



# A twenty-four-year exploratory spatial data analysis of Lyme disease incidence rate in Connecticut, USA

Abolfazl Mollalo,<sup>1</sup> Jason K. Blackburn,<sup>1-3</sup> Lillian R. Morris,<sup>1-4</sup> Gregory E. Glass<sup>1,3</sup>

<sup>1</sup>Department of Geography, University of Florida, Gainesville, FL; <sup>2</sup>Spatial Epidemiology and Ecology Research Laboratory, Department of Geography, University of Florida, Gainesville, FL; <sup>3</sup>Emerging Pathogens Institute, University of Florida, Gainesville, FL; <sup>4</sup>Environmental Epidemiology Section, Office of Environmental Public Health Sciences, Washington State Department of Health, Olympia, WA, USA

## Abstract

Despite efforts to control Lyme disease in Connecticut, USA, it remains endemic in many towns, posing a heavy burden. We examined changes in the spatial distribution of significant spatial clusters of Lyme disease incidence rates at the town level from 1991 to 2014 as an approach for targeted interventions. Lyme disease data were grouped into four discrete time periods and incidence rates were smoothed with Empirical Bayes estimation in GeoDa. Local clustering was measured using a local indicator of spatial autocorrelation (LISA). Elliptic spatial scan statistics (SSS) in different shapes and directions were also performed in SaTScan. The accuracy of these two cluster detection methods was assessed and compared for sensitivity, specificity, and overall accuracy. There was significant clustering during each period and significant clusters persisted predominantly in western and eastern parts of the state. Generally, the SSS method was more sensitive, while LISA was more specific with higher overall accuracy in identifying clusters. Even though the location of clusters changed over time, some towns were persistently (across all four periods) identified as clusters in LISA and their neighbouring towns (three

of four periods) in SSS suggesting these regions should be prioritised for targeted interventions.

## Introduction

Lyme disease (LD), a tick-borne, bacterial, zoonotic infection, remains a serious challenge for public health. The disease is distributed globally, predominantly in temperate portions of the Northern Hemisphere such as Europe, Canada and USA (Oliver *et al.*, 2014). In the United States (US), the geographical distribution of LD is primarily confined to the north-eastern and mid-western areas (Pepin *et al.*, 2012). Past studies have shown that in these areas, LD is caused by *Borrelia burgdorferi sensu stricto*. The pathogen is mainly transmitted to humans during blood meals by the bite of infected blacklegged ticks (*Ixodes scapularis*) with white footed mice (*Peromyscus leucopus*) serving as the primary reservoir for this bacteria (Anderson, 1989). The disease is the most common vector-borne disease in U.S. with an estimated average number of 30,000 new cases every year (Radolf *et al.*, 2012); however, the genuine number is likely much higher. The mean incidence rate of LD in the top 13 U.S. states with the highest incidence rate during 2005-2009 progressively rose from 29.6±10.6 per 100,000 in 2005 to 49.6±15.5 per 100,000 in 2009 (Pepin *et al.*, 2012). At the same time, in 11 states with the lowest incidence rate, the mean incidence developed from 1.3±0.7 to 2.3±1.7 per 100,000 individuals (Pepin *et al.*, 2012). Although this common zoonotic disease rarely leads to death, it can cause severe symptoms related to skin, joints and heart in addition to anxiety and depression if untreated (CDC, 2016). LD can also be a socio-economic burden to society.

In recent years, exploratory spatial data analyses (ESDA) to describe spatial patterns of LD has increased significantly as a strategy to improve our understanding of disease transmission and risk. Several recent studies from different parts of the U.S. have examined the spatial pattern of LD using ESDA. For example, Kugeler and colleagues (2015) applied circular scan statistics to detect high-risk counties of LD in the U.S. from 1993 to 2012. They showed that the number of counties with high incidence of LD successively increased from 69 (1993-1997) to 130 (1998-2002) and further to 197 (2003-2007) and 260 (2008-2012) counties, respectively (Kugeler *et al.*, 2015). In Texas, which is a non-endemic LD area, Szonya *et al.* (2015) applied global and local Moran's *I*-tests to determine the distribution and location of possible clusters, respectively, with respect to the spatial distribution of LD at the county level (2000-2011). They observed a clustered distribution with a high incidence cluster in central parts of the

Correspondence: Abolfazl Mollalo, Department of Geography, University of Florida, 3141 Turlington Hall, P.O. Box 117315, Gainesville, FL 32611-7315, USA.  
Tel: +1-352-226-3768.  
E-mail: abolfazl@ufl.edu

Key words: Accuracy assessment; Cluster detection; Exploratory spatial data analysis; GIS; Lyme disease.

Acknowledgements: the authors are grateful to the anonymous reviewers and Prof. Robert Bergquist for their time and effort spent in reviewing the paper and for their professional comments and suggestions.

Funding: GG was supported by the University of Florida Preeminence Initiative.

©Copyright A. Mollalo *et al.*, 2017  
Licensee PAGEPress, Italy  
Geospatial Health 2017; 12:588  
doi:10.4081/gh.2017.588

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

state, mainly in a cross-timbers eco-region (Szonyi *et al.*, 2015). In Virginia, Li *et al.* (2014) utilized the Empirical Bayes smoothing (EBS) method on census tract LD cases with the aim of lessening random variations, especially in censuses with small populations. Then, they applied space-time scan statistic and found a primary cluster in northern Virginia which had experienced population growth and urban-sub-urban improvements between 2008 and 2011 (Li *et al.*, 2014).

The town of Lyme in Connecticut was the first spot that LD was recognized in the U.S. The initial cluster in 1976 was observed in children (Steere *et al.*, 1977). Since then, in spite of all endeavours conducted by the Connecticut Department of Public Health (CTDPH) to control the disease, it remains endemic with substantial morbidity rates. Although LD is a well-investigated epidemiological subject in Connecticut, historical changes in patterns of disease have been minimally studied (Cromley *et al.*, 1993). Additionally, even though geographical information systems (GIS) is a useful tool to study infectious diseases (Moore and Carpenter, 1999), powerful GIS-based studies of LD from this region are insufficient for prioritizing counties for intervention. Thus, the main objective of this study is to use the combination of GIS and ESDA to better describe changes in the spatial pattern of LD, supposing that the reported passive cases of LD represent a spatially random subsample of the disease in the state. Our specific research questions included: 1) what is the spatial distribution (random/ dispersed/ clustered) of LD incidence over the past 24 years?; 2) If clustered, where have the clusters/hotspots occurred? 3) Can we avoid or minimize the effects of edges in the spatial scan statistic technique?; and 4) Which cluster detection technique (LISA/spatial scan statistic) is more sensitive, specific and accurate?

## Materials and Methods

### Data collection and preparation

We used passively reported indigenous LD cases over a period of 24 years from 1991 to 2014 throughout the state of Connecticut. We retrieved data from the CTDPH containing yearly counts and rates of LD at the town level. The CTDPH has a well-established LD surveillance system operating since 1987. Reports of cases were based on the National Surveillance Case Definition from the Centers for Disease Control and Prevention (CDC) for LD (CDC, 1996, 2008, 2011). Data were geocoded and grouped into four equal intervals (each period included six years: 1991–1996, 1997–2002, 2003–2008 and 2009–2014) to further explore clustering, possible clusters and how hotspots had changed.

Administrative boundaries of towns were obtained from the Map and Geographic Information Center (MAGIC) of Connecticut GIS data using the shapefile format (<http://magic.lib.uconn.edu/>). Similarly, annual population statistics were downloaded from CTDPH (<http://www.ct.gov/dph/site/default.asp>). The study area and the names of the towns mentioned in this paper are shown in Figure 1.

### Global clustering

We applied global clustering techniques to statistically evaluate whether the existing pattern of LD incidence was random, clustered, or dispersed. We used the global Moran's  $I$  statistic (Moran, 1950) to measure spatial autocorrelation using GeoDa software

version 1.6.7 (Anselin, 2004). The null hypothesis assumes that there is no spatial pattern among the incidence of LD in different towns (*i.e.* complete spatial randomness) (O'Sullivan and Unwin, 2014). This statistic employs a covariance term between each town and its neighbours as follows (Mitchell, 1999):

$$I = \frac{N \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^N (x_i - \bar{x})^2} \quad \text{Eq. 1}$$

$$S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij} \quad \text{Eq. 2}$$

where  $x_i$  and  $x_j$  are incidences of LD in the  $i^{\text{th}}$  and  $j^{\text{th}}$  towns, respectively;  $N$  the aggregate number of towns; and  $w_{ij}$  the spatial neighbourhood weight for towns  $i$  and  $j$  generated based on the first-order Queen's contiguity which shares all common points including boundaries and vertices. The generated spatial weight is used as a criterion for recognising neighbours of each town. The weight is defined taking into account adjacent neighbours and written as:

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are adjacent neighbours} \\ 0 & \text{otherwise} \end{cases} \quad \text{Eq. 3}$$

The Moran's  $I$  index varies between -1 and +1, with 0 showing spatially random distribution, while negative values indicate dispersed distributions and positive values for clustered distributions. We assessed significance of the index using both the Z-score and P value.

### Local clustering

#### Spatial smoothing

The global clustering techniques provide information about the overall distribution of LD (random, clustered or dispersed), but we were also interested in identifying local clusters. First, we applied the EBS routine to account for variation in town sizes and populations. Contrasts in population size among the spatial areal units (*i.e.* towns of Connecticut) may lead to variance instability and spurious outliers (Anselin, 2004). This is due to the observed raw rate in spatial areal units with small population being profoundly affected by small changes of adding or removing few cases. Thus, crude rates might not reflect underlying risk compared with other areal units with large populations. EBS provides a solution to avoid this type of possible bias as it adjusts the estimated risk toward the global mean to reduce variance instability (Clayton and Kaldor, 1987); areas with low population are adjusted more than areas with larger populations. Since there was a considerable differences in areas of some towns (*e.g.*, Derby and New London are approximately 5 mi<sup>2</sup> while Woodstock and New Milford cover more than 60 mi<sup>2</sup>) and also population size of towns (*e.g.*, Union and Canaan have about 1,000 people, whereas New Haven and Bridgeport have more than 120,000 individuals) applying the EBS is justifiable. We calculated spatial weights for each time interval using the first-order Queen's contiguity. EBS smoothed rates were employed in local cluster detection analyses.

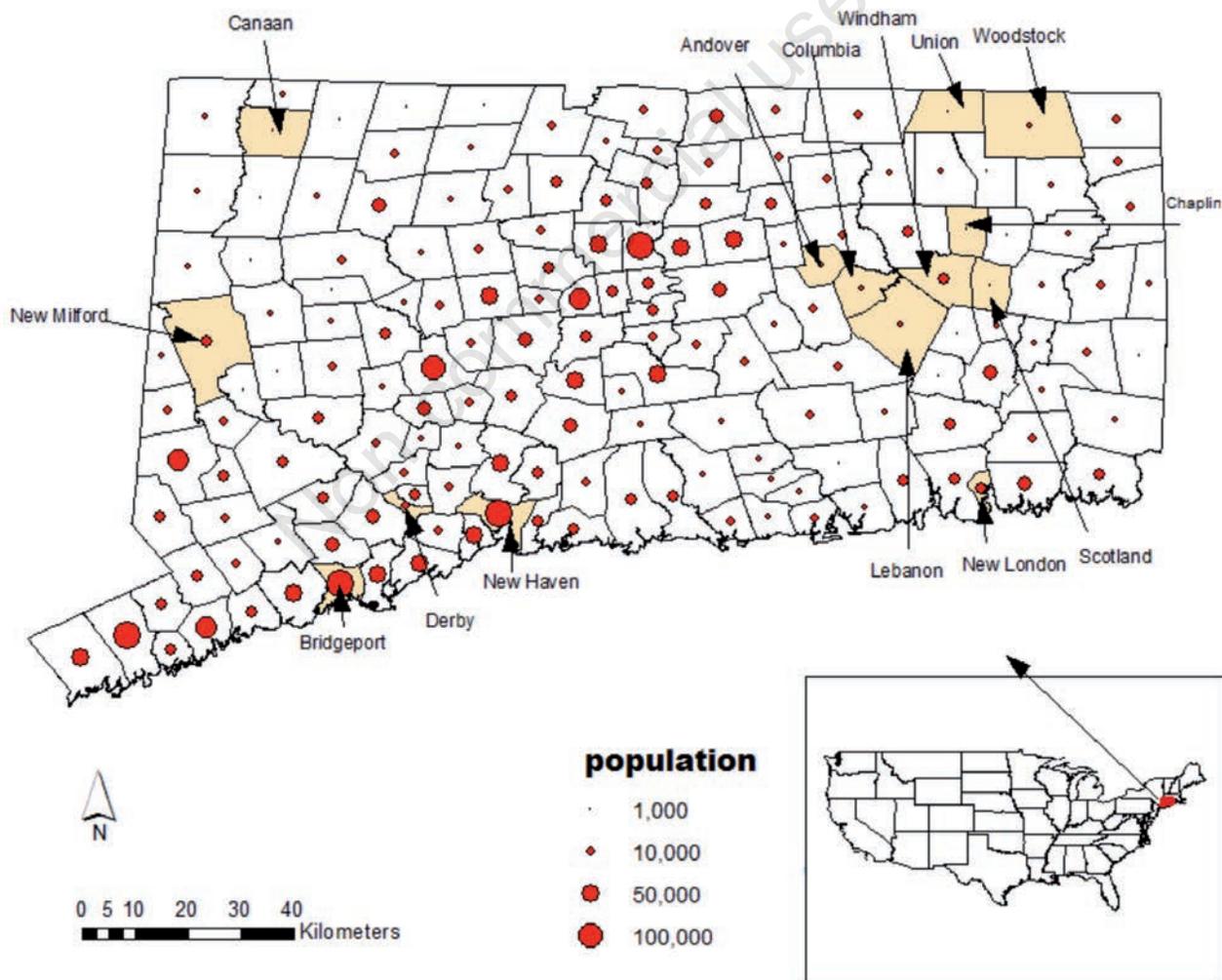
### Local Moran's $I$

To detect local clusters of the LD rates smoothed by EBS, we applied Anselin's local indicator of spatial autocorrelation (LISA) statistics (Kulldorff, 2010). LISA identifies hotspots (towns with a high incidence surrounded by high incidences); coldspots (towns with a low incidence surrounded by low incidences) and outliers (towns with a high incidence surrounded by low incidences, or towns with a low incidence surrounding by high incidences). We used GeoDa (<https://geodacenter.github.io/>) for LISA analyses. We set the number of permutation tests to 999 and 95% significance level ( $P < 0.05$ ). We mapped significant clusters using ArcGIS 10.2 (ESRI, Redlands, CA, USA).

### Spatial scan statistics

We were interested in comparing LD clusters identified by LISA and the Spatial Scan Statistic (SSS), for which we used an ellipsoidal, moving window situated on the centroid of each town so that at any point the window incorporated different sets of neighbours. At each position, the radii of ellipse was set to vary

continuously from 0 to a maximum that never included more than half of the total population at risk. If the window contained the centroid of the neighbouring towns, then that whole town was included. This procedure produces a very large number of ellipsoidal windows and each one can be a possible cluster of LD with different set of neighbours. The ratio of the length of longest to the shortest axis of the ellipse was 1.5, 2, 3, 4 or 5. For each shape, a different number of angles (*i.e.* the angle between the horizontal east-west line and the longest axis of the ellipse) of the ellipse were also tested. For each ellipse a likelihood ratio statistic was computed based on the number of observed and expected cases within and outside the ellipse. The null hypothesis (*i.e.* LD incidence is equal inside and outside of the window) was tested against the alternative hypothesis that the risk was elevated within the ellipse. The likelihood ratio is reported by P values calculated based on the Monte Carlo simulation approach which finds the maximum likelihood ratio over the entire study region. The ellipse with the maximum likelihood was signified the most likely (primary) cluster. This approach also detected secondary clusters which were addi-



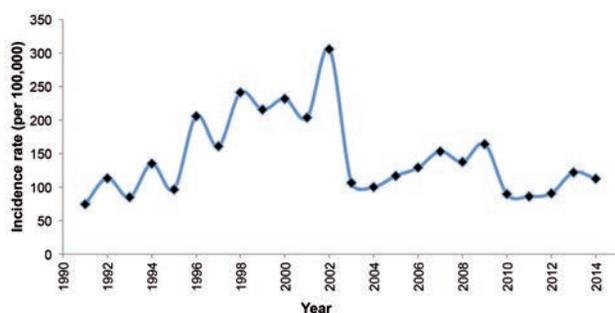
**Figure 1.** Geographic location of Connecticut, USA, its towns and approximate populations. The names on the map show the location of the towns that are mentioned in this paper.

tional ranked clusters that had high likelihood ratios but did not overlap the primary cluster (Kulldorff, 2010).

Three sets of data were built for the analysis based on discrete Poisson probability model in SaTScan software version 9.4.2 (Kulldorff, 2010). The datasets were: *Case file* representing the annual number of cases of LD for each town ( $n=169$ ) from 1991 to 2014; *Coordinate file* was the two-dimensional Cartesian coordinates of centroid of each town; and *Population file* was the population size of each town. We used the mean population and mean number of cases per time interval. To scan the study area we applied two criteria: *no geographic overlap between clusters* and 5% of the population as the maximum to search for hotspots with the aim of comparing the results with LISA's high-high clusters as the LISA analysis only considered the neighboring towns. Also in another run, we used *no geographic overlap between clusters* and 50% population at risk to investigate whether the results depended on the primary settings. To ensure statistical power, the number of Monte Carlo replications was set to 999 and only clusters at 99% confidence interval were considered ( $P<0.01$ ).

#### Accuracy assessments

Table 1 shows a 2x2 confusion matrix used to compare the performance of LISA and SSS (for 5% and 50% of populations at risk) to detect hotspots, sensitivity, specificity and overall accuracy (Fielding and Bell, 1997). In this case, sensitivity measures how well the cluster detection techniques correctly detected the presence of LD hotspots, whereas specificity provides a measure of how well the techniques correctly identified the absence of LD hotspots. Overall accuracy shows the ability of cluster detection techniques to identify true positive and true negative LD hotspots. Locations of the detected clusters with both LISA and SSS techniques were compared with the location of true clusters as defined by Birnbaum *et al.* (1996) that *true clusters explain fewer than 5% of all reported clusters*. Therefore the top 5% towns with regard to high LD incidence were considered as *True Hotspots*. We calculated these statistics for each technique in each period, separately.



**Figure 2.** Temporal trend of Lyme disease incidence throughout Connecticut, USA, from 1991 to 2014. Black dots show the incidence rate for each year and the blue line represents a scatter with smooth lines.

## Results

There were 54,478 reported human LD cases from 1991 to 2014, out of which 10,328 cases (19.0%) occurred in first period (1991-1996), 20,234 cases (37.1%) in second period (1997-2002), 11,210 cases (20.6%) in third period (2003-2008) and 12,706 cases (23.3%) in the last period (2009-2014). The annual incidence of LD ranged between 84.7 (in 1993) and 305.6 (in 2002) cases per 100,000 individuals with a mean of 144.8 cases per 100,000 people (Figure 2). The P values for the global Moran's *I* statistic were close to zero which reject the null hypothesis of complete spatial randomness (CSR) for all time periods (Table 2). The index values ranged from 0.55 to 0.71, which indicates significant clustering. Although, we detected clustering with both raw and smoothed methods, and in all four periods, the variation in rates required EBS before running LISA. Based on the results of EB smoothing technique, LISA found High-High clusters (hotspots), which varied for each study period. Results of LISA showed that in the first and last period, the hotspots were completely restricted to the towns in the East. In other words, for the first period, 24 towns and for the last period 30 towns were identified as the hotspot in eastern Connecticut. The towns in the West were more influenced in the second and third periods. The number of towns in the West identified as hotspots in the second period was 8, while 7 towns were hotspot in the East of Connecticut. In addition, in the third period, 10 towns in the West and 6 towns in the East were identified as hotspots. It should be noted that applying EBS before running LISA, reduces the likelihood of detecting a false cluster in low-population areas where the cases were detected. But comparison of the detected clusters by raw and EBS methods in LISA showed small differences with regard to the location of the detected clusters.

Spatial scan statistics with 5% and 50% of the populations at risk with no geographic cluster overlaps identified clusters with high incidence rates for each period. Except for the second period, The SSS results revealed that the most likely cluster predominant-

**Table 1.** Comparison of local indicator of spatial autocorrelation and SaTScan clusters by confusion matrix and its derivatives statistics.

Actual hotspots (top 5% incidence rates)	Identified hotspots (LISA or SaTScan)	
	Presence	Absence
Presence	TP	FN
Absence	FP	TN

TP, true positive; FN, false negative; FP, false positive; TN, true negative. Sensitivity= $(TP)/(TP+FN)$ ; Specificity= $(TN)/(TN+FP)$ ; Overall accuracy= $(TP+TN)/(TP+FN+FP+TN)$ .

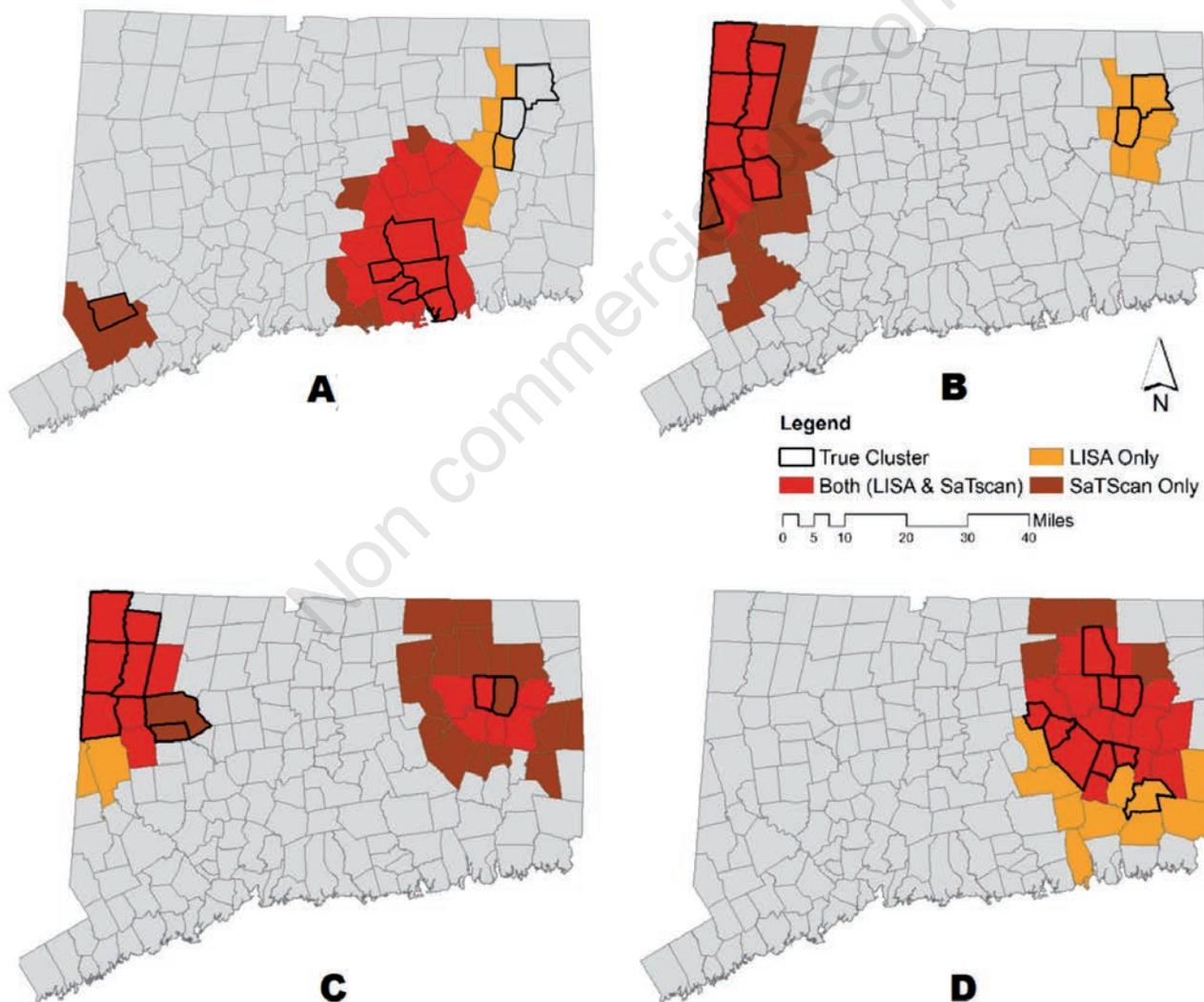
**Table 2.** Results of the global Moran statistics of Lyme disease incidence rate for Connecticut, USA, 1991-2014.

Years	Index	Z-score	P value	Type of distribution
1991-1996	0.71	16.05	$\approx 0$	Clustered
1997-2002	0.59	13.54	$\approx 0$	Clustered
2003-2008	0.55	12.73	$\approx 0$	Clustered
2009-2014	0.59	13.28	$\approx 0$	Clustered

ly occurred in the eastern region of the state, while the secondary cluster occurred in towns in the West (Figure 3). Primary and secondary hotspots were observed in different locations when 5% of the population at risk was investigated using SSS for comparison with LISA. For the first period, the primary cluster occurred in the eastern parts of the state and included 22 towns and 792 cases. The risk of LD incidence within the primary cluster was 7.67 times greater than outside the cluster. The secondary cluster occurred in the western region and contained 5 towns and 152 cases. During the second period, the primary cluster occurred only in the eastern parts of the state and included 23 towns and 1,208 cases. The risk of LD incidence within this cluster was 3.40 times higher than outside. During the 2003-2008 period, the disease largely affected the eastern region with 22 towns and 738 cases as

compared to the western parts with 10 towns and 145 cases. The relative risks of primary and secondary clusters were 3.33 and 9.18, respectively. In the last period, only eastern parts of the state showed statically significant clustering (no secondary cluster). The primary cluster contained 24 towns and 942 cases. The risk of LD incidence in this cluster was 3.69 times more than in other areas (Table 3).

Comparison of the results of accuracy assessments of the spatial scan statistics at 5% and 50% of the populations at risk, shows that sensitivity of this method increases with an increment of the population defined to be at risk. However, increasing the population at risk also leads to a decrease in the specificity of the results. In addition, LISA tended to have higher specificity and had the highest overall accuracy (Figure 4).



**Figure 3.** Locations of spatial clusters of Lyme disease incidence in Connecticut, USA, based on the true-cluster definition with local indicator of spatial autocorrelation and spatial scan statistics methods targeting 5% of the population at risk, for the period 1991-1996 (A), 1997-2002 (B), 2003-2008 (C), and 2009-2014 (D).

## Discussion

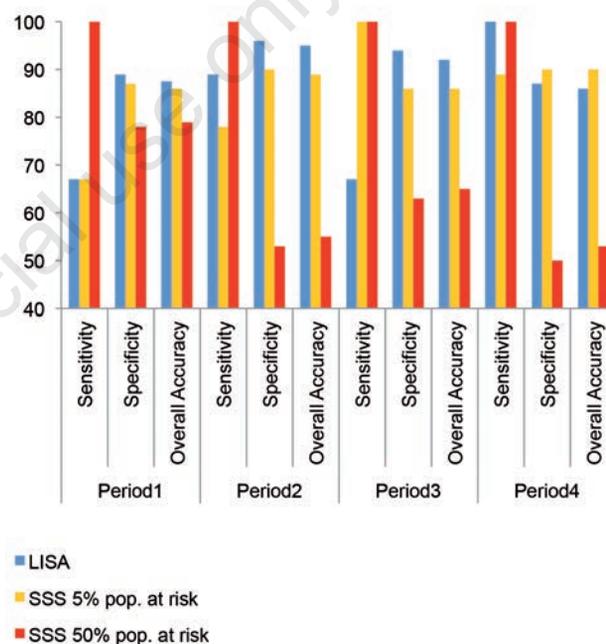
This retrospective study examined the spatial structure of LD incidence distribution in Connecticut based on 24 years of reported data with the aim of describing the spatial distribution and the changes that have occurred with regard to the disease. It differs from most other studies in the region by not focusing on local studies of the pathogen, reservoir or vector and their associations with the environment. Instead, it focuses on the changing patterns of documented human disease occurrences. We assumed that the number of affected human cases corresponded to the density of *Ixodes scapularis* in Connecticut (Connally *et al.*, 2006; Mather *et al.*, 1996; Stafford *et al.*, 1998). In addition, it should be noted that ticks have limited capabilities to move to new areas because of their small size (Ogden *et al.*, 2008). Therefore, one of the means to fight against the risk of the disease in the study area would be targeted intervention in the areas (towns in this case) that were constantly affected with a very high morbidity rate. Targeted control, planning and management of the disease can assist with resource allocation to the towns with persistent high incidence rates resulting in time and costs savings.

Results of this study confirm and extend the findings of Xue *et al.* (2015) in Connecticut. They analysed yearly clusters of LD and showed that the distribution was clustered and that this clustering occurred in western and eastern Connecticut with few cases in the central region supposing the yearly incidence weighted-geographic mean analysis represents the clusters of the case distributions. According to the paper, the epidemic of LD reached equilibrium in 2007 in the western parts of Connecticut, while this happened in 2009 in the East. Comparison of the results shows that periods 1-3 included epidemic conditions while period 4 contained equilibrium condition. Therefore, actual/sudden shifts in the locations of clusters occurred roughly before equilibrium (period 1-3); and it may be that epidemic conditions affected both western and eastern parts of the state; while equilibrium conditions only occurred in the eastern regions.

This study identified towns with high LD incidence rates using the LISA test and spatial scan statistics approaches. By intersecting the locations of the clusters identified by LISA some towns, including Chaplin, Windham and Scotland, were persistently detected to have high-rate clusters. Likewise, neighbours of these towns including Andover, Columbia and Lebanon, were recognized as having high-rate clusters in 3 out of the 4 periods of study (Figure 1). This indicates that these areas were almost persistently affected by a high incidence of LD during the 24 years of study and therefore they deserve closer consideration.

In this study, we used an elliptical window in the scan rather

than highly applied circular shape used in other studies for several reasons. First, according to previously published papers, the elliptic shape had better power and precision compared with circular one and also follows more accurately certain geographic features with varying shapes and directions (Root *et al.*, 2009). Another reason was to avoid edge effects at the borders of the study area. When circles are used for scanning the study area, particularly when the circle should centre on the centroid of border towns, considerable parts of it inevitably covers the neighbouring state (*i.e.* Rhode Island in the East, Massachusetts in the North, Long Island Sound to the South and New York in West, where states are also LD-endemic areas where data were not accessible) and the true number of cases would be underestimated. The use of ellipses with different shapes and directions helps to reduce this problem to some degree.



**Figure 4.** Comparison of local indicator of spatial autocorrelation and spatial scan statistics with 5% and 50% of the population at risk of Lyme disease in Connecticut, USA, for each period from 1991 to 2014. Sensitivity, specificity and overall accuracy expressed as %.

**Table 3.** Characteristics of Lyme disease spatial clusters detected by spatial scan statistics with 5% of the population at risk throughout Connecticut, USA, for the period 1991-2014.

Cluster	Year	Observed cases (n)	Expected cases (n)	Ratio (Obs/Exp)	P value	RR	Log likelihood ratio
Most likely (East)	1991-1996	792	171.26	4.62	<0.0001	7.67	736.02
Secondary (West)	1991-1996	152	32.26	4.71	<0.0001	5.07	120.17
Most likely (East)	1997-2002	1,208	474.52	2.55	<0.0001	3.40	496.50
Most likely (East)	2003-2008	738	305.46	2.42	<0.0001	3.33	284.31
Secondary (West)	2003-2008	145	16.96	8.55	<0.0001	9.18	187.61
Most likely (East)	2009-2014	942	376.77	2.50	<0.0001	3.69	400.80

RR, relative risk. P values were obtained from the Monte Carlo hypothesis test.

There were similarities in the locations of the spatial hotspots identified by SaTScan and LISA even though these methods apply different methodologies to detect hotspots. It should be noted that we smoothed the incidence rate for the LISA cluster detection method, while this is not available for spatial scan statistics. The results were also in agreement with the findings of Barro *et al.* (2015), who compared different cluster detection techniques including Getis-Ord  $G_i^*$  statistic, a multidirectional optimal ecotop-based algorithm (AMOEBa) and the spatial scan statistic for identifying hotspots of human cutaneous anthrax in Georgia for point data. They also found that SSS was more sensitive (due to augmentation in the quantity of true positives) but less specific (by increment of the population at risk because of a declining number of true negative clusters). Here, the results were highly dependent on defining the weight matrix in LISA or the percentage of the population at risk in spatial scan statistics.

The most important limitation of this study is attributed to the data reported that were used in this study. As indicated by the CDC, surveillance data are subject to under-reporting and misclassification in highly endemic areas such as Connecticut; however, this problem would not be very severe in light of well-designed LD surveillance system in this state. Additionally, as long as there is simply a spatially random thinning of reported cases, the overall spatial analysis should not be affected. However, if there is persistent over- or under-reporting in selected regions, it may be difficult to recognize the impact of such biases. Another factor influencing the results is that the definition of LD has changed several times: from using a two-test approach for laboratory affirmation (CDC, 1996) to the addition of *probable* and *suspect* categories with less strict criteria (CDC, 2007) and subtle changes in the specification of *confirmed* cases (CDC, 2011). However, empirically (Figure 3), these changes do not seem to substantially have altered the areas of high-risk clusters. Therefore, although the analyses conducted were based on data reported by CTDPH, the findings of this study should be interpreted with some caution.

## Conclusions

The findings of this study can be regarded as a basis for generating hypotheses about underlying risk components. For instance, visual comparison of the locations of towns that were never identified as cluster areas (Figure 3) and the digital elevation model (DEM) of the study area suggest that low-risk towns are located in low-lying areas; thus it seems that people who live at moderate or higher altitudes in Connecticut are at a higher risk. This would be consistent with the earliest epidemiologic investigations (Steere *et al.*, 1977) that identified increased risk in areas away from the seashore. Thus, altitude as a proxy of other environmental factors can uncover the reasons of the spatial distribution of the disease in the study area. Moreover, except for Windham, all of the other towns identified as persistent clusters had populations lower than 8,000 people showing that persistent towns as clusters occurred in less populated areas. Therefore, further studies should incorporate other environmental factors that have significant influence on the spatial and temporal patterns of the disease.

## References

- Anderson J, 1989. Ecology of Lyme disease. *Conn Med* 53:343-6.
- Anselin L, 1995. Local indicators of spatial association – LISA. *Geogr Anal* 27:93-115.
- Anselin L, 2004. Exploring spatial data with GeoDa™: a workbook. University of Illinois, Urbana-Champaign, IL, USA.
- Barro AS, Kracalik IT, Malania L, Tsertsvadze N, Manvelyan J, Imnadze P, Blackburn JK, 2015. Identifying hotspots of human anthrax transmission using three local clustering techniques. *Appl Geogr* 60:29-36.
- Birnbaum D, Jacquez G, Waller L, Grimson R, Wartenberg D, 1996. The analysis of disease clusters, part I: state of the art. *Infect Control Hosp Epidemiol* 17:319-27.
- CDC, 1996. Lyme disease (*Borrelia burgdorferi*): 1996 case definition. Available from: <https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/1996/>
- CDC, 2008. Lyme disease (*Borrelia burgdorferi*): 2008 case definition. Available from: <https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/2008/>
- CDC, 2011. Lyme disease (*Borrelia burgdorferi*): 2011 case definition. Available from: <https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/2011/>
- CDC, 2016. Lyme disease. Available from: [http://www.cdc.gov/lyme/signs\\_symptoms](http://www.cdc.gov/lyme/signs_symptoms)
- Clayton D, Kaldor J, 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671-81.
- Connally NP, Ginsberg HS, Mather TN, 2006. Assessing peridomestic entomological factors as predictors for Lyme disease. *J Vector Ecol* 31:364-70.
- Cromley EK, Cartter ML, Mrozinski RD, Ertel S-H, 1998. Residential setting as a risk factor for Lyme disease in a hyperendemic region. *Am J Epidemiol* 147:472-7.
- Fielding AH, Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38-49.
- Kugeler KJ, Farley GM, Forrester JD, Mead PS, 2015. Geographic distribution and expansion of human Lyme disease, United States. *Emerg Infect Dis* 21:1455-7.
- Kulldorff M, 2015. SaTScan user guide for version 9.0. Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, MA, USA.
- Li J, Kolivras KN, Hong Y, Duan Y, Seukep SE, Prisley SP, Campbell JB, Gaines DN, 2014. Spatial and temporal emergence pattern of Lyme disease in Virginia. *Am J Trop Med Hyg* 91:1166-72.
- Mather TN, Nicholson MC, Donnelly EF, Matyas BT, 1996. Entomologic index for human risk of Lyme disease. *Am J Epidemiol* 144:1066-9.
- Mitchell A, 1999. The ESRI guide to GIS analysis: geographic patterns and relationships: ESRI, Inc., Redlands, CA, USA.
- Moore DA, Carpenter TE, 1999. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev* 21:143-61.
- Moran PA, 1950. Notes on continuous stochastic phenomena. *Biometrika* 37:17-23.
- Ogden NH, St-Onge L, Barker IK, Brazeau S, Bigras-Poulin M, Charron DF, Francis CM, Heagy A, Lindsay L, Maarouf A, Michel P, 2008. Risk maps for range expansion of the Lyme disease vector, *Ixodes scapularis*, in Canada now and with climate change. *Int J Health Geogr* 7:24-39.

- Oliver J, Lin T, Gao L, Clark K, Banks C, Durden L, James AM, Chandler FW Jr, 2003. An enzootic transmission cycle of Lyme borreliosis spirochetes in the southeastern United States. *P Natl Acad Sci* 100:11642-5.
- O'Sullivan D, Unwin D, 2014. *Geographic information analysis*. John Wiley and Sons, New York, NY, USA.
- Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, Barbour, AG, Hamer S, Diuk-Wasser MA, 2012. Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking *Ixodes scapularis* nymphs in the eastern United States. *Am J Trop Med Hyg* 86:1062-71.
- Radolf JD, Caimano MJ, Stevenson B, Hu LT, 2012. Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nat Rev Microbiol* 10:87-99.
- Root ED, Meyer RE, Emch ME, 2009. Evidence of localized clustering of gastroschisis births in North Carolina, 1999-2004. *Soc Sci Med* 68:1361-7.
- Stafford KC, Cartter ML, Magnarelli LA, Ertel S-H, Mshar PA, 1998. Temporal correlations between tick abundance and prevalence of ticks infected with *Borrelia burgdorferi* and increasing incidence of Lyme disease. *J Clin Microb* 36:1240-4.
- Stafford KC, Magnarelli LA, 1993. Spatial and temporal patterns of *Ixodes scapularis* (Acari: Ixodidae) in southeastern Connecticut. *J Med Entomol* 30:762-71.
- Steere AC, Hardin JA, Malawista SE, 1977. *Erythema chronicum migrans* and Lyme arthritis: cryoimmunoglobulins and clinical activity of skin and joints. *Science* 196:1121-2.
- Steere AC, Malawista SE, Snyderman DR, Shope RE, Andiman WA, Ross MR, Steele FM, 1977. An epidemic of oligoarticular arthritis in children and adults in three Connecticut communities. *Arthritis Rheum* 20:7-17.
- Szonyi B, Srinath I, Esteve-Gassent M, Lupiani B, Ivanek R, 2015. Exploratory spatial analysis of Lyme disease in Texas—what can we learn from the reported cases? *BMC public health* 15:924-32.
- Xue L, Scoglio C, McVey DS, Boone R, Cohnstaedt LW, 2015. Two introductions of Lyme disease into Connecticut: a geospatial analysis of human cases from 1984 to 2012. *Vector Borne Zoonotic Dis* 15:523-8.

Non commercial use only