Lyme Disease in New Jersey

Spatiotemporal Variation in County Incidence from 2008-2018

Orion Chen

GEOG 495

June 7, 2020

Table of Contents

Abstract

Existing research on Lyme disease is numerous but has not arrived at a commonly used form of spatiotemporal analysis, nor has it visited New Jersey as a geography of interest. The lack of healthcare funding and resources allocated to Lyme disease within the state poses an additional challenge. I use SaTScan, a space-time scan implementation by Martin Kulldorff, building off similar exploratory studies to understand Lyme disease variance within New Jersey counties from 2008 to 2018. I seek to answer the question: what are the spatiotemporal patterns of Lyme disease in New Jersey counties from 2008 to 2018? Lyme disease is most commonly observed in the northwestern region of the state, while it is least commonly observed in the northeastern region of the state. Various outlier counties regardless of geographical patterns are also observed. Such results are evidenced both in incidence rate maps and in Kulldorff's space-time scan. A lack of greater patterns observed points to a need for greater specificity in temporal and/or spatial accuracy, and/or an expansion of scope to incorporate multiple states. Nonetheless, the contribution of this work as an exploratory project provides a baseline upon which future research into underlying factors can be developed.

Introduction

Lyme disease has been known for decades within the United States as a persistent biological disorder affecting primarily the Northeastern region. Despite its well-documented history, academics have not arrived at a commonly used method by which to understand its spatiotemporal emergence. The variety of space-time methods used to analyze Lyme disease are applied to few case studies and lack generalizability and cross-comparison with other studies across multiple geographic areas. Furthermore, little literature has been published concerning variance patterns in the state of New Jersey, despite a considerable amount of cases occurring there. Addressing this lack of knowledge, then, is the primary focus of my work, which seeks to answer the question: what are the spatiotemporal patterns of Lyme disease in New Jersey counties from 2008 to 2018? I first contextualize the biological foundations of the disease, its underlying factors, and existing research on it in my background research. I then explain my project's data processing and analysis in the methods section, using Excel and ArcMap to process initial datasets into use for a SaTScan-based Poisson scan. The results are discussed and visualized in detail in the results section. The immediate implications of these statistics are discussed in the discussion section, particularly regarding curious phenomena within the information output. Finally, fundamental data and analysis issues are tackled in the conclusion and recommendations section, linking this project's work to future research pathways to better situate Lyme disease incidence within the larger sphere of ecological understanding.

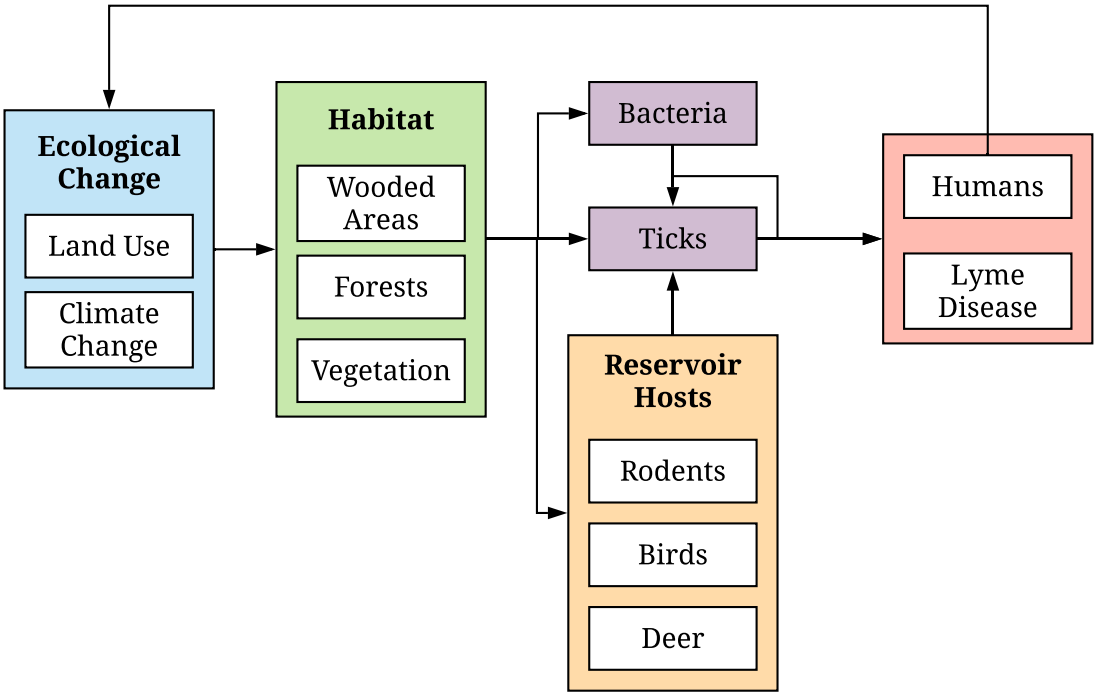
Lyme disease (LD), known alternatively as Lyme borreliosis (World Health Organization, n.d.), is one of many zoonotic vector-borne diseases impacting human populations. Its causative pathogen is a strain of bacteria, *Borrelia* (Bunikis et al., 2004); *Borrelia burgdorferi* is the most common species observed in North America (Ozdenerol, 2015), while *Borrelia afzelii*, *Borrelia garinii*, and *Borrelia valaisiana* are more prevalent in Europe and Asia (Etti et al., 2003). *Borrelia* bacteria reside primarily in the *Ixodes* genus, the largest genus of the *Ixodidae* family of hard ticks (Nava et al., 2017). *Ixodes scapularis* and *Ixodes pacificus* are the most common transmission vectors in North America, while *Ixodes ricinus* is most common in Europe and *Ixodes persulcatus* is most common in Asia (Ozdenerol, 2015). These ticks feed mostly on smaller vertebrates like mice as juveniles, shifting their reservoir hosts to larger animals like deer as they reach adulthood (Wood & Lafferty, 2013). As generalized feeders, however, ticks can feed on birds and other mammals, "an adaptation that maintains tick populations and LD transmission during the crash phase of mouse population cycles" (Wood & Lafferty, 2013, pp. 239-240). When ticks feed on humans, they transfer *Borrelia* bacteria over, thereby causing LD as the pathogens interact with human immune systems. This interaction, and modern LD, have been characterized as an unintentional side effect of ecological proximity, as "humans are accidental dead-end hosts" (Ozdenerol, 2015, p. 15182). *B. burgdorferi*, for example, has existed in the Northeast and Midwest regions of North America for at least several thousands of years pre-Columbian "before the emergence of modern Lyme disease" (Hoen et al., 2009, p. 15013). This emergence is attributed to multiple "independent emergence events" of the original bacterial strain "likely as a consequence of reexpansions of multiple vector tick populations" (Hoen et al., 2009, p. 15016). Almost all LD cases occur in the Northern Hemisphere, most being reported in the United States and Europe (Hubálek, 2009). Within the States, most cases are reported in the Northeastern and Midwestern areas, specifically coastal and riparian zones (Pepin et al., 2012; Ozdenerol, 2015).

LD produces numerous and diverse symptoms of varying significance. Affected cases typically suffer from headache and fatigue, which are observed in most tick-borne diseases at large (Centers for Disease Control and Prevention, 2019a); joint pain is occasionally observed. The most prominent symptom is a red, circular rash called erythema migrans that may develop into additional lesions if left untreated. Serious, but rare, cases suffer from musculoskeletal, nervous system, and cardiovascular symptoms. Antibiotics, oral or otherwise, are typically prescribed for LD cases (National Institute of Allergy and Infectious Diseases, 2019), but occasionally result in 'Post-treatment Lyme disease syndrome', a generalized condition encompassing various auto-immune response correlated with LD incidence (Centers for Disease Control and Prevention, 2019c). LD prevention involves risk avoidance, by promoting awareness of potential exposure locations to ticks such as wooded and grassy areas, and by using tick repellant and checking for ticks on skin (Centers for Disease Control and Prevention, 2019b). No vaccine has been developed yet; Willadsen (2006) points to a lack of existing research on tick biochemistry and immunology informing potential developments. Other proposed ideas have involved biocontrol (the use of biological agents to control tick populations), parasiticides, and integrated pest management.

As a disease, LD presents not just a public health issue, but a financial strain: a review of LD economic analysis papers by Mac et al. (2019) cited Zhang et al.'s (2006) estimate of LD costing the US healthcare system 292 million USD annually. This figure may be compared against West Nile virus, which costs 778 million USD over 13 years (Staples et al., 2014), and Zika virus, costing 500 million USD annually (Lee et al., 2017). If treated, the mitigated cost of LD has been estimated at around $4,466 (1992) USD (Meltzer et al., 1999). Addressing LD, then, has a significant monetary incentive.

As stated before, LD is a modern phenomenon. Since its initial identification as Lyme arthritis in Lyme, Connecticut, in the mid-1970s after observations of clustered juvenile rheumatoid arthritis (Steere et al., 1977; Borchers et al., 2015), LD has spread "to affect large areas of the Northeast and Midwest" (Hoen et al., 2009, p. 15013). Such a phenomenon has been attributed in large part to changes in land use (such as increasing levels of suburbanization into wooded areas) creating 'peri-urban' and 'peri-domestic' ecologies that have placed humans in close proximity to tick vectors (Gubler et al., 2001; Ozdenerol, 2015). Climate change has also been commonly theorized as expanding tick ranges, as areas not currently endemic "may become so due to climate change" that impacts significant criteria like "warming, humidity, and moisture" which "are important factors in the biology of tick survival and dispersion" (Ozdenerol, 2015, pp. 15193-15194). Ogden et al. (2009), for example, theorize that warmer temperatures and tick dispersion by migratory birds have contributed to the emergence of LD in Canada due to the expansion of the range of *I. scapularis*. These understandings, along with the ecological components of LD outlined above, form the basis of my conceptual framework (Figure 1), which links these patterns together under a unified structure.

Figure 1: A conceptual framework of the project that provides the ecological basis of LD existence and life cycle patterns.



A spatial and/or temporal understanding of LD emergence patterns has been studied for decades, but still remains a topic of considerable opportunity. Ozdenerol's 2015 review, characterized as "the first comprehensive review on spatiotemporal patterns of LD and the use of geospatial technologies to study interfacing populations of the reservoir hosts, vectors and humans" (p. 15185), overviews various spatial, temporal, and spatiotemporal studies of LD, both within the United States and internationally. Considerable biological work has been undertaken, based on either field data or *a priori* ecological studies, however are often hard to generalize upwards "to longer timer periods or large spatial scales" owing to field surveys being typically collected over small geographic areas (Ozdenerol, 2015, p. 15187). The studies Ozdenerol outlines concerning spatiotemporal analysis are diverse in methods, but also point to a lack of agreed-upon understanding in the choice of a singular (or set of) analysis that can be used to understand LD patterns across multiple geographies. The publication of relatively new exploratory analyses, as Szonyi et al. (2015) did with Texas, also point to a lack of common understanding in LD pattern dynamics at large. Works on LD incidence have been conducted since 1997 (Kitron & Kazmierczak, 1997), but all use different methodologies that make cross-study comparison difficult (Li et al., 2014; Waller et al., 2007; Zeman & Benes, 2014). This is not to say a heterogeneity of analytic methods is detrimental, but rather that the existing body of literature is not mature enough to have developed commonly used methods of analysis. One method, space-time scan, has been utilized on occasion (Li et al., 2014; Kugeler et al., 2015), an occurrence I will revisit in my methodology. New Jersey, in particular, is a geographic area of apparent disinterest in current literature. Searches on Google Scholar, EBSCOhost, and PubMed Central concerning "Lyme disease New Jersey" have returned few LD studies done specifically in New Jersey. Of note is that New Jersey lacks statewide surveillance of tick-born pathogens since 2000-2001 (Egizi et al., 2018). Subsequent discontinuations of tick collections in 2002 and

deer check stations in 2012 appear to indicate an overall downscaling of LD-related public health surveillance in the state, making this work increasingly important in light of the existing budget shortfalls and limited resources.

Methods

While there are numerous ecological factors at play in how LD might vary spatially and temporally, my project focuses only on incidence rates (IRs) for New Jersey counties, as an exploratory analysis of LD, as a matter of expedience.

County-level IRs were calculated using decennial population counts from the U.S. Census Bureau (2000, 2010) and recorded Lyme disease cases from the Centers for Disease Control and Prevention (CDC) (n.d.c). Then, population counts by county were linearly interpolated using the following equations:

$$\text{CHANGE} = \frac{\text{POPULATION}_{2010} - \text{POPULATION}_{2000}}{10}$$

$$\text{POPULATION}_{20XX} = \text{POPULATION}_{2000} + \text{CHANGE} \cdot XX$$

For each county, their census-enumerated populations for 2000 and 2010 (POPULATION_{2000} and POPULATION_{2010}, respectively) are used to estimate a yearly CHANGE in their population. Then, that county's estimated population for a given year $20XX$ (between 2008 and 2018) is calculated with the POPULATION_{20XX} equation, where XX is between 8 and 18. The CHANGE calculations are enumerated in Appendix A.

Then, IRs (per 100,000 population) for each county were calculated using the above population estimates in conjunction with yearly CDC case numbers using the following equation:

$$\text{IR}_{20XX} = \frac{\text{LDCASES}_{20XX}}{\text{POPULATION}_{20XX}} \cdot 100000$$

For each county, LDCASES_{20XX} represents the recorded number of LD cases in the year $20XX$. Similarly. the subscript $20XX$ in IR_{20XX} denotes the year corresponding to the IR.

These equations were implemented using Microsoft Excel, where data was manually imported into a workbook then processed with the use of Excel-based equations.

A space-time scan statistic was used to understand the spatiotemporal variation of LD. A space-time scan takes point data with spatial and temporal dimensions, then creates a shaped window centered at each point. From there, a space-time scan repeatedly resizes the size of the windows over a set number of permutations, "noting the number of observed and expected observations inside the window at each location" (Kuldorff, 2018, p. 6). As Kulldorff (2018) states:

> The space-time scan statistic is defined by a cylindrical window with a circular (or elliptic) geographic base and with height corresponding in time. The base is defined exactly as for the purely spatial scan statistic, while the height reflects the time period of potential clusters. The cylindrical window is then moved in space and time, s that for each possible geographical location and size, it also visits each possible time period. In effect, we obtain an infinite number of overlapping cylinders of different size and shape,

jointly covering the entire study region, where each cylinder reflects a possible cluster. (p. 7)

I used Martin Kuldorff's implementation of the space-time scan statistic through the software SaTScan (v9.6). In particular, I utilized the Poisson statistic, which counts the number of occurrences of an event in a given unit of analysis, presuming that each event occurs independently and therefore creates clustering based on the probability that event distribution is significantly beyond what would be observed solely due to random chance. Kulldorff chooses an inhomogeneous Poisson process (1997), where:

"The probability of n_G number of points in the study area is" (p. 1486):

$$\frac{e^{-p\mu(Z)-q(\mu(G)-\mu(Z))}[p\mu(Z)+q(u(G)-\mu(Z))]^{n_G}}{n_G!}$$

"The density function $f(x)$ of a specific point being observed at location x is" (p. 1486):

$$\begin{cases} \dfrac{p\mu(x)}{p\mu(Z)+q(\mu(G)-\mu(Z))} & \text{if } x \in Z \\ \dfrac{q\mu(x)}{p\mu(Z)+q(\mu(G)-\mu(Z))} & \text{if } x \notin Z \end{cases}$$

These equations inform the "test statistic λ of the likelihood ratio test": (p. 1487)

$$\lambda = \frac{\sup_{Z \epsilon \mathcal{Z}} L(Z)}{\frac{e^{-n}G}{n_G!}\left(\frac{n_G}{\mu(G)}\right)^{n_G}\prod_{x_i}\mu(x_i)}$$

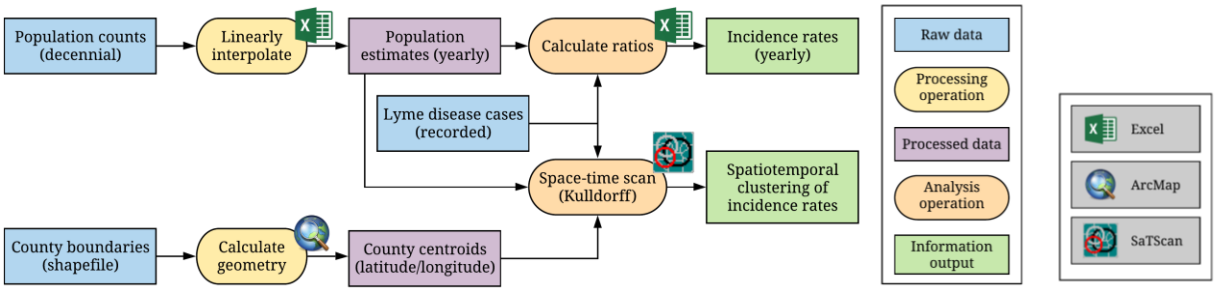
To undergo SaTScan-based Poisson analysis, the software requires population data ($\mathrm{POPULATION}_{20XX}$), case data ($\mathrm{LDCASES}_{20XX}$), and location data of these datasets. To calculate location data, I processed New Jersey county boundary shapefiles (New Jersey Office of Information Technology, 2020) in ArcMap (v10.7.1) using a customized version of Esri's Calculate Geometry tutorial (Esri, n.d.). Using ArcMap's built-in Calculate Geometry field calculator, I set each county centroid's longitude to its x-coordinate and its latitude to its y-coordinate.

The population, case, and location datasets were inputted into SaTScan. Most of the settings were left on default. The "Type of Analysis" was chosen to be a retrospective space-time analysis, using the Poisson probability model. High and low rates were chosen to be scanned for, and data was aggregated by a length of 1 year. The maximum spatial cluster size is 25% of the population at risk (total county population), and the maximum temporal cluster size is 50% of the study period. The 25% and 50% parameters were chosen after permuting through a few variations to see if there were major inconsistencies arising from different maxima that might indicate an inappropriate choice of values. A 25% maximum spatial cluster size, in particular, was chosen to minimize the outsize variation in cluster size potentially caused when counties

with small populations artificially increase circle diameter. A circle was used as the scan window, similar to other space-time studies (Li et al., 2014; Kugeler et al., 2015). The number of Monte Carlo replications was set to 999. The criteria for reporting secondary (hierarchical) clusters was set to include no geographic overlap only.

All the processing and analysis was performed on a thin client (consumer laptop). Figure 2 shows the workflow implemented for this project.

Figure 2: A workflow diagram showing the data and methods used, and outputs created for this project.



Results

By and large, IRs show some variation over time in New Jersey at the county level, but appear to be stable relative to other counties rankings-wise. Relatively high IRs are consistently observed in the northwestern part of the state, as Sussex, Warren, and Hunterdon Counties have IRs ranging from mid-150s to almost and past 300. Relatively low IRs are observed in the northeastern corner around Bergen, Essex, Hudson, and Union Counties with IRs mostly under 30. Heterogeneity is observed in the southern part of the state, with adjacent counties often displaying wildly different IRs. Burlington County, for example, has IRs ranging from 90.9906 in 2009, to 39.6583 in 2018. Its neighbor, Camden County, however, sees a general increase in IRs from an abnormally low 0.3901 in 2008 to 30.3419 in 2018. Atlantic County, curiously, appears to have a general decrease in IRs, moving from 45.5304 in 2008 to 25.6721 in 2018. A map showing counties with their names is provided for convenience in Appendix B. Figure 3 shows a three-year period of IRs, from 2013 to 2015, as a visual representation of the IR variation present in New Jersey.

A lack of large-scale clustering patterns at the county level, is observed with respect to Kulldorff's space-time scan. Only two non-zero clusters are observed: one high-IR cluster from 2008 to 2012 with radius 40.54km is centered around Warren County, while a low-IR cluster from 2010 to 2014 is centered around Essex County with radius 15.02km (Table 2). Such results appear to support the earlier observations of raw county-level IRs, however no significant spatial-LD-expansion-over-time clustering is observed. SaTScan also outputs three zero-radius low-IR clusters in Camden, Middlesex, and Monmouth Counties with varying time intervals, indicating isolated pockets of extremely low IRs.

Table 1: Incidence rates per 100,000 population of each county in New Jersey.

County	Incidence Rates per 100,000 (year)										
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Atlantic	45.5304	50.3031	45.1650	44.4448	34.7735	23.8309	35.6453	39.9234	32.3201	22.7629	25.6721
Bergen	19.3137	31.1179	18.3402	24.3603	20.1250	26.4424	23.2071	30.0345	30.4016	32.9414	25.1650
Burlington	62.6597	90.9906	54.8209	51.6323	57.2937	46.2379	42.7136	55.9163	60.3529	48.6632	39.6583
Camden	0.3901	1.1692	1.7521	17.3108	17.1006	25.0449	21.9185	26.7432	23.8145	30.7565	30.3419
Cape	36.6312	40.9119	27.7592	44.4408	42.5965	27.1552	12.5997	28.5009	20.1639	23.4737	33.2561
Cumberland	63.3050	125.1187	61.8236	53.1834	55.3494	37.4916	37.2481	53.0445	60.0586	60.2850	49.0119
Essex	15.5236	26.6264	18.6232	18.7739	15.4724	16.7719	14.6135	21.1773	20.3040	26.6339	21.5140
Gloucester	35.5158	52.2943	38.8500	45.9456	46.7779	33.8504	41.0958	41.9541	38.5791	39.7668	34.5834
Hudson	2.7018	9.3393	6.3065	4.3970	4.0668	7.6341	6.8282	8.3474	7.6989	11.0435	3.9725
Hunterdon	298.2444	347.6545	251.6576	341.1249	159.6963	172.7354	67.9945	346.6916	158.8923	334.3348	251.0548
Mercer	64.3986	69.3270	43.1090	58.9533	38.1428	37.9810	38.3570	66.5084	69.6875	58.8019	43.5225
Middlesex	6.7676	22.6400	16.5461	27.0891	19.2262	19.3291	26.3872	34.2977	35.8294	37.6918	17.9568
Monmouth	41.2838	65.0371	53.3012	56.3392	53.2053	58.5916	68.9806	83.0826	76.9439	85.8121	78.7619
Morris	82.8101	130.5939	97.3031	96.6668	80.7348	111.8472	81.4207	113.6481	106.6241	127.8264	92.5623
Ocean	26.4449	33.1578	43.1867	38.0703	34.5940	35.2194	38.4853	47.5884	44.4836	48.5123	46.0986
Passaic	31.2757	50.5992	35.3134	34.4318	34.1500	35.6521	37.5423	48.0964	40.1155	32.3689	29.1604
Salem	66.9476	68.2822	45.3975	98.0954	79.7688	78.0523	28.4423	55.2387	47.6467	66.8241	85.8992
Somerset	48.0750	79.7886	63.0712	52.4477	76.3768	64.9096	43.1364	76.6896	63.7138	81.9646	65.3676
Sussex	238.1193	291.7549	267.3098	295.1113	222.9113	230.1153	170.5176	203.5385	170.6888	216.5748	160.4236
Union	12.5537	16.4454	12.4884	16.5460	14.0926	14.4261	10.8840	19.6880	17.4353	20.1366	16.7986
Warren	239.2010	270.2052	172.9658	257.9654	178.2742	192.6416	97.1275	240.5673	216.1057	347.5723	214.6074

Figure 3: A three-year time series of maps showing Lyme disease incidence rates in New Jersey by county from 2013 to 2015. Incidence rates aredisplayed as a 10-category quantile, bounded by the minimum and maximum incidence rates observed during this time period.

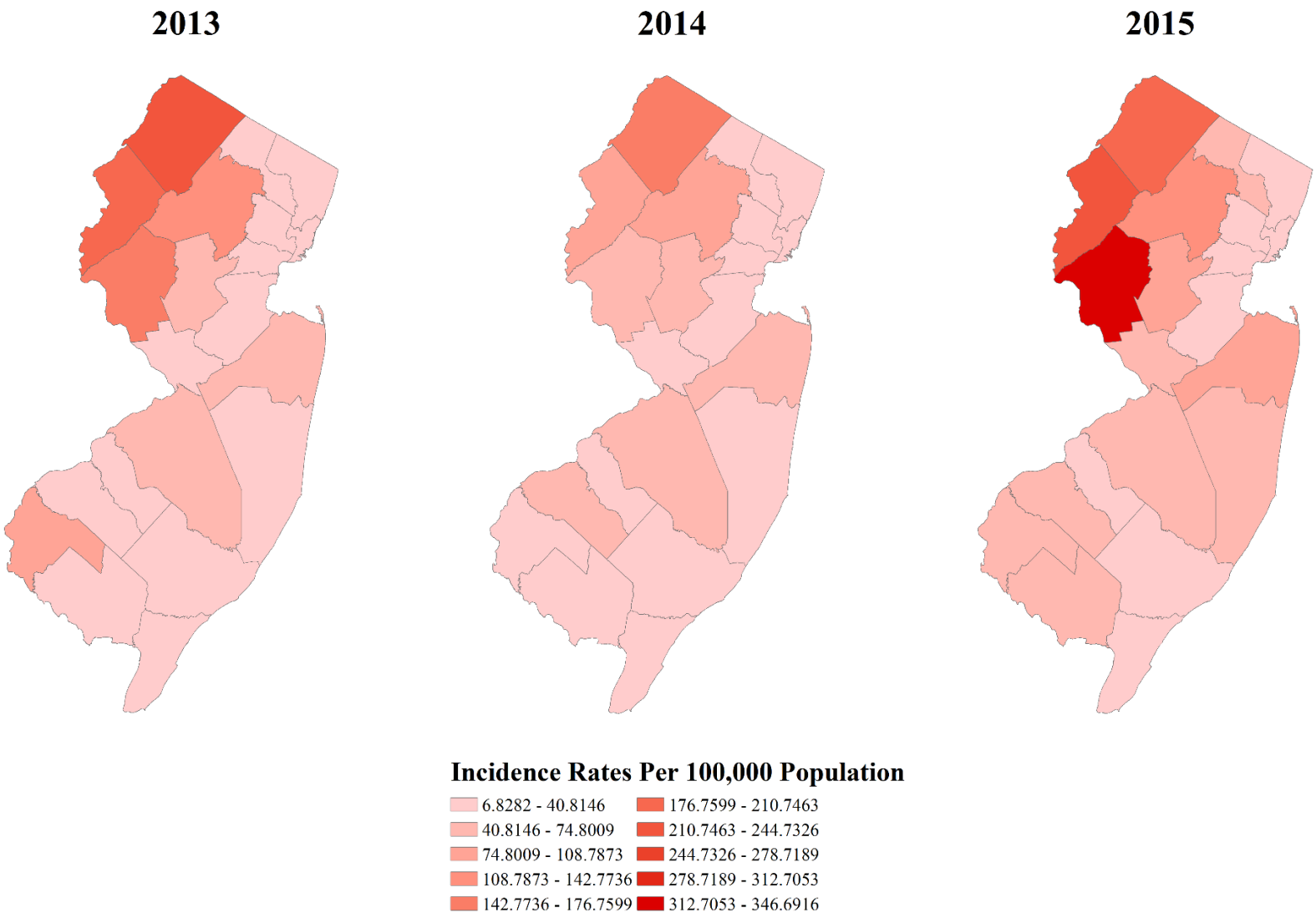


Figure 4: A map of space-time scan clustering at the county level in New Jersey, from 2008 to 2018.

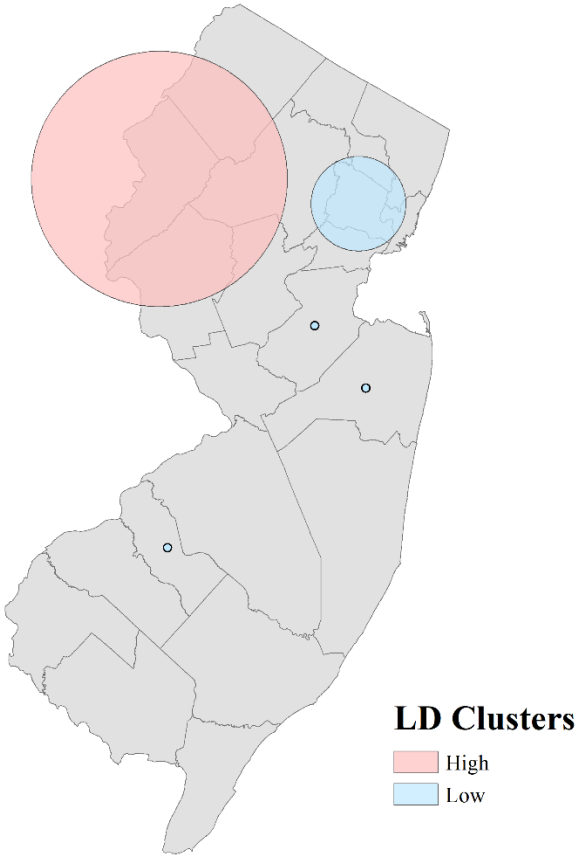


Table 2: Statistics for all clusters derived from county-level space-time scan clustering in New Jersey.

Cluster Location	Warren County	Essex County	Camden County	Middlesex County	Monmouth County
Radius (km)	40.54	15.02	0	0	0
Time Frame	2008-2012	2010-2014	2008-2012	2008-2012	2014-2018
Population	890,525	1,963,616	515,076	827,783	634,907
Observed Cases	7,373	1,215	194	749	2,517
Expected Cases	2,033.03	4,533.88	1,188.47	1,874.15	1478.55
Annual Cases per 100,000	167.7	12.4	7.6	18.5	78.7
Observed / Expected	3.63	0.27	0.16	0.4	1.7
Relative Risk	4.14	0.25	0.16	0.39	1.74
Log Likelihood Ratio	4502.539459	1850.460782	653.958989	452.63033	313.006986
P-Value	$< 1\times10^{-17}$	$< 1\times10^{-17}$	$< 1\times10^{-17}$	$< 1\times10^{-17}$	$< 1\times10^{-17}$

Discussion

Incidence rate maps provide useful information to public health practitioners by visualizing where LD hotspots are within the state. The abnormally high IRs in Hunterdon, Sussex, and Warren Counties, and clustering around Warren County, point to high-risk areas health officials should focus their efforts to best address LD. The yearly variation observed in numerous counties may also point to other factors at play: Camden County's consistent IR increase, and Atlantic County's IR decrease, are two counties where patterns are ostensibly observed but do not necessarily show any meaningful results given the current scope of the methodology. LD variance is heterogeneous between and within counties, potentially indicating that a county-level analysis masks more granular spatial variation in LD emergence.

The space-time scan appears to support patterns seen from individual IR maps (a high cluster in the northwest, with low clusters in the northeast and for individual low-IR counties) but generally lacks meaningful spatiotemporal variation concerning adjacent counties, such as an initial LD cluster expanding outwards as with Li et al. (2014). As a whole, the geographic and temporal distance between clusters and the presence of multiple zero-radius clusters provide no visible patterns to link any of the clusters together meaningfully; they exist seemingly independent of each other without any overarching structure that connects them across space or time.

Conclusion and Recommendations

The value of the results of this project are mixed. While the IR table and maps may prove to be of some value to practitioners, that work is relatively straightforward. Undergoing more complex spatiotemporal analysis through SaTScan has not resulted in the display of significant LD variance patterns, an issue likely coming from lack of data. Counties are considerably large spatial units of analysis, and aggregating LD cases upwards into county centroids can introduce significant degradation of spatial accuracy when scoped only to a single state. Kugeler et al. (2015) also used county-level data, but for the entire Northeastern region encompassing numerous states, which shows variance better due to the inclusion of more data points. This may also account for the fact that tick environments are not bounded solely by county boundaries, and thus better represent the ecology on a larger scale. Similarly, Li et al.'s (2014) work with Virginia census tracts strongly shows how initial LD events in northern Virginia expand southward over time. The use of more spatially and/or temporally granular data (such as by census tract, or by month) leads to more data points, with higher dimensional accuracy, which may then lead to more accurate analysis. Temporally analyzing by month can also get at seasonal variations in LD (Grassly & Fraser, 2006); SaTScan has a "seasonal scan statistic" option that could be used for this purpose.

A few other data concerns also arise from my work. Li et al. (2014) note that smoothed maps provide more data continuity than raw rate maps, eliminating sparseness, and are therefore "more valuable in identifying trends in emergence" (p. 1169). Utilizing a kernel function as they did may help process LD case data into a more substantial form. Using larger time intervals may also prove useful in eliminating low-level data variation: for example, Kugeler et al. (2015) used 5-year intervals "to minimize the influence of travel-associated cases and short-term changes in surveillance practices" (p. 1455). However, expanding the scope of LD analysis beyond 2008 introduces a data quality issue. The CDC began systemic surveillance of LD in 1982 (Schmid et al., 1985) but did not introduce standardized reporting criteria for state health departments until 1991 (Centers for Disease Control and Prevention, n.d.a). Therefore, the use of data before 1990 must be heavily moderated, because they would be based on non-uniform reporting standards. Changing national classifications by the CDC in 1995, 1996, 2008, 2011, and 2017 (Centers for Disease Control and Prevention, n.d.b) may also introduce reporting inconsistency. While most of them are very minor, the 2008 definition change potentially altered case numbers. Before, a case could be confirmed simply by the presence of an erythema migrans rash, or an LD symptom with supplementary laboratory work (Li et al., 2014). Post-2008 case reporting, however, required that LD exposure/rashes must be accompanied by its emergence in an already endemic county. Accounting for these reporting changes remains a significant challenge in analyzing decadal LD variance. This discussion of cases must also not neglect underreporting present in the field: Waller et al. (2007) state that

> Surveillance data for Lyme disease can be problematic for a number of reasons: recognition by health-care providers of the signature symptoms over the study period, the associated potential lack of accurate diagnoses, imprecise serologic results, uneven case detection, reporting biases, and difficulty relating location of report to location of exposure. (p. 85)

For example, LD case underreporting in New York was attributed to laboratory follow-up testing on presumptive provider-diagnosed cases, and incomplete case information that does not meet the CDC case standards for inclusion (White et al., 2018). Szonyi et al. (2015) hypothesize that the contradictions between existing literature and their results may be resolved by understanding how surveillance case definition interpretations vary in different jurisdictions "due to unequal knowledge and practices of physicians, resulting in regional over- or under-reporting bias" (p. 5) and how cases are reported by patients' county of residence rather than the location where the disease was likely acquired. How to account for such issues remains an open question.

The use of different population numbers may also slightly change results. A linear interpolation assumes that all population-affecting factors in New Jersey remain unchanged over time; work by Weden et al. (2015) have pointed out the limits to such estimates. Using more complex models accounting for multiple factors may result in more accurate estimate county populations.

SaTScan itself does not escape scrutiny either. The choice of percentages as maximum cluster sizes is, in many ways, arbitrary and not based on statistical evidencing. Han et al. (2016) propose the use of the Gini coefficient as a superior window size selection method that can, among other things, "determine when it makes more sense to report a collection of smaller non-overlapping clusters versus a single large cluster containing all of them" (p. 1). The use of different distance structures aside from circles can introduce significant differences in the results as well, given that circles presume that each case is related to all nearby cases uniformly by distance only. LD spatiotemporal analysis through a space-time scan could also be achieved through other methods that engage with non-circular distance structures, potentially better illuminating emergence patterns by relying on different sets of assumptions based more upon statistical reasoning (Takahashi et al., 2008; Yao et al., 2011; Tango et al., 2011). For example, Duczmal et al. (2011) used Thiessen polygon buffers to optimize a minimum spanning tree algorithm for dengue fever in Brazil. Broader research that better interrogates the relationship between LD and land use may even necessitate a shift away from space-time scan as a viable method. For example, Stevenson (2019) used a zero-inflation gamma distribution model to map out the correlation between LD and land cover in northern Virginia, while Jackson (2005) explored LD and forest fragmentation in Maryland by bounding geographical units with road networks, accounting for actual physical boundaries limiting tick travel as opposed to political boundaries like counties which may not have real-world equivalents in the local ecologies.

At its core, this project relies on fundamentally symptomatic data. Incidence patterns are simply the symptom of various causes; describing what those patterns are helps others compare different potential causes to them, such as land use. Such analysis could provide further insight into the epidemiological patterns of Lyme disease, ultimately informing public health response efforts to improve health conditions of residents in high-risk areas. This exploratory research does not make any claims as to underlying factors influencing LD, but does contribute towards an understanding of existing LD patterns in the state of New Jersey, and furthers the understanding of SaTScan's strengths and limitations by bringing it to an new geography.

Bibliography

Literature

Borchers, A. T., Keen, C. L., Huntley, A. C., & Gershwin, M. E. (2015). Lyme disease: A rigorous review of diagnostic criteria and treatment. *Journal of Autoimmunity, 57*, 82–115. https://doi.org/10.1016/j.jaut.2014.09.004

Bunikis, J., Garpmo, U., Tsao, J., Berglund, J., Fish, D., & Barbour, A. G. (2004). Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents Borrelia burgdorferi in North America and Borrelia afzelii in Europe. *Microbiology, 150*(6), 1741–1755. https://doi.org/10.1099/mic.0.26944-0

Centers for Disease Control and Prevention. (2019a, January 10). *Symptoms of tickborne illness*. https://www.cdc.gov/ticks/symptoms.html

Centers for Disease Control and Prevention. (2019b, May 1). *Prevent Lyme disease*. https://www.cdc.gov/features/lymedisease

Centers for Disease Control and Prevention (2019c, November 8). *Post-treatment Lyme disease syndrome.* https://www.cdc.gov/lyme/postlds/index.html

Centers for Disease Control and Prevention (n.d.a). *Lyme disease surveillance and available data*. https://www.cdc.gov/lyme/stats/survfaq.html

Centers for Disease Control and Prevention (n.d.b). *Lyme disease (*Borrelia burgdorferi*).* https://wwwn.cdc.gov/nndss/conditions/lyme-disease/

Duczmal, L. H., Moreira, G. J., Burgarelli, D., Takahashi, R. H., Magalhães, F. C., & Bodevan, E. C. (2011). Voronoi distance based prospective space-time scans for point data sets: A dengue fever cluster analysis in a southeast Brazilian town. *International Journal of Health Geographics, 10*, 29. https://doi.org/10.1186/1476-072X-10-29

Egizi, A., Roegner, V., Faraji, A., Healy, S. P., Schulze, T. L., & Jordan, R. A. (2018). A historical snapshot of Ixodes scapularis-borne pathogens in New Jersey ticks reflects a changing disease landscape. *Ticks and Tick-Borne Diseases, 9*(2), 418–426. https://doi.org/10.1016/j.ttbdis.2017.12.009

Esri (n.d.). *How to: Find the centroid of polygons using Calculate Geometry*. https://support.esri.com/en/technical-article/000009381

Grassly, N. C., & Fraser, C. (2006). Seasonal infectious disease epidemiology. *Proceedings of the Royal Society B: Biological Sciences, 273*(1600), 2541–2550. https://doi.org/10.1098/rspb.2006.3604

Gubler, D. J., Reiter, P., Ebi, K. L., Yap, W., Nasci, R., & Patz, J. A. (2001). Climate variability and change in the United States: Potential impacts on vector- and rodent-borne diseases. *Environmental Health Perspectives, 109*(Supplement 2), 223–233.

Hoen, A. G., Margos, G., Bent, S. J., Diuk-Wasser, M. A., Barbour, A., Kurtenbach, K., & Fish, D. (2009). Phylogeography of Borrelia burgdorferi in the eastern United States reflects multiple independent Lyme disease emergence events. *Proceedings of the National Academy of Sciences of the United States of America, 106*(35), 15013–15018. https://doi.org/10.1073/pnas.0903810106

Hubálek, Z. (2005). North Atlantic weather oscillation and human infectious diseases in the Czech Republic, 1951–2003. *European Journal of Epidemiology, 20*(3), 263. https://doi.org/10.1007/s10654-004-6518-3

Jackson, L. E. (2005). *The relationship of land-cover pattern to Lyme disease* [Doctoral dissertation, University of North Carolina at Chapel Hill]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.508.8757&rep=rep1&type=pdf

Kitron, U., & Kazmierczak, J. J. (1997). Spatial analysis of the spatial distribution of Lyme disease in Wisconsin. *American Journal of Epidemiology*, *145*(6), 558–566. https://doi.org/10.1093/oxfordjournals.aje.a009145

Kugeler, K. J., Farley, G. M., Forrester, J. D., & Mead, P. S. (2015). Geographic distribution and expansion of human Lyme disease, United States. *Emerging Infectious Diseases*, *21*(8), 1455–1457. https://doi.org/10.3201/eid2108.141878

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

Kulldorff, M. (2018, March). SaTScan™ user guide for version 9.6. https://www.satscan.org/cgi-bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download

Lee, B. Y., Alfaro-Murillo, J. A., Parpia, A. S., Asti, L., Wedlock, P. T., Hotez, P. J., & Galvani, A. P. (2017). The potential economic burden of Zika in the continental United States. *PLOS Neglected Tropical Diseases*, *11*(4), e0005531. https://doi.org/10.1371/journal.pntd.0005531

Li, J., Kolivras, K. N., Hong, Y., Duan, Y., Seukep, S. E., Prisley, S. P., Campbell, J. B., & Gaines, D. N. (2014). Spatial and temporal emergence pattern of Lyme disease in Virginia. *The American Journal of Tropical Medicine and Hygiene*, *91*(6), 1166–1172. https://doi.org/10.4269/ajtmh.13-0733

National Institute of Allergy and Infectious Diseases (2018). *Lyme Disease Antibiotic Treatment Research*. https://www.niaid.nih.gov/diseases-conditions/lyme-disease-antibiotic-treatment-research

Mac, S., da Silva, S. R., & Sander, B. (2019). The economic burden of Lyme disease and the cost-effectiveness of Lyme disease interventions: A scoping review. *PLOS One*, *14*(1), e0210280. https://doi.org/10.1371/journal.pone.0210280

Meltzer, M. I., Dennis, D. T., & Orloski, K. A. (1999). The cost effectiveness of vaccinating against Lyme disease. *Emerging Infectious Diseases*, *5*(3), 321–328. https://doi.org/10.3201/eid0503.990302

Ogden, N. H., Lindsay, L. R., Morshed, M., Sockett, P. N., & Artsob, H. (2009). The emergence of Lyme disease in Canada. *CMAJ*, *180*(12), 1221–1224. https://doi.org/10.1503/cmaj.080148

Ozdenerol, E. (2015). GIS and Remote Sensing Use in the Exploration of Lyme Disease Epidemiology. *International Journal of Environmental Research and Public Health*, *12*(12), 15182–15203. https://doi.org/10.3390/ijerph121214971

Pepin, K. M., Eisen, R. J., Mead, P. S., Piesman, J., Fish, D., Hoen, A. G., Barbour, A. G., Hamer, S., & Diuk-Wasser, M. A. (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. *The American Journal of Tropical Medicine and Hygiene*, *86*(6), 1062–1071. https://doi.org/10.4269/ajtmh.2012.11-0630

Schmid, G. P., Horsley, R., Steere, A. C., Hanrahan, J. P., Davis, J. P., Bowen, G. S., Osterholm, M. T., Weisfeld, J. S., Hightower, A. W., & Broome, C. V. (1985). Surveillance

of Lyme disease in the United States, 1982. *The Journal of Infectious Diseases*, *151*(6), 1144–1149. https://doi.org/10.1093/infdis/151.6.1144

Staples, J. E., Shankar, M. B., Sejvar, J. J., Meltzer, M. I., & Fischer, M. (2014). Initial and long-term costs of patients hospitalized with West Nile Virus disease. *The American Journal of Tropical Medicine and Hygiene*, *90*(3), 402–409. https://doi.org/10.4269/ajtmh.13-0206

Steere, A. C., Malawista, S. E., Snydman, D. R., Shope, R. E., Andiman, W. A., Ross, M. R., & Steele, F. M. (1977). An epidemic of oligoarticular arthritis in children and adults in three connecticut communities. *Arthritis & Rheumatism*, *20*(1), 7–17. https://doi.org/10.1002/art.1780200102

Stevenson, M. N. (2019). *The Effects of Land Cover on the Spatial Distribution of Lyme Disease in Northern Virginia Since 2005* [Masters dissertation, Virginia Polytechnic Institute and State University]. https://vtechworks.lib.vt.edu/handle/10919/89374

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*(1), 924. https://doi.org/10.1186/s12889-015-2286-0

Takahashi, K., Kulldorff, M., Tango, T., & Yih, K. (2008). A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics*, *7*(1), 14. https://doi.org/10.1186/1476-072X-7-14

Tango, T., Takahashi, K., & Kohriyama, K. (2011). A space–time scan statistic for detecting emerging outbreaks. *Biometrics*, *67*(1), 106–115. https://doi.org/10.1111/j.1541-0420.2010.01412.x

Waller, L. A., Goodwin, B. J., Wilson, M. L., Ostfeld, R. S., Marshall, S. L., & Hayes, E. B. (2007). Spatio-temporal patterns in county-level incidence and reporting of Lyme disease in the northeastern United States, 1990–2000. *Environmental and Ecological Statistics*, *14*(1), 83. https://doi.org/10.1007/s10651-006-0002-z

Weden, M. M., Peterson, C. E., Miles, J. N., & Shih, R. A. (2015). Evaluating linearly interpolated intercensal estimates of demographic and socioeconomic characteristics of U.S. counties and census tracts 2001–2009. *Population Research and Policy Review*, *34*(4), 541–559. https://doi.org/10.1007/s11113-015-9359-8

White, J., Noonan-Toly, C., Lukacik, G., Thomas, N., Hinckley, A., Hook, S., & Backenson, P. B. (2018). Lyme disease surveillance in New York State: An assessment of case underreporting. *Zoonoses and Public Health*, *65*(2), 238–246. https://doi.org/10.1111/zph.12307

WHO | Lyme Borreliosis (Lyme disease). (n.d.). WHO; World Health Organization. Retrieved April 29, 2020, from https://www.who.int/ith/diseases/lyme/en/

Willadsen, P. (2006). Tick control: Thoughts on a research agenda. *Veterinary Parasitology*, *138*(1), 161–168. https://doi.org/10.1016/j.vetpar.2006.01.050

Wood, C. L., & Lafferty, K. D. (2013). Biodiversity and disease: A synthesis of ecological perspectives on Lyme disease transmission. *Trends in Ecology & Evolution*, *28*(4), 239–247. https://doi.org/10.1016/j.tree.2012.10.011

Yao, Z., Tang, J., & Zhan, Fb. (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. https://doi.org/10.1186/1476-072X-10-23

Zeman, P., & Benes, C. (2014). Peri-urbanisation, counter-urbanisation, and an extension of residential exposure to ticks: A clue to the trends in Lyme borreliosis incidence in the Czech Republic? *Ticks and Tick-Borne Diseases, 5*(6), 907–916. https://doi.org/10.1016/j.ttbdis.2014.07.006

Zhang, X., Meltzer, M. I., Peña, C. A., Hopkins, A. B., Wroth, L., & Fix, A. D. (2006). Economic impact of Lyme disease. *Emerging Infectious Diseases, 12*(4), 653–660. https://doi.org/10.3201/eid1204.050602

Data

United States Census Bureau (2000). Total population [Comma Separated Values]. https://data.census.gov/cedsci/table?g=0400000US34.050000&t=Populations%20and%20People&tid=DECENNIALSF12000.P001&hidePreview=true&vintage=2000&layer=VT_2018_050_00_PY_D1&cid=S0101_C01_001E&y=2000&d=DEC%20Summary%20File%201

United States Census Bureau (2010). Total population [Comma Separated Values]. https://data.census.gov/cedsci/table?g=0400000US34.050000&t=Populations%20and%20People&tid=DECENNIALSF12010.P1&hidePreview=false&vintage=2010&layer=VT_2018_050_00_PY_D1&cid=S0101_C01_001E&d=DEC%20Summary%20File%201

Centers for Disease Control and Prevention (n.d.c). Lyme disease surveillance and available data. [Comma Separated Values]. https://www.cdc.gov/lyme/resources/LD-Case-Counts-by-County-00-18.csv

New Jersey Office of Information Technology (2020). County boundaries of NJ [Shapefile]. https://njogis-newjersey.opendata.arcgis.com/datasets/5f45e1ece6e14ef5866974a7b57d3b95_1

Appendix A

A table of the yearly population change estimates, per New Jersey county.

County	Yearly Change in Population
Atlantic	2199.7
Bergen	2099.8
Burlington	2534
Camden	472.5
Cape	-506.1
Cumberland	1046
Essex	-966.4
Gloucester	3361.5
Hudson	2529.1
Hunterdon	636
Mercer	1575.2
Middlesex	5969.6
Monmouth	1507.9
Morris	2206.4
Ocean	6565.1
Passaic	1217.7
Salem	179.8
Somerset	2595.4
Sussex	509.9
Union	1395.8
Warren	625.5

Appendix B

A map of New Jersey counties, with each county labeled by name.