ENVIRONMETRICS
Environmetrics 2004; **15**: 797–810
Published online 19 August 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/env.662

Likelihood-based tests for localized spatial clustering of disease

Ronald E. Gangnon[1]*,† and Murray K. Clayton[2]

[1]*Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, 610 N. Walnut Street, Madison, Wisconsin 53726, U.S.A.*
[2]*Departments of Statistics and Plant Pathology, University of Wisconsin–Madison, 1210 W. Dayton Street, Madison, Wisconsin 53706, U.S.A.*

SUMMARY

Numerous methods have been proposed for detecting spatial clustering of disease. Two methods for likelihood-based inference using parametric models for clustering are the spatial scan statistic and the weighted average likelihood ratio (WALR) test. The spatial scan statistic provides a measure of evidence for clustering at a specific, data-identified location; it can be biased towards finding clusters in areas with greater spatial resolution. The WALR test provides a more global assessment of the evidence for clustering and identifies cluster locations in a relatively unbiased fashion using a posterior distribution over potential clusters. We consider two new statistics which attempt to combine the specificity of the scan statistic with the lack of bias of the WALR test: a scan statistic based on a penalized likelihood ratio and a localized version of the WALR test. We evaluate the power of these tests and bias of the associated estimates through simulations and demonstrate their application using the well-known New York leukemia data. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: Bayes factor; disease clustering; hypothesis testing; multiple testing; Poisson; simulation

1. INTRODUCTION

Disease clustering studies are typically approached as hypothesis testing problems. The null hypothesis of no clustering, i.e. a common rate of disease across the study region, is tested against an alternative hypothesis of clustering. The hypothesis of clustering is very broad, and many definitions of clustering have appeared in the literature. Besag and Newell (1991) and Kulldorff (1998) have classified testing procedures into three classes based on characteristics of the clustering hypothesis: (i) tests for focused clustering; (ii) general tests for global clustering; and (iii) general tests for localized clusters. Here, we consider only general tests for localized clusters based on likelihood ratio test statistics. For these general tests, the location of the potential cluster cannot be specified in advance, and the goal of the study is to determine whether the disease rate is elevated in one of a large number of potential clusters.

Kulldorff and Nagarwalla (1995) and Kulldorff (1997) proposed using a scan statistic based on a likelihood ratio test from a Poisson model, allowing for essentially any arbitrary collection of potential

*Correspondence to: R. E. Gangnon, Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, 610 N. Walnut Street, Madison, Wisconsin 53726, U.S.A.
†E-mail: ronald@biostat.wisc.edu

Received 12 July 2002
Accepted 15 January 2004

clusters. For every potential cluster, one calculates the likelihood ratio statistic for testing the null hypothesis of a common rate in all cells against the alternative of two rates—one rate for cells outside the cluster and a second higher rate for cells inside the cluster. The potential cluster associated with the largest likelihood ratio is the estimated cluster. Significance levels are evaluated through simulation.

As an alternative to the scan statistic, Gangnon and Clayton (2001) proposed using the weighted average of the likelihood ratios (WALR) as a test statistic. The WALR statistic has an appealing Bayesian interpretation as an approximation to the marginal likelihood ratio (or Bayes factor) of the composite one cluster model to the no clustering model. If the null hypothesis is rejected, one can exploit this Bayesian interpretation of the statistic to identify cluster locations and estimate cell-specific risks using the posterior distribution on clusters. Gangnon and Clayton (2001) demonstrate that the posterior estimates associated with the WALR test can satisfy a criterion for unbiased cluster detection and, in most settings, the WALR test is more powerful than the scan statistic. However, the posterior estimates associated with the WALR test need not always identify a single area of clustering associated with the rejection of the null hypothesis.

In this article, we consider two new general testing procedures for localized clusters based on the likelihood: the penalized scan statistic and the weighted average likelihood ratio scan statistic. In both cases, we attempt to define procedures that combine the specificity of the scan statistic with the unbiasedness of the WALR statistic. In Section 2, we discuss likelihood-based cluster detection methods, both old and new. In Section 3, we evaluate both the bias of the estimators associated with these tests and the power of these four likelihood-based tests via simulation. In Section 4, we apply the new testing procedures to the well-known New York leukemia data set. In Section 5, we provide some brief concluding remarks.

2. LIKELIHOOD-BASED TESTS

We begin by defining some terminology and a basic statistical model. We consider situations in which the study region is divided into N subregions, or cells. For each cell i, we observe O_i, the number of cases of disease, and n_i, the population at risk in cell i. We assume a Poisson model for the data, i.e. $O_i \sim \text{Poisson}(\rho_i n_i)$, where ρ_i is the disease rate in cell i. This Poisson model allows for the inclusion of covariate effects by replacing n_i with the expected case count E_i and reinterpreting ρ_i as a standardized incidence ratio.

We are interested in testing the null hypothesis of a constant rate of disease across all cells, i.e. $H_0 : \rho_1 = \rho_2 = \cdots = \rho_N$. For the methods we consider here, we consider a family of possible alternative hypotheses defined in terms of a collection of potential clusters or subsets of the N cells. For each potential cluster, we consider the two-parameter Poisson model which assigns $\rho_i = \lambda_1$ if cell i belongs to the cluster and $\rho_i = \lambda_0$ if cell i does not belong to the cluster.

For each potential cluster, we then calculate the ratio of the likelihood under the alternative hypothesis $\lambda_1 \neq \lambda_0$ to the likelihood under the null hypothesis $\lambda_1 = \lambda_0$, which is equivalent to the global null hypothesis $\rho_1 = \rho_2 = \ldots \rho_N$. The likelihood ratio for cluster k, denoted by LR_k, equals

$$\left(\frac{O_c/n_c}{O_t/n_t}\right)^{O_c} \left\{\frac{(O_t - O_c)/(n_t - n_c)}{O_t/n_t}\right\}^{O_t - O_c}$$

where O_c (n_c) is the case (respectively, population) count inside cluster c and O_t (n_t) is the total case (respectively, population) count in the study region.

To make this discussion concrete, we will now describe a specific set of potential clusters, although we note that the methods discussed are all suitable for any arbitrary collection of potential clusters. In particular, we consider circular clusters centered at the cell centroids as potential clusters. The radius of the circles varies continuously from zero up to a fixed maximum radius, $r_{\max}$. If the centroid of a cell falls within the circle, then the whole cell is included in the cluster. Since there are only a finite number of cells, there will only be a finite number of clusters about each cell centroid. To identify these clusters, let $0 = r_{i,(1)} < r_{i,(2)} < \cdots < r_{i,(m_i)} \leq r_{\max}$ be the ordered distances from the centroid of cell i to the centroids of all cells, truncated at $r_{\max}$. (If two or more centroids are equidistant from the centroid i, the common distance is only listed once.) Then, the distinct potential clusters about cell i are circles of radii $r_{i,(1)}, r_{i,(2)}, \ldots, r_{i,(m_i)}$. We refer to the cluster centered at the centroid of cell i of radius $r_{i,(j)}$ as cluster i,j for $j = 1, 2, \ldots, m_i$ and $i = 1, 2, \ldots, N$.

Several other methods for defining a set of potential clusters have been described in the literature, and the methods described in this article can be applied with any of them. Kulldorff and Nagarwalla (1995) define a similar collection of clusters, but instead of using a single maximum radius, they define a separate maximum radius for each centroid so that the largest circle includes no more than a fixed percentage (e.g. 50%) of the population. Turnbull *et al.* (1990) and Besag and Newell (1991) define more limited sets of clusters using a single population or case radius, respectively, about each centroid.

2.1. The scan statistic

For a general test for localized clusters, Kulldorff and Nagarwalla (1995) and Kulldorff (1997) suggested using the maximum likelihood ratio over all potential clusters, $\mathrm{LR}_{\max} = \max_{i,j} \mathrm{LR}_{i,j}$, as a test statistic. The distribution of this statistic under the null hypothesis of no clustering can be found through Monte Carlo simulation, conditional on the total number of cases observed. If the null hypothesis is rejected, the cluster associated with $\mathrm{LR}_{\max}$ (effectively, the maximum likelihood estimate of the true cluster) may be viewed as the estimated cluster location. This MLE cluster is best viewed as only an approximate cluster location, because the many clusters overlapping the MLE cluster will likely also be associated with large likelihood ratios.

2.2. The weighted average likelihood ratio (WALR) statistic

As an alternative to the scan statistic, Gangnon and Clayton (2001) proposed using the weighted average of the likelihood ratios (WALR) statistic, $\mathrm{WALR} = \sum_{i=1}^{N} \sum_{j=1}^{m_i} w_{i,j} \mathrm{LR}_{i,j}$, where $w_{i,j} \geq 0$ is a known weight associated with cluster i,j. As with the $\mathrm{LR}_{\max}$ test, in our applications, the null distribution of the WALR test is found by Monte Carlo simulation. As noted by Gangnon and Clayton (2001), the WALR statistic has an appealing Bayesian interpretation as approximating the marginal likelihood ratio (or Bayes factor) of the composite one cluster model to the no clustering model. If we view the $w_{i,j}$ as a prior distribution on the space of potential clusters, the WALR statistic is proportional to the approximate marginal likelihood of the composite one cluster model. In this sense, WALR is a reasonable measure of the evidence for a single cluster model in the data.

Gangnon and Clayton (2001) more fully exploit the Bayesian interpretation of the weights to interpret the data if the null hypothesis is rejected. Specifically, after rejection, one conditions on the composite one cluster model being correct and formally treats the $w_{i,j}$ as a prior distribution on the clusters. The posterior probability of the true cluster being cluster i,j is approximately $p_{i,j} = w_{i,j}\mathrm{LR}_{i,j}/\mathrm{WALR}$. Using these probabilities, we can average over the single cluster models to obtain estimates of the probability that a cell belongs to the cluster and the cell-specific disease rates.

These statistics pinpoint the general location of a single cluster, indicate our uncertainty about the exact composition of the cluster, and quantify the risks associated with the cluster.

The WALR procedure can be applied using any set of weights. In some settings, prior information regarding cluster locations may be available and should be used to develop weights for the specific problem. However, in many, if not most, practical settings, there is either little prior information regarding cluster locations or one wants to select a set of weights which do not favor clusters in one area over clusters in another. In such settings, we would recommend the following set of weights, which is motivated by a simple scheme for selecting a cluster 'uniformly' from the available clusters. To select a cluster, we first select a point from a uniform distribution over the study area and make the centroid of the cell to which the point belongs the cluster center. The radius is then selected at random from a uniform distribution on $[0, r_{\max}]$. Then the weight for cluster i,j, $w_{i,j}$, is defined to be the probability that cluster i,j is selected by this procedure. Note that

$$w_{i,j} = \frac{a_i}{A} \cdot \frac{r_{i,j+1} - r_{i,j}}{r_{\max}}$$

where a_i is the area of cell i, A is the area of the study region and $r_{i,m_i+1} = r_{\max}$. In this article, we will restrict our consideration to the use of these weights.

2.3. The penalized scan statistic

The principal advantage of the scan statistic is its specificity; that is, it identifies a single cluster responsible for the rejection of the null hypothesis. However, as noted by Gangnon and Clayton (2001), the MLE cluster associated with the scan statistic favors the detection of clusters in areas with a large number of geographically small cells over the detection of clusters in areas with a small number of geographically large cells. One approach to minimizing or eliminating this effect is to maximize a penalized version of the LR. An obvious candidate for a penalty is the 'uniform selection' weight $w_{i,j}$ described above, which has the desirable property of penalizing clusters in areas with geographically large cells less severely than clusters in areas with geographically small cells.

We propose using the penalized scan statistic, $\mathrm{PLR}_{\max} = \max_{i,j} w_{i,j} \mathrm{LR}_{i,j}$, as a test for clustering. As with the other statistics, the null distribution is found by simulation under the null hypothesis. If the null hypothesis is rejected, the cluster associated with $\mathrm{PLR}_{\max}$ (from a Bayesian point of view, the maximum a posteriori estimate of the true cluster) may be viewed as an estimate of the true cluster. Alternatively, one could sensibly combine the WALR statistic for testing with the maximum a posteriori cluster for estimation.

2.4. The weighted average likelihood ratio scan (WALRS) statistic

A disadvantage of the WALR statistic is its lack of specificity. It does not identify a single cluster responsible for the rejection of the null hypothesis. In practice, the approximate posterior described earlier may not even identify a single region of clustering. As an alternative, we propose the weighted average likelihood ratio scan (WALRS) statistic,

$$\mathrm{WALRS} = \max_k \frac{\sum_{i=1}^{N} \sum_{j=1}^{m_i} w_{i,j} \mathrm{LR}_{i,j} I_{\{k \in \mathbf{c}_{i,j}\}}}{\sum_{i=1}^{N} \sum_{j=1}^{m_i} w_{i,j} I_{\{k \in \mathbf{c}_{i,j}\}}}$$

where $I_{\{k\in\mathbf{c}_{i,j}\}}$ is 1 if cell k belongs to cluster i,j and 0 otherwise. The WALRS statistic is constructed by first calculating a WALR statistic for each of the N cells using the set of potential clusters that contain the cell and then finding the maximum of these WALR statistics. The WALRS statistic localizes our assessment of the evidence for clustering by considering only clusters containing the same cell. In this regard, the WALRS statistic differs from a traditional scan statistic (Glaz and Balakrishnan, 1999; Glaz *et al.*, 2001) since the scan is over sets of clusters, not the individual clusters.

As with the WALR statistic, the WALRS statistic has an appealing Bayesian interpretation as an approximate marginal likelihood ratio (or Bayes factor) for a composite one cluster model to the no clustering model. If we view the $w_{i,j}$ as a prior distribution on the space of potential clusters, the WALRS statistic is proportional to the approximate marginal likelihood of a one cluster model containing cell k, where cell k is the cell associated with the maximum. In this sense, WALRS is a reasonable measure of the evidence for a single cluster model in the data.

We more fully exploit our Bayesian interpretation of the weights to interpret the data if the null hypothesis is rejected. Specifically, after rejection, we assume that the composite one cluster model associated with the WALRS statistic is correct and formally treat the $w_{i,j}$ as a prior distribution on the clusters. Conditional on this one cluster model associated with cell k, the posterior probability of the true cluster being cluster i,j is approximately $p^k_{i,j} = w_{i,j}\mathrm{LR}_{i,j}I_{\{k\in\mathbf{c}_{i,j}\}}/\mathrm{WALRS}$. Using these probabilities, we can average over the single cluster models to obtain estimates of the probability that a cell belongs to the cluster and the cell-specific disease rates, as was discussed for the WALR statistic. As an alternative, one could use the WALR statistic for testing and the restricted posterior for focused estimation of cluster location and risk.

2.5. *Assessment of multiple clusters*

All of the statistics described above are designed to assess the significance of a single cluster model. (The full WALR posterior may provide some indirect evidence of multiple clusters.) To assess the significance of multiple clusters, we note that estimates of the location and risk associated with the single cluster are easily obtained for each testing procedure. Thus, we can evaluate the significance of a second cluster by replacing the populations at risk n_i with expected case counts E_i from the selected one cluster model or posterior distribution. We suggest testing in this setting using a closed sequential testing procedure, stopping if the test statistic fails to achieve a prespecified significance level, say 5% or 1%.

3. SIMULATION RESULTS: BIAS AND POWER

To assess the merits of the tests (and estimators) based on $\mathrm{LR}_{\max}$, $\mathrm{PLR}_{\max}$, WALR and WALRS discussed above, we consider two properties of the procedures: unbiased cluster detection and power. Previous simulation results in Gangnon and Clayton (2001) explore these properties for the $\mathrm{LR}_{\max}$ and WALR statistics. Here, we extend the simulations to include the new statistics and estimators proposed in this article and expand the power calculations to incorporate false rejections.

3.1. *Unbiased cluster selection*

Unbiased cluster detection is, strictly speaking, a property of an estimator of cluster location, not of a testing procedure. An estimator of the cluster is said to be unbiased (Gangnon and Clayton, 2001) if, under the null hypothesis of no clustering, each cell in the study region has an equal chance of

belonging to the estimated cluster. By requiring an even distribution of the 'estimated cluster' under the null hypothesis, one ensures a fair, or 'unbiased', assessment of the evidence for clustering in an area. (Note that this notion of unbiasedness is different from the traditional definition of an unbiased test given in Lehmann, 1959.)

To evaluate the bias of the estimators associated with these tests, we utilized the structure of the New York leukemia data. The New York leukemia data set consists of data on leukemia incidence between 1978 and 1982 in eight counties in upstate New York: Broome, Cayuga, Chenango, Cortland, Madison, Onondaga, Tioga and Tompkins. The two largest cities in the study region are Syracuse in Onondaga county and Binghamton in Broome county. The eight-county region is divided into 790 cells, either census blocks or census tracts. For each cell, the population at risk, count of leukemia cases, and geographic centroid are available. In our analysis, cell regions are imputed using the Dirichlet tessellation of the centroids (Gangnon and Clayton, 2000). Additional background information on the New York leukemia data is available in Waller *et al.* (1994).

We simulated 1000 data sets under the null hypothesis, using totals of 200, 400, 600, 800 and 1000 cases. The total of 600 cases is comparable to the overall number of cases (592) in the New York data. For these analyses, we considered all circles centered at the cell centroids with radii less than or equal to 20 km as potential clusters. For each simulated data set, we found the MLE cluster, the maximum a posteriori cluster, and the full (WALR) or restricted (WALRS) posterior distributions. For each cell, we then found the proportion of the simulations in which the cell belonged to the MLE cluster or the maximum a posteriori cluster, the mean probability that the cell belonged to a cluster based on the full (WALR) or restricted (WALRS) posteriors, and the proportion of simulations in which the posterior probability that the cell belonged to a cluster based on the restricted (WALRS) posterior exceeded 1/2.

Figure 1 displays the results of the simulations for the 600 case scenario on a map of the New York data. Results for other case totals were quite similar. For an unbiased test, this graphic will show a uniform intensity. Greater intensity contrasts between light and dark colors indicate greater bias in the test, with dark areas being more likely to belong to the estimated cluster and light areas being less likely to belong.

The simulation results are quite striking. The estimators based on the full (WALR) and restricted (WALRS) posteriors show fairly uniform color across the study region, indicating relatively little bias. Both the MLE cluster and the maximum a posteriori cluster show strong evidence of bias, although the nature of the bias differs greatly between two estimators. In Gangnon and Clayton (2001), the MLE cluster shows bias towards clusters in the Syracuse area over clusters in rural areas, likely due to the large number of geographically small cells in urban areas. The maximum a posteriori cluster appears to overcompensate for the urban bias in the MLE cluster. There is a pronounced move away from clusters in urban areas such as Syracuse and Binghamtom and towards clusters in rural areas. In some sense, the MLE cluster and the maximum a posteriori cluster are complementary; the MLE cluster favors clusters in areas with much overlap (geographically small cells) and the maximum a posteriori cluster favors cells in areas with little overlap (geographically large cells). On a regular grid of cells, the bias in either estimator would likely be minimal, and the two estimators would be essentially equivalent.

3.2. Power and false detections

We next examined the power of the four tests to detect clusters in simulated data sets. For our simulations, we introduced a single circular cluster centered at a cell centroid into the data set. A total of 15 clusters including a variety of populations and geographic locations were used. For each cluster,

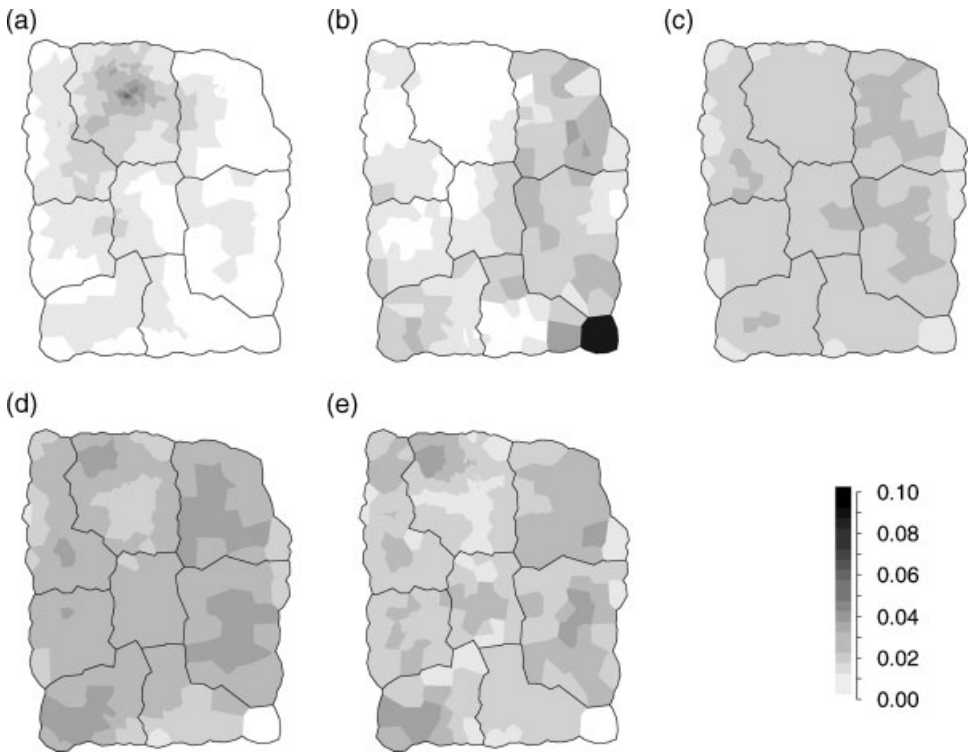

Figure 1. Estimated cluster membership probabilities based on 1000 simulations under the null hypothesis with a total of 600 cases: (a) Proportion of simulations in which cell belonged to the MLE cluster; (b) proportion of simulations in which cell belonged to the maximum a posteriori estimated cluster divided by the maximum such proportion; (c) mean probability that cell belonged to a cluster using the full (WALR) posterior; (d) mean estimated probability that cell belonged to a cluster using the restricted (WALRS) posterior; (e) proportion of simulations in which the posterior probability that the cell belonged to the cluster based on the restricted (WALRS) posterior exceeded 1/2

we simulated 1000 data sets with positive (higher rate than background) clustering and 1000 data sets with negative (lower rate than the background) clustering. The rate outside the cluster was fixed at 5 cases per 10 000 people. Rate ratios (RRs) were assigned to the clusters to produce asymptotic log-likelihood ratios of 8, 9 or 10 using the formulas in Gangnon and Clayton (2000). This method of assigning RRs was designed to ensure reasonable (not exceptionally high or low), yet variable power to detect each cluster. The 15 clusters (and RRs) used for this power study are displayed on a map of the New York data in Figure 2.

The 95th percentiles of the null distribution of the natural logarithm of each test statistic were estimated using 1000 simulations for five values of the total numbers of cases (200, 400, 600, 800 and 1000). The 95th percentiles of the null distribution of the natural logarithm of each test statistic for other values of the total number of cases were obtained by interpolation.

In assessing the power of any localized cluster detection test, one needs to think about the potentially serious error of correctly rejecting the null hypothesis while incorrectly identifying the cluster location. In this situation, we have essentially committed two errors (detecting a false cluster and failing to detect the true cluster). Obviously, it is desirable for any testing/estimation procedure to

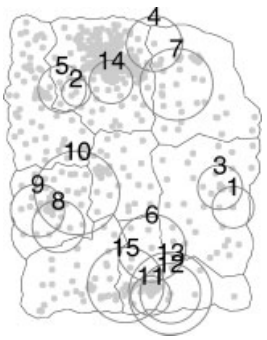

Figure 2. Map of the centroids of cells in New York data. Gray dots indicate cell centroids. Circles indicate clusters used for power simulations. Text indicates cluster number used in Tables 1 and 2

minimize the rate of these 'false rejections'. For the LR_{max} and PLR_{max} tests, a cluster identification is classified as *incorrect* if the estimated cluster (MLE or MAP, respectively) fails to contain even a single cell belonging to the true cluster. For the WALR and WALRS tests, a cluster identification is classified as *incorrect* if no cell belonging to the true cluster has a posterior probability ≥ 0.5 using the restricted WALRS posterior. For the WALR test, this corresponds to the modal area of clustering in the posterior overlapping the true cluster.

The estimated power and false rejection rates based on 1000 simulations for the four tests are provided in Table 1 for positive clustering and in Table 2 for negative clustering.

Summarizing the results, the WALR and WALRS tests provide at least competitive power in most settings. The WALR test frequently produces the highest power. However, the power loss with the

Table 1. Estimated power and false rejection rates of the LR_{max}, PLR_{max}, WALR and WALRS tests for detecting *positive* clusters based on 1000 simulations for each cluster. Cluster information provided includes the coordinates of cluster center (km from centroid of region), its radius and population, and rate ratio used for simulations. **Bold text** indicates the test with largest power for each cluster. *Italics text* indicates tests with power not significantly different from the test with the largest power at a nominal 5% level using McNemar's test

	Cluster information					Power (false rejection rate) of tests							
No.	*X*	*Y*	Radius	Population	RR	LR_{max}		PLR_{max}		WALR		WALRS	
1	44.00	−29.32	9.84	8638	3.62	45.5	(4.3)	*58.2*	(3.6)	*58.0*	(2.9)	**58.3**	(2.6)
2	−29.51	25.23	6.68	9244	3.29	19.1	(3.8)	**22.8**	(4.9)	21.4	(4.9)	20.9	(4.1)
3	37.87	−19.84	10.55	18 062	2.75	55.5	(4.1)	66.7	(1.7)	**72.4**	(2.5)	*71.6*	(2.0)
4	7.17	48.29	13.10	30 325	2.31	54.2	(2.9)	60.2	(2.0)	**62.0**	(2.6)	*61.9*	(2.3)
5	−35.98	26.87	11.31	40 248	2.26	*73.5*	(1.8)	64.7	(2.3)	**74.6**	(2.6)	*74.5*	(2.5)
6	5.89	−48.61	16.03	45 639	2.06	45.8	(3.0)	*61.8*	(2.1)	**62.2**	(3.0)	*61.5*	(2.4)
7	17.48	29.01	17.06	47 658	2.06	62.0	(2.4)	68.3	(1.3)	**76.9**	(2.0)	74.4	(1.5)
8	−37.73	−38.44	12.39	57 789	2.02	74.9	(1.9)	75.6	(1.6)	**85.6**	(1.5)	84.6	(1.3)
9	−47.49	−30.82	12.60	60 663	2.07	79.8	(1.7)	82.5	(1.5)	**92.5**	(0.8)	*92.2*	(0.7)
10	−29.02	−22.49	19.85	109 096	1.63	49.9	(3.6)	45.5	(2.7)	**61.2**	(3.1)	59.3	(2.6)
11	5.52	−71.26	9.41	125 472	1.73	77.0	(2.8)	77.1	(2.1)	**87.1**	(1.8)	*86.7*	(1.2)
12	14.65	−70.15	14.21	134 588	1.73	81.2	(3.1)	87.6	(1.1)	**90.8**	(1.9)	*90.6*	(1.5)
13	14.65	−70.15	19.57	167 627	1.52	52.5	(5.0)	58.5	(2.6)	*71.6*	(2.8)	**71.7**	(2.3)
14	−13.10	30.14	10.28	195 099	1.58	**85.8**	(0.8)	54.2	(2.6)	81.3	(1.6)	79.6	(1.2)
15	−6.27	−66.09	17.91	201 191	1.52	62.4	(5.1)	64.4	(2.3)	**77.3**	(2.8)	*76.6*	(2.3)

Table 2. Estimated power and false rejection rate of the LR_{max}, PLR_{max}, WALR and WALRS tests for detecting *negative* clusters based on 1000 simulations for each cluster. Cluster information provided includes the coordinates of cluster center (km from centroid of region), its radius and population, and rate ratio used for simulations. **Bold text** indicates the test with largest power for each cluster. *Italics text* indicates tests with power not significantly different from the test with the largest power at a nominal 5% level using McNemar's test

	Cluster information					Power (false rejection rate) of tests							
No.	*X*	*Y*	Radius	Population	RR	LR_{max}		PLR_{max}		WALR		WALRS	
1	44.00	−29.32	9.84	8638	0.01	0.8	(4.9)	2.5	(4.4)	*3.1*	(4.4)	**3.7**	(4.0)
2	−29.51	25.23	6.68	9244	0.01	**1.0**	(5.4)	*0.4*	(4.7)	*0.5*	(4.4)	0.2	(4.7)
3	37.87	−19.84	10.55	18 062	0.01	55.0	(4.3)	76.4	(2.9)	*83.2*	(2.3)	**83.9**	(2.0)
4	7.17	48.29	13.10	30 325	0.07	82.6	(1.6)	**86.0**	(1.3)	*85.5*	(1.1)	*85.2*	(1.2)
5	−35.98	26.87	11.31	40 248	0.15	**83.8**	(1.0)	75.1	(2.0)	*82.7*	(1.8)	82.3	(1.6)
6	5.89	−48.61	16.03	45 639	0.24	43.5	(3.6)	**60.3**	(2.2)	*59.6*	(3.4)	58.1	(3.0)
7	17.48	29.01	17.06	47 658	0.24	65.8	(1.7)	71.7	(2.0)	**82.5**	(2.2)	79.1	(1.7)
8	−37.73	−38.44	12.39	57 789	0.32	65.6	(2.0)	68.0	(2.0)	**82.3**	(2.3)	81.4	(1.8)
9	−47.49	−30.82	12.60	60 663	0.24	90.2	(0.6)	91.2	(0.5)	*97.1*	(0.4)	**97.4**	(0.4)
10	−29.02	−22.49	19.85	109 096	0.49	53.6	(3.4)	49.2	(2.4)	**64.9**	(3.0)	63.5	(2.3)
11	5.52	−71.26	9.41	125 472	0.53	49.5	(3.3)	48.2	(5.0)	**61.4**	(4.4)	*61.2*	(3.6)
12	14.65	−70.15	14.21	134 588	0.43	86.5	(1.7)	91.9	(1.0)	*95.7*	(0.9)	**95.9**	(0.7)
13	14.65	−70.15	19.57	167 627	0.57	52.7	(2.6)	58.0	(2.2)	**71.1**	(2.3)	*71.0*	(1.7)
14	−13.10	30.14	10.28	195 099	0.53	**85.2**	(0.2)	52.4	(3.5)	80.1	(2.0)	79.4	(1.5)
15	−6.27	−66.09	17.91	201 191	0.57	63.3	(3.4)	65.3	(2.3)	**80.5**	(2.7)	*80.4*	(2.1)

WALRS test, although occasionally 'statistically significant' using McNemar's test, is typically quite small. The LR_{max} and PLR_{max} tests are not consistently competitive with the other two tests in these simulations. In the few settings where either of these tests shows the best power (e.g. Cluster 14 for the LR_{max} test and Cluster 6 for the PLR_{max} test), the WALR and WALRS tests provide comparable power, while in other settings (e.g. Cluster 13 for both tests), neither LR_{max} nor PLR_{max} is competitive with the WALR and WALRS tests. To be fair, these simulations do not include many clusters in the areas in which the bias of the MLE or MAP cluster is greatest, where we would expect substantial power advantages for the LR_{max} or PLR_{max} test, as was seen in the simulations in Gangnon and Clayton (2001).

The definitions of incorrect cluster identification provided above are restricted to obvious, indisputable errors. Some of the 'correct' cluster identifications could be very poor estimates for the true cluster. However, we find it reassuring that the rate of indisputable errors is quite low for all four tests. Obviously, the choice of a cut-off value of 0.5 for the posterior was arbitrary. However, the results are not meaningfully affected by the choice of this cut-off. Even if we defined a rejection to be incorrect when the posterior modal cell does not belong to the true cluster (i.e. we used an extreme cut-off value of 1.0), the increase in false rejection rates would be quite modest ($<2.0\%$) in most cases. (The exceptional cases are Clusters 3, 8 and 9, where the posterior modal cell frequently falls just outside the border of the true cluster.) Thus, the conclusions drawn in Tables 1 and 2 would be substantially unchanged. Alternatively, if we ignore the issue of correct cluster identifications and simply count rejections of the null hypothesis, our conclusions regarding the relative power of the tests, both for specific cases and in general, would be unchanged.

For general use, we would recommend the WALR test. For estimation, we would recommend using either the full (WALR) or restricted (WALRS) posterior. The WALRS posterior has the advantage of

restricting consideration to clustering in a single area. If prior knowledge indicates that clustering in a particular area (or areas) was of special concern, we recommend developing a tailored set of weights based on that prior knowledge for that specific application rather than use a testing procedure based on a biased estimator.

4. EXAMPLE: NEW YORK LEUKEMIA DATA

To assess the evidence for localized clustering in the New York leukemia data, we considered all circles centered at the cell centroids with radii less than or equal to 20 km as potential clusters. We tested the null hypothesis with the four test statistics discussed earlier—LR_{max}, PLR_{max}, WALR and WALRS. We conducted each test at the 5% significance level, rejecting the null hypothesis if the observed test statistic was larger than the 95th percentile of the null distribution based on 1000 simulations under the null hypothesis.

The observed values of the test statistics, on the log scale, were 13.05 for LR_{max}, 2.00 for the PLR_{max}, 4.96 for WALR and 8.11 for WALRS. All four test statistics were clearly significant at the 5% level. (The critical values for the test statistics were 9.76, −0.02, 1.83 and 4.87, respectively.) The estimated clusters associated with each of the four test statistics are displayed in Figure 3.

The analyses for the LR_{max} and WALR test replicate the analyses in Gangnon and Clayton (2001), although a different set of simulations and a different significance level were used. Using the LR_{max}, we find statistically significant evidence for a cluster in Broome county. On the map, cells belonging to the circle corresponding to the most likely cluster are drawn in black. There is some evidence for other 'statistically significant' clusters in Cortland county (high rate) and in Onondaga county (low rate).

Using PLR_{max}, we observe qualitatively similar results to the LR_{max} results. We again find evidence of a high rate cluster in Broome county and secondary clusters in Cortland county and in Onondaga county. On the map, cells belonging to the circle corresponding to the maximum a posteriori estimate of the cluster are drawn in black. Despite the qualitative similarity with the results of the LR_{max} test (both tests find clusters in the same general area of Broome County), there are substantial differences in the estimates. The estimated cluster associated with PLR_{max} is substantially larger than the MLE cluster associated with LR_{max}. These differences between the findings of the two tests reflect the impact of the penalty term.

The Bayesian estimates of cluster membership probabilities associated with the WALR and WALRS statistics provide two methods of evaluating the same data. The posterior probabilities associated with the restricted (WALRS) posterior reflect our uncertainty about the exact composition of the possible cluster in Broome County. The probabilities indicate strong evidence for a small core set of cells that definitely belong to such a cluster, but there is substantial uncertainty about the exact composition of the cluster. In fact, there is some support in the data for a very extensive cluster covering most of Broome County.

The probabilities associated with the full (WALR) posterior reflect the relative strengths of evidence for three possible areas of clustering. There is strong support for a cluster in Broome county; the 'posterior probability' of that cluster is 62%. There is additional evidence for two clusters—a cluster in Onondaga county (probability 23%) and a cluster in Cortland county (probability 11%). The possible clusters in Broome and Cortland counties have higher rates than the background, while the possible cluster in Onondaga county appears to have a lower rate than the background.

We now perform a formal assessment of the significance of multiple clusters using the closed sequential testing procedure outlined in Section 2.5. For this discussion, we only consider the LR_{max}

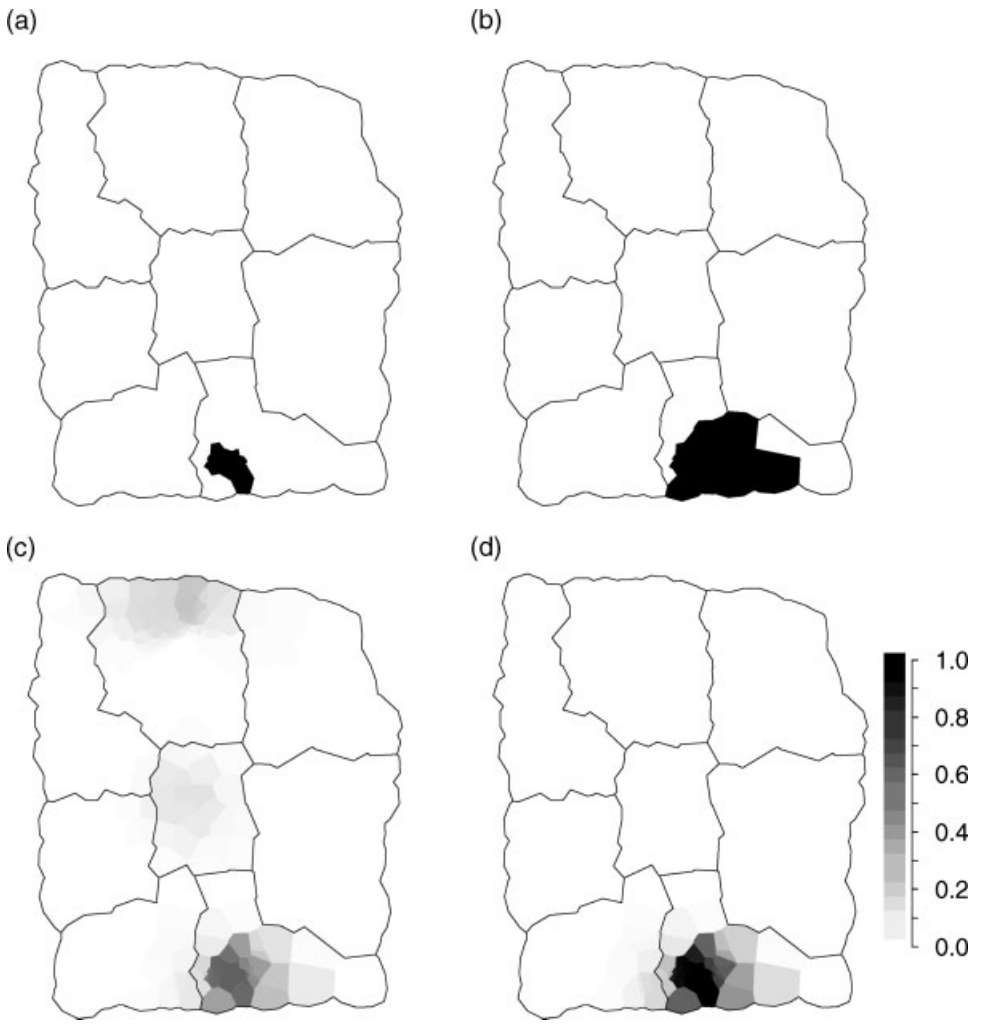

Figure 3. (a) Cells belonging to most likely cluster, i.e. the cluster associated with LR_{max}, drawn in black; (b) cells belonging to maximum a posteriori estimated cluster, i.e. the cluster associated with PLR_{max}, drawn in black; (c) posterior cell-specific probabilities of belonging to a single cluster based on the full (WALR) posterior; (d) posterior probabilities of cells belonging to a single cluster containing cell 7012720 (coordinates 5.52, −71.26) based on the restricted (WALRS) posterior

and WALR tests. For the LR_{max} test, estimation is based on the MLE cluster; for the WALR test, estimation is based on the restricted (WALRS) posterior to localize the estimates to a single area of clustering.

Figure 4 displays the estimated risks associated with the MLE clusters for 1, 2, 3 or 4 cluster models; the *p*-values for the LR_{max} test are <0.001, 0.010, 0.026 and 0.305. Using a 5% significance level, we would favor the model with three clusters. The first two clusters identified are the clusters in Broome and Cortland county discussed previously; the third cluster is a new area of clustering in the Syracuse area associated with an elevated leukemia risk.

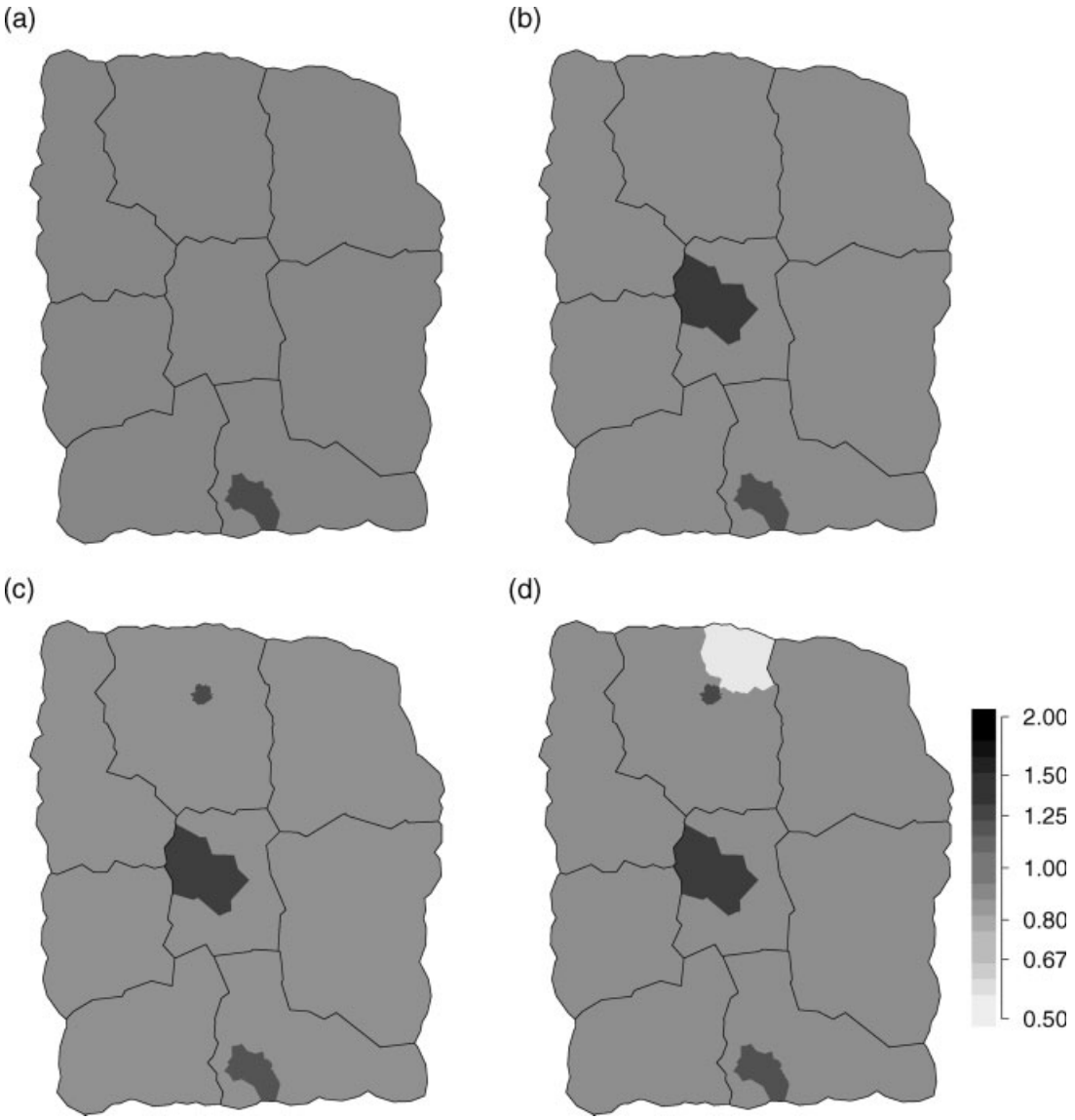

Figure 4. Estimated disease-rates-based stepwise selection of 1, 2, 3 and 4 cluster models based on LR_{max} test and MLE cluster. Rates are given relative to the overall leukemia rate of 5.5 cases per 10 000 persons: (a) 1 cluster model, $p < 0.001$; (b) 2 cluster model, $p = 0.010$; (c) 3 cluster model, $p = 0.026$; (d) 4 cluster model, $p = 0.306$

Figure 5 displays the estimated risks associated with the restricted (WALRS) posterior for 1, 2, 3 or 4 cluster models; the *p*-values for the WALR test are 0.001, 0.004, 0.072 and 0.502. Using a 5% significance level, we would favor the model with two clusters; using a 10% significance level, we would favor the model with three clusters. Again, the first two clusters identified are the clusters in Broome and Cortland county discussed previously. The third cluster identified is an area in Onondaga county associated with a lowered risk of leukemia.

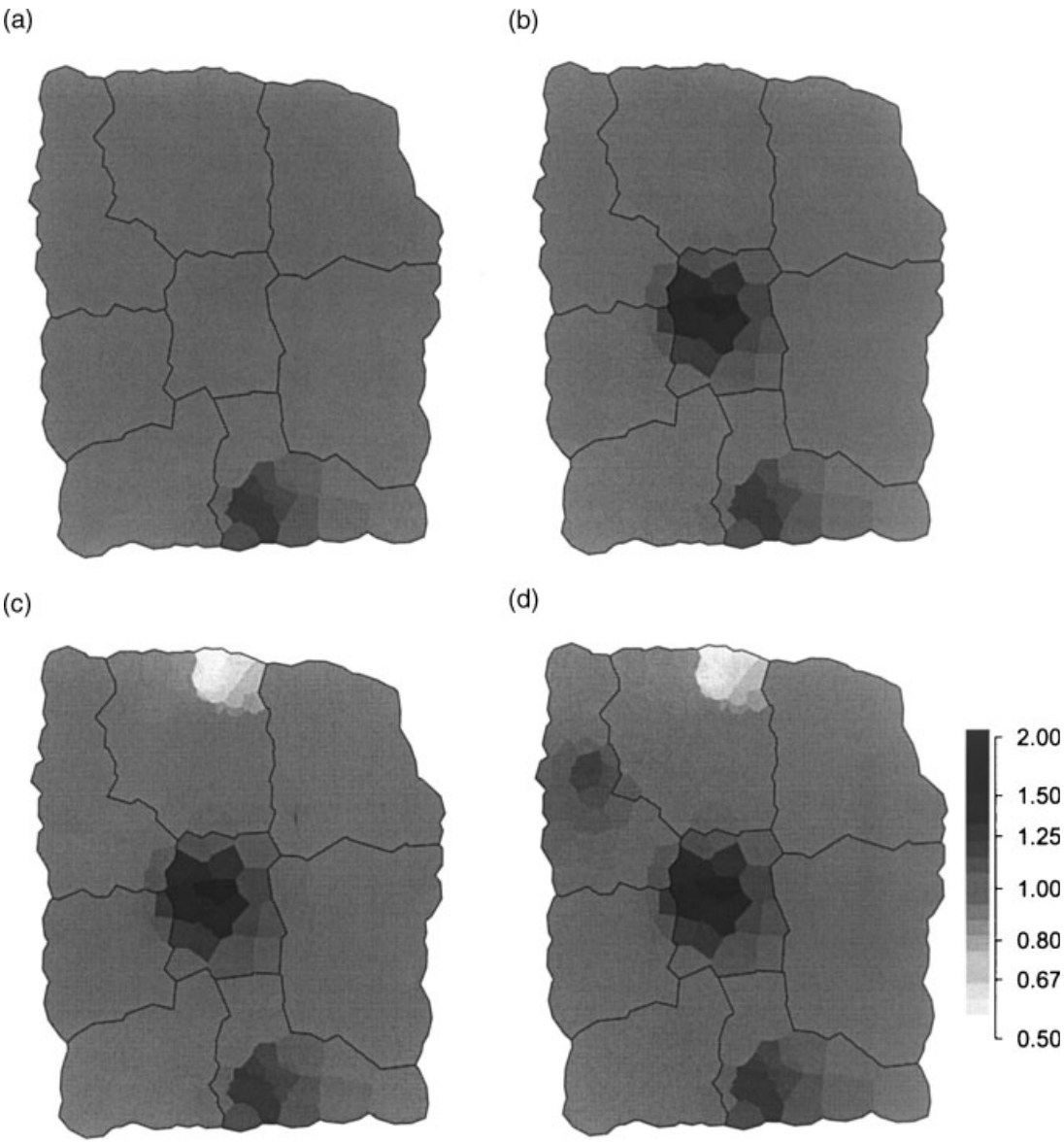

Figure 5. Estimated disease-rates-based stepwise selection of 1, 2, 3 and 4 cluster models based on WALR test and restricted posterior. Rates are given relative to the overall leukemia rate of 5.5 cases per 10 000 persons: (a) 1 cluster model, $p = 0.001$; (b) 2 cluster model, $p = 0.004$; (c) 3 cluster model, $p = 0.072$; (d) 4 cluster model, $p = 0.502$

The contrasts between the results of the LR_{max} and the WALR test are striking. Both procedures identify the same first two clusters, but differ on the choice of the third. The MLE procedure identifies a higher risk cluster in the Syracuse area, while the restricted posterior identifies a lower risk cluster north of Syracuse.

5. CONCLUSIONS

For a global test for localized clusters, either the weighted average likelihood ratio (WALR) test or the weighted average likelihood ratio scan (WALRS) statistic provides an attractive alternative to the maximum likelihood ratio (scan) test. Both tests appear to have superior power compared to the scan statistic and, with the proper choice of weights, the associated posterior estimates are relatively unbiased. Both tests maintain a relatively low rate of obviously false cluster identifications. In future work, we will explore the usefulness of alternative measures of the accuracy of cluster identifications as a tool for improving our assessment of the power of cluster detection tests.

The penalized scan statistic discussed here does not achieve our goal of correcting for the bias in the scan statistic, or perhaps more accurately achieves our goal too well. The penalized statistic shows a strong bias towards clusters in areas with geographically large cells. Thus, we could not recommend the penalized scan statistic for general use. In future work, we will continue to explore potential adjustments to the scan statistic (e.g. a new penalty term) and/or other methods to produce an unbiased point estimator of cluster locations.

Estimated cluster membership probabilities from the full (WALR) posterior can be very useful in assessing the relative support in the data for different areas of clustering, while estimated cluster membership probabilities from the restricted (WALRS) posterior can quantify the uncertainty about the exact composition of a possible cluster. By quantifying our level of uncertainty about cluster locations, both probabilities have substantial advantages over the single 'approximate cluster locations' associated with scan statistics.

In practice, we would utilize the WALR statistic for testing and both the full (WALR) and restricted (WALRS) posteriors for estimation. The WALR statistic provides a global assessment of the evidence for clustering. Estimates based on the full (WALR) posterior reflect large-scale uncertainty about the location of the cluster, while the restricted (WALRS) posterior reflects uncertainty about the specific composition of a cluster in a specific area. Estimates from the restricted posterior also allow us to assess the significance of multiple areas of clustering using a simple closed testing procedure.

REFERENCES

Besag J, Newell J. 1991. The detection of clusters of rare diseases. *Journal of the Royal Statistical Society, Series A* **154**: 143–155.

Gangnon RE, Clayton MK. 2000. Bayesian detection and modeling of spatial disease clustering. *Biometrics* **56**: 922–935.

Gangnon RE, Clayton MK. 2001. A weighted average likelihood ratio test for spatial clustering of disease. *Statistics in Medicine* **20**: 2977–2987.

Glaz J, Balakrishnan N (eds). 1999. *Scan Statistics and Applications*. Boston: Birkhäuser.

Glaz J, Naus JI, Wallenstein S. 2001. *Scan Statistics*. New York: Springer-Verlag.

Kulldorff M. 1997. A spatial scan statistic. *Communications in Statistics, Part A* **26**: 1481–1496.

Kulldorff M. 1998. Statistical methods for spatial epidemiology: tests for randomness. In *GIS and Health for Europe*, Gatrell A, Løytønen M (eds). Taylor & Francis: London; 49–62.

Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters: detection and inference. *Statistics in Medicine* **14**: 799–810.

Lehmann EL. 1959. *Testing Statistical Hypotheses*. John Wiley & Sons: New York.

Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. 1990. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology* **132**: S136–S143.

Waller LA, Turnbull BW, Clark LC, Nasca P. 1994. Spatial pattern analyses to detect rare disease clusters. In *Case Studies in Biometry*, Lange N, Ryan L, Billiard L (eds). John Wiley & Sons: New York; 3–22.

