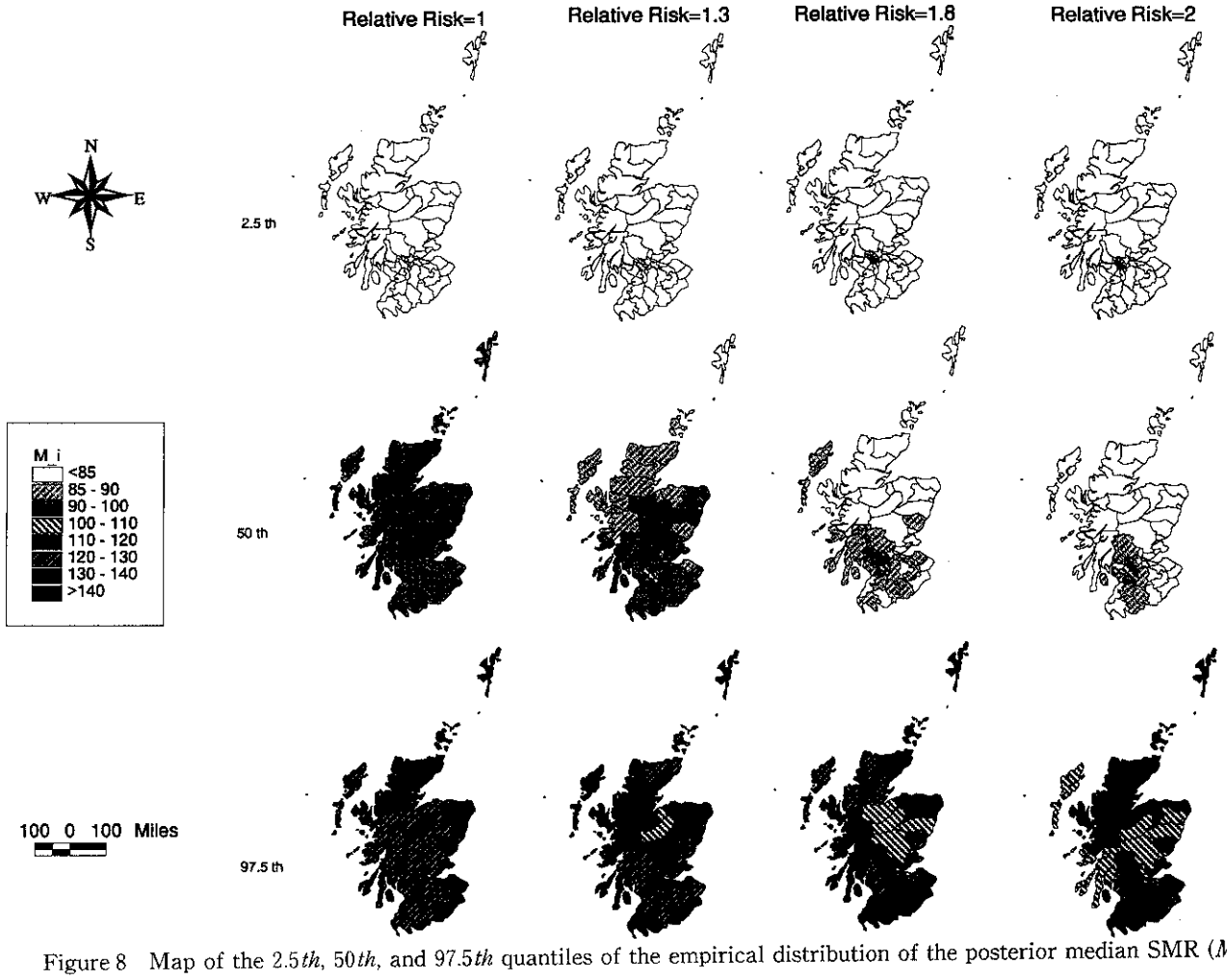


Figure 8 Map of the 2.5*th*, 50*th*, and 97.5*th* quantiles of the empirical distribution of the posterior median SMR (M_i).

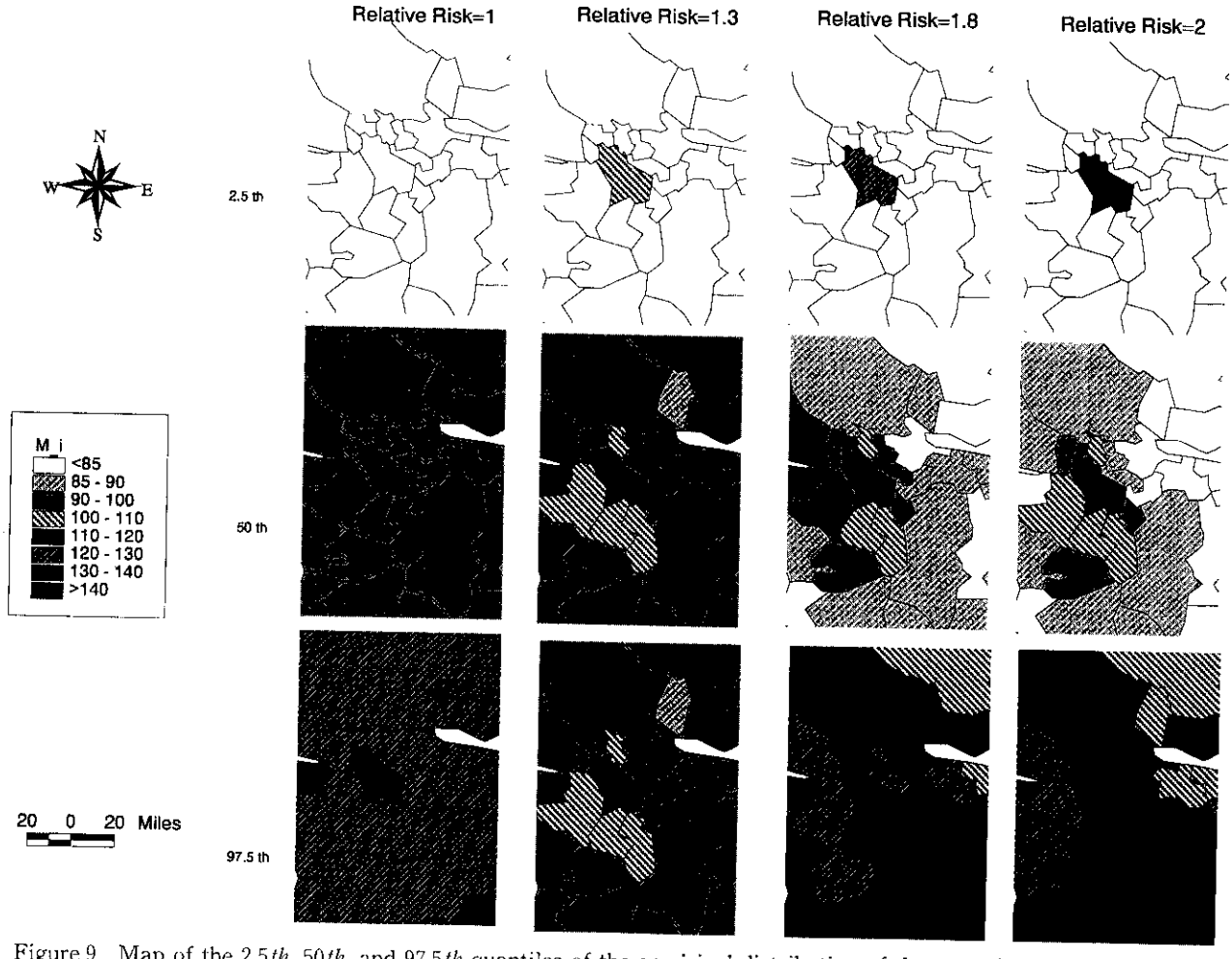
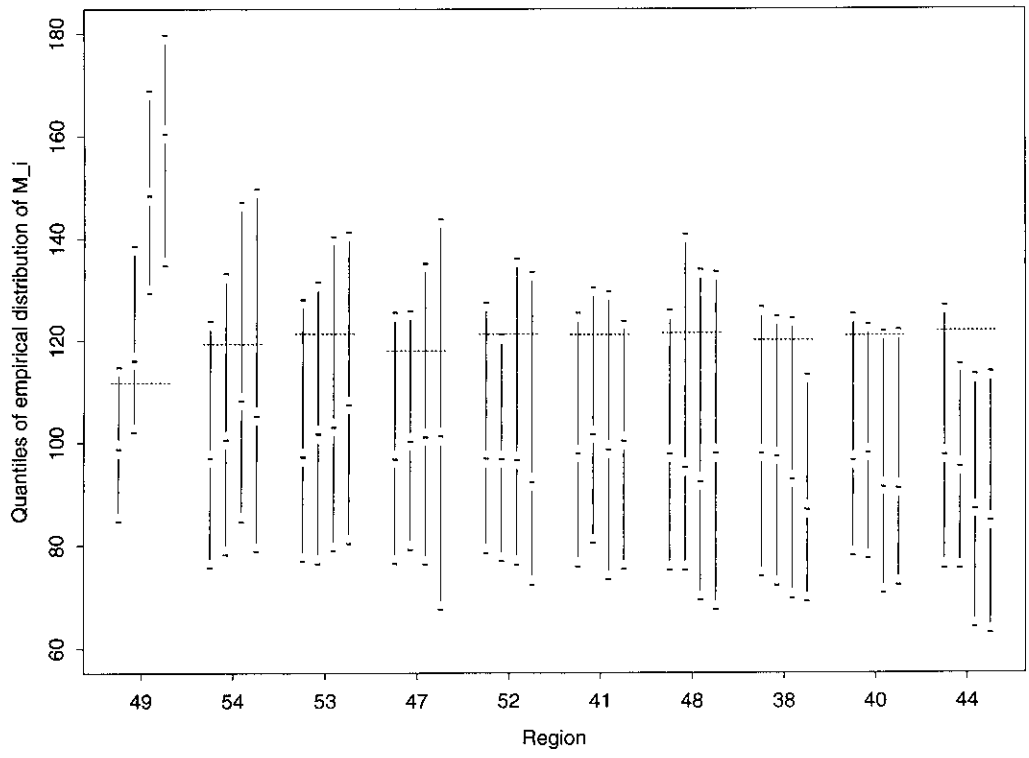


Figure 9 Map of the 2.5*th*, 50*th*, and 97.5*th* quantiles of the empirical distribution of the posterior median SMR (M_i) in the regions surrounding the simulated focus.

tities of interest for making maps such as Figures 8 and 9. What is immediately evident from Figures 8 and 9 reaffirms the point of the preceding paragraph; i.e., increasing levels of SMR associated with increasing risk are highly localized about the focus of our exposure, even though we do not specify the location of the focus in the Bayesian model formulation. The top row indicates the lower 2.5*th* percentiles of the empirical distribution of M_i for $i=1, \ldots, 56$, the middle row the 50*th* percentiles, and the bottom row the 97.5*th* percentiles. The top row indicates that only the estimate of M_i for $i=49$ offers strong evidence of increased risk. In most applications of the Bayesian approach, investigators map only the median values (middle row of Figures 8 and 9). In contrast, we note that the map of lower percentiles offers a clearer picture of the actual cluster center (highest increased risk) than the map of median values, especially for relatively modest increases in risk. We also note that our use of the multinomial alternative hypothesis results in lower M_i's when i denotes areas away from the focus, since the multinomial hypothesis requires allocation of a fixed number of cases. To increase incidence near the focus, cases become less likely than expected in areas away from the focus. This effect does not occur if one allows the total number of cases to vary between simulated data sets.

Figure 10 demonstrates the discriminatory ability of the Bayesian approach by plotting the empirical distribution quantiles of M_i for region 49 and its 9 adjacent neighbors under four levels of increased risk. We order regions by increasing distance from 49 from left to right. The dashed lines indicate the 95th percentile of the sampling distribution of each M_i, $i=49$, 54, 53, 47, 52, 41, 48, 38, 40, and 44. (Coincidentally, these values are all near a SMR of 120, a value suggested by collaborating epidemiologists as the smallest 'interesting' increase in SMR!) Although many of the closest neighbors' posterior SMRs increase above 100 with increasing risk, Glasgow (region 49) sees by far the largest jump, and is the only one clearly above the tail of the null sampling distribution of M_i.



Fgure 10 2.5*th*, 50*th*, and 97.5*th* quantiles of the empirical distribution of the posterior median SMR (M_i) for the simulated focus (region 49) and its nine adjacent neighbors in order of increasing distance from the focus. Each group represents, from left to right, the interval obtained from a relative risk specified at the focus of 1.0, 1.3, 1.8, and 2.0. The dashed line corresponds to the 95*th* percentile of the empirical distribution under the null hypothesis.

5 Discussion

In results not shown here and as expected, the focused score test appears to have better power (as defined above) to detect clustering than M_i when the exposure is correctly specified. However, in routine application, the 'true' clustering model is never known. The results above indicate that, if the form of the exposure is mis-specified (but close to the true exposure-distance relationship), the power of the score tests is comparable to the 'power' (as defined above) of M_i when the centroid of region i corresponds to the focus. If the focus *location* is mis-specified in the focused score test, there is a risk in concluding that cases are significantly clustered around the wrong focus. This lack of discrimination can be minimized by decreasing the distance (from the focus) at which regions receive large weight (i.e. high exposure areas). We could decrease the value of τ, or consider different parametric forms of the weight vector, one that decreases more quickly with distance. However, both of these modifications can decrease the power of the focused score test to detect clustering when the focus location is correctly specified. Similarly, the 'power' of M_i will change if the focus location does not correspond to a district centroid, e.g. the method may have low 'power' versus a focus near the boundary between geographically large regions. Additional simulation studies clarifying these issues are the subject of future research.

Our results indicate that even though the hierarchical model 'smooths' small area disease rates by allowing spatial similarity, such rate estimate stabilization does not come at the cost of ability to detect aberrant observations. The 'power' to detect unusually high SMRs is comparable to that of the score test even for low, environmentally tenable, levels of risk. This along with the Bayesian model's ability to discriminate between regions and the richness of the results obtained ('smoothed' maps of all SMRs with associated posterior probability intervals), makes the Bayesian hierarchical an attractive approach for the spatial analysis of regional disease data. Arguing against the use of the Bayes approach is its computational burden, which is far greater than required by the focused score test. If one desires a 'quick and dirty' test of clustering about a focus, the focused score test should provide good results

as long as the range of the exposure function is considered so as to minimize the chances of detecting significant clustering about the wrong focus. If however, one wishes to map stabilized estimates and search for areas of observed increased risk across the study area, and one has access to software like BUGS, the hierarchical Bayes approach offers a wider choice of inferential opportunities without the loss of sensitivity to local clustering.

Acknowledgements

This research was supported in part by the National Institute of Environmental Health Sciences grant 1-R01-ES07750 (LAW), and National Institutes of Health training grant T32-AI07442 (EGH). The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIEHS or NIH.

References

1) A.B. Lawson and L.A. Waller. A review of point pattern methods for spatial modelling of events around sources of pollution. *Environmetrics*, 7: 471-487, 1996.
2) P. Elliott, M. Martuzzi, and G. Shaddick. Spatial statistical methods in environmental epidemiology: A critique. *Statistical Methods in Medical Research*, 4: 137-159, 1995.
3) R.J. Marshall. A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society, Series A*, 154: 421-441, 1991.
4) J. Besag and J. Newell. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society, Series A*, 154: 143-155, 1991.
5) L.A. Waller and A.B. Lawson. The power of focused tests to detect disease clustering. *Statistics in Medicine*, 14: 2291-2308, 1995.
6) T. Tango. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine*, 14: 2323-2334, 1995.
7) D. Clayton and L. Bernardinelli. *Geographical and Environmental Epidemiology*, chapter Bayesian methods for mapping disease risk. Oxford University Press, Oxford, 1992.
8) N.G. Best, R.A. Arnold, A. Thomas, L.A. Waller, and E. M. Conlon. To appear in *Bayesian Statistics 6*, chapter Bayesian models for spatially correlated disease and exposure data. Oxford University Press, Oxford, 1999.
9) D. Clayton and J. Kaldor. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43: 671-681, 1987.
10) N.A.C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, Inc., 1993.
11) L.A. Waller, B.W. Turnbull, L.C. Clark, and P. Nasca. Chronic disease surveillance and testing of clustering of disease and exposure: application to leukemia incidence and tce-contaminated dumpsites in upstate New York. *Environmetrics*, 3: 281-300, 1992.
12) A.B. Lawson. On the analysis of mortality events associated with a prespecified fixed point. *Journal of the Royal Statistical Society Series A*, 156: 363-377, 1993.
13) J. Besag, J.C. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43: 1-109, 1991.
14) J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36: 192-236, 1974.
15) D.J. Spiegelhalter, A. Thomas, N. Best, and W.R. Gilks. Bugs: Bayesian inference using Gibbs sampling, version 0.50. Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University, 1995a.
16) D.J. Spiegelhalter, A. Thomas, N. Best, and W.R. Gilks. Bugs examples, version 0.50. Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University, 1995b.
17) A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7: 457-511, 1992.

本論文では，危険物廃棄施設のように，事前にある疾病のリスク増大が疑われる特定された施設の周辺における疾病集積性の検出のための二つのアプローチを比較している．最初の方法は，スコア検定である．この方法は，簡単に計算でき，かつ仮説検定のパラダイムに基づいた方法である．もう一つの方法は，地域差に変量模型を導入した階層的Bayesianモデルから導出される地域別SMR（標準化死亡比）の推定値の疾病地図に基づく方法である．この方法は，スコア検定より計算上の苦労は大きいものの，仮説検定特有の「集積性があった，なかった」という単純な2値の結論よりは，はるかに柔軟な推測が可能であると主張している．本論文では，両者のアプローチの比較をFrequentist（頻度論者，Bayesianとの対比）流の方法により検出力を比較した．後者の検出力は，Bayesianモデルで計算されたSMR（事後分布の中央値）が，集積性が無いという帰無仮説の下で計算される95パーセンタイルを越える割合として計算している．シミュレーションの結果，ほとんど類似の検出力を示したが，真の集積性の位置を推定する精度はスコア検定よりやや優れていると結論している．