2010

The Incidence of Lyme Disease in Southeastern Pennsylvania

A Geospatial Analysis

The spread of Lyme disease is a spatial problem, concerned with the geographic distribution of deer, the I. scapularis (the black-legged or deer tick) and human population. This paper examines the incidence of Lyme disease and its putative relationship with human population density, deer population, age of housing, and human development into previously vegetated areas– in this case the variable change in impervious cover. This analysis focuses on Lyme disease in southeastern Pennsylvania. Statistical analysis of spatial patterns using a measure of spatial autocorrelation indicated that counties with most human Lyme cases and deer population were clustered in parts of southeastern Pennsylvania.

Maggie Cawley
Johns Hopkins University
12/10/2010
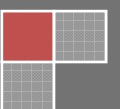
I.   Introduction

Lyme disease is a tick-borne disease shared between animals and people (a zoonosis) caused by infection with the spirochete, Borrelia burgdorferi. I. scapularis, the black-legged or deer tick, is the vector in the eastern United States. Typical habitats of this tick species are wooded, brushy, or overgrown grassy areas that are favorable for deer and the tick's rodent hosts. The number of annually reported cases of Lyme disease in the United States has increased about 25-fold since reporting began in 1982. In the United States, the disease is mostly localized to the northeastern, mid-Atlantic, and upper Midwest regions, and to several counties in northwestern California. Cases are scattered throughout Pennsylvania with the highest incidence of disease being located in the south eastern parts of the state. From 2003-2007, an average of 3,700 Lyme disease cases were reported annually in Pennsylvania.[1]

Deer are the principal maintenance hosts for adult deer ticks, and the presence of deer appears to be a prerequisite for the establishment of the deer tick in any area. There has been an explosive repopulation in the eastern United States by white-tailed deer during recent decades that has been linked to the spread of I. scapularis ticks and of Lyme disease in this region.[2]

The objective of this paper is to explore the applicability of GIS-based spatial analytical procedures in determining a correlation between the instance of Lyme disease, deer population and human development. I hypothesize that that the instance of Lyme disease increases with an increase in deer harvest and impervious cover into previously vegetated areas.

This paper is organized into six sections; a brief introduction, a review of selected literature on the subject, a description of the study area and research background, the analysis methodology, results of the analyses, and concluding remarks.

II.   Literature Review

A review of literature on Lyme disease and spatial analysis was performed to provide background on Lyme disease and to avoid the duplication of efforts. The ideas presented by previous studies have been considered in developing a hypothesis for this analysis.   Three of these literature reviews are provided below.

> A. A Climate-Based Model Predicts the Spatial Distribution of the Lyme Disease Vector Ixodesscapularis in the United States Author(s): John S. Brownstein, Theodore R. Holford, Durland Fish Source: Environmental Health Perspectives, Vol. 111, No. 9 (Jul., 2003), pp. 1152-1157Published by: Brogan & Partners Stable URL: http://www.jstor.org/stable/3435502 Accessed: 03/11/2010 18:08
>
> The authors of this study were interested in understanding of the spatial distribution of the black-legged tick, Ixodes scapularis, as a component in assessing human risk for Lyme disease in the United States. The purpose was to develop a spatially predictive logistic model for the species in the continental US to improve the previous vector distribution map of future tick habitat. The authors used the seasonality of temperature and humidity to explore the role of

---

[1] PA Department of Health Lyme disease Fact Sheet September 11, 2009

[2] Centers for Disease Control and Prevention. Recommendations for the use of Lyme disease vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR 1999;48 No. RR-7):5.

climate upon the tick species. The spatially modeled relationship between I. scapularis presence and large-scale environmental data provided a suitability model that revealed essential environmental determinants of habitat suitability, predicted emerging areas of Lyme disease risk, and generated the future pattern of I. scapularis across the United States. The analysis suggested the importance of climatic extremes and variation in vapor pressure as major indicators of habitat suitability.

B. Increasing Habitat Suitability in the United States for the Tick That Transmits Lyme Disease: A Remote Sensing Approach Author(s): Agustín Estrada-Peña Source: Environmental Health Perspectives, Vol. 110, No. 7 (Jul., 2002), pp. 635-640Published by: Brogan & Partners Stable URL: http://www.jstor.org/stable/3455508Accessed: 03/11/2010 17:45

The author of this study, Agustín Estrada-Peña, looked at the spread of the Ixodes scapularis tick species, one of the vectors of Lyme disease, into the United States. The modeling approach used the recorded distribution of this species and remotely sensed bioclimatic factors over the United States to establish the changes of habitat for this tick since 1982. The purpose was to detect the areas with factors adequate to support tick colonization and determine if the tick species expanded into areas with adequate habitat suitability in the period 1982 to 2000. The result of the study was that the tick did expand into these areas, and that the increase in winter temperature and in vegetation vitality is a key to tick habitat being suitable or unsuitable.

C. Lyme Disease and Conservation Author(s): Howard S. Ginsberg Source: Conservation Biology, Vol. 8, No. 2 (Jun., 1994), pp. 343-353 Published by: Blackwell Publishing for Society for Conservation Biology Stable URL: http://www.jstor.org/stable/2386458

The author states that the influence of Lyme disease can be manifested through its effects on wild populations and concomitant changes in natural communities and through potentially adverse effects of tick control on natural systems. In this essay the author discusses the possible influence of these effects on conservation efforts. I intend to use the author's research as background information for my analysis. I am not fully aware of all aspects of Lyme disease and tick habitat and this article provides some background on earlier findings of human development and ticks.

III.   Study Area and Background

The incidence of Lyme disease is examined at the county level. Six counties were chosen in southeastern Pennsylvania based on the reported rate of Lyme disease; York, Montgomery, Berks, Lancaster, Delaware, and Chester (Figure 1). As shown in Table 1, the rate of incidence in the southeastern region is the highest of all Pennsylvania regions since Lyme was made reportable in 1987.
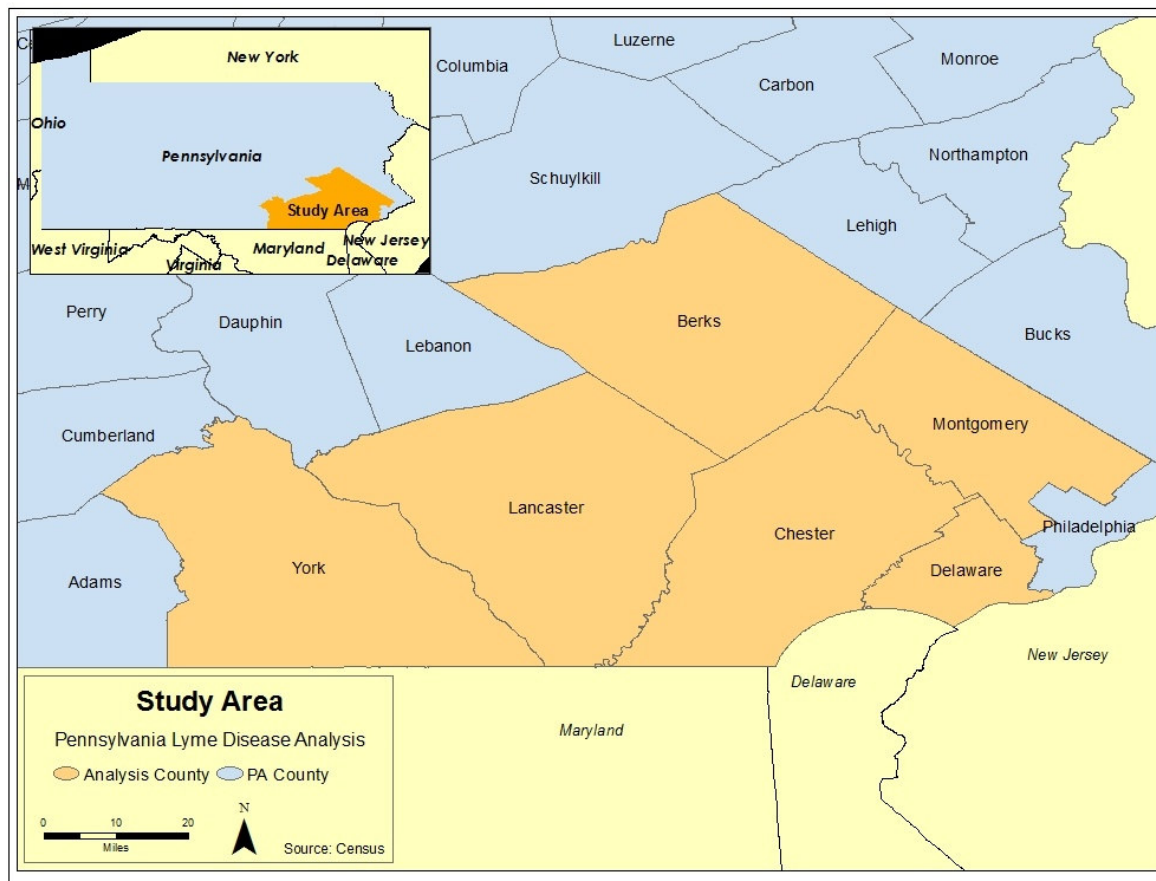
Figure 1: Study Area

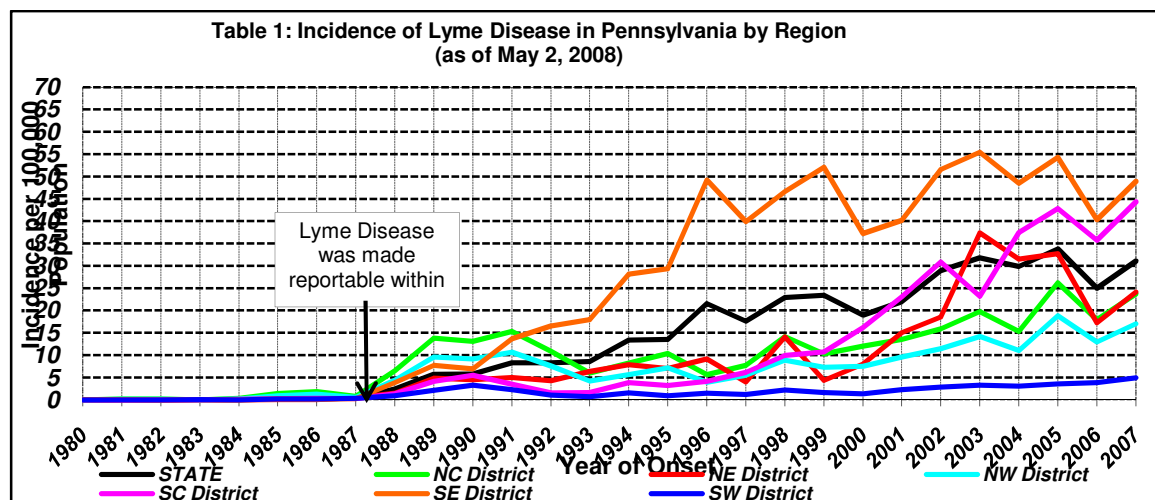Table 1: Incidence of Lyme Disease in Pennsylvania by Region (as of May 2, 2008)

Figure 2 shows the most recent distribution of the incidence of Lyme for each Pennsylvania County as provided by the PA DPH. The counties with the highest rates appear to be clustered in the southeastern corner of the state.
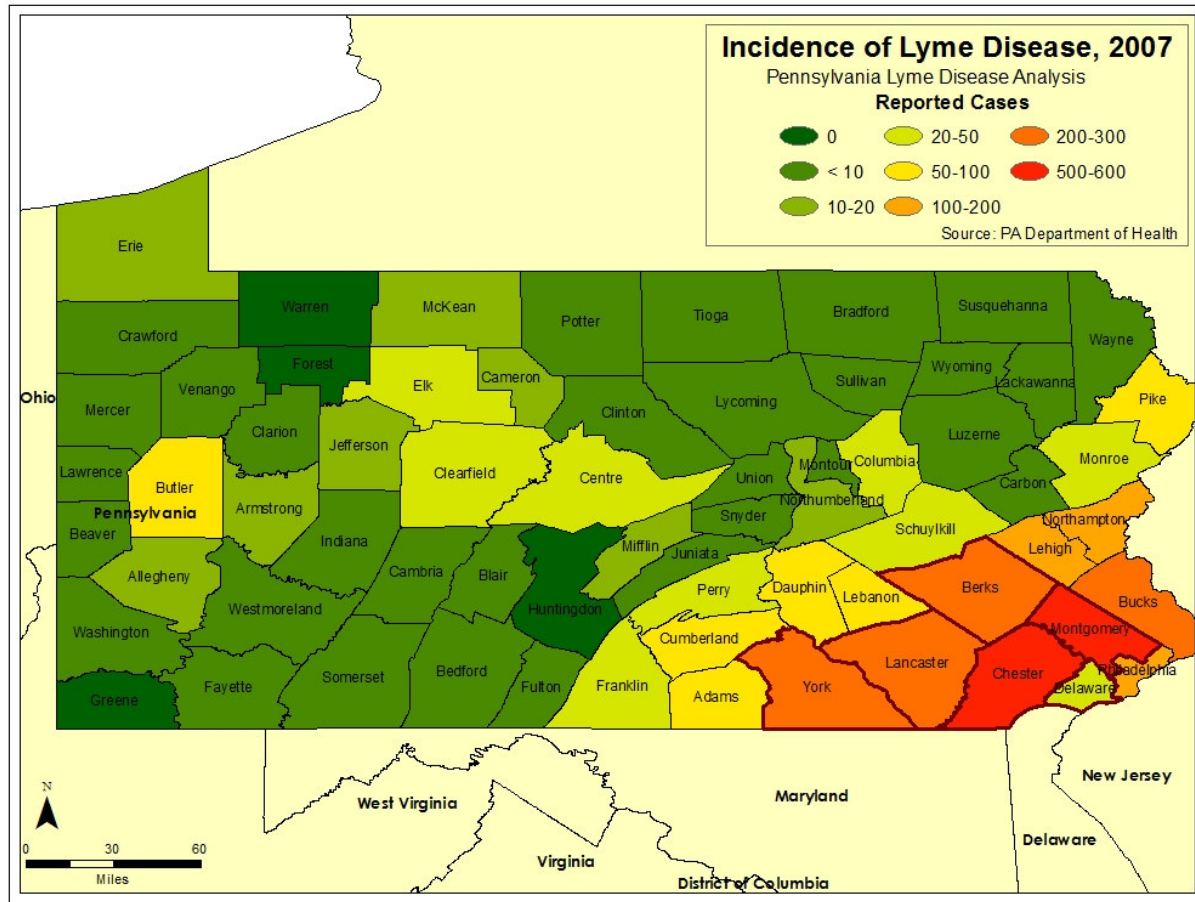
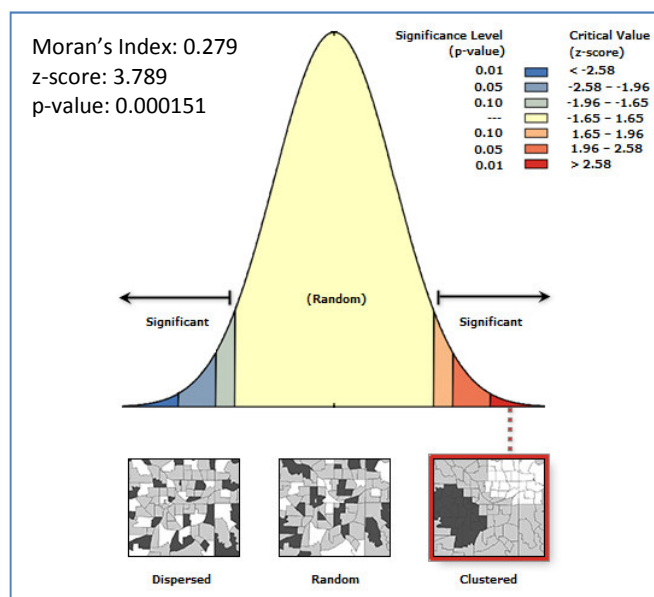**Figure 2: Incidence of Lyme Disease 2007**

Figure 3: Moran's I using change in Lyme 1990-2000

The Moran's Index was calculated using GIS to display the degree of clustering among the counties with high rates of Lyme. The model produced a Moran's index of 0.279 that reveals positive spatial autocorrelation and a z-score of 3.789 that reveals significant clustering (Figure 3). An Ordinary Least Squares (OLS) regression was then performed using the change in population density in Pennsylvania from 1990 to 2000 as the independent variable and the change in Lyme disease cases during the same period as the dependent variable and produced the following results:

Number of Observations: 67   Akaike's Information Criterion (AICc) [2]: 791.779069
Multiple R-Squared [2]:   0.238039   Adjusted R-Squared [2]:   0.226317

The R-squared, which measures the proportion of the variation in the dependent variable which is accounted for by the variation in the model, shows that population density accounts for 23.8 percent of the variation of Lyme disease within this model. To increase the R-squared, this analysis will provide an additional variable when looking at the study area in subsequent sections. The results of the OLS were modeled to produce a new Moran's Index which shows that the OLS accounted for the spatial clustering in the original dataset (Figure 4).

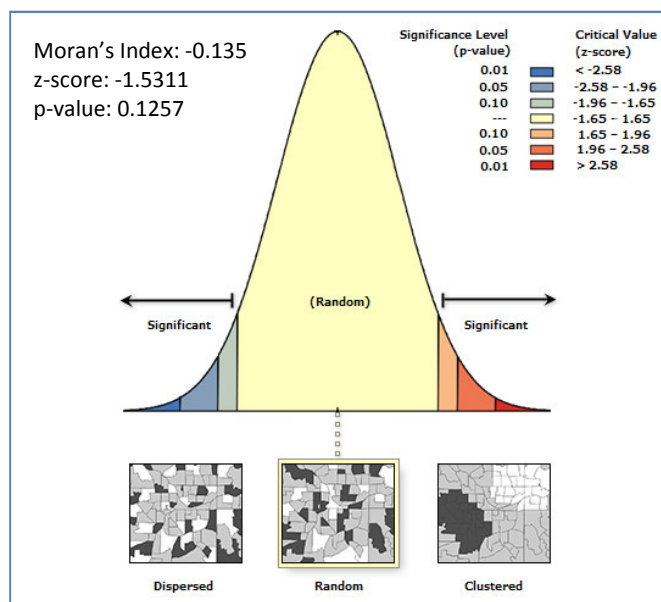**Figure 4: Moran's I using residual results from OLS**

To study area was chosen based on the OLS results that revealed that that correlation was present among counties with higher rates of Lyme. This paper looks at six adjacent counties in southeastern Pennsylvania that exhibit a range in standard deviation and change in number of Lyme cases; Delaware, Chester, Montgomery, Berks, Lancaster, and York (Figure 5).
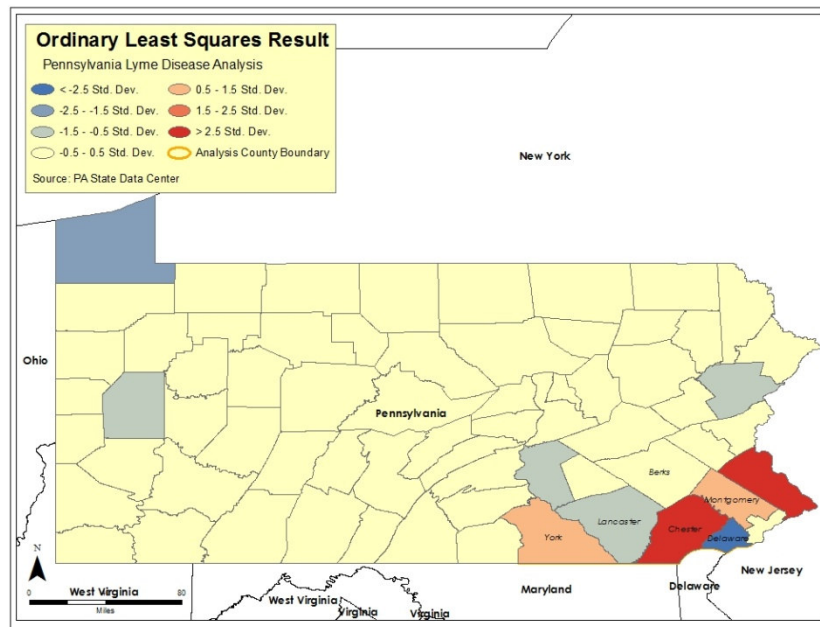
6

Figure 5: OLS Results showing cluster

IV.   Analysis and Methodology

Once the counties for analysis were chosen, a GIS and spatial statistics were used to associate county-level data Lyme disease case distribution with deer population distribution, human population density, and land cover to explain the distribution of Lyme disease in southeastern Pennsylvania, focusing on the years 1990 and 2000. The data used for the analyses are shown in Tables 2 and 3.

The selection of units of analysis was dictated by the availability of data. Lyme data is available by county from the Pennsylvania Department of Health. Population density data by county were sourced from the Pennsylvania State Data Center, deer harvest data was available from the Pennsylvania Game Commission on a county basis until 2001, median housing age was sourced from the 1990 and 2000 U.S. Census, and the change in impervious cover was sourced from the Natural Lands Trust.3 Land cover was available from United States

Table 2: Data 1990				
County	Reports of Lyme	Deer Harvest	Median Housing Age	Population Density
Berks	10	13,404	1953	389
Chester	39	7,652	1968	495
Delaware	121	1,496	1952	2,867
Lancaster	11	4,156	1961	429
Montgomery	102	2,795	1959	1,392
York	12	10,171	1961	372
Sources: PA Game Commission, PA Dept. of Health, National Land Trust, US Census				

Table 3: Data 2000				
County	Reports of Lyme	Deer Harvest	Median Housing Age	Population Density
Berks	66	17,465	1958	435
Chester	628	10,347	1973	573
Delaware	29	2,113	1954	2,990
Lancaster	30	6,493	1968	495
Montgomery	388	4,159	1963	1,552
York	156	14,717	1967	422
Sources: PA Game Commission, PA Dept. of Health, National Land Trust, US Census				

Geological Survey, but due to multiple data processing errors the change in impervious cover was used to represent the spread of human development. For acreage calculation purposes, these data were reformatted from a raster base to shapefiles, clipped to be specific for each county, merged based on

---

[3] http://www.natlands.org/

individual values and then calculated. A follow-on study could provide an additional analysis of specific land cover and vegetation types and the relationship with Lyme.

A. Spatial Autocorrelation

Spatial autocorrelation measures the extent to which the instance of a variable in one place, in this case Lyme disease, makes it probable that there will be an instance of Lyme disease in a neighboring county. The Moran's Index is used to compare the value of Lyme at any one location with the value at all other locations, measuring the randomness of values. Moran's I ranges from — 1 to +1, and equals 0 when there is no spatial autocorrelation (no effect of distance on the distribution of a variable). Increased positive or negative values indicate a stronger spatial component.

The Moran's Index was calculated for the 1990 and 2000 Lyme case data using first order polygon contiguity and Euclidean distances. The results are as follows:

Global Moran's I Summary – Lyme 1990
Moran's Index: 0.250952
Expected Index: -0.200000
Variance: 0.052679
Z-score: 1.964769
P-value: 0.049441

Global Moran's I Summary – Lyme 2000
Moran's Index: -0.366546
Expected Index: -0.200000
Variance: 0.083796
Z-score: -0.575340
P-value: 0.565062

A Moran's Index of 0.25 from 1990 data indicates a positive spatial autocorrelation of counties with regard to Lyme disease confirming the apparent clustering observed in Figure 2. This positive result indicates that the variations in Lyme disease cases have some dependency on the distance between counties. In contrast, the Moran's Index of -0.36 for the 2000 data indicates a negative spatial autocorrelation, or a tendency toward dispersion.

The Z-score represents simply standard deviations. The analysis returned a z-score of 1.96 standard deviations for 1990 and a -0.57 for 2000. The p-value is the probability that the observed spatial pattern was created by some random process. The p-value is very small for the 1990 data, which means it is very unlikely that the observed spatial pattern is the result of random processes, so you can reject the null hypothesis, which states that feature values are randomly distributed across the study area. With the 2000 data, there is a higher chance that the observed pattern is the result of random processes.

B. Geographic Weighted Regression

The Moran's Index showed us that the variables are not independent, therefore a Geographic Weighted Regression (GWR) was used to model the determinants of Lyme disease in the selected Pennsylvania counties. GWR was chosen over performing an Ordinary Least Squares (OLS) regression due to the small number of cases (6 counties). OLS provides a more accurate analysis when there are at least thirty cases.

To support this decision, a GWR was run using county-wide population density and change in Lyme data, the same data used in the OLS earlier in the study. The results are compared in Table 4.

The GWR produced an AICc of 786, which is lower

Table 4: OLS and GWR Results Comparison		
Result	OLS	GWR
Akaike's Information Criterion (AICc)	791.779	786.303
Adjusted R-Squared	0.2263	0.2974
R-Squared	0.238	0.329

8

than the OLS result, and an R-squared of .329 which is significantly higher than the OLS result of 0.238. This comparison reveals that a GWR model is more suited for this analysis and is therefore used for the study area data.

Initial data exploration was done to visualize the spatial variation in each of the variables. Figure 6 provides a visual comparison of the dependent variable, Lyme disease in 1990, and population density, median housing age, and deer harvest counts. The 1990 data show some clear patterns in the population density variable with higher values in Montgomery and Delaware counties. This may be attributed to a greater chance that Lyme disease would be reported in these areas given the higher population density. Figure five compares these same variables against Lyme disease in 2000. The 2000 data show a greater correlation between median housing age and deer population.

The GWR utilized the number of Lyme cases per county as the dependent variable (x) and the two independent variables (y) used were the annual deer harvest and population density. Median age of housing and change in impervious cover were not utilized by this analysis tool due to data format differences. The GWR was run initially with 1990 population density as the sole independent variable and Lyme cases as the dependent variable. Results are provided in Table 5.

The R2 measures the proportion of the variation in the dependent variable which is accounted for by the variation in the model, and the possible values range from 0 to 1. Values closer to 1 indicate that the model has a better predictive performance. The R2 can often be increased by adding variables, so the adjusted r2 is often reported – the adjustment takes into account the number of independent variables in the model and reflects model parsimony. Using one y value the adjusted R2 is 0.807, which tells the observer that 80% of the variance of Lyme is explained by the population density. The addition of the second variable, deer harvest, increases the R2 value to 0.83 and improves the reliability and performance of the model. Sigma, the estimated standard deviation for the residuals, is also reduced when using two y variables.

The 2000 date were analyzed the same as the 1990 data. Results are provided in Table 5. Figure 7 provides a visual comparison of the variables. The addition of the second y variable, deer population, provides a similar result to the 1990 data; the adjusted R2 increase to 0.83 and both the Sigma and AICc are reduced. The AICc, or corrected Akaike Information Criterion, is used as a measure of goodness of fit. The goodness of fit serves as an index of the spatial dependency—a good fit indicates strong spatial autocorrelation. In this GWR, the AICc can be used to compare the two models, and for both years the AICc is lower when two y variables are provided in the model.

Table 5: 1990 GWR Results		
Result	1 Variable	2 Variables
Residual Squares	1907.9956	529.7019
Effective Number	2.0037	4.7423
Sigma	21.8505	20.5231
AICc	69.6277	-334.0355
R2	0.8465	0.9573
R2Adjusted	0.8079	0.8305

Table 6: 2000 GWR Results		
Result	1 Variable	2 Variables
Residual Squares	1907.9956	529.7019
Effective Number	2.0030	4.742
Sigma	21.8505	20.5231
AICc	69.6276	-334.03551
R2	0.8464	0.9573
R2Adjusted	0.8078	0.8305

Incidence of Lyme Disease, 1990
Pennsylvania Lyme Disease Analysis
Reported Cases
- 0
- < 10
- 10-20
- 20-40
- 102
- 121
Source: PA Department of Health

Population Density, 1990
Pennsylvania Lyme Disease Analysis
1990 Density (pop/sq.mile)
- 11 - 129
- 130 - 313
- 314 - 869
- 870 - 2867
- 2868 - 11088
- Analysis County Boundary
Source: PA State Data Center

Median Housing Age, 1990
Pennsylvania Lyme Disease Analysis
- 1952
- 1953
- 1959
- 1961
- 1968
Source: U.S. Census 1990

Deer Harvest,1990
Pennsylvania Lyme Disease Analysis
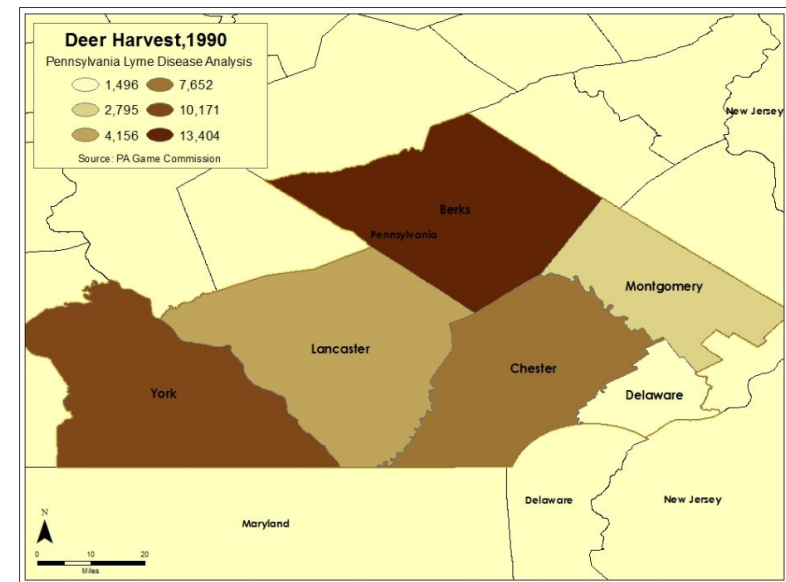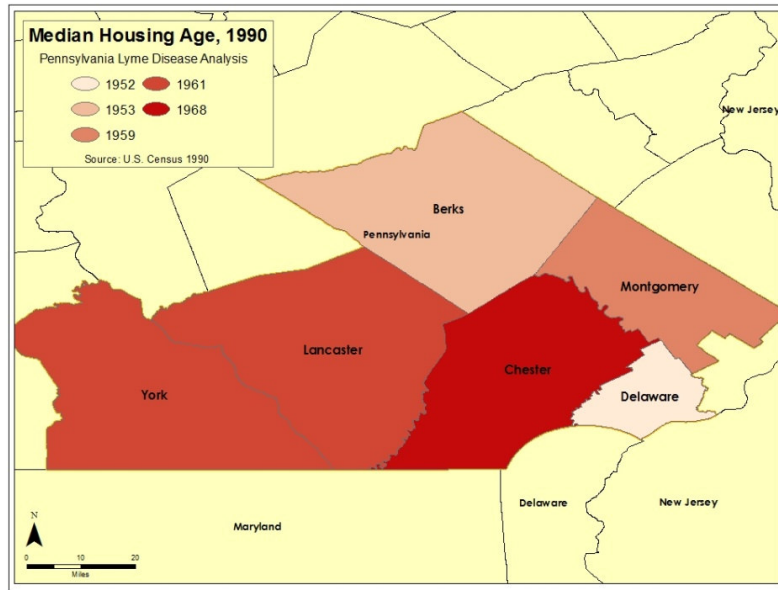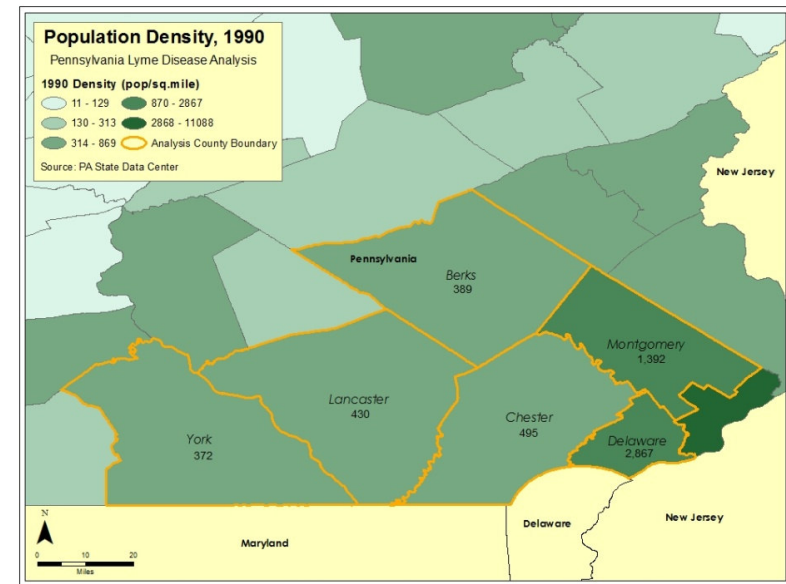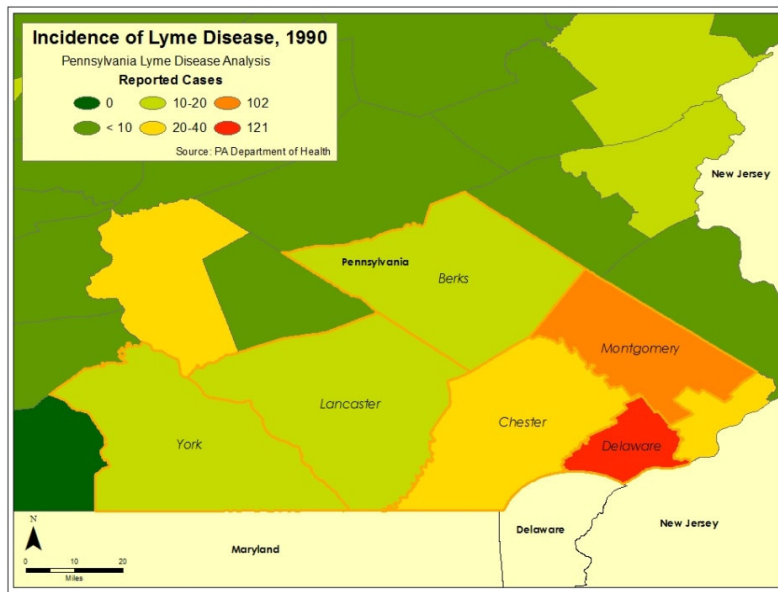- 1,496
- 2,795
- 4,156
- 7,652
- 10,171
- 13,404
Source: PA Game Commission

Figure 6: Initial Variable Exploration, 1990

Incidence of Lyme Disease, 2000
Pennsylvania Lyme Disease Analysis
Reported Cases
- 0
- < 10
- 10-20
- 20-40
- 40-70
- 156
- 388
- 439
- 628
Source: PA Department of Health

Population Density, 2000
Pennsylvania Lyme Disease Analysis
2000 Density (pop/sq.mile)
- 12 - 263
- 264 - 984
- 985 - 1,755
- 1,756 - 2,990
- 11,234
- Analysis County Boundary
Source: PA State Data Center

Median Housing Age, 2000
Pennsylvania Lyme Disease Analysis
- 1954
- 1958
- 1963
- 1967
- 1968
- 1973
Source: U.S. Census 2000

Deer Harvest, 2000
Pennsylvania Lyme Disease Analysis
- 2,113
- 4,159
- 6,493
- 10,347
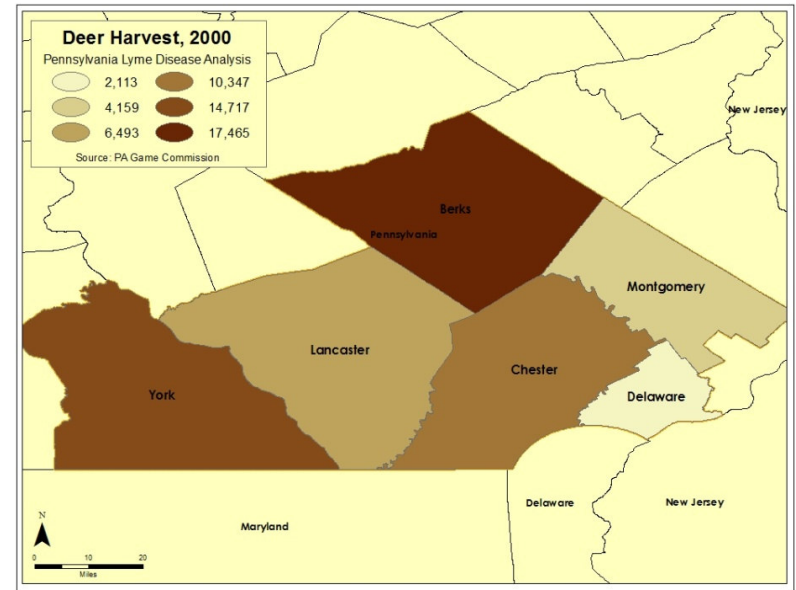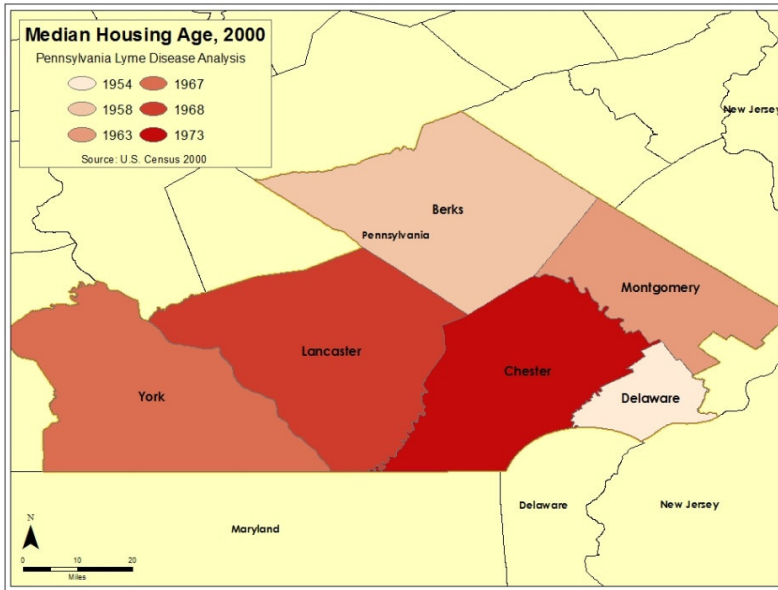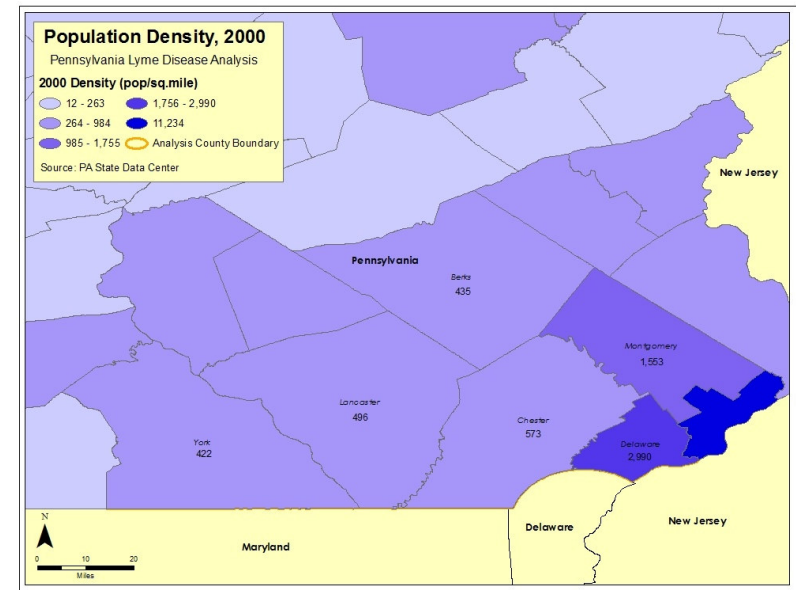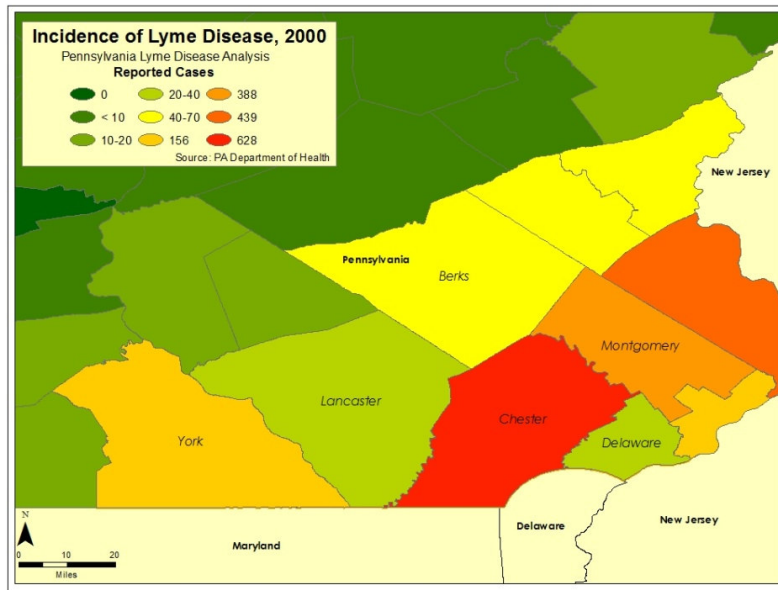- 14,717
- 17,465
Source: PA Game Commission

Figure 7: Initial Variable Exploration, 200

11

C. Additional Analysis

In addition to GIS analysis, the percent change of the variables was also examined. Table 7 provides the percent change from 1990 to 2000 of deer harvest, Lyme cases, impervious cover, and population density. Chester County has the highest rate of increase in both Lyme disease and population density, which supports the results from the GWR. The lower percentage of increase in impervious cover could potentially mean that the increases that did occur were in areas with high deer populations.

County	Reports of Lyme	Deer Harvest	Percent of land with more than a 4% increase in impervious cover	Population Density
Berks	560%	30%	7%	12%
Chester	1510%	35%	13%	16%
Delaware	-76%	41%	60%	4%
Lancaster	173%	56%	21%	15%
Montgomery	280%	49%	20%	11%
York	1200%	45%	17%	13%

Table 7: Percent Change 1990-2000

Sources: PA Game Commission, PA Dept. of Health, National Land Trust, US Census

Figure 8 provides a visual of the increase of impervious surfaces in study area from 1985-2000. Both the table and the map table would suggest that the relationship between Lyme and impervious cover would need to be further examined; Delaware County has the largest percentage of land that shows an increase in impervious cover but actually shows a decrease in reports of Lyme disease. The use of impervious cover as an indicator of vegetation loss is a weakness of the study, as much suburban development does not necessarily increase imperviousness significantly. Also, the Lyme disease data are reported at the county level. If addresses were used in the study it could be more precisely determined whether or not these two variables are correlated.
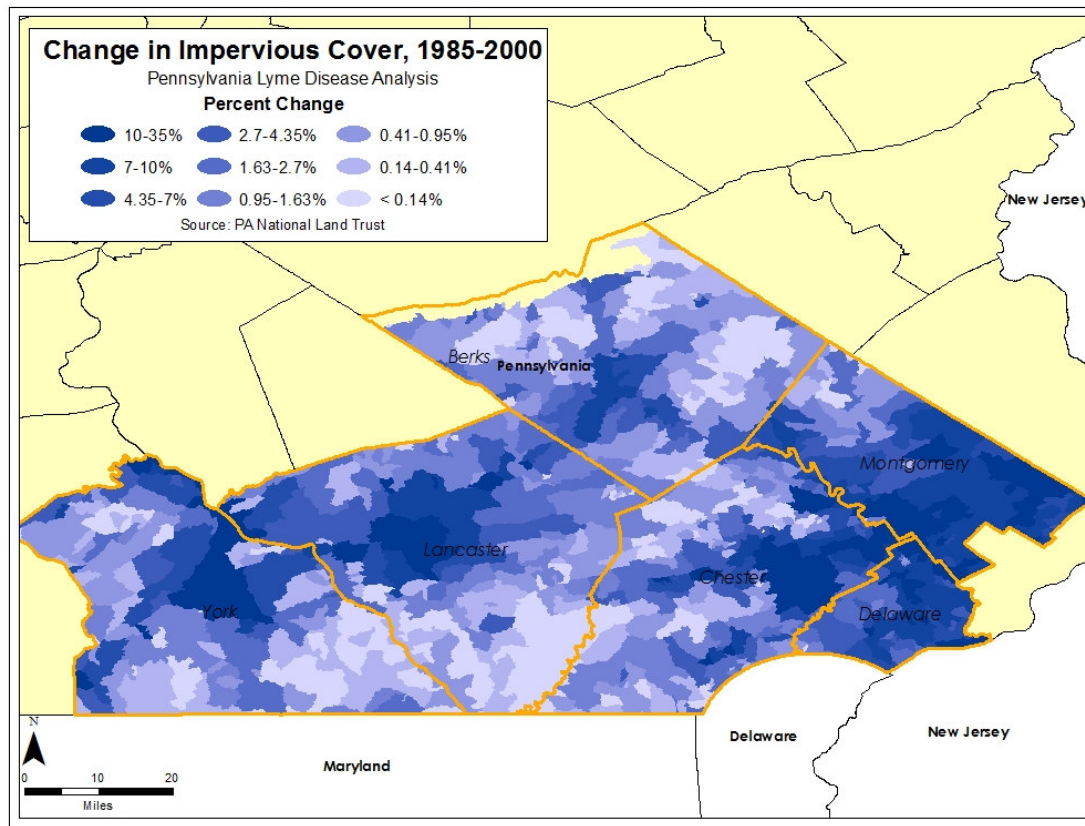
Figure 8: Impervious Cover Change 1985-2000

V. Results and Discussions

I hypothesized that that the instance of Lyme disease would increase with an increase in deer harvest and impervious cover into previously vegetated areas. Lyme disease did show a relationship with increased deer habitat, but the relationship with vegetation needs further study. The 2000 data showed Chester County with the highest rate of Lyme and the youngest median age of housing. This would suggest a correlation between new housing and increases in Lyme which could also be analyzed further.

The Moran's Index provided conflicting results, showing positive spatial autocorrelation with the 1990 data and negative spatial autocorrelation with the 2000 data. This could be explained by reporting methods and knowledge of the disease outside of more densely populated areas, or it could be due to over-reporting in a one or more counties. Chester County has a much higher rate than the surrounding counties in 2000, which could account for the inconsistent results.

The GWR indicated that the variables of population density and deer population explained some of the variation in Lyme disease cases. An increase or change in variables could provide a different result and perhaps increase the R-squared to provide greater explanation.

VI.   Concluding Remarks

The application of GIS to the study of Lyme disease in southeastern Pennsylvania could lead the observer to a few important conclusions. First, the hypothesis that Lyme disease is not spatially random is rejected, as clusters were detected. Those counties with the highest rates of Lyme tended to be neighbors, and also show higher levels of deer population and population density particularly for the 1990 data. The 2000 data showed a negative correlation, potentially based on the fact that Chester County showed such an increase in Lyme that is was an outlier (as revealed in a GeoDa exercise). This revelation suggests that further scrutiny of Chester County might be necessary to determine what is driving such a dramatic increase in Lyme disease cases.

A more robust analysis could be performed with data for a smaller unit area than the county level. Because cases were assigned to the county level, the spatial pattern found indicates relatively large geographic areas (clusters of counties) as being higher reporting areas for Lyme disease rather than specific parts of a county. The limited study area proved prohibitive in the number and types of analyses that could be done, and expanding the study area to include more counties in Pennsylvania and Maryland might provide a broader look at the relationship of Lyme to human development. Additional information about the degree of human activity in areas with higher deer population may provide valuable in a follow-on analysis as well.

According to the CDC, both underreporting and over diagnosis are common with Lyme disease.[4] Moreover, there is still no consensus on the precise geographic distribution of Lyme disease in the United States because of increased human case surveillance, over diagnosis, under-reporting, and human travel.[5] These factors could have a large role to play in correctly deciphering the explanatory variables for an increased rate of Lyme disease. This spatial analysis provides justification for further consideration of local patterns of spatial autocorrelation, and a more detailed study that includes a greater number of independent variables.

---

[4] Centers for Disease Control and Prevention. Recommendations for the use of Lyme disease vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR 1999;48 No. RR-7):5.

[5] A Climate-Based Model Predicts the Spatial Distribution of the Lyme Disease Vector Ixodesscapularis in the United States Author(s): John S. Brownstein, Theodore R. Holford, Durland Fish Source: Environmental Health Perspectives, Vol. 111, No. 9 (Jul., 2003), pp. 1152-1157 Published by: Brogan & Partners Stable URL: http://www.jstor.org/stable/3435502