Applied Geography 60 (2015) 29–36

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Identifying hotspots of human anthrax transmission using three local clustering techniques

Alassane S. Barro [a], Ian T. Kracalik [a], Lile Malania [b], Nikoloz Tsertsvadze [b], Julietta Manvelyan [b], Paata Imnadze [b], Jason K. Blackburn [a, *]

[a] Spatial Epidemiology & Ecology Research Laboratory, Department of Geography and Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA
[b] National Center for Disease Control and Public Health, Tbilisi, Georgia

ARTICLE INFO

Article history:
Available online

Keywords:
Cluster morphology
Local spatial clusters
Human cutaneous anthrax
Georgia
Cluster sensitivity analyses
Cluster specificity analyses

ABSTRACT

This study compared three local cluster detection methods to identify local hotspots of human cutaneous anthrax (HCA) transmission in the country of Georgia where cases have been steadily increasing since the dissolution of the Soviet Union. Recent reports have indicated that the disease has reached historical levels in 2012 highlighting the need for better informed policy recommendations and targeted control measures. The purpose of this paper was to identify spatial clusters of HCA to aid in the implementation of targeted public health interventions. At the same time, we compared the utility of different statistical tests in identifying hotspots. We used the Getis-Ord ($G_i^*(d)$), a multidirectional optimal ecotope-based algorithm (AMOEBA) – a cluster morphology statistic, and the spatial scan statistic in SaTScan™. Data on HCA cases from 2000 to 2012 at the community level were aggregated to an 8 × 8 km grid surface and population data from the Global Rural and Urban Mapping Project (GRUMP) were used to calculate local incidence. In general, there was agreement between tests in the locations of HCA hotspots. Significant local clusters of high HCA incidence were identified in the southern, eastern and western regions of Georgia. The $G_i^*(d)$ and spatial scan statistics appeared more sensitive but less specific than the AMOEBA algorithm. The scan statistic identified larger geographic areas as hotspots of transmission. In general, the spatial scan statistic and $G_i^*(d)$ performed well for spatial clusters with lower incidence rates, whereas AMOEBA was well suited for defining local spatial clusters of higher HCA incidence. In resource constrained areas, efficient allocation of public health interventions is crucial. Our findings identified hotspots of HCA that can be used to target public health interventions such as livestock vaccination and training on proper outbreak management. This paper illustrates the benefits of evaluating statistical approaches for defining disease hotspots and highlights differences in these clustering approaches applicable beyond public health studies.

Introduction

Generally, spatial clustering statistics identify if, where, and the spatial scale at which the spatial distribution of observed phenomena significantly differs from an expectation of the distribution of those phenomena under complete spatial randomness. When applied to disease, clustering can be defined as an excess of reported cases in space, time, or both space and time (hotspots) (Jacquez, Waller, Grimson, & Wartenberg, 1996) or regions with fewer than expected cases (cold spots); though these same statistics can be applied broadly across spatial datasets. Broadly, these statistical methods can be categorized into global, local, and focal clustering. Global clustering tests, including the Moran's *I* (Moran, 1950) and the Ripley's *K* (Ripley, 1977) (to name a few commonly applied in the literature), are used to evaluate whether events are clustering over a study area, and in both cases the spatial scale at which clustering is maximized. For example, Kracalik et al. (2013) employed a spatial correlogram to plot Moran's *I* value measured across a range of distance thresholds to illustrate the level of clustering of anthrax cases in Georgia. Likewise, O'Brien et al. (1999) described methods to define the maximum scale of clustering for Ripley's *K* plots, such as comparing the difference

* Corresponding author.
E-mail address: jkblackburn@ufl.edu (J.K. Blackburn).

http://dx.doi.org/10.1016/j.apgeog.2015.02.014

between the observed and expected values across distances. These techniques do not identify specific regions within a study area where events are clustering, whereas local statistics can identify the geographic position and spatial scale of clusters (Auchincloss, Gebreab, Mair, & Diez Roux, 2012).

Local statistics encompass some of the commonly used distance-based statistics in spatial epidemiology such as the $G_i^*(d)$ statistics of Getis and Ord (Getis & Ord, 1992; Ord & Getis, 1995), and the spatial scan statistics (Kulldorff, 1997). For example, Kao, Getis, Brodine, and Burns (2008) used the $G_i^*(d)$ statistic to identify spatial and temporal clusters of Kawasaki syndrome in San Diego, California. DeGroote, Sugumaran, Brend, Tucker, and Bartholomay (2008) used the $G_i^*(d)$ to elucidate patterns of West Nile Virus (WNV) incidence in Iowa. In the latter study, the authors used different distance thresholds to identify the spatial extent of transmission, finding that disease rates were clustered in the western half of the state. Similarly, the spatial scan statistic has been used to identify high risk clusters of pulmonary non-tuberculosis mycobacterium in counties across the USA (Adjemian et al., 2012). Focal statistics, on the other hand, may be used to explore clustering of a disease around a specific location such as an environmental pollutant (Jacquez, 2008) and may be modifications of these local measures. For example, Clennon et al. (2004) evaluated local clusters of human schistosomaisis in relation to specific water features using a modification of the Getis-Ord statistic.

More recently, morphological spatial cluster detection techniques have been developed to capture the shape of spatial clusters. Some of these techniques include the following: a multidirectional optimum ecotope-based algorithm (AMOEBA) (Aldstadt & Getis, 2006), the maxima-likelihood-first algorithm and the non-greedy growth algorithm (Yao, Tang, & Zhan, 2011), the flexibly shaped spatial scan statistic (FleXScan) (Tango & Takahashi, 2005), and the cluster morphology analysis (CMA) (Jacquez, 2009). These techniques employ search pattern algorithms to identify the shape of clusters. For example, AMOEBA employs a search algorithm that begins with a random seed on the landscape (in a given spatial unit) and employs the Getis-Ord statistic to calculate a local statistic. It then moves to a near neighbor and determines if that neighbor increases the test statistic from the previous cell, if it does, the cluster grows in that direction and the process iterates growing the statistic across cells as they contribute to the cluster. This approach differs from a traditional Getis-Ord statistic, which is traditionally employed with a search defined by a circle of radius d (set in map units) or using a contiguity matrix defining neighbor connections based on a weights matrix. AMOEBA has been used to identify cluster shapes of socioeconomic and physical environmental factors to identify continuous groupings or neighborhoods of characteristics (Weeks, Getis, Hill, Agyei-Mensah, & Rain, 2010), and for the detection of significant local clusters of livestock anthrax in Kazakhstan (Kracalik et al., 2012).

In this current study, we apply local clustering techniques to search for cluster of anthrax. Anthrax is a growing veterinary and public health concern in the country of Georgia. Recent reports have indicated a dramatic increase in the incidence of human cutaneous anthrax (HCA) while livestock cases remain under-reported (Kracalik, Malania, et al., 2014). The causative agent of the disease, *Bacillus anthracis*, is a soil-borne bacterium with a remarkable ability to survive in the environment for long periods of time (Hugh-Jones & Blackburn, 2009). Studies have shown that the geographic distribution of the bacterium is limited by a combination of environmental characteristics including soil pH, several soil minerals, soil moisture, and temperature (Blackburn, McNyset, Curtis, & Hugh-Jones, 2007; Griffin, Petrosky, Morman, & Luna, 2009; Griffin et al., 2014; Hugh-Jones & Blackburn, 2009; Kracalik et al., 2012; Smith et al., 2000). Human transmission of anthrax is generally a direct result of coming into contact with infected animals or contaminated materials, hence control of the disease in humans is dependent upon targeting control efforts in animals. Previous research in the neighboring country of Azerbaijan has suggested that public health interventions such as anthrax livestock vaccination, and proper outbreak management can reduce the occurrence of human cases (Kracalik, Abdullayev, et al., 2014). Ideally, areas with a high incidence of livestock anthrax would be targeted for control measures, however, in Georgia reporting is anthropocentric, relying heavily on the dissemination of human reporting (Kracalik, Malania, et al., 2014). To achieve more effective levels of disease management, a recent study (Kracalik, Malania, et al., 2014) has suggested that in resource constrained environments identifying hotspots of transmission may allow for a better allocation of public health services.

As a case study comparing multiple spatial clustering approaches, we examine the distribution of HCA in the country of Georgia. Given the countries limited resources, identifying hotspots of anthrax transmission may allow for better allocation of public health services, such as livestock vaccination. The availability of high resolution (community-level) HCA data provides an opportunity to identify hotspots of human transmission while comparing the utility of three different statistical methods: the Getis-Ord $G_i^*(d)$, AMOEBA, and the spatial scan statistic in SaTScan™. At the same time, we comment on the differences in these tests that may be useful for guiding future exploratory studies, including those unrelated to public health.

Methods

Data processing

We used data on HCA cases reported at the community-level from 2000 to 2012 in Georgia. A GIS database of 171 communities with at least one reported HCA case was constructed in ArcGIS v10 for the time period 2000 to 2012. Locations of communities reporting human anthrax were geocoded following Kracalik et al. (2013) (Fig. 1). Here we aggregated the total cases per grid cell using an 8×8 km spatial resolution. This size allowed us to test for local spatial clusters with each statistic proposed. The calculation the AMOEBA statistic is computationally intensive and would not complete cluster detection at resolutions smaller than 8×8 km in less than 10 days per iteration. Kracalik et al. (2012) also encountered a similar limitation when comparing the performance of AMOEBA and $G_i^*(d)$ in evaluating the spatial patterns of livestock in Kazakhstan. Grid cells were generated for the entire country of Georgia using the 'genshapes' command in the Geospatial Modeling Environment (Beyer, 2012). HCA cases were aggregated to the grid surface and each of the grid cells was considered the unit of analysis for each given statistical test.

A population count grid for the year 2000 at the spatial resolution 30 arc-second (~1 km) for the country of Georgia was downloaded from the Global Rural-Urban Mapping Project (GRUMP) website (http://sedac.ciesin.columbia.edu/data/collection/grump-v1) to derive human population estimates per grid cell using the zonal statistics routine in ArcGIS. Cumulative HCA incidence was calculated by dividing the total number of anthrax cases in each cell grid by the estimated median year population (2007). The population for the median year was derived following Kracalik et al. (2013).

Smoothing crude incidence rates has been suggested to stabilize variability in disease incidence rates caused by variations in the numerators (e.g. numbers of cases) and denominators (such as population at risk) (Kafadar, 1996). Empirical Bayes smoothing (EBS) is one such method for stabilizing incidence rates before

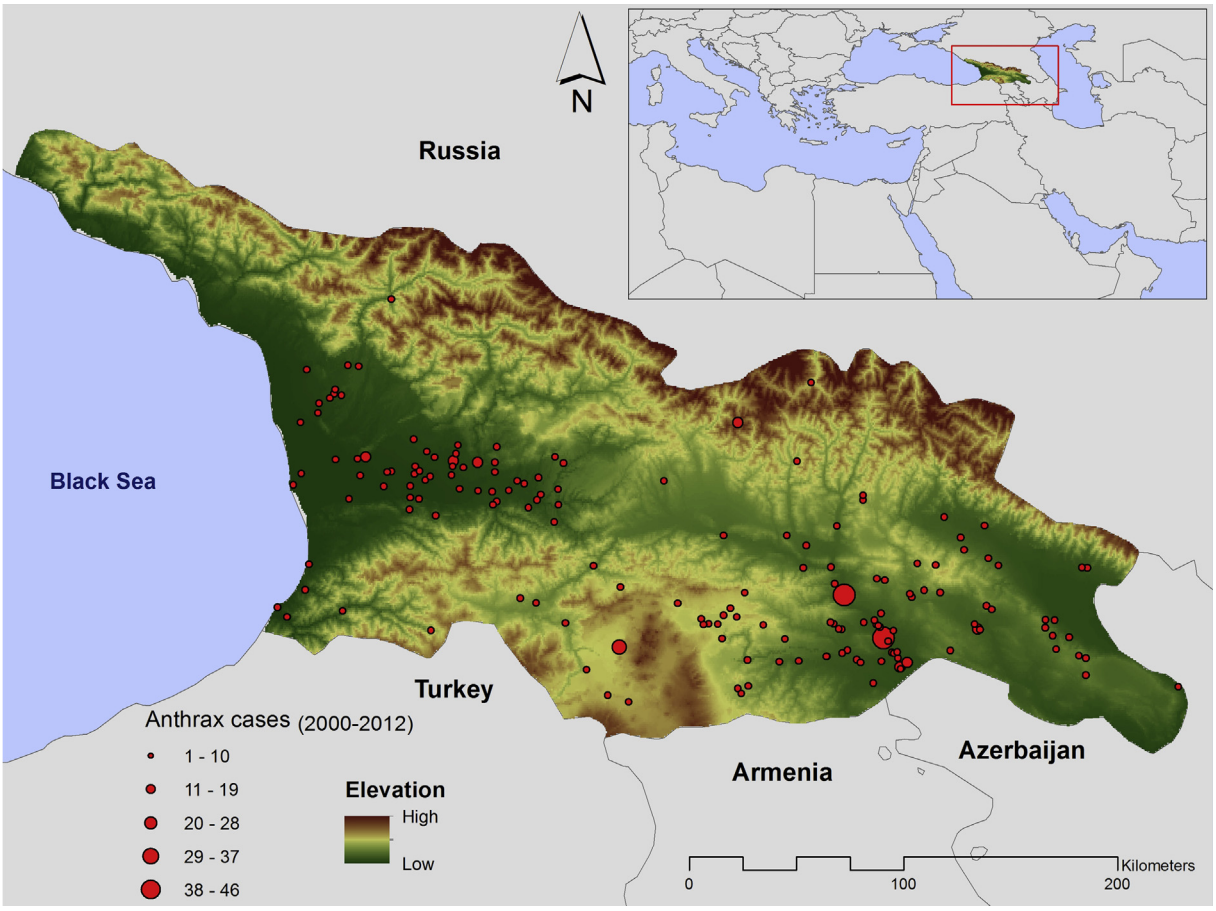

Fig. 1. The distribution of human cutaneous anthrax (HCA) cases per village from 2000 to 2012 in the country of Georgia.

mapping (Devine, Louis, & Halloran, 1994; Leyland & Davies, 2005). Variation in HCA was expected in small spatial units, in particular where case numbers were relatively high and the population size relatively low. To account for this in Georgia, EBS was performed in openGeoDa (Anselin & McCann, 2009). Smoothed rates per grid cell were used for the Getis-Ord $G_i^*(d)$ and AMOEBA statistic tests (see below). SaTScan only accepts discrete data for the number of disease cases and the population at risk per geographic location as inputs and derives rates as part of the modeling process (see below).

Spatial analyses

Getis-Ord $G_i^*(d)$ statistic

The Getis-Ord $G_i^*(d)$ statistic was run using ArcGIS software (v. 10.1, Esri, Inc. Redlands, California, USA) to identify spatial clusters of high HCA incidence. The $G_i^*(d)$ statistic is defined as (Ord & Getis, 1995):

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j - w_i^*\bar{x}}{S\left\{\left[\left(nS_{1i}^*\right) - w_i^{*2}\right]\Big/(n-1)\right\}^{\frac{1}{2}}}$$

where $\bar{\boldsymbol{x}}$ is the mean of all human anthrax outbreaks within the country; $\boldsymbol{S}$ is the standard deviation; $\boldsymbol{n}$ is the total number of grid cells; $\boldsymbol{w_{ij}}$ is a binary weights matrix used to determine the spatial structure and association among locations in the dataset; if the distance from a neighbor $\boldsymbol{j}$ to the feature $\boldsymbol{i}$ is within the distance $\boldsymbol{(d)}$, then $\boldsymbol{w_{ij}} = \boldsymbol{1}$; otherwise $\boldsymbol{w_{ij}} = \boldsymbol{0}$; $w_i^* = \sum_j w_{ij}(d)$; $S_{1i}^* = \sum_j w_{ij}^2$.

The Getis-Ord $G_i^*(d)$ statistic repeated with d set at 5–100 km in 5 km intervals. The statistic is interpreted using standardized z-scores to define hotspots of the variable of interest. Positive z-scores indicate high values of the variable and low z-scores suggest low values within a specified distance d of feature i (Getis & Ord, 1992). Clusters of high HCA incidence were defined at a particular spatial distance if the z-score was greater than 3.18 ($p \leq 0.001$) and the highest at that distance compared to the other shorter spatial distances. To illustrate this, let us consider the $G_i^*(d)$ values of a single grid cell (expressed as the z-scores) at 20 km, 40 km and 60 km to be 3.75, 4.35, and 3.25 respectively. The cell will be defined as a member of a spatial cluster will be defined at 40 km since the highest z-score is observed at that distance. This approach was suggested in Ord and Getis (1995) and Getis, Morrison, Gray, and Scott (2003) to identify the distance at which spatial autocorrelation is greatest. This is defined as the critical distance d_c (Getis & Aldstadt, 2010; Getis & Griffith, 2002; Kracalik et al., 2012).

AMOEBA statistic

The second method employed was AMOEBA (Aldstadt & Getis, 2006). AMOEBA is a clustering algorithm that uses $G_i^*(d)$ values to identify morphological spatial clusters (or ecotopes) of high or low values (Aldstadt & Getis, 2006; Duque, Aldstadt, Velasquez, Franco, & Betancourt, 2011). The algorithm calculates a $G_i^*(d)$ value for each cell i and for contiguous neighbor j. Then an iterative multidirectional search identifies every neighboring cell j that maximizes (either positively or negatively) the $G_i^*(d)$ value for cell i. If the value at i and a set of neighbors j is greater than the $G_i^*(d)$ value for cell i alone, then j, or j neighbors, is/are included to the ecotope (Aldstadt

& Getis, 2006; Duque et al., 2011). This iterative process is repeated until additional cells fail to increase the absolute value of the $G_i^*(d)$ statistic. The AMOEBA clustering algorithm was performed with an $\alpha = 0.001$ to define significance. AMOEBA allows of one of three parameters for limiting the maximum size of a cluster: a Bonferroni adjustment, the false discovery rate (FDR) and the core cutoff (which is defined by a weights matrix). Kracalik et al. (2012) reported an increase in the number of spatial units designated as "outside of a cluster", or units not identified as high or low. For this study, we applied the core cut-off to limit the cluster size. We used a distance-based weights matrix of 25 km to define spatial weights (equivalent to three neighboring 8 × 8 km cells). The weights matrix was constructed in SpaceStat (http://www.biomedware.com). All neighbor grids within the threshold distance were assigned a weight of 1, and those outside a weight of 0.

The circular spatial scan statistic

Third, we used the spatial scan statistic implemented in SaTScan™ (Kulldorff, 1997) to detect spatial clusters of high HCA incidence. We employed the retrospective space-only Poisson model, as it is appropriate for case and population data. The space-only method employs circular moving windows of varying diameter, each varying up to a maximum size of a user-defined population at risk. Each circle is considered a potential candidate cluster. We used five maximum spatial cluster sizes (10, 20, 30, 40, and 50% of the population at risk) to identify different spatial cluster sizes. For each window, SaTScan™ uses a Monte Carlo simulation to test the null hypothesis that there was not an elevated risk of human anthrax. The maximum number of replications for Monte Carlo simulation was set to 999. The likelihood function under the Poisson assumption for a specific window is proportional to (Kulldorff, 2011):

$$\left(\frac{c}{E[c]}\right)^{c}\left(\frac{C-c}{C-E[c]}\right)^{C-c} I()$$

where $\boldsymbol{C}$ is the total number of anthrax outbreaks; $\boldsymbol{c}$ is the observed number of anthrax outbreaks within the window; $\boldsymbol{E[c]}$ is the expected number of anthrax outbreaks within the window under the null hypothesis; $\boldsymbol{I()}$ is an indicator function. When SaTScan is set to scan only for clusters with high incidence rates, $\boldsymbol{I()}$ is equal to 1 when the window has more anthrax cases than the expected under the null hypothesis.

The likelihood function is maximized over all window locations and sizes. The window with the greatest maximum likelihood value constitutes the most likely cluster. The statistic was run using the total observed anthrax cases and the median population per grid cell for the time period 2000 to 2012.

Sensitivity analyses

Definition of true spatial clusters

For this study, we aimed to evaluate differences in the three statistics for detecting clusters of HCA by deriving measures of sensitivity and specificity. Toward this, we needed to provide a clear definition of true spatial clusters. Spatial clusters have been defined as an excess of cases in space, time, or space and time (Jacquez et al., 1996) or "an aggregation of cases in an identifiable subpopulation" (Wartenberg, 2001). In this study, we used smoothed incidence rates to identify true spatial clusters. Because there is no clear definition of the magnitude of clusters, we considered all grid cells with smoothed incidence rates greater than or equal to the 95th percentile as true spatial clusters, considering the suggestion of Jacquez et al. (1996) that "true clusters explain fewer than 5% of all reported clusters."

Sensitivity, specificity and accuracy

We used sensitivity analyses to evaluate the performance of each statistical test to correctly detect true spatial clusters characterized by incidence rates greater than the 95th percentile (Tables 1 and 2). The proportion of grid cells that were correctly identified as true spatial clusters was calculated following Fielding and Bell (Fielding & Bell, 1997)

$$\boldsymbol{S} = \frac{\boldsymbol{a}}{(\boldsymbol{a}+\boldsymbol{c})}$$

where $\boldsymbol{S}$ is the sensitivity of a method; $\boldsymbol{a}$ is the total number of true positive spatial clusters, and $\boldsymbol{c}$ is the total number of false negative spatial clusters.

Additionally, the specificity test ($\boldsymbol{Sp}$), which represents the proportion of grid cells that were correctly identified as true negative spatial clusters, and the accuracy test ($\boldsymbol{A}$) illustrating the proportion of true positive and negative spatial clusters for each cluster detection method were defined by

$$\boldsymbol{Sp} = \frac{\boldsymbol{d}}{(\boldsymbol{b}+\boldsymbol{d})}$$

where $\boldsymbol{Sp}$ is the specificity of a method; $\boldsymbol{d}$ is the total number of true negative spatial clusters, and $\boldsymbol{b}$ is the total number of false positive spatial clusters.

$$\boldsymbol{A} = \frac{(\boldsymbol{a}+\boldsymbol{d})}{(\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c}+\boldsymbol{d})}$$

where $\boldsymbol{A}$ represents the accuracy of a cluster detection method.

The proportion of false positive ***(Fp)*** and false negative ***(Fn)*** spatial clusters were calculated as

$$\boldsymbol{Fp}(\%) = \frac{\boldsymbol{b}}{\boldsymbol{b}+\boldsymbol{d}} \times 100$$

$$\boldsymbol{Fn}(\%) = \frac{\boldsymbol{c}}{\boldsymbol{a}+\boldsymbol{c}} \times 100$$

Results

Spatial clusters identified by $G_i^*(d)$, AMOEBA and the spatial scan statistic are presented in Fig. 2. Hotspots of HCA were present across Georgia. In general, the three statistics showed a pattern of clustering in the east, west and south of Georgia. However, there were noticeable differences between each of the three methods.

The AMOEBA clusters presented a similar spatial pattern to that defined by $G_i^*(d)$. On the other hand, the spatial scan statistic identified a larger portion of the landscape as part of significant clusters located in the eastern, western and southern parts of Georgia. Only the primary and secondary clusters obtained by the spatial scan statistic at the maximum spatial cluster sizes ≤50% of the population at risk are presented since the results were similar with clusters at lower sizes of the population at risk (≤40%). The spatial scan statistic identified the greatest number of cells associated with statistically significant local spatial clusters of HCA incidence (primary + secondary clusters: n = 152), followed by the $G_i^*(d)$ statistic (n = 118 cells) and AMOEBA, which detected the fewest HCA clusters (n = 4 individual cells; Fig. 2). AMOEBA did not identify any clusters with more than one grid cell. The $G_i^*(d)$ statistic revealed that the geographic extent of significant spatial clusters expanded toward the east with the increase of spatial distances.

Table 1
Sensitivity analysis parameters and methods of computation for evaluating spatial clustering techniques to evaluate patterns of human cutaneous anthrax in Georgia.

	Actual clusters				Measures				
		Positive	Negative	Total	Sensitivity	Specificity	Accuracy	False positive	False negative
Predicted clusters	Positive	TP (a)	FP (b)	a + b	a/(a + c)	b/(b + d)	(a + d)/(a + b + c + d)	b/(b + d)	c/(a + c)
	Negative	FN (c)	TN (d)	c + d					
	Total	a + c	b + d	a + b + c + d					

The spatial scan statistic had the highest sensitivity value (S = 58.93%) followed by $G_i^*(d)$ and AMOEBA (Table 2). On the other hand, AMOEBA was more accurate (A = 95.67%) and specific (Sp = 100%) than $G_i^*(d)$ and the spatial scan statistic. Furthermore, AMOEBA did not detect false positive spatial clusters and had the highest percentage of false negative spatial clusters followed by the spatial scan statistic and $G_i^*(d)$. Fig. 3 presents the mean incidence rates of all true spatial clusters for each spatial statistic. Higher sensitivity values were inversely associated with lower mean incidence rates.

Discussion

Anthrax has re-emerged as a public health threat in Georgia. Previous studies have suggested that targeting hotspots of HCA transmission may allow for the more efficacious use of limited resources such as livestock vaccination (Kracalik, Malania, et al., 2014). Properly identifying spatial clusters can aid in targeting the distribution of such resources. In this study, we tested the performance of three different local spatial cluster detection methods to identify spatial clusters of high HCA incidence. Our findings show that while HCA reports were widely distributed across the country, high incidence was clustered on the landscape. In general, HCA was clustered in the southeast and in western Georgia, consistent with previous research linking areas of persistence to environmental and anthropogenic factors (Kracalik et al., 2013). Clustering in the east corresponded to livestock migration routes, while in the west clusters were associated with agricultural croplands.

The clustering of anthrax identified in our study has several explanations. First, human cases are often a direct result of handling sick animals, which have been shown to cluster in relation to environmental factors such as alkaline soils (high pH) across multiple landscapes and ecosystems (Kracalik et al., 2012; Smith et al., 2000). Second, human cases are often related to agricultural activities such as herding or working in abattoirs (Turnbull, Böhm, Hugh-Jones, & Melling, 2008). Interestingly, in Georgia cases were, in general, clustered in two distinct geographic areas: one associated with animal migrations in close proximity to urban areas and one in croplands. These findings suggest there may be two different epidemiologic patterns of transmission related to geography. While there were distinct differences in the results of the three spatial statistics used here, the clustered identified by each support this hypothesis.

A major finding was related to differences in the sensitivity of each method relative to its ability to identifying true spatial clusters. Clusters defined by $G_i^*(d)$ had high sensitivity out to ~40 km, after this distance cell assignment to false clusters increased, suggesting that long distances decrease the likelihood of true cluster membership. This finding is in line with previous research that suggested transmission of HCA occurs across relatively short distances (Chakraborty et al., 2012; Kracalik et al., 2013). In a separate study of anthrax in west Texas, Blackburn, Curtis, Hadfield, and Hugh-Jones (2014) found similar patterns of highly localized clusters (defined with the Getis-Ord statistic and a short d_c) of biting flies that may promote transmission in white-tailed deer, *Odocoileus virginianus,* on that landscape. In this study, AMOEBA had high specificity but only identified clusters with very high incidence, which is in agreement with a recent study of livestock anthrax in Kazakhstan (Kracalik et al., 2012). Aldstadt and Getis (2006) also reported similar findings that AMOEBA was likely to detect clusters of high values based on simulated data. On the other hand, the spatial scan statistic had a high sensitivity when both primary and secondary clusters were considered, but tended to overestimate the cluster limits, as documented elsewhere (Vazquez-Prokopec, Spillmann, Zaidenberg, Gürtler, & Kitron, 2012). Spatial agreement between the three statistics indicated that all AMOEBA clusters were identified by at least one of the other two tests (Fig. 4). In contrast, each $G_i^*(d)$ and SaTScan™ identified cells not identified by either of the other two tests. Clusters identified by each AMOEBA and one other statistic may be considered as areas to prioritize for intervention efforts including targeted livestock vaccination campaigns and educational programs to inform the public about anthrax risk in communities within these clusters.

Differences between tests are likely due to varying assumptions of distribution and model parameters, resulting in the different spatial clustering patterns observed in Fig. 2. For example, the spatial scan statistic and the $G_i^*(d)$ (as implemented here) searches for spatial clusters with circular windows. SaTScan™ varies the circular window size from 0 to a maximum limit set by the user. The $G_i^*(d)$ was limited to fixed search radii from 5 to 100 km. In contrast, the AMOEBA algorithm uses a multidirectional search approach to identify spatial clusters. Another important factor influencing the spatial pattern of the identified clusters is the variation in how thresholds were determined for each test (Openshaw & Taylor, 1979).

Based on our sensitivity analysis, there was evidence to support the suggestion of Aldstadt and Getis (2006) that AMOEBA is more sensitive to grid cells with higher incidence rates compared to the spatial scan statistic, where spatial clusters had overall lower mean incidence rates (Table 2) (Aldstadt & Getis, 2006). Jacquez (2009) compared the statistical power of several cluster detection methods and found that the spatial scan statistic had a higher power when compared to $G_i^*(d)$. We drew similar conclusions, although we used different parameter settings in SaTScan™. The

Table 2
Sensitivity measures for $G_i^*(d)$, AMOEBA, and the spatial scan statistics when examining patterns of human cutaneous anthrax in Georgia.

Tests	Positive		Total	Negative		Total	Measures				
	TP (a)	FP (b)		FN (c)	TN (d)		Sensitivity (%)	Specificity (%)	Accuracy (%)	FP (%)	FN (%)
AMOEBA	4	0	4	52	1144	1196	7.14	100	95.67	0	92.86
$G_i^*(d)$	19	99	118	37	1045	1082	33.93	91.35	88.67	8.65	66.07
SaTScan	33	119	152	23	1025	1048	58.93	89.6	88.17	10.4	41.07

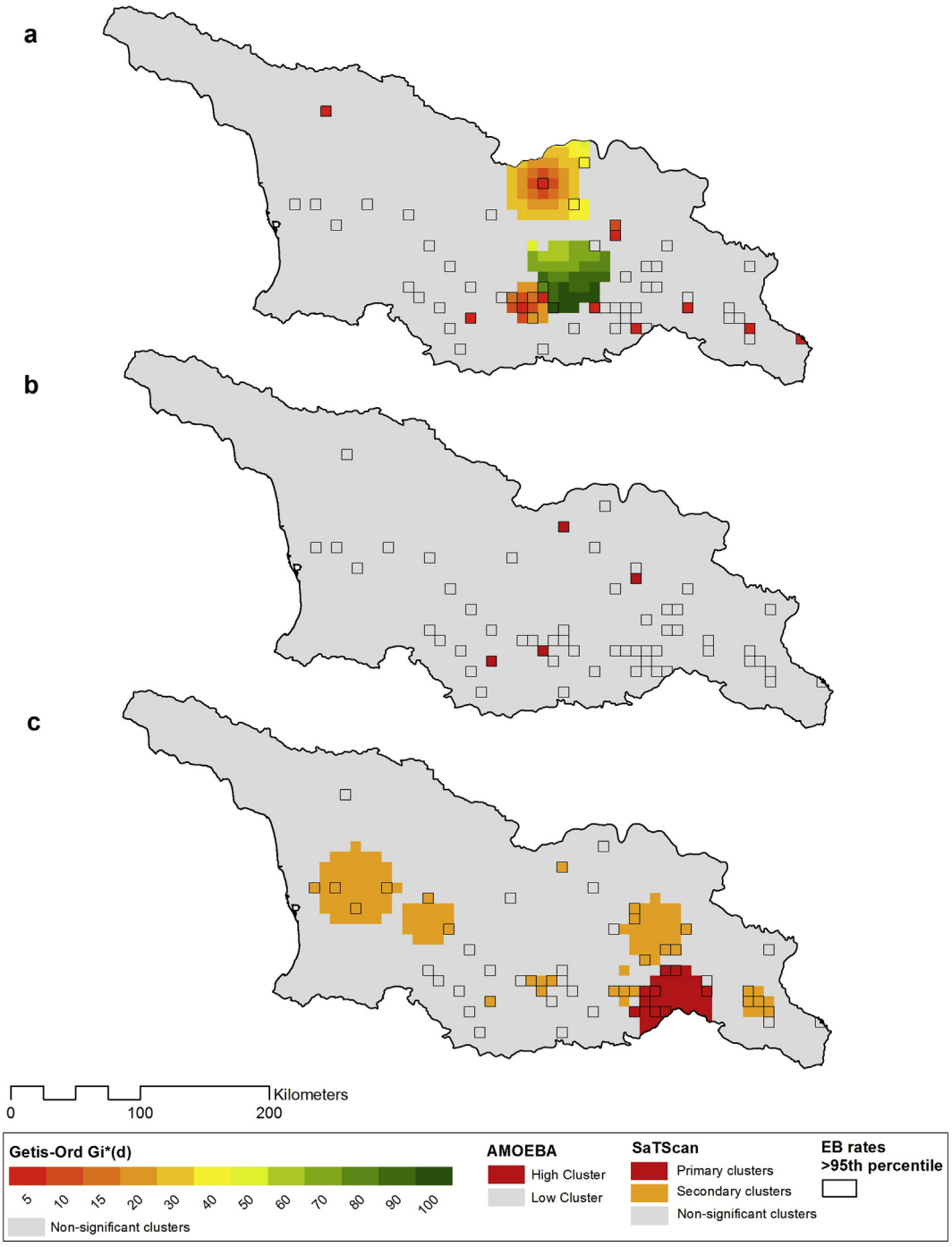

Fig. 2. Spatial clusters of human cutaneous anthrax (HCA) incidence (based on 8 × 8 km cell) in Georgia from 2000 to 2012 using three spatial cluster detection methods: a) $G_i^*(d)$; b) AMOEBA; and c) the spatial scan statistic. Open cells represent *true clusters* based on the definition of the upper 95th percentile of HCA incidence rates per cell.

sensitivity and specificity of a cluster detection method, the percentage of false negative and false positive spatial clusters, can guide the selection of a suitable cluster detection method in a given study. In our analysis, the spatial scan statistic had the highest overall sensitivity; hence, one would preferably suggest this approach for the detection of true spatial clusters, although the mean rates of disease within spatial clusters were lower. Conversely, in spatial cluster studies, there is a tendency to prefer false negative results to avoid false cluster alarms in community residents (Wartenberg, 2001). This preference might make AMOEBA appealing, as it had the highest percentage of false negative and no false positive spatial clusters. Although it has been demonstrated elsewhere, with simulated data points, that the sensitivity to detect a spatial cluster decreased at coarser resolutions (Ozonoff, Jeffery, Manjourides, White, & Pagano, 2007), this study suggests that the sensitivity of a cluster detection method rather depends on the nature of the spatial data to hand. However, sensitivity and specificity will be affected by the definition of true spatial clusters.

Some limitations of this study pertain to the computational time of AMOEBA and the definition of cluster shapes by the spatial scan statistic. AMOEBA has a high computational time, limiting the

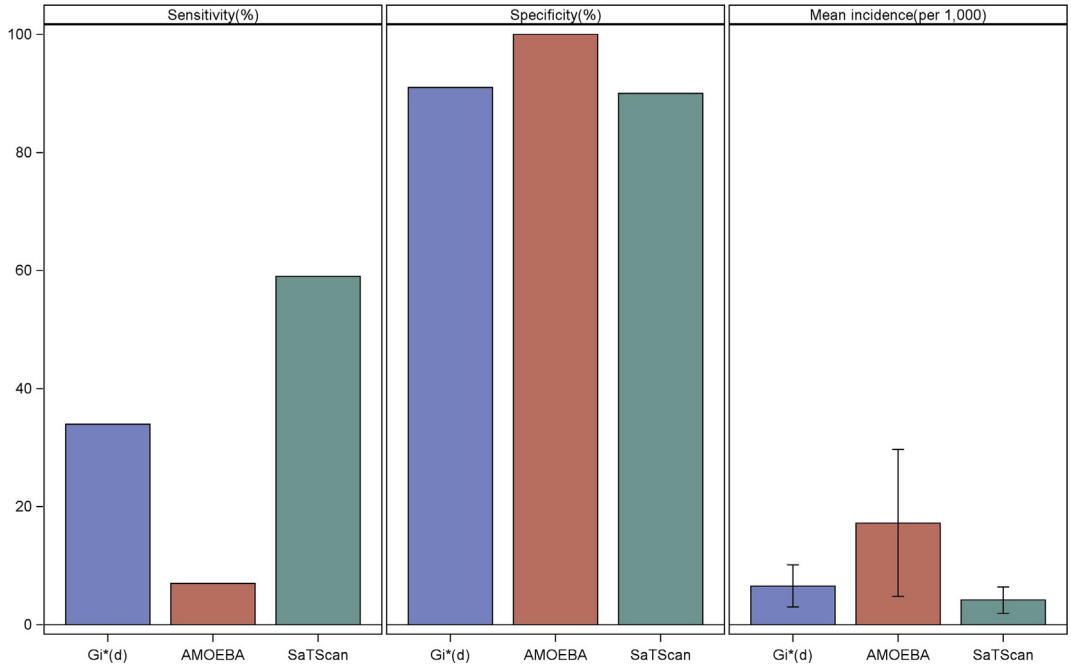

Fig. 3. Sensitivity, specificity and mean incidence rates for clusters of human cutaneous anthrax (HCA) identified by $G_i^*(d)$, AMOEBA, and the spatial scan statistic.

spatial scale of the analyses. Furthermore, we used smoothed incidence rates to detect spatial clusters of human anthrax incidence rates, using $G_i^*(d)$ and AMOEBA. However, the discrete Poisson model in SaTScan™ requires cases and population counts for each location, and does not produce smoothed incidence rates. Therefore, there might be a bias in the comparison of the sensitivity of the three statistics. Despite these limitations, spatial clusters identified in this study can inform targeted implementation of surveillance and control measures for HCA in Georgia. Additionally, these results may guide others in selecting cluster methods for exploratory studies that extend beyond this single disease or epidemiology.

In reality, researchers are often faced with challenging decisions when selecting the appropriate statistical tests for defining these clusters. Inherently, spatial statistics are complicated by issues of aggregation and defining spatial relationships between spatial units (e.g. weights matrices), and setting model parameters. Here we illustrate that the selection of a test could also be informed by *a*

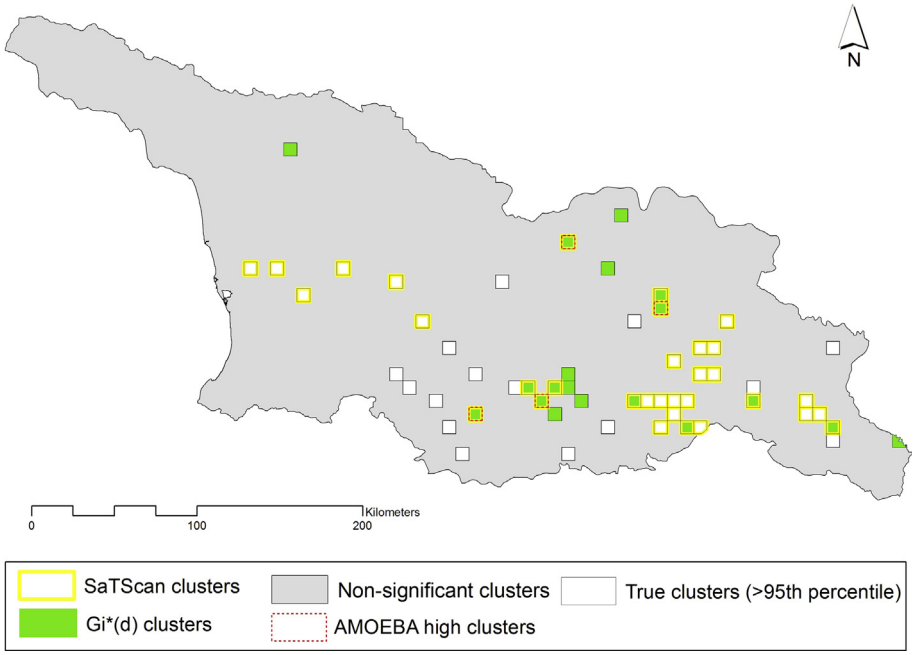

Fig. 4. Spatial agreement between human cutaneous anthrax (HCA) clusters detected by $G_i^*(d)$, AMOEBA, and the spatial scan statistic. Overlapping spatial clusters between the three statistics are presented in green, red and yellow polygons. Open cells with black outlines represent *true clusters* based on the definition of the upper 95th percentile of HCA incidence rates per cell. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

priori knowledge on the desired level of sensitivity or specificity; though we acknowledge that the definition of a true cluster needs careful consideration. In this study, we illustrate how each of three popular spatial statistics identify cluster of disease incidence from real HCA reporting in Georgia. These tests differed in their identification of clusters, with AMOEBA identifying the fewest clusters and only those with high disease incidence. We suggest that multiple tests be employed and that careful consideration be given to whether investigators are interested in identifying regions of disease occurrence or only areas of highest incidence.

Acknowledgments

This work was funded by the U.S. Defense Threat Reduction Agency through the Cooperative Biological Engagement Program in Georgia under the GG-18 Project. Thanks to L.S. Smith, K.H. Bagamian, and N. Royal for assistance in mapping anthrax locations.

References

Adjemian, J., Olivier, K. N., Seitz, A. E., Falkinham, J. O., Holland, S. M., & Prevots, D. R. (2012). Spatial clusters of nontuberculous mycobacterial lung disease in the United States. *American Journal of Respiratory and Critical Care Medicine, 186*(6), 553–558.

Aldstadt, J., & Getis, A. (2006). Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis, 38*(4), 327–343.

Anselin, L., & McCann, M. (2009). OpenGeoDa, open source software for the exploration and visualization of geospatial data. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 550–551). ACM.

Auchincloss, A. H., Gebreab, S. Y., Mair, C., & Diez Roux, A. V. (2012). A review of spatial methods in epidemiology, 2000–2010. *Annual Review of Public Health, 33*, 107–122.

Beyer, H. L. (2012). *Geospatial modelling environment (version 0.6. 0.0)*. Software retrieved, 07–05.

Blackburn, J., Curtis, A. J., Hadfield, T., & Hugh-Jones, M. E. (2014). Spatial and temporal patterns of anthrax in white-tailed deer, *Odocoileus virginianus*, and hematophagous flies in west Texas during the summertime anthrax risk period. *Annals of the Association of American Geographers, 104*(5), 939–958. http://dx.doi.org/10.1080/00045608.2014.914834.

Blackburn, J. K., McNyset, K. M., Curtis, A., & Hugh-Jones, M. E. (2007). Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. *The American Journal of Tropical Medicine and Hygiene, 77*(6), 1103–1110.

Chakraborty, A., Khan, S. U., Hasnat, M. A., Parveen, S., Islam, M. S., Mikolon, A., et al. (2012). Anthrax outbreaks in Bangladesh, 2009–2010. *The American Journal of Tropical Medicine and Hygiene, 86*(4), 703–710.

Clennon, J. A., King, C. H., Muchiri, E. M., Kariuki, H. C., Ouma, J. H., Mungai, P., et al. (2004). Spatial patterns of urinary schistosomiasis infection in a highly endemic area of coastal Kenya. *The American Journal of Tropical Medicine and Hygiene, 70*(4), 443–448.

DeGroote, J., Sugumaran, R., Brend, S., Tucker, B., & Bartholomay, L. (2008). Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA. *International Journal of Health Geographics, 7*(1), 19.

Devine, O. J., Louis, T. A., & Halloran, M. E. (1994). Empirical Bayes methods for stabilizing incidence rates before mapping. *Epidemiology, 5*(6), 622–630.

Duque, J. C., Aldstadt, J., Velasquez, E., Franco, J. L., & Betancourt, A. (2011). A computationally efficient method for delineating irregularly shaped spatial clusters. *Journal of Geographical Systems, 13*(4), 355–372.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation, 24*(1), 38–49.

Getis, A., & Aldstadt, J. (2010). Constructing the spatial weights matrix using a local statistic. In *Perspectives on spatial data analysis* (pp. 147–163). Springer.

Getis, A., & Griffith, D. A. (2002). Comparative spatial filtering in regression analysis. *Geographical Analysis, 34*(2), 130–140.

Getis, A., Morrison, A. C., Gray, K., & Scott, T. W. (2003). Characteristics of the spatial pattern of the dengue vector, Aedes aegypti, in Iquitos, Peru. *The American Journal of Tropical Medicine and Hygiene, 69*(5), 494.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*(3), 189–206.

Griffin, D., Petrosky, T., Morman, S., & Luna, V. (2009). A survey of the occurrence of *Bacillus anthracis* in North American soils over two long-range transects and within post-Katrina New Orleans. *Applied Geochemistry, 24*(8), 1464–1471.

Griffin, D. W., Silvestri, E. E., Bowling, C. Y., Boe, T., Smith, D. B., & Nichols, T. L. (2014). Anthrax and the geochemistry of soils in the contiguous United States. *Geosciences, 4*(3), 114–127.

Hugh-Jones, M., & Blackburn, J. (2009). The ecology of *Bacillus anthracis*. *Molecular Aspects of Medicine, 30*(6), 356–367.

Jacquez, G. M. (2008). Spatial cluster analysis. In *The handbook of geographic information science* (pp. 395–416).

Jacquez, G. M. (2009). Cluster morphology analysis. *Spatial and Spatio-Temporal Epidemiology, 1*(1), 19–29.

Jacquez, G., Waller, L., Grimson, R., & Wartenberg, D. (1996). The analysis of disease clusters, Part I: state of the art. *Infection Control and Hospital Epidemiology*, 319–327.

Kafadar, K. (1996). Smoothing geographical data, particularly rates of disease. *Statistics in Medicine, 15*(23), 2539–2560.

Kao, A. S., Getis, A., Brodine, S., & Burns, J. C. (2008). Spatial and temporal clustering of Kawasaki syndrome cases. *The Pediatric Infectious Disease Journal, 27*(11), 981.

Kracalik, I., Abdullayev, R., Asadov, K., Ismayilova, R., Baghirova, M., Ustun, N., et al. (2014). Changing patterns of human anthrax in Azerbaijan during the post-soviet and preemptive livestock vaccination eras. *PLoS Neglected Tropical Diseases, 8*(7), e2985.

Kracalik, I. T., Blackburn, J. K., Lukhnova, L., Pazilov, Y., Hugh-Jones, M. E., & Aikimbayev, A. (2012). Analysing the spatial patterns of livestock anthrax in Kazakhstan in relation to environmental factors: a comparison of local (Gi*) and morphology cluster statistics. *Geospatial Health, 7*(1), 111–126.

Kracalik, I. T., Malania, L., Tsertsvadze, N., Manvelyan, J., Bakanidze, L., Imnadze, P., et al. (2013). Evidence of local persistence of human anthrax in the country of georgia associated with environmental and anthropogenic factors. *PLOS Neglected Tropical Diseases, 7*(9), e2388.

Kracalik, I., Malania, L., Tsertsvadze, N., Manvelyan, J., Bakanidze, L., Imnadze, P., et al. (2014). Human cutaneous anthrax, georgia 2010–2012. *Emerging Infectious Diseases, 20*(2), 261.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and Methods, 26*(6), 1481–1496.

Kulldorff, M. (2011). *SaTScan user guide for version 9.0*.

Leyland, A. H., & Davies, C. A. (2005). Empirical Bayes methods for disease mapping. *Statistical Methods in Medical Research, 14*(1), 17–34.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1/2), 17–23.

O'Brien, D. J., Kaneene, J. B., Getis, A., Lloyd, J. W., Rip, M. R., & Leader, R. W. (1999). Spatial and temporal distribution of selected canine cancers in Michigan, USA, 1964–1994. *Preventive Veterinary Medicine, 42*(1), 1–15.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Sciences, 21*, 127–144.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis, 27*(4), 286–306.

Ozonoff, A., Jeffery, C., Manjourides, J., White, L., & Pagano, M. (2007). Effect of spatial resolution on cluster detection: a simulation study. *International Journal of Health Geographics, 6*(1), 52.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 172–212.

Smith, K., DeVos, V., Bryden, H., Price, L., Hugh-Jones, M., & Keim, P. (2000). *Bacillus anthracis* diversity in Kruger National Park. *Journal of Clinical Microbiology, 38*(10), 3780.

Tango, T., & Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics, 4*(1), 11.

Turnbull, P., Böhm, R., Hugh-Jones, M. E., & Melling, J. (2008). *Guidelines for the surveillance and control of anthrax in humans and animals* (4th ed.). World Health Organization http://www.who.int/csr/resources/publications/anthrax_webs.pdf.

Vazquez-Prokopec, G. M., Spillmann, C., Zaidenberg, M., Gürtler, R. E., & Kitron, U. (2012). Spatial heterogeneity and risk maps of community infestation by Triatoma infestans in rural Northwestern Argentina. *PLOS Neglected Tropical Diseases, 6*(8), e1788.

Wartenberg, D. (2001). Investigating disease clusters: why, when and how? *Journal of the Royal Statistical Society: Series A (Statistics in Society), 164*(1), 13–22.

Weeks, J. R., Getis, A., Hill, A. G., Agyei-Mensah, S., & Rain, D. (2010). Neighborhoods and fertility in Accra, Ghana: an AMOEBA-based approach. *Annals of the Association of American Geographers, 100*(3), 558–578.

Yao, Z., Tang, J., & Zhan, F. B. (2011). Detection of arbitrarily-shaped clusters using a neighbor-expanding approach: a case study on murine typhus in South Texas. *International Journal of Health Geographics, 10*(1), 23.