Biometrics 68, 174–182
March 2012
DOI: 10.1111/j.1541-0420.2011.01643.x

Local Multiplicity Adjustment for the Spatial Scan Statistic Using the Gumbel Distribution

Ronald E. Gangnon

Departments of Biostatistics and Medical Informatics and Population Health Sciences, University of Wisconsin–Madison, Madison, Wisconsin 53726, U.S.A.
email: ronald@biostat.wisc.edu

Summary. The spatial scan statistic is an important and widely used tool for cluster detection. It is based on the simultaneous evaluation of the statistical significance of the maximum likelihood ratio test statistic over a large collection of potential clusters. In most cluster detection problems, there is variation in the extent of local multiplicity across the study region. For example, using a fixed maximum geographic radius for clusters, urban areas typically have many overlapping potential clusters, whereas rural areas have relatively few. The spatial scan statistic does not account for local multiplicity variation. We describe a previously proposed local multiplicity adjustment based on a nested Bonferroni correction and propose a novel adjustment based on a Gumbel distribution approximation to the distribution of a local scan statistic. We compare the performance of all three statistics in terms of power and a novel unbiased cluster detection criterion. These methods are then applied to the well-known New York leukemia dataset and a Wisconsin breast cancer incidence dataset.

Key words: Bonferroni adjustment; Gumbel distribution; Multiple comparisons; Permutation test; Poisson; Scan statistic.

1. Introduction

The spatial scan statistic (Kulldorff, 1997) is a widely used tool for cluster detection. The spatial scan statistic is based on the simultaneous evaluation of the statistical significance of the maximum likelihood ratio test statistic over a large collection of potential clusters. Although most commonly used with circular potential clusters (Kulldorff and Nagarwalla, 1995), many other possibilities have been considered, including ellipses (Kulldorff et al., 2006), rectangles (Neill and Moore, 2004), irregularly shaped connected regions (Patil and Taillie, 2004; Tango and Takahashi, 2005; Duczmal, Cançado, and Takahashi, 2008), and even all possible subregions (Neill, 2008). We will focus on the circular scan statistic, although the issues discussed are applicable to all of these settings.

Local multiplicity refers to the number of potential clusters that overlap each location. In most, if not all, realistic cluster detection problems, there will be substantial variation in local multiplicity across space. For example, using circular clusters with a fixed maximum geographic radius or fixed maximum population radius, areas with fine geographic resolution will have many more potential clusters than areas with coarse geographic resolution.

Differences in the extent of local multiplicity across a study region impact the operating characteristics of the spatial scan statistic, the so-called "local multiplicity problem" (Gangnon, 2010). Under the null hypothesis of constant disease risk, the selection of the most likely cluster is biased in favor of clusters in areas with more extensive local multiplicity (Gangnon and Clayton, 2001, 2004). Further, the spatial scan statistic has greater power for single clusters in areas with more extensive local multiplicity than for otherwise identical clusters in areas with lesser local multiplicity (Gangnon and Clayton, 2001, 2004; Waller, Hill, and Rudd, 2006; Gangnon, 2010). Thus, the spatial scan statistic overstates the evidence for clustering in areas with fine geographic resolution and understates the evidence for clustering in areas with coarse geographic resolution.

To account for variations in local multiplicity, Gangnon (2010) proposed the Bonferroni locally adjusted spatial scan (BLASS) statistic using a nested Bonferroni correction based on the number of clusters in a random neighborhood of each cluster. To avoid double counting in the correction, the BLASS procedure requires (1) the identification and exclusion of duplicate representations of the same cluster, e.g., circles with different centers that contain identical sets of cell centroids, from the set of potential clusters and (2) a random partitioning of the set of potential clusters to create nonoverlapping neighborhoods. The first task is nontrivial and tedious, and the second is unintuitive.

As an alternative, we propose the Gumbel locally adjusted spatial scan (GLASS) statistic, which uses a novel adjustment based on Gumbel approximations to the distributions of local scan statistics for each cluster. The GLASS statistic is unaffected by duplicate cluster representations and overlapping cluster neighborhoods, but it requires a small additional set of simulations to estimate the local Gumbel parameters. We evaluate the performance of the GLASS statistic relative to the other statistics in terms of both power and a novel unbiased cluster detection criterion.

2. The Spatial Scan Statistic

We consider situations in which aggregated data for N administrative regions or cells within the study area are observed. 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 cell i, E_i is the expected number of cases of disease in cell i in the absence of clustering and $\mathbf{x}_i = (x_{1i}, x_{2i})$ is the geographic centroid of cell i. The expected number of cases, E_i, may reflect the overall disease rate applied to the regional population at risk or may be fitted values from a nonspatial Poisson regression model incorporating the effects of individual and/or regional covariates. We assume that y_i are independent Poisson random variables with mean $\rho_i E_i$.

For any subset $\mathbf{Z}$ of the study region, we consider the following model for ρ_i: $\log(\rho_i) = \alpha_{\mathbf{Z}} + \theta_{\mathbf{Z}}\delta_{\mathbf{Z}}(\mathbf{x_i})$, where $\delta_{\mathbf{Z}}(\mathbf{x_i}) = 1$ if $\mathbf{x}_i \in \mathbf{Z}$ and $\delta_{\mathbf{Z}}(\mathbf{x_i}) = 0$ otherwise, $\alpha_{\mathbf{Z}}$ is the log disease risk for locations outside $\mathbf{Z}$ and $\theta_{\mathbf{Z}}$ is the log relative risk for locations inside $\mathbf{Z}$. If $\mathbf{Z}$ is not a cluster, $\theta_{\mathbf{Z}} = 0$; if $\mathbf{Z}$ is a cluster, $\theta_{\mathbf{Z}} \neq 0$. We consider the two-sided alternative, clusters with elevated or reduced risk, rather than the one-sided alternative, clusters with elevated risk only. The issues discussed here arise regardless of the choice of alternative.

The evidence for $\mathbf{Z}$ as a cluster is given by the log-likelihood ratio test statistic for H_0: $\theta_{\mathbf{Z}} = 0$ versus $H_A : \theta_{\mathbf{Z}} \neq 0$,

$$\mathrm{LR}_{\mathbf{Z}} = y(\mathbf{Z}) \log\left\{\frac{y(\mathbf{Z})}{E(\mathbf{Z})}\right\} + \{y_{tot} - y(\mathbf{Z})\} \log\left\{\frac{y_{\mathrm{tot}} - y(\mathbf{Z})}{E_{\mathrm{tot}} - E(\mathbf{Z})}\right\},$$

where $y(\mathbf{Z}) = \sum_{i=1}^N y_i \delta_{\mathbf{Z}}(\mathbf{x_i})$ is the number of cases inside $\mathbf{Z}$, $E(\mathbf{Z}) = \sum_{i=1}^N E_i \delta_{\mathbf{Z}}(\mathbf{x_i})$ is the expected number of cases inside $\mathbf{Z}$, $y_{\mathrm{tot}} = \sum_{i=1}^N y_i$, and $E_{\mathrm{tot}} = \sum_{i=1}^N E_i$. Without loss of generality, we will assume $y_{\mathrm{tot}} = E_{\mathrm{tot}}$. Using standard asymptotic results, we can obtain the nominal p-value $p_{\mathbf{Z}} = P(X^2 > 2\mathrm{LR}_{\mathbf{Z}})$, where X^2 is a χ_1^2 random variate.

For the set of potential clusters, we consider (closed) circular regions centered at the observed locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, with radii ranging from 0 up to a fixed maximum radius, $r_{\max}$. To identify the m_s unique clusters centered at $\mathbf{x}_s$ for $s = 1, 2, \ldots, N$, we let $0 = r_{s,1} < r_{s,2} < \cdots < r_{s,m_s} \leq r_{\max}$ be the (unique) ordered distances from $\mathbf{x}_s$ to all locations, truncated at $r_{\max}$. We denote the set of locations inside a circular cluster centered at $\mathbf{x}_s$ with radius r_{st} by $\mathbf{Z}_{st}$, the associated likelihood ratio test statistic by LR_{st}, and the associated nominal p-value by p_{st} for $t = 1, 2, \ldots, m_s$; $s = 1, 2, \ldots, N$. Although we use a specific set of potential clusters, the method described here are applicable to any discrete set of potential clusters.

Kulldorff and Nagarwalla (1995) and Kulldorff (1997) proposed the spatial scan statistic, i.e., the maximum likelihood ratio test statistic over all potential clusters $\mathrm{LR}_{\max} = \max_{s,t} \mathrm{LR}_{st}$, as a global cluster detection test statistic. The global p-value is found by comparing $LR_{\max}$ with its simulated null distribution. For Poisson data, the null distribution, conditional on y_{tot} and the observed locations, is multinomial and free of unknown parameters. One can also obtain adjusted p-values for each potential cluster by comparing LR_{st} with the simulated null distribution of $\mathrm{LR}_{\max}$.

3. Sample Datasets

We consider three different geographic datasets. The first is synthetic, the others are real. Detailed descriptions of the datasets are given below.

30 × 30 Square Grid: The 30 × 30 square grid is the unit square divided into 900 cells in a regular grid of 30 rows and 30 columns. The expected number of cases, E_i, is identical for all cells. The set of 12,586 potential clusters consists of circular clusters centered at the 900 cells with radii ranging from 0 up to 1/6 unit.

New York Leukemia Data: The New York leukemia dataset (Waller et al., 1994) describes leukemia incidence between 1978 and 1982 in eight counties in upstate New York. There are a total of 592 incident leukemia cases. The two largest cities are Syracuse in the north and Binghamton in the south. The region, approximately 136 km from north to south and 115 km from east to west, is divided into 789 cells with non-coincident centroids for which the population at risk, count of incident leukemia cases, and geographic centroid are available. A total of 179,904 potential circular clusters centered at the 789 distinct cell centroids with radii ranging from 0 up to 20 km were used.

Wisconsin Breast Cancer Data: The Wisconsin breast cancer dataset (Gangnon and Clayton, 2007) describes female breast cancer incidence in 1990 for the state of Wisconsin, which is approximately 500 km from north to south and 420 km from east to west. There are a total of 2407 incident breast cancer cases. The three largest cities in Wisconsin are Milwaukee in the southeast, Madison in the south and Green Bay in the northeast. For each of the 716 zip code areas, geographic centroids were obtained from ArcView, incident breast cancer cases were obtained from the Wisconsin Cancer Registry, and age-specific female population counts (in 5 year intervals) were obtained from the Census Bureau. Expected numbers of breast cancer cases for each zip code were calculated using indirect standardization. A total of 27,843 potential circular clusters centered at the 716 zip code centroids with radii ranging from 0 up to 50 km were used.

4. Variation in Local Multiplicity and Unbiased Cluster Detection

In most settings, the number of potential clusters to be evaluated will vary greatly across the study region, with more potential clusters in higher resolution, urban areas and fewer potential clusters in lower resolution, rural areas. To illustrate the variation in local multiplicity, we display the numbers of potential clusters overlapping each cell for the New York leukemia data and the Wisconsin breast cancer data in Figure 1. For the New York map, the median number of clusters overlapping a cell is 25,880 and the maximum is 117,600, whereas the minimum is only 15. For the Wisconsin map, the median number of cluster overlapping a cell is 638 and the maximum is 4756, whereas the minimum is only 13. Similar variation in local multiplicity is seen using circles with a fixed maximum population radius.

The underlying multiple comparisons problem is more severe in areas with thousands or tens of thousands of overlapping clusters than in areas with only 10 overlapping clusters, and evaluation of statistical significance should account for this phenomenon. One could avoid or minimize local

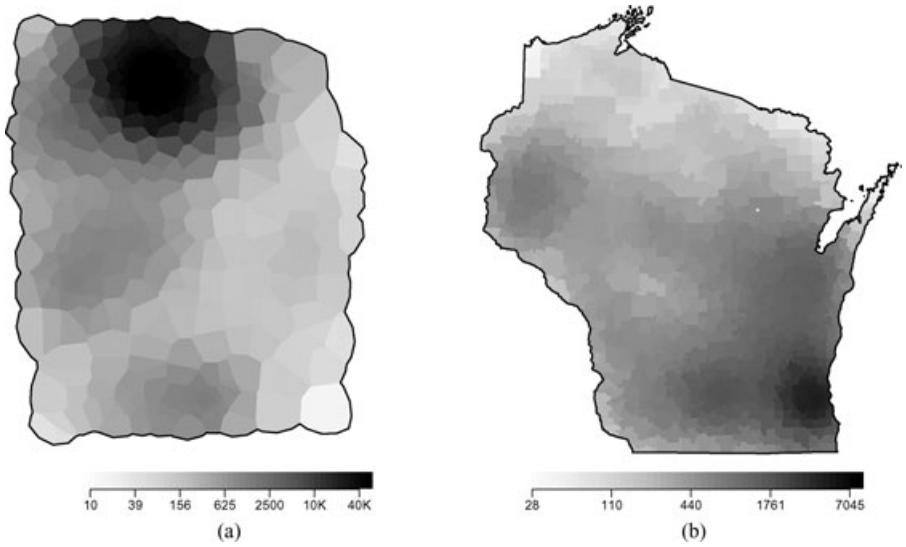

Figure 1. Number of potential clusters overlapping each cell for (a) census tracts/blocks in an eight-county region of New York and (b) zip codes in Wisconsin. Potential clusters are circles with maximum radii of 20 km and 50 km, respectively.

variation in multiplicity by using more restrictive sets of clusters. However, the set of potential clusters should be chosen based on the underlying research questions, not artificial methodological restrictions.

Variation in local multiplicity results in variation in the operating characteristics of the spatial scan statistic across the study region (Gangnon and Clayton, 2001, 2004; Gangnon, 2010). Under the null hypothesis, the spatial scan statistic is more likely to select clusters in areas with fine geographic resolution cells than clusters in areas with coarse geographic resolution (Gangnon and Clayton, 2001, 2004). For comparable clusters, the spatial scan statistic has higher power in urban areas and lesser power in rural areas (Gangnon, 2010). Thus, the spatial scan statistic overstates the evidence for clustering in urban areas relative to rural areas.

To quantify the impact of variations in local multiplicity on the operating characteristics of cluster detection tests, we propose an *unbiased cluster detection* criterion applicable when the test statistic is associated with a specific cluster, e.g., the most likely cluster for the spatial scan statistic.

Unbiased cluster detection is a refinement of the earlier concept of unbiased cluster selection (Gangnon and Clayton, 2001). A cluster detection test satisfies unbiased cluster *selection* if, under the null hypothesis, each cell in the study region has an equal probability of belonging to the selected cluster. This criterion is a property of the cluster selection procedure; it is unaffected by the test results.

In contrast, unbiased cluster detection further incorporates the hypothesis test. A cluster detection test satisfies unbiased cluster *detection* at significance level α if, under the null hypothesis, each cell in the study region has an equal probability of belonging to the selected cluster, conditional on rejection of the global hypothesis test at the α level. This is an intuitive notion of a fair test, similar to unbiased split selection for tree-based models (Loh and Shih, 1997).

We conducted simulation studies using the three different geographic setups and underlying populations described previously: the 30×30 square grid, the New York leukemia data, and the Wisconsin breast cancer data. For each cell, we estimated the proportion of simulations in which the cell belongs to the detected cluster, e.g., the selected cluster when the null hypothesis is rejected at the 5% level, from the 50,000 most extreme of 1,000,000 simulations.

In Figure 2a, we map the probability, under the null hypothesis, of belonging to the selected cluster conditional on the rejection of the null hypothesis at the 5% level for the spatial scan statistic with the 30×30 grid. Grayscale colors indicate the magnitude of deviation from the geometric mean on a logarithmic scale, black lines separate regions with probabilities above and below the geometric mean, and $+/-$ symbols indicate whether the probability is above/below the geometric mean. If the test strictly satisfied the unbiased cluster detection criterion, all cells would be white; greater contrast between light and dark indicates greater bias.

There is little evidence for bias in cluster detection within the interior of the grid, but some evidence for edge effects. Cells on the edge are less likely to belong to the detected cluster (mean detection probability 0.0127) than cells in the interior (mean detection probability 0.0311).

This bias corresponds to the variation in local multiplicity. In the interior, cells belong to the same number of clusters, and there is no bias in cluster detection. On the edge, cells belong to relatively few clusters as circles centered outside the grid cannot be evaluated due to the absence of data outside the study region. The resulting local deficit in the number of clusters results in a lower probability of belonging to the detected cluster.

In Figure 3a, we map the probability, under the null hypothesis, of belonging to the selected cluster conditional on rejection of the null hypothesis at the 5% level for the spatial scan statistic with the New York leukemia dataset. Substantial bias is evident, with cells near Syracuse being more likely to belong to the detected cluster than other cells. The lack of apparent bias in favor of clusters near Binghamton, an

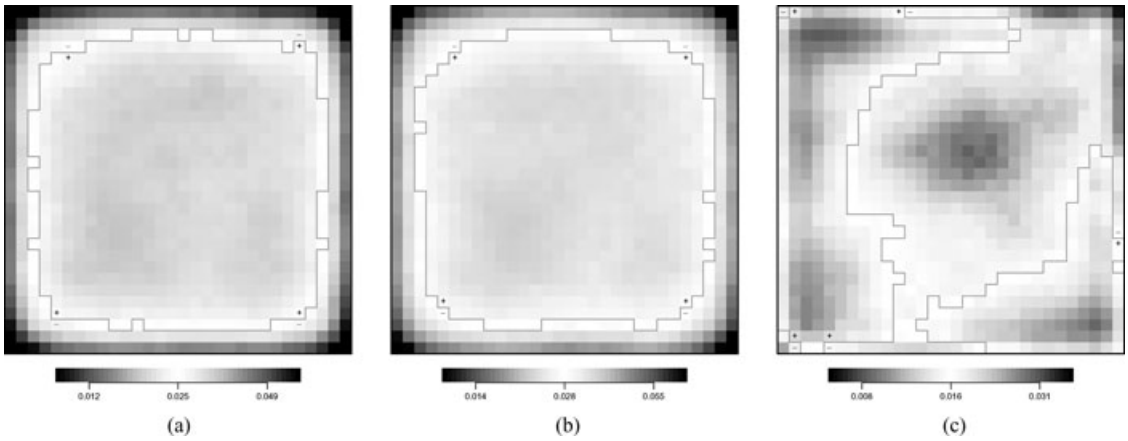

Figure 2. Probability, under the null hypothesis of constant risk, of belonging to the estimated cluster conditional on the global p-value being less than 0.05 for the (a) scan, (b) BLASS, and (c) GLASS procedures for a 30 × 30 regular grid. Grayscale indicates magnitude of deviation from the geometric mean, black lines separate adjacent regions with probabilities above and below the geometric mean, and +/ − symbols indicate areas with probability above/below the geometric mean. Estimated probabilities are based on the 50,000 most extreme of 1,000,000 simulations under the null hypothesis.

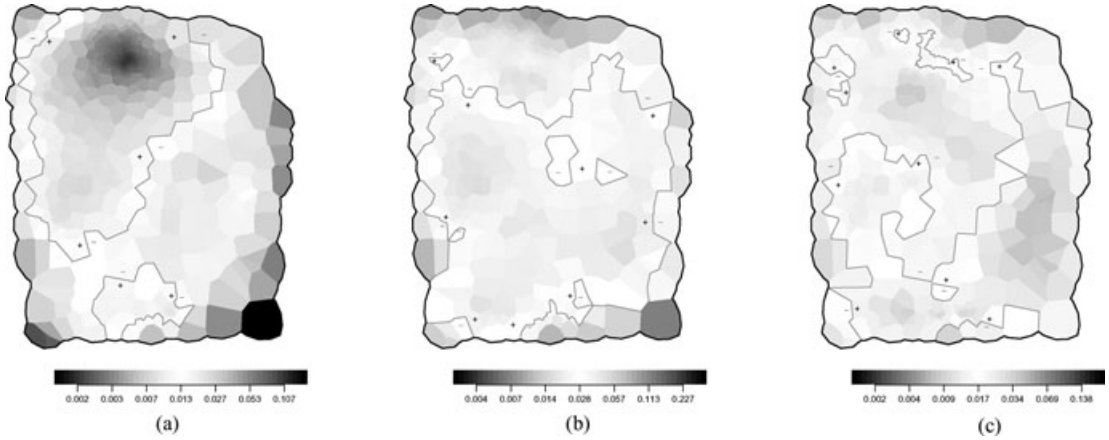

Figure 3. Probability, under the null hypothesis of constant risk, of belonging to the estimated cluster conditional on the global p-value being less than 0.05 for the (a) scan, (b) BLASS, and (c) GLASS procedures for census tracts/blocks in an eight-county region of New York. Grayscale indicates magnitude of deviation from the geometric mean, black lines separate adjacent regions with probabilities above and below the geometric mean, and +/ − symbols indicate areas with probability above/below the geometric mean. Estimated probabilities are based on the 50,000 most extreme of 1,000,000 simulations under the null hypothesis.

area with relatively poor geographic resolution (census tracts instead of census blocks) and relatively high population, is compelling evidence that bias in cluster detection is truly due to variation in local multiplicity rather than variation in population.

The bias in cluster detection largely matches the local multiplicity in Figure 1a. The mean detection probability is 0.0120 for cells in the lowest quartile of local multiplicity, 0.0246 for the second quartile, 0.0612 for the third quartile, and 0.0941 for the top quartile. There is also an edge effect. Cells on the edge are less likely (mean detection probability 0.0105) to belong to the detected cluster than cells in the interior (mean probability 0.0507).

In Figure 4a, we map the probability, under the null hypothesis, of belonging to the selected cluster conditional on rejection of the null hypothesis at the 5% level for the spatial scan statistic for the Wisconsin breast cancer dataset. Substantial bias is evident, with cells in the southeastern portion of the state (roughly the triangle formed by Milwaukee, Madison, and Green Bay) being more likely to belong to the detected cluster than cells elsewhere, especially those in the north.

As with the New York data, the bias in cluster detection largely matches the local multiplicity in Figure 1b. The mean detection probability is 0.0111 for cells in the lowest quartile of local multiplicity, 0.0169 for the second quartile, 0.0234 for the third quartile and 0.0304 for the top quartile. There are also edge effects. Cells on the border are less likely (mean detection probability 0.0121) to belong to the detected cluster than cells in the interior (mean detection probability 0.0223).

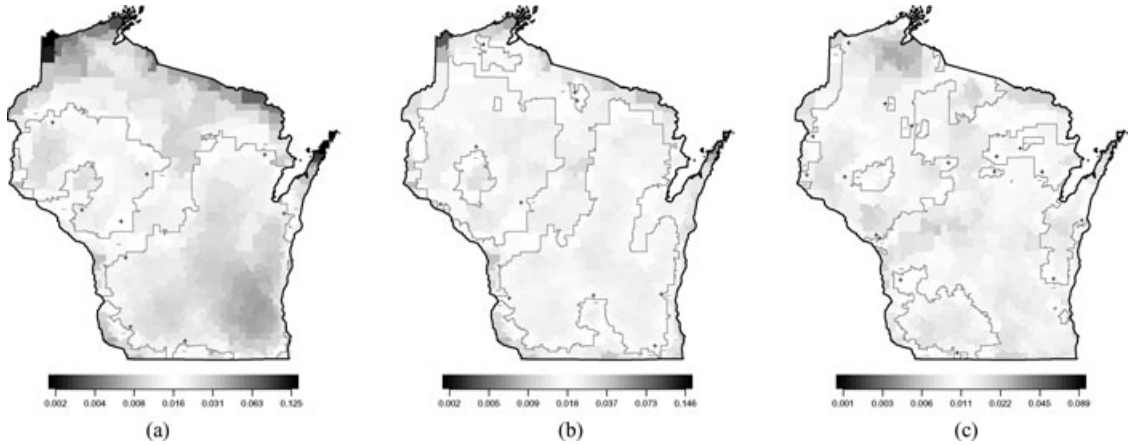

Figure 4. Probability, under the null hypothesis of constant risk, of belonging to the estimated cluster conditional on the global p-value being less than 0.05 for the (a) scan, (b) BLASS, and (c) GLASS procedures for zip codes in Wisconsin. Grayscale indicates magnitude of deviation from the geometric mean, black lines separate adjacent regions with probabilities above and below the geometric mean, and $+/-$ symbols indicate areas with probability above/below the geometric mean. Estimated probabilities are based on the 50,000 most extreme of 1,000,000 simulations under the null hypothesis.

5. Local Multiplicity Adjustments

We present two modifications of the spatial scan statistic intended to reduce or eliminate bias in cluster detection. The first modification, previously proposed by Gangnon (2010), applies a Bonferroni-type adjustment to the nominal p-value for each cluster. We refer to the resulting cluster detection test statistic as the BLASS statistic. To avoid double counting of potential clusters in the Bonferroni correction, the BLASS procedure requires (1) the identification and exclusion of duplicate representations of the same cluster and (2) a conceptual random partitioning of the set of potential clusters to create nonoverlapping neighborhoods.

We next propose a novel modification based on a Gumbel approximation to the distributions of the maximum likelihood ratio test statistics over neighborhoods for each cluster. We refer to the resulting test statistic as the GLASS statistic. The GLASS statistic is unaffected by either duplicate cluster representations or overlapping cluster neighborhoods, but it requires a small set of additional simulations to estimate the parameters of the local Gumbel distributions.

5.1 *Bonferroni Locally Adjusted Spatial Scan Statistic*

Gangnon (2010) proposed the BLASS statistic, which is motivated by a two-stage Bonferroni adjustment to account for local multiplicity. At the first stage, unique representations of potential clusters, previously identified in a preprocessing step, are partitioned into N groups, one for each location. Clusters are randomly assigned to groups based on the selection of a random location inside each cluster. If we denote the assigned group for cluster $\mathbf{Z}_{st}$ by $g(\mathbf{Z}_{st})$ and the number of clusters assigned to group g by $m\{g\}$, the adjusted p-value for cluster $\mathbf{Z}_{st}$ is $Nm\{g(\mathbf{Z}_{st})\}p_{st}$, which is dependent on the random assignment of the clusters to groups. To avoid this, $m\{g(\mathbf{Z}_{st})\}$ is replaced with its expected value over all possible random assignments,

$$M_{st} = \mathrm{E}[m\{g(\mathbf{Z}_{st})\}] = \sum_{g=1}^{N} M_g \delta_{\mathbf{Z}_{st}}(\mathbf{x}_g)/|\mathbf{Z}_{st}| + (1 - 1/|\mathbf{Z}_{st}|),$$

where $M_g = \sum_{s,t} \delta_{\mathbf{Z}_{st}}(\mathbf{x}_g)/|\mathbf{Z}_{st}|$ is the expected number of clusters in group g for $g = 1, 2, \ldots, N$ and $|\mathbf{Z}_{st}|$ is the number of locations inside cluster $\mathbf{Z}_{st}$. The adjusted p-value for cluster $\mathbf{Z}_{st}$ is $NM_{st}p_{st}$.

The minimum of these adjusted p-values, $p^B_{\min} = \min_{s,t} NM_{st}p_{st}$, the BLASS statistic, serves as a global test statistic. The global p-value is obtained by comparing $p^B_{\min}$ with its simulated null distribution as described previously for the spatial scan statistic. The cluster-specific adjustment factors M_{st} account for local variations in the numbers and/or overlap of potential clusters. One can also obtain cluster-specific adjusted p-values, by comparing $p^B_{\min}(\mathbf{x}) = \min\{NM_{st}p_{st}\colon \mathbf{x} \in \mathbf{Z}_{st}\}$ with the simulated null distribution of $p^B_{\min}$.

5.2 *Gumbel Locally Adjusted Spatial Scan Statistic*

As an alternative, we propose the GLASS statistic. The GLASS procedure is motivated by the observation that the null distribution of the spatial scan statistic can be well approximated by the Gumbel distribution (Abrams, Kleinman, and Kulldorff, 2010).

To develop the GLASS statistic, we apply the Gumbel approximation to a series of local scan statistics, one for each potential cluster. Define the neighborhood for cluster Z_{st}, $\mathfrak{N}(Z_{st})$ to be the set of potential clusters that overlap (share at least one cell in common with) the given cluster, e.g., $\mathfrak{N}(Z_{st}) = \{Z_{jk}, k = 1, 2, \ldots, m_j, j = 1, 2, \ldots, N : Z_{jk} \cap Z_{st} \neq \phi\}$. The local scan statistic for cluster Z_{st}, $\mathrm{LR}_{\max}(Z_{st})$, is obtained by scanning over the neighborhood $\mathfrak{N}(Z_{st})$, e.g., $\mathrm{LR}_{\max}(Z_{st}) = \max_{j,k}\{\mathrm{LR}_{jk} : Z_{jk} \in \mathfrak{N}(Z_{st})\}$.

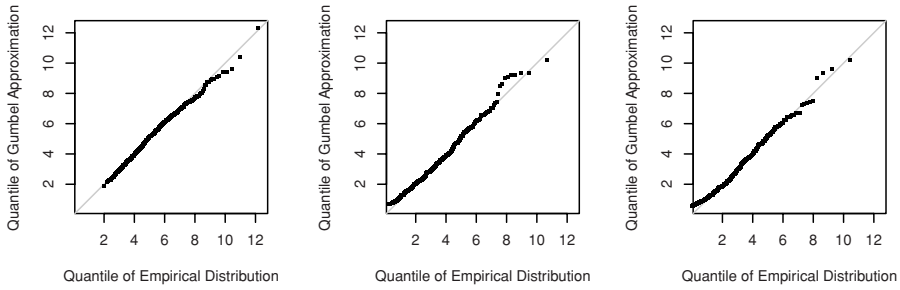

Figure 5. Q–Q plots of the empirical distribution of the local scan statistic against the Gumbel approximation for the local scan statistics with (a) 1799th, (b) 180th, and (c) 18th most poorly fitting (99th, 99.9th, and 99.99th percentiles, respectively) Gumbel approximations in terms of the Cramèr-von Mises criterion among the 179,904 local scan statistics for the New York dataset.

This neighborhood definition (all clusters sharing at least one cell in common with the current cluster) is chosen to maintain a total computational complexity of $O(N^2)$ for the GLASS statistic using circular clusters. One can first find the maximum likelihood ratio statistic for all potential clusters overlapping each cell (one value per cell) in $O(N^2)$ steps and then find the maximum of the statistics for the cells belonging to each cluster in an additional $O(N^2)$ steps, maintaining an overall complexity of $O(N^2)$. Other neighborhood definitions would result in a total complexity of $O(N^4)$.

We approximate the null distribution of the local scan statistic for cluster Z_{st} using the Gumbel distribution (Johnson, Kotz, and Balakrishnan, 1995) with cumulative distribution function,

$$F(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\beta}\right)\right\},$$

where μ is the location parameter and β is the scale parameter. For each local scan statistic $\mathrm{LR}_{\max}(Z_{st})$, we estimate the parameters of its Gumbel distribution by the method of moments from relatively small samples from its null distribution. If we denote the sample mean and standard deviation from the sample as $\bar{x}(Z_{st})$ and $s(Z_{st})$, respectively, the method of moments estimates of the Gumbel parameters are $\hat{\beta}(Z_{st}) = s(Z_{st})\sqrt{6}/\pi$ and $\hat{\mu}(Z_{st}) = \bar{x}(Z_{st}) - 0.5772\hat{\beta}(Z_{st})$ (Johnson et al., 1995).

To verify the accuracy of the Gumbel approximation of the local scan statistic, we obtained 1000 simulations from the null distributions of the local scan statistic for all 179,904 clusters used for the New York data. After ranking these samples in terms of the goodness of fit of the Gumbel approximation to the empirical distribution using the Cramèr-von Mises criterion (Anderson, 1962), we present Q–Q plots for three clusters with poor goodness of fit (ranks 1799, 180, and 18, respectively) in Figure 5. The good appearance of these Q–Q plots selected from the worst available fits strongly supports the adequacy of the Gumbel approximation for the distribution of local scan statistics.

For any cluster Z_{st}, we calculate an adjusted p-value for cluster Z_{st} as $p^G_{st} = P\{G > \mathrm{LR}_{st}|\hat{\mu}(Z_{st}), \hat{\beta}(Z_{st})\}$, where G is a Gumbel random variate with parameters $\hat{\mu}(Z_{st}), \hat{\beta}(Z_{st})$. The minimum of these adjusted p-values, $p^G_{\min} = \min_{s,t} p^G_{st}$, is the GLASS statistic. The global p-value is obtained by comparing $p^G_{\min}$ with its simulated null distribution. The Gumbel parameters account for local variation in the number and overlap of the potential clusters.

In addition to a global p-value, one can also obtain cluster-specific adjusted p-values for the GLASS procedure, by comparing p^G_{st}, respectively, with the simulated null distribution of $p^G_{\min}$.

6. The Impact of Local Multiplicity Adjustment

6.1 *Bias*

To evaluate the impact of local multiplicity adjustment on bias in cluster detection, we revisit the simulation studies described in Section 4. For each cell, we estimated the proportions of simulations in which the cell belongs to the detected cluster from the 50,000 most extreme of 1,000,000 simulations under the null hypothesis. For the GLASS statistic, we used a separate set of 1000 simulations under the null hypothesis to estimate the parameters of the local Gumbel approximations.

In Figure 2b and c, we map the probability, under the null hypothesis, of belonging to the selected cluster conditional on the rejection of the null hypothesis at the 5% level for the BLASS and GLASS statistics with the 30×30 grid. The map for the BLASS statistic is essentially identical to the map for the spatial scan statistic. The mean detection probability is 0.0163 for cells on the edge with 0.0336 for cells in the interior. In contrast, the GLASS statistic shows little, if any, evidence of bias and minimal edge effects. The lack of consistent behavior across the four quadrants of the symmetric grid is indicative of random variation in these samples rather than systematic bias. The mean detection probabilities for cells on the edge and cells in the interior are very similar (0.0151 and 0.0137, respectively).

In Figure 3b and c, we map the probability, under the null hypothesis, of belonging to the selected cluster conditional on rejection of the null hypothesis at the 5% level for the BLASS and GLASS statistics with the New York leukemia data. Both adjustments produce fairly uniform detection probabilities in the interior, largely, but not entirely, eliminating the bias in cluster detection. For BLASS, the mean detection probabilities by quartile (lowest to highest) of local multiplicity, e.g., number of clusters overlapping the cell, are 0.0316, 0.0284,

Table 1
Mean and median power (relative risk 4 of radius 5, 10, and 15 km centered at the cell centroids within each county for the New York dataset using a global 5% level test

Average									
				Scan		BLASS		GLASS	
County	Cells	local multiplicity	Radius	Mean	Median	Mean	Median	Mean	Median
			5	0.1	0.0	0.2	0.0	0.6	0.0
Chenango	38	149	10	7.9	0.0	9.7	0.2	13.7	0.6
			15	34.3	13.4	39.4	25.8	43.8	31.7
			5	1.0	0.0	1.2	0.0	2.4	0.0
Tioga	34	415	10	35.5	1.8	36.2	2.4	39.0	6.0
			15	64.0	89.8	66.4	92.7	67.6	93.5
			5	38.1	0.0	27.9	0.0	39.9	0.0
Cortland	38	713	10	58.3	98.8	54.8	76.3	58.9	99.1
			15	75.6	100.0	74.0	100.0	76.6	100.0
			5	75.0	100.0	75.2	100.0	76.7	100.0
Broome	55	935	10	86.5	100.0	86.2	100.0	86.9	100.0
			15	93.7	100.0	95.2	100.0	94.8	100.0
			5	43.5	0.3	42.2	0.8	43.7	0.7
Tompkins	51	939	10	68.3	100.0	69.6	100.0	69.3	100.0
			15	94.5	100.0	95.4	100.0	94.7	100.0
			5	33.2	0.0	30.8	0.0	33.3	0.0
Cayuga	75	1384	10	45.7	20.8	44.3	17.3	46.5	27.8
			15	76.0	100.0	77.6	100.0	76.3	100.0
			5	0.4	0.0	0.4	0.0	2.1	0.0
Madison	41	2044	10	29.2	3.8	29.4	5.9	33.7	7.3
			15	68.4	91.4	66.4	79.5	72.4	95.3
			5	82.6	100.0	81.2	100.0	82.2	100.0
Onondaga	457	73090	10	95.8	100.0	95.8	100.0	95.7	100.0
			15	98.7	100.0	98.8	100.0	98.6	100.0

0.0189, and 0.0222; for GLASS, the corresponding values are 0.0206, 0.0188, 0.0153, and 0.0182. Both adjustments also reduce the edge effects with GLASS having a greater impact than BLASS. For GLASS, the mean detection probability is 0.0149 for cells on the edge versus 0.0184 for cells in the interior. For BLASS, the comparable values are 0.0193 and 0.0257.

In Figure 4b and c, we map the probability, under the null hypothesis, of belonging to the selected cluster conditional on rejection of the null hypothesis at the 5% level for the BLASS and GLASS statistics with the Wisconsin breast cancer data. Similar to our observations with the New York data, both adjustments produce fairly uniform detection probabilities in the interior, e.g., little bias in cluster detection. For BLASS, the mean detection probabilities by quartile of local multiplicity are 0.0170, 0.0113, 0.0202, and 0.0186. For GLASS, the corresponding values are 0.0127, 0.0113, 0.0114, and 0.0103. GLASS virtually eliminated edge effects, whereas BLASS had little, if any, impact on these edge effects. For GLASS, the mean detection probability is 0.0108 for cells on the edge versus 0.0112 for cells in the interior. For BLASS, the comparable values are 0.0129 and 0.0199, not much changed from the values seen for the spatial scan statistic (0.0121 and 0.0223).

6.2 *Power*

We next evaluate the impact of these local multiplicity adjustments on power to detect specific clusters. Here, we define power as the probability, under a specified single cluster alternative within the set of potential clusters, that the cluster-specific adjusted p-value for that single cluster is less than 0.05. Many other definitions of power for cluster detection tests have been considered in the literature, incorporating different notions of sufficiently correct cluster identification into the power calculation (Gangnon and Clayton, 2004; Waller et al., 2006; Gangnon, 2010). The different definitions of power have little, if any, impact on assessments of comparative performance of the same test across the study region or of different tests at the same location. An advantage of the definition of power used here is that simulations are only needed under the null hypothesis to identify the critical values for the cluster-specific hypothesis test. Given the critical value for a given cluster, power can be found using exact or approximate binomial probability calculations, allowing us to evaluate power for much larger collections of clusters than is feasible via simulation.

Using the null simulations described in Sections 4 and 6, we identified the cluster-specific critical values associated with the global 5% level test for each test for the New York data. We considered all possible single cluster alternatives with a cluster relative risk of 4. We report mean and median power for clusters of sizes 5, 10, and 15 km centered within each county in Table 1. Power is universally higher for clusters with larger radii and generally higher for clusters in urban areas (Broome and Onondaga counties) relative to rural areas. GLASS has increased power (sometimes minimal) for all cluster sizes in counties with relatively low local multiplicity with substantial increases in power for larger clusters in

Chenango, Madison, and Tioga Counties. The loss of power in Onondaga County, the Syracuse area with the highest local multiplicity, is minimal. BLASS has similar impacts on power, but the results are less consistent.

6.3 *Computational Speed*

The major determinant of total runtime is the number of simulations used for the randomization test. For GLASS, we have specified the number of simulations for estimation of the local Gumbel parameters as a fraction (0.10) of the number used for the randomization test. In future work, we will explore the impact of different choices for this fraction. With 10,000 simulations for the randomization test, the Monte Carlo standard error for global p-values near the significance threshold of 0.05 is 0.002 for all three methods.

For a large number of simulated datasets, GLASS was shown to have very similar runtime (typically 20–40% slower) than the standard scan statistic approach and ran almost twice as fast as BLASS. The relatively long runtime for BLASS is due to the computational requirements in evaluating tail probabilities of the chi-square distribution for every potential cluster in each simulation. For GLASS, such explicit evaluations of the comparable tail probabilities for the Gumbel distribution can be avoided as it is a location-scale family.

7. Data Analyses

7.1 *New York Leukemia Data*

We assessed the evidence for clustering in the New York leukemia data using all three tests. We considered circular clusters centered at the cell centroids with radii less than or equal to 20 km. Using all three tests, there is strong evidence for a primary cluster of elevated risk in the city of Binghamton ($p = 0.003$ for the spatial scan, $p = 0.002$ for BLASS, and $p = 0.003$ for GLASS), moderate evidence for secondary clusters of elevated risk in Cortland County ($p = 0.050$, $p = 0.024$, and $p = 0.024$, respectively), and reduced risk north of Syracuse in Onondaga County ($p = 0.018$, $p = 0.050$, and $p = 0.043$, respectively). The spatial scan statistic finds weak evidence for an additional cluster of reduced risk northwest of Syracuse in Onondaga County ($p = 0.084$).

Although similar conclusions are drawn from all three tests, differences in the p-values, e.g., the strengths of evidence for clustering, reflect the differing consequences of local multiplicity on the three test statistics. For example, the spatial scan statistic finds stronger evidence (smaller p-values) for the clusters near Syracuse and lesser evidence (larger p-values) for the cluster in Cortland County than BLASS or GLASS.

7.2 *Wisconsin Breast Cancer Data*

We assessed the evidence for clustering in the Wisconsin breast cancer incidence data using all three tests. We considered circular clusters centered at the cell centroids with radii less than or equal to 50 km. Using any of the three tests, there is at most weak evidence for clustering ($p = 0.073$ for the spatial scan, $p = 0.11$ for BLASS, and $p = 0.072$ for GLASS). The primary cluster for the spatial scan and BLASS consists of three zip codes of lowered risk in Fond du Lac County ($p = 0.082$ for GLASS), whereas the primary cluster for GLASS is a single zip code of lowered risk in Kenosha County ($p = 0.15$ for the spatial scan and $p = 0.15$ for BLASS). All three statistics also find very weak evidence for a single zip code of elevated risk in Kenosha County ($p = 0.20$ for the spatial scan, $p = 0.18$ for BLASS and $p = 0.077$ for GLASS). The spatial scan statistic finds very weak evidence for an extremely large area of elevated risk west of Milwaukee ($p = 0.16$).

Although there is, at most, weak evidence for clustering in the Wisconsin breast cancer data, differences in the p-values for the three tests clearly reflect the differing impacts of local multiplicity. For example, the spatial scan statistic finds some evidence (smaller p-values) in suburban Milwaukee, an area of very high levels of local multiplicity, and lesser evidence (larger p-values) in a more rural area (Fond du Lac County) and near the state border and Lake Michigan (Kenosha County) than BLASS or GLASS. In addition, differences in the p-values for the two clusters in Kenosha County using BLASS or GLASS reflect greater reductions in the edge effect with GLASS.

8. Conclusions

Both the BLASS statistic and the GLASS statistic are attractive alternatives to the spatial scan statistic that can account for the local variation in the extent of the multiplicity problem. The BLASS statistic evaluates the evidence for a given cluster (likelihood ratio statistic) relative to the number of unique clusters within a defined neighborhood of the cluster of interest with a Bonferroni adjustment. The GLASS statistic evaluates the evidence for a given cluster (likelihood ratio statistic) relative to the (approximate Gumbel) distribution of the maximum likelihood ratio statistic over all clusters within a neighborhood of the cluster of interest.

The same benchmark for rejecting the null hypothesis, a global p-value less than 0.05, was used for all three procedures. For each procedure, one first calculates the global test statistic, $p_{\min}$ for the spatial scan statistic, $p^B_{\min}$ for the BLASS statistic or $p^G_{\min}$ for the GLASS statistic. The specified global test statistic is then compared to its simulated null distribution to obtain a single global p-value.

Both the BLASS and GLASS statistics outperformed the spatial scan statistic in terms of unbiased cluster detection. With the spatial scan statistic, locations in urban areas with relatively high geographic resolution are more likely to belong to the detected cluster than locations in rural areas with low geographic resolution; there is also a clear edge effect such that locations on the edge of the study region are less likely to belong to a detected cluster. Both the BLASS and GLASS statistics minimize the bias in the interior of the study region; the GLASS statistic also appears to reduce the impact of edge effects.

Of the two local multiplicity adjustments, we recommend the GLASS statistic over the BLASS statistic because of reduced edge effects; the persistence of edge effects with BLASS likely reflects an undercounting of the true local multiplicity for clusters in the center of the study region relative to clusters on its edge. Furthermore, to avoid double counting potential clusters in the correction, the BLASS procedure requires (1) the identification and exclusion of duplicate representations of the same cluster, e.g., circles with different centers that

contain identical sets of cell centroids, from the set of potential clusters and (2) a conceptual random partitioning of the set of potential clusters to create nonoverlapping neighborhoods. In contrast, the GLASS statistic requires only a slight modification of the traditional scan statistic using a small set of additional simulations under the null hypothesis to estimate the parameters of the Gumbel approximations to the local scan statistic.

The adjustments for local multiplicity discussed here utilize two different definitions for the neighborhood of a cluster. For the BLASS statistic, the neighborhoods of clusters are random, but nonoverlapping, e.g., each cluster belongs to exactly one neighborhood. For the GLASS statistic, the neighborhoods of clusters are fixed, but overlapping. Adaptation of the BLASS statistic to the GLASS neighborhoods would be straightforward, but the overlapping neighborhoods might undermine the justification for the local Bonferroni adjustment. Adaptation of the GLASS statistic to the BLASS neighborhoods or to neighborhoods based on other measures of overlap is conceptually straightforward, but it would greatly increase the computational burden relative to the spatial scan statistic or the GLASS statistic with the current neighborhoods from $O(N^2)$ to $O(N^4)$. We plan to explore possible algorithms for efficiently implementing the GLASS statistic with more general neighborhood definitions in future work.

The BLASS and GLASS statistics depend on the spatial structure of the underlying dataset only through the cluster memberships of the observed locations. The adaptation of these methods to case–control data is straightforward, although the performance in that setting needs to be evaluated. Adaptations to other scan statistics for which the potential clusters are enumerated, e.g., the flexible spatial scan statistic (Tango and Takahashi, 2005) or the elliptic spatial scan statistic (Kulldorff et al., 2006), is conceptually straightforward, but future work is required to make these adjustments computationally feasible. Adaptations to scan statistics that do not enumerate the potential clusters, e.g., the upper level set scan statistic (Patil and Taillie, 2004), the implementation of the spatial scan statistic using a branch-and-bound algorithm (Neill and Moore, 2004), the multiobjective scan statistic (Duczmal et al., 2008), and the linear-time subset scan statistic (Neill, 2008), are more difficult and will be an important area of future work.

Acknowledgements

I thank Amy Trentham-Dietz, Jane McElroy, and John Hampton for their assistance with the Wisconsin breast cancer dataset. I also thank Tom Cook, Murray Clayton, and two anonymous referees for their helpful comments on this article.

References

Abrams, A. M., Kleinman, K., and Kulldorff, M. (2010). Gumbel based p-value approximations for spatial scan statistics. *International Journal of Health Geographics* **9,** 61.

Anderson, T. W. (1962). On the distribution of the two-sample Cramèr-von Mises criterion. *Annals of Mathematical Statistics* **33,** 1148–1159.

Duczmal, L., Cançado, A. L. F., and Takahashi, R. H. C. (2008). Geographic delineation of disease clusters through multi-objective optimization. *Journal of Computational and Graphic Statistics* **17,** 243–262.

Gangnon, R. E. (2010). Local multiplicity adjustments for spatial cluster detection. *Environmental and Ecological Statistics* **17,** 55–71.

Gangnon, R. E. and Clayton, M. K. (2001). A weighted average likelihood ratio test for spatial clustering of disease. *Statistics in Medicine* **20,** 2977–2987.

Gangnon, R. E. and Clayton, M. K. (2004). Likelihood-based tests for detecting spatial clustering of disease. *Environmetrics* **15,** 797–810.

Gangnon, R. E. and Clayton, M. K. (2007). Cluster detection using Bayes factors from overparameterized models. *Environmental and Ecological Statistics* **14,** 69–82.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Volume 2, 2nd edition. New York: Wiley.

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

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

Kulldorff, M., Huang, L., Pickle, L., and Duczmal, L. (2006). An elliptic spatial scan statistic. *Statistics in Medicine* **25,** 3929–3943.

Loh, W.-Y. and Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica Sinica* **7,** 815–840.

Neill, D. B. (2008). Fast and flexible outbreak detection by linear-time subset scanning. *Advances in Disease Surveillance* **5,** 48.

Neill, D. B. and Moore, A. W. (2004). Rapid detection of significant spatial clusters. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '04), 256–265. New York, NY: ACM Press.

Patil, G. P. and Taillie, C. (2004). Upper level set scan statistic for detection arbitrarily shaped hotspots. *Environmental and Ecological Statistics* **11,** 183–197.

Tango, T. and Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* **4,** 11.

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

Waller, L. A., Hill, E. G., and Rudd, R. A. (2006). The geography of power: Statistical performance of tests of clusters and clustering in heterogeneous populations. *Statistics in Medicine* **25,** 853–865.

Received March 2010. *Revised April* 2011.
Accepted May 2011.