The association between the incidence of Lyme disease in the USA and indicators of greenness and land cover

Sydney Westra [a,*], Mark S. Goldberg [a,b], Kamel Didan [c]

[a] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
[b] Department of Medicine, McGill University, Montreal, Canada
[c] Department of Biosystems Engineering, Remote Sensing / Spatial Analysis – GIDP Program, University of Arizona, Tucson, AZ, USA

ARTICLE INFO

ABSTRACT

Lyme disease (LD) is the most common vector-borne illness in the USA. Incidence is related to specific environmental conditions such as temperature, metrics of land cover, and vertebrate species diversity. To determine whether greenness, as measured by the Normalized Difference Vegetation Index (NDVI), and other selected indices of land cover were associated with the incidence of LD in the northeastern USA for the years 2000–2018, we conducted an ecological analysis of incidence rates of LD in counties of 15 "high" incidence states and the District of Columbia for 2000–2018. Annual counts of LD by county were obtained from the US Centers for Disease Control and values of NDVI were acquired from the Moderate Resolution Imaging Spectroradiometer instrument aboard Terra and Aqua Satellites. County-specific values of human population density, area of land and water were obtained from the US Census. Using quasi-Poisson regression, multivariable associations were estimated between the incidence of LD, NDVI, land cover variables, human population density, and calendar year. We found that LD incidence increased by 7.1% per year (95% confidence interval: 6.8–8.2%). Land cover variables showed complex non-linear associations with incidence: average county-specific NDVI showed a "u-shaped" association, the standard deviation of NDVI showed a monotonic upward relationship, population density showed a decreasing trend, areas of land and water showed "n-shaped" relationships. We found an interaction between average and standard deviation of NDVI, with the highest average NDVI category; increased standard deviation of NDVI showed the greatest increase in rates. These associations cannot be interpreted as causal but indicate that certain patterns of land cover may have the potential to increase exposure to infected ticks and thereby may contribute indirectly to increased rates of LD. Public health interventions could make use of these results in informing people where risks may be high.

1. Introduction

The US Centers for Disease Control and Prevention (CDC) reports that there are more than 300,000 cases of Lyme disease (LD) each year in the USA, making it the most prevalent vector-borne disease (CDC, 2019). LD is transmitted through the bite of a tick that has been infected with *Borrelia burgdorferi* spirochaetes (CDC, 2019). In the northeastern USA, *Ixodes scapularis*, commonly referred to as the black-legged tick, is the most abundant vector for LD (CDC, 2019).

Ticks become infected with *Borrelia burgdorferi* by feeding on an organism infected with this spirochaete. Small rodents are the primary sources of infection, with the white-footed mouse being the principal reservoir in eastern North America (Allan et al., 2003). Other species of

mammals and birds may also act as reservoirs, such as eastern chipmunks (*Tamias striatus*) and American robins (*Turdus migratorius*) (Mather, 1993). An infected tick life stage can then transmit the bacterium in its next blood meal. LD is seasonal, with high rates of infections occurring in spring and summer, especially during the first week of July (Schwartz et al., 2017). The different tick life stages (larvae, nymphs, adults) are active in specific temperature ranges (Alonso-Carné et al., 2015).

One would expect on average that the trend in the abundance of ticks, and thus the incidence of LD, should decrease from high (e.g. urban or built environments) to lower human population density areas in rural and semi-rural areas because ticks are often found in wooded and brushy areas where their main food sources live (such as deer, rabbits, birds) (Jobe et al., 2007; Hamer et al., 2012; VanAcker et al., 2019). Those areas

---

Table 1
Distribution of county-specific values of counts and rates of Lyme disease, NDVI and selected other variables, 2000–2018 (580 counties, 15 states, 11,020 observations).

Variable	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	Standard deviation
Number of cases of Lyme disease	0	1	6	46.1	40	1720	104.1
Rates of Lyme disease (per 100,000 population)	0	< 0.1	11.6	40.2	48.4	1581.2	73.2
Total human population	2196	20,368	44,324	141,210	122,142	2,611,232	271,081.4
Human population density (per km^2)	1.0	14.9	32.0	290.9	117.3	27,681.8	1480.5
Average NDVI	0.22	0.75	0.81	0.78	0.85	0.92	0.10
Standard deviation of NDVI	< 0.01	0.03	0.05	0.06	0.08	0.33	0.04
Total land area (km^2)	19.4	914.6	1353.5	1652.0	2021.1	17,278.0	1515.2
Area of water (km^2)	0.05	9.8	31.2	197.7	98.9	4890.6	520.6
Rural area (km^2)	0	792.0	1231.9	1,537.0	1881.4	17,253.8	1534.6

Notes: The following eight counties were excluded from the variance, standard deviation, 25%, 50%, 75% of NDVI: Charlottesville City, VA; Colonial Heights City, VA; Frankline City, VA; Fredericksburg City, VA; Galax City, VA; Hopewell City, VA; Martinsville City, VA; and Winchester City, VA.
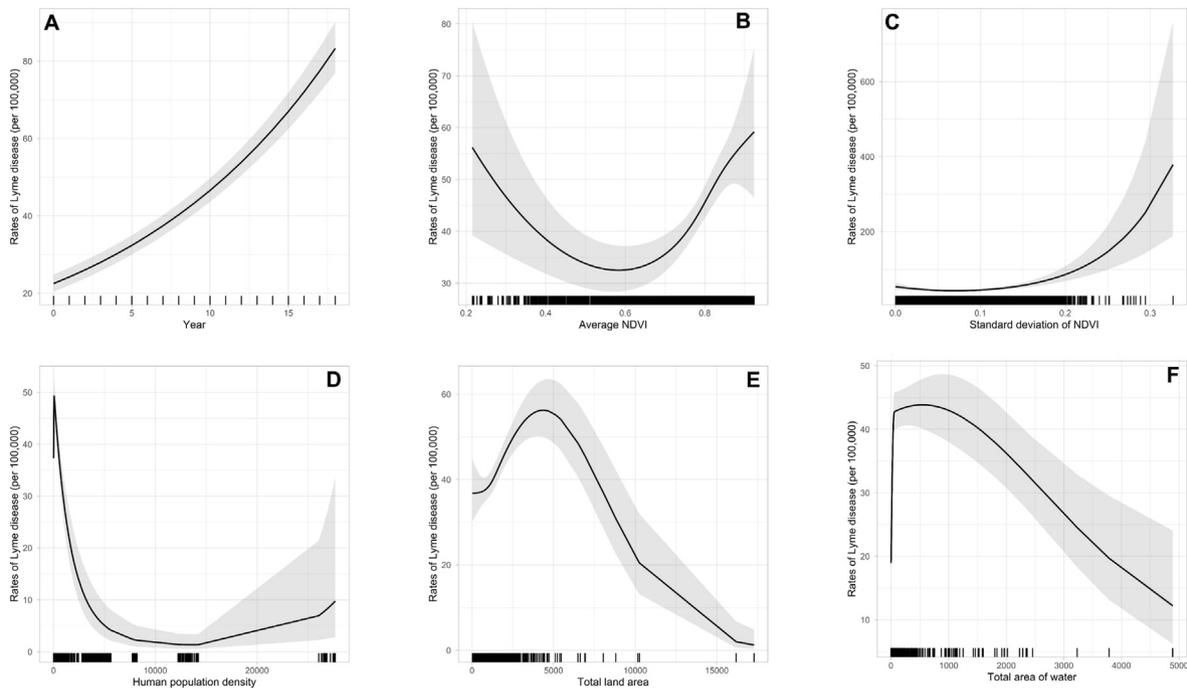

Fig. 1. Predicted rates of Lyme disease from a quasi-Poisson regression model that included year as a linear term, and natural cubic splines for average NDVI, the standard deviation of NDVI, population density, area of water, and total area of the county for the years 2000–2018, including 580 counties in 15 states (11,020 observations). Three degrees of freedom were used for all variables except for standard deviation of NDVI ($df = 2$). The solid line represents the predicted value and the grey area surrounding the line represents the pointwise 95% confidence interval. The rug plot (tick marks) on the x-axis represents the location of the data points. **A** Response function for year. **B** Response function for average NDVI. **C** Response function for standard deviation of NDVI. **D** Human population density. **E** Total land area of the county. **F** Total area of water.
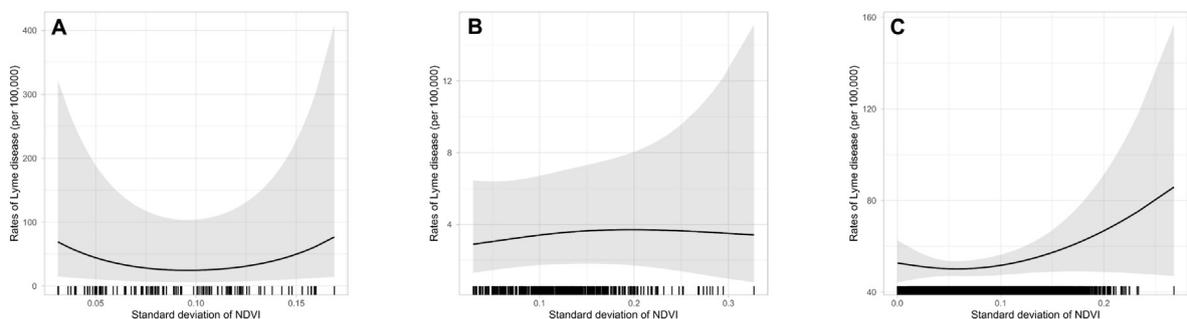

Fig. 2. Interaction analysis of standard deviation of NDVI and rates of Lyme disease. Interaction between average NDVI and standard deviation of NDVI adjusted for other covariables. Average NDVI was split into 3 categories. Category 1 includes average NDVI from 0 to 0.4. Category 2 includes average NDVI from 0.4 to 0.6. Category 3 includes average NDVI from 0.6 to 1.0. For each category of average NDVI, the fitted response curves of the interaction analysis are shown. **A** Lyme disease rates and standard deviation of NDVI. Category 1 (0–0.4). **B** Lyme disease rates and standard deviation of NDVI. Category 2 (0.4–0.6). **C** Lyme disease rates and standard deviation of NDVI. Category 3 (0.6–1.0).
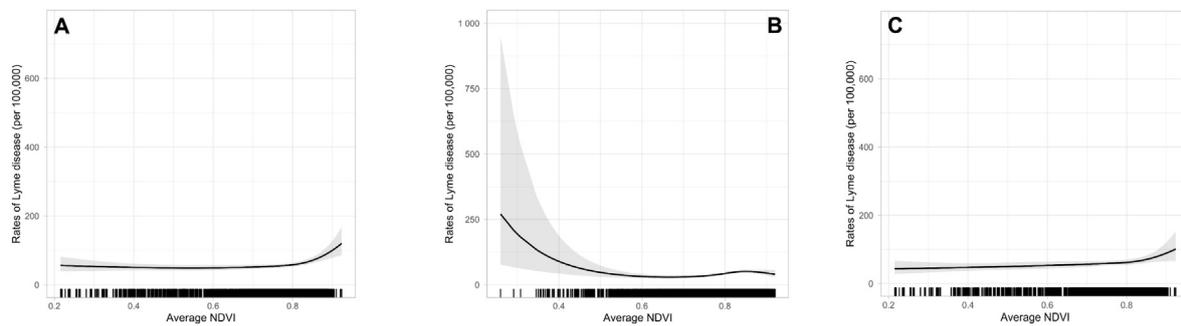
Fig. 3. Subgroup analyses of the associations between rates of Lyme disease and average NDVI and according to calendar year in which rates began to become underestimated (2013) as well as in different states. **A** Lyme disease rates and average NDVI for the years 2000–2012. **B** Lyme disease rates and average NDVI for the years 2013–2018. **C** Lyme disease rates and average NDVI for the years 2013–2018 but excluding New York State, Maryland, Minnesota, and Connecticut.

with a mixture of habitats that include forest, various types of open and mixed areas, as well as wildland-urban interfaces should have lower tick densities (Diuk-Wasser et al., 2021). Wood and Lafferty (2013) suggested that biodiversity of vertebrates played an important role in infections when comparing rates in urban, semi-urban, to rural areas but were not convinced of the effect of biodiversity within forests. They also suggested non-linear associations between the incidence of LD and land cover whereby the incidence of LD increased with the degree of forest cover until a certain level of biodiversity becomes protective ("n-shaped curve").

Metrics of land cover may be interpreted as indirect measures of the density of infected ticks (Kotchi et al., 2021) but may also represent the potential for human exposures to infected ticks, and so it is plausible that potential exposures would increase with increasing density of green landscapes including forests (Wood and Lafferty, 2013; Kilpatrick et al., 2017; Diuk-Wasser et al., 2021). Infections may also occur in urbanites when they travel to areas of higher densities of infected ticks, but infections can also occur in urban areas (Jobe et al., 2007; Hamer et al., 2012; VanAcker et al., 2019).

Remote sensing has been proposed as a tool for assessing vegetation cover (Gabriele-Rivet et al., 2017). In particular, the Normalized Difference Vegetation Index (NDVI) (Huete et al., 2002; Tucker et al., 2005) provides a measurement of vegetation canopy "greenness" (Myneni et al., 1995), as it measures a combination of leaf chlorophyll, leaf area, canopy cover, and structure. This is quantified through the ratio of the absorption of red light (R) by photosynthesis and reflectance of near-infrared by leaves (Tucker, 1979). NDVI is calculated as (NIR − R)/(NIR + R). Values of NDVI range from -1.0 to +1.0, where very low and negative values indicate an absence of vegetation, as found in areas of barren soils, snow, or water, and built environments in urban areas, and high values correspond to dense vegetation, as found in temperate or tropical forests (Trishchenko et al., 2002). A study using remotely sensed greenness and wetness in Westchester County that compared two communities in one county of New York State (Dister et al., 1997) reported higher incidence rates of LD in greener and wetter residential properties. With the considerations above, our objective was to determine whether the incidence of LD in the northeastern part of the USA was associated with NDVI and other metrics of land cover.

2. Materials and methods

2.1. Incidence of Lyme disease

Counts of new cases of LD by county were obtained from the Centers for Disease Control for the years 2000–2018 (CDC, 2019). Surveillance of LD was established in the early 1990s, but we used counts starting in 2000 because this was the first year of available remote sensing data from MODIS (Justice et al., 2002). Since 1991, local and state health departments are required to report confirmed cases to the CDC (Schwartz

et al., 2017).

Over 95% of all LD cases in the USA occur in the 15 states classified as having high incidence rates (at least 10 confirmed cases per 100,000 persons for the prior three reporting years) (CDC, 2019). We thus included in our analyses Connecticut, Delaware, Maine, Maryland, Massachusetts, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Virginia, West Virginia, Wisconsin, and the District of Columbia.

2.2. Normalized Difference Vegetation Index (NDVI)

We used high fidelity preprocessed NDVI time series available at the Climate Modelling Grid at a resolution of 0.05 degrees (Didan and Barreto, 2016a). NDVI data covering the period 2000–2016 was acquired from the Land Processes Data Active Archive Center (Didan and Barreto, 2016a) from the Moderate Resolution Imaging Spectroradiometer (Shu et al., 2020). MODIS is an instrument on board both Terra and Aqua satellites, operated by NASA since 2000 (Justice et al., 1998). MODIS acquires quasi-daily multispectral surface reflectance images of the Earth at 250 m, 500 m, and 1 km resolutions.

The full NDVI time series is a 35-year-long multi-sensor NDVI record at 5.6 km resolution (Climate Modelling Grid derived from the remote sensing instruments AVHRR (1981–1999), SPOT (1998–2002) and MODIS (2000–2016) (Didan and Barreto, 2016b). This high-fidelity time series is free from spatial gaps that result from clouds and atmospheric contaminants, like heavy aerosols. We obtained the NDVI time series from the Vegetation Index and Phenology Laboratory at the University of Arizona.

This time series was generated from the finer 250 m resolution MODIS daily observations. Surface reflectance observations were calibrated, corrected for atmosphere attenuation, aerosols, and screened for clouds (Justice et al., 1998). The daily surface reflectance were then filtered, refined and composited into multiday value-added measures of vegetation greenness (Huete et al., 2002, 2010), such as NDVI, and any remaining spatial gaps resulting from cloud screening were filled-in using a simple moving window regression or with long-term average data if the gaps were persistent for long periods of time (Didan and Barreto, 2016b). These resulting high fidelity data records are used for long-term vegetation-climate interactions, detecting changes, modelling, and other derivative studies like public health (Tourre et al., 2008; Jamison et al., 2015).

The resulting NDVI images provide a spatially explicit measure of vegetation canopy greenness, health, and productivity over space and time (Myneni and Williams, 1994) and serves as a proxy for many structural and functional ecosystem parameters (Krishnaswamy et al., 2009).

We used NDVI data only for the month of July as it corresponds to the time of high ecosystem productivity and peak incidence of LD (Schwartz et al., 2017). For July of each year, we spatially aggregated the pixel data

at the county level and generated multiple statistical parameters that represent the distribution of NDVI within the county (Supplementary file 1: Fig. S1 shows the geographical distribution in the continental USA of maximum NDVI for July 2000.)

Because the NDVI data record ends at 2016 and the CDC LD data end in 2018, we extended the NDVI record through years 2017 and 2018 using a simple regression model for average and the standard deviation of NDVI. This latter statistic may be interpreted as indicating fragmented areas of greenness within a county. This prediction model uses a county-specific linear regression in which we regressed the natural logarithm of NDVI *versus* time to generate predicted county-specific values for the missing two years.

2.3. County-specific characteristics

Human population counts at the county level were obtained from the US Census (USA Census Bureau, 2020) downloaded as county populations for the years 2000–2010 and 2010–2018. The population for the year 2010 was used from the county population for the 2010–2018 dataset, as new estimates incorporated the most up-to-date data for this year. In addition, we obtained from the census annual estimates of total land area, area of water, and we computed population density (USA Census Bureau, 2020).

2.4. Statistical analysis

Counts of LD were not characterized by any personal characteristics, such as age and sex. As the outcome data are counts (or equivalently rates), the statistical model should follow approximately a Poisson distribution (Cameron and Trivedi, 2013). We thus used generalized linear models (McCullagh and Nelder, 1989) in which the natural logarithm of rates (or counts with a fixed offset for the logarithm of the population) was regressed against a linear combination of covariables. The counts were over-dispersed, so we used quasi-likelihood Poisson and negative binomial models to account for extra Poisson variability.

The variables included in the models were calendar year, human population density, total land area, area of water, average NDVI, and the standard deviation of NDVI. These variables were computed for each year from 2000 to 2018. Weather variables which vary on short time scales, such as days, were not included because we only had annual counts of LD.

We made use of natural cubic splines to model the functional form of each covariable. These smoothers have a tunable parameter (number of knots or degrees of freedom, *df*) that allows for a different level of smoothness, with a higher number of knots or degrees of freedom allowing for more variability (less smoothing). We used the following algorithm to select the degrees of freedom: we modelled each covariable separately, adjusting for calendar year, and investigated the pattern of the graphs from 2 *df* to 6 *df*. The final value for the number of *df* was assessed visually by comparing the graphs and selecting the lowest *df* that revealed the pattern without having unnecessary variability that would likely be due to random fluctuations. We then modelled all of the covariables together; we used these multivariable models as the main results of the analysis, where we present graphs of marginal effects (Lüdecke, 2018). As well, because the patterns were complex, to provide quantitative estimates of rates, we computed rate ratios between selected cut points for each covariable, adjusting for all of the other covariables (Cao et al., 2006).

2.5. Sensitivity analyses

We also conducted additional analyses using various cut points of the human population of each county, and we conducted separate analyses of each of the 14 states (the District of Columbia was excluded because it comprises only one area).

We investigated interactions between selected variables, and we assessed these interactions using the likelihood ratio test as well as through fitted marginal graphs. As it is exceedingly difficult to graph the interaction of two nonlinear functions, we categorized one variable and then fitted the fully adjusted model on the subset of data points defined by the different values of the categorical variable.

In addition, because there are reports of under-reporting of LD in many states (Schwartz et al., 2017; Rutz et al., 2018; Schiffman et al., 2018; White et al., 2018) we conducted a number of additional analyses excluding certain states as well as certain years.

3. Results

3.1. Main analyses

Table 1 shows the distribution of the variables included in the analysis (2000–2018 for the selected states). The area of some counties was too small to measure the variability of NDVI, and thus these counties were excluded from the substantive analyses (193 observations).

The counties included in our analyses varied dramatically in area and population density and this type of heterogeneity is a cardinal characteristic of this administrative unit. The ratio of the population size between the largest to the smallest county was 1200 and that of the total land area was 78. Rates of LD (per 100,000 persons) varied from 0 to 1581.2 with a mean of 40.2 and standard deviation of 0.73. Counts of LD did not follow a Poisson distribution: the mean number of incident cases was 46.1 (range of 0–1720) and the variance was 10,837 (ratio of variance to mean of 235), implying considerable overdispersion (Table 1).

Average county-specific NDVI also varied considerably, from a very low value of 0.22 to a maximum of 0.92 (dense forest cover). There was also considerable variability in NDVI across counties, as shown by the distribution of its standard deviation. The predictions of NDVI to 2017 and 2018 were similar to the observed values for 2000–2016 (Supplementary file 1: Fig. S2).

We present results for the quasi-Poisson model, as the results with the negative binomial model were similar (see Supplementary file 1: Fig. S3). The fitted response curves (Lüdecke, 2018) for the model that included all variables are shown in Fig. 1. The estimate of the dispersion parameter in the quasi-Poisson model was 0.0012, close to the crude value. Calendar year was modelled using a linear function, the standard deviation of NDVI was modelled using 2 *df*, and all other variables were modelled using 3 *df*.

The predicted response function for calendar year, adjusted for all the other variables in the model, is shown in Fig. 1A. In this plot, each tick on the rug plot represents a value for year, with each county contributing 19 values to the analyses. Although we fitted a linear term for year, the reason the graph shows some curvature is because the predicted values were computed by taking their exponent. The estimated annual increase in rates was 7.1% (95% CI: 6.8–8.2%).

The marginal response patterns for average NDVI are shown in Fig. 1B where we found a non-monotonic function for rates of LD whereby adjusted rates decreased between about 0.2 to about 0.6 and then increased from 0.6 to 1.0. The rate ratio for an increase of average NDVI from 0.2 to 0.6 was 0.56 (95% CI: 0.38–0.82), showing that higher average NDVI values conferred lower rates of LD and the rate ratio from 0.6 to 0.85 was 1.60 (95% CI: 1.37–1.88), showing a 60% increase of rates of LD for this increment of average NDVI.

For the standard deviation of NDVI (Fig. 1C), we found a monotonic increase in the rate ratio, with a modest change from 0 to 0.2 and then increasing more dramatically, although the confidence intervals were wide at the upper end. Namely, for greater variability in NDVI, related partly to land fragmentation from cities, water, roads, etc., the rate ratio between a value of 0 to 0.2 was 1.61 (95% CI: 1.25–2.07) and from 0.2 to 0.3 the rate ratio was 3.12 (95% CI: 2.14–4.56).

The predicted marginal relationship for human population density is shown in Fig. 1D. Most counties had population densities under 10,000 per km^2 with three clusters at higher values representing larger urban centers. These three clusters represented three counties and the different

population densities in each cluster represented values for different years, hence the 95% confidence bands were wide. The upward line at zero is an artifact of seven counties having very low population densities (< 2 per km^2). Rates decreased up to a population density of about 12,000 and then increased slowly. The rate ratio from a population density of 500 persons per km^2 up to 10,000 was 0.04 (95% CI: 0.02–0.10) and from 10,000 to 30,000 the rate ratio was 9.49 (95% CI: 1.84–49.02). Counties with the highest population densities such as New York County or Philadelphia County led to considerable uncertainty at the high end.

The marginal response functions for total land area and area of water, respectively, are shown in Fig. 1E, F. Fifteen counties had a total area greater than about 5000 km^2 and 35 counties had areas of water greater than 1000 km^2. For total land area, we found another non-monotonic function increasing at the low values of the land area until about 5000 km^2 and then decreasing. The rate ratio from a minimum of 20 until 4500 km^2 was 1.53 (95% CI: 1.22–1.91) and from 4500 to 10,000 km^2 the rate ratio was 0.39 (95% CI: 0.26–0.58).

For area of water, rates increased at lower values and then dropped off rapidly; for example, from 10 to 100 km^2 the rate ratio was 1.69 (95% CI: 1.56–1.83) and from 100 to 2000 km^2 the rate ratio was 0.84 (95% CI: 0.72–0.99). Although there are few observations for larger areas of water, the decreasing trend covers the entire range.

3.2. Sensitivity analyses

We investigated the interaction between average NDVI and the standard deviation of NDVI. The likelihood ratio test adding in the interaction term between these two variables indicated an important synergistic effect. We then categorized average NDVI into three groups (0–0.4; > 0.4–0.6; and > 0.6–1.0) and we estimated the adjusted response curves for the standard deviation of NDVI for each separate category of average NDVI (Fig. 2). Although we found a "u-shaped" response in rates of LD with increasing standard deviation of NDVI in the lowest category of average NDVI (0 to 0.4), the wide confidence intervals do not admit any meaningful interpretation. As shown in Fig. 2, rates for LD were constant in the second category of average NDVI (> 0.4–0.6) and increased marginally in the third category of average NDVI (> 0.6–1.0).

We excluded counties with populations greater than 100,000 and populations greater than 381,000 but they exhibited functional forms and rate ratios comparable to the main analysis (Supplementary file 1: Fig. S4).

State-specific analyses for the two indices of NDVI are shown in Supplementary file 1: Figs. S5 and S6. The patterns for NDVI differ between states and many states exhibit different patterns from the average shown above. The state-specific patterns are more difficult to interpret as there are few counties in each state, leading to considerable variability in these associations.

We conducted a series of separate analyses for different time periods in which the definitions used for LD changed, namely 2000–2008, 2008–2011, and 2011–2018. Because of the fewer numbers of observations, variability was increased but the general patterns found in the main analyses were similar (results not shown).

We also conducted analyses splitting the time period according to when public health departments began to experience difficulties in recording cases (around 2013) (Rutz et al., 2018; Schiffman et al., 2018; White et al., 2018). The response patterns for average NDVI split between the time periods 2000–2012 and 2013–2018 are shown in Fig. 3. For the period prior to 2013, we found a flat response until around average NDVI of 0.8 (Fig. 3A) whereas for the latter period we found higher rates below values of 0.4 and then subsequently a flat response for larger values (Fig. 3B). We hypothesized that the change in pattern observed in this latter analysis may have been due to serious underreporting in certain states (New York, Maryland, Minnesota and Connecticut) (Schwartz et al., 2017; Lukacik et al., 2018) so we excluded these states (Fig. 3C) and indeed recovered the pattern seen in Fig. 3A for the earlier period

(2000–2012).

4. Discussion

4.1. Findings

We found that the reported incidence of LD increased by 7.1% per year for the period 2000–2018, consistent with other reports (Murphree Bacon et al., 2008; Kugeler et al., 2015; Schwartz et al., 2017).

The definition of LD changed in the years 2008 and 2011. Before 2008, a confirmed case of LD was "a case with erythema migrans or a case with at least one late manifestation that is laboratory confirmed" (CDC, 2019). From 2008 to 2011, a confirmed case was defined as "a case of erythema migrans with a known exposure, or a case of erythema migrans with laboratory evidence of infection and without a known exposure or a case with at least one late manifestation that has laboratory evidence of infection" (CDC, 2019). From 2011 to 2017, the definition of a confirmed case remained the same, but the laboratory criteria for diagnosis became more specific as the presence of antibodies for *Borrelia burgdorferi* were used (CDC, 2019). Although these modifications were minor, these changes may have influenced the number of reported cases.

Part of the secular increase was due to improvements in reporting but also likely reflects increased rates of infection (Murphree Bacon et al., 2008; Kugeler et al., 2015). There has been under-reporting (Schwartz et al., 2017; Rutz et al., 2018; Schiffman et al., 2018; White et al., 2018) because public health departments could not deal with the volume of cases, healthcare providers could not order tests to detect LD and, thus, many cases were missed because the complete clinical information was not available. For example, White et al. (2018) found that over a three-year period, three counties in New York State failed to report 20% of LD cases through standard surveillance. Our reported response functions could be biased if under-reporting was also associated with the variables in the model. The sensitivity analyses shown in Fig. 3 indicated that this was the case, especially after 2013, but that the overall pattern for average NDVI that we found (Fig. 1) likely is correct.

We found complex relationships in the functional forms of rates of LD and various indicators of greenness and land cover. We found that for increasing values of average NDVI from 0 to 0.6 lower rates of LD were observed at higher values of average NDVI, and for increasing values of average NDVI from 0.6 to 1.0 we found increases in disease rates. Low values of average NDVI, from 0.2 to 0.3 reflect areas in which there is relatively sparse vegetation (Weier and Herring, 2000). This may include areas such as urban centers, unvegetated areas, or sparsely vegetated meadows. For example, our data showed the average NDVI value for Kings County, NY, which comprises Brooklyn, NY, had an average NDVI value of 0.25 in 2002 which is comparable, for example, to desert-like regions.

Low values of average NDVI should imply that vegetation is largely absent, which should in theory translate to an absence of habitats for ticks and the vertebrate species that carry ticks. Indeed, living in urbanized areas is sometimes thought to be protective of LD (Glass et al., 1995), but our model found the highest rates of LD in areas with little vegetation and areas with dense vegetation. Higher rates of LD in areas of little vegetation may be partly the result of travel-related behaviour, such as when a person acquires LD elsewhere, and the infection is reported in their urban/residential county. However, studies have shown that most cases of LD are acquired around people's homes, such as in one's lawn or backyard (Falco and Fish, 1988; Maupin et al., 1991; Smith et al., 2001). Furthermore, our sensitivity analysis suggests that after excluding large cities we observed the same functional form, where the highest rates of LD were associated with the lowest and highest levels of average NDVI (Supplementary file 1: Fig. S4B). This suggests that travel-related behaviour may not fully explain the relationship between higher rates of LD in areas of little vegetation.

The incidence of LD in the USA cities is understudied and is likely more prevalent in cities than thought previously (Jobe et al., 2007;

Hamer et al., 2012; VanAcker et al., 2019). VanAcker et al. (2019) found that 17 out of 24 parks in New York City contained the tick *Ixodes scapularis*, and of all the ticks sampled, an average of 26.6% were infected with the LD spirochaete. Cities contain a high density of people, making it especially important to be aware of the expansion of this zoonotic disease into cities. Additionally, lower to moderate levels of NDVI averaged in a county reflect a mixture of residential/urban areas and a small amount of forested area, such as the edge of forests. Many of these areas may be in proximity to forest edge habitats that may increase the interaction between disease vectors and humans and thus increase risk (Brownstein et al., 2005; Jackson et al., 2006a).

The diversity of vertebrate species (DeLong, 1996) may affect the ecology and epidemiology of vector-borne zoonoses through a process called the "dilution effect" (Ostfeld and Keesing, 2000) which posits that a greater variety of species dilutes the proportion of infected rodents for ticks to feed on. A loss of biodiversity can indirectly increase risk of disease, as species carrying the LD spirochaete are predominantly small rodents that are better able to survive and adapt to loss and change of habitats. Although biodiversity is difficult to measure across large spatial scales, indirect measures that relate to land cover have been used (Bawa et al., 2002; Grantham et al., 2008). The hypothesized dilution effect implies that the greater diversity of species in communities may dilute the number of infected white-footed mice and as a result reduce the incidence of LD (Ostfeld and Keesing, 2000). Habitat fragmentation and destruction of forests can be an important driving force in the loss of many species in a community (LoGiudice et al., 2003). The dilution effect has been referenced (LoGiudice et al., 2003) as a means in which biodiversity can ameliorate LD risk, but in few studies using direct metrics of biodiversity has this hypothesis been studied or confirmed (Wood and Lafferty, 2013).

As average NDVI increased in the range 0.6–0.8 (which corresponds to highly vegetated counties), higher rates of LD were observed. This finding is compatible with various studies in the northeastern United States which suggest that the risk of LD increases with increasing forestation (Glass et al., 1995; Kitron and Kazmierczak, 1997; Brownstein et al., 2005). A review on spatial patterns and environmental correlates of human cases of LD by Killilea et al. (2008) found that the only environmental variable consistently associated with increased LD risk was the presence of forests. However, possible non-linear relationships between forest coverage and LD were not assessed in the studies included in this review (Kitron and Kazmierczak, 1997; Orioski et al., 1998). Our model is consistent with previous findings where higher rates of LD were associated with greater forestation (Glass et al., 1995; Kitron and Kazmierczak, 1997; Brownstein et al., 2005).

We also found increased rates of LD with higher values of the standard deviation of NDVI. Higher variability in NDVI reflects fragmentation of landscapes and/or the presence of large urban to natural interfaces. Landscape likely contributes to the incidence of LD through the reservoir species it supports, namely the white-footed mouse (Falco and Fish, 1988; Brownstein et al., 2005). White-footed mice can tolerate fragmented habitats while many other species cannot (LoGiudice et al., 2003). This is a possible explanation for higher disease rates in areas with a higher standard deviation of NDVI. As it has been found that most LD cases are contracted peri-domestically, the landscape can further contribute to LD risk through promoting the interaction between infected ticks and humans. An area with an increasing variation of NDVI may have a greater area of interface between residential property and forest edge. This is supported by several studies showing that fragmented landscapes are associated with either greater disease rates or acarological risk (Allan et al., 2003; Brownstein et al., 2005; Tran and Waller, 2013; Diuk-Wasser et al., 2021). It is likely important to consider educating people living in or interacting with areas of highly fragmented vegetation.

The interaction analysis (Fig. 2) of average NDVI and standard deviation of NDVI suggests that the relationship between standard deviation of NDVI and rates of LD at low levels of average NDVI have too much variability to make firm conclusions. At moderate levels of average NDVI,

increases in standard deviation of NDVI showed constant rates of LD. The highest levels of average NDVI showed slight increases of rates of LD with increasing standard deviation of NDVI, which is similar to the finding of the main analysis. This may suggest that fragmentation of forests within counties of high levels of NDVI have higher rates of LD, which may provide support of the dilution effect.

We found increased rates of LD with counties with lower population densities, compatible with a previous study (Seukep et al., 2015), and also consistent with our findings for average and the standard deviation of NDVI as one would expect that more sparsely populated areas would have more forested areas and a greater presence of infected ticks. There was a suggestion, however, of increased rates in high density human populations that may be due to the presence of ticks in city parks or to people being infected while traveling. As LD continues to expand into cities, we may see a shift in this relationship between population density and LD rates.

Rates of LD increased with the greater total land area until 5000 km^2. The size of the county has not been found previously to predict LD risk (Jackson et al., 2006b), but larger counties likely indicate a higher probability they harbor more infected ticks and hence higher rates of LD. For example, Maine has both the highest rates of LD and relatively large counties which contribute to the finding that the relationship between county area size and LD risk is greater. On the other hand, the largest counties exhibited rates that diminished with land area. The decrease in rates of LD in counties with land areas above 5000 km^2 occurred in 15 counties: six in Maine, eight in Minnesota, and one in New York State. Collectively, they have a median population of 15,000, a median population density of 6.1 per km^2, and a rural area that is about 46% of the total. These lower rates are unlikely due to lower densities of ticks infected with *Borrelia burgdorferi* spirochaetes (e.g. 50% of deer ticks tested positive in St. Lawrence County, NY) but could be due to human populations better protecting themselves.

Finally, as water area increased, LD rates initially increased in smaller areas, and steadily decreased afterward. Naturally, in areas with large tracts of water, where there is sparse to no vegetation, there will be no viable habitat for the ticks or their hosts, and as we expect, there may be lower rates of disease.

4.2. Methodological aspects

The strengths of this study include the utilization of the MODIS sensor that allowed us to include a long time series of observations in this analysis. These spaceborne platforms provide quasi-daily observations about the Earth's surface and enable novel data fusion techniques and insights into how this disease is moderated by environmental factors. Our study considered only the month of July, when vegetation is at its peak, while minimizing NDVI noise, like clouds, atmosphere contaminants, and maximizing illumination. Instruments like MODIS have been providing consistent and accurate measures, like NDVI, a proxy of vegetation health and productivity, at a resolution of 250 m. We have used coarser spatial resolution NDVI data (5.6 km) that were derived from the native finer 250 m resolution observations making our estimates of county-specific variability reasonably accurate and robust. We note here that some counties were simply too small to estimate variability in NDVI and were omitted from the multivariable analyses.

We used standard statistical models for rates of diseases, namely the quasi-Poisson and negative binomial regression models that account for the variability in LD counts and rates, and both models yielded similar results. Moreover, we did not assume one value for any of the covariables but allowed them to vary by year over the 19 years of observation, thereby reducing measurement error. Lastly, we did not assume any specific functional form for the associations between LD and the covariables, such as linear relationships, and by doing so we reduced bias from mis-specifying the regression model.

Our study is limited by the complex processes underlying the epidemiology of LD. As mentioned above, cases of LD are quantified based on

their county of residence, which assumes that the person contracting LD will visit the medical centre in closest proximity to where they acquired LD. As well, where LD is endemic, it is frequently acquired at home from activities in the lawn or backyard (Falco and Fish, 1988; Connally et al., 2006). Our data, however, cannot be used to establish the place of infection. As well, we did not have information on patients' specific characteristics, such as age, sex, and income.

We could not investigate the effects of age because we did not have counts of LD by age. But, because the age distribution of LD is bimodal (Mead, 2015), it is possible that different relationships by age may exist, that could be related to behaviour-related differences and age-related differences in susceptibility and care-seeking behaviors (Mead, 2015).

We focused only on land cover, the habitat of the vector and hosts, while numerous other factors contribute to the dynamics of the infection such as changes in temperature and human interactions within ecosystems. We did not account for temperature as we only had annual data for LD. As well, temperature does not vary dramatically between areas; for example, the mean maximum temperature in July is about 26 °C in Maine and 30 °C in Delaware.

Furthermore, we were only able to incorporate reported cases in the analysis, as LD is underreported and misdiagnosed in the USA (Schwartz et al., 2017; Rutz et al., 2018; Schiffman et al., 2018; White et al., 2018). The increased number of cases reported throughout our study period can be partly explained by true increases in rates, improved techniques to detect LD, and the changing definition of what constitutes a positive case of LD.

4.3. Study contribution

The present analysis contributes to the literature by confirming prior findings about LD dynamics while providing a simple framework for studying key parameters of the spatial dynamics of LD (Killilea et al., 2008). Our findings show clear non-linear relationships for many variables which may prove useful in designing future studies. One of the best-suited epidemiological designs for making causal inferences is the case-control study in which one can design detailed investigations of factors related to exposure in cases and appropriately selected controls (non-cases) (Orioski et al., 1998; Connally et al., 2009).

We found associations with land cover variables that can be utilized in predicting exposure and risk of LD and may suggest where it may be important to focus on prevention and early detection. Our results suggest that both urban areas (such as cities with large parks) along with suburban and natural settings may be critical areas of infection. We also found associations between increased rates of LD with higher values of the variability of NDVI within counties. Public health education programmes would focus not just on rural areas but also on urban and suburban areas and those people traveling to natural settings in which there is the likelihood of infection, including those living in fragmented areas. Additionally, people may be less likely to be vigilant about LD infection whilst walking through a city park rather than those traveling to natural settings. As LD continues to expand into cities, future studies on LD should also have a strong focus on its dynamic in urban areas along with rural and natural settings.

5. Conclusions

The relationships uncovered and reported in the present analysis cannot be interpreted as causal, rather, they indicate that certain patterns of land cover may increase or decrease exposure to infected ticks and thereby may contribute indirectly to increased rates. Only properly conducted cohort or case-control studies can be used to make causal inferences. Generally, our study points to a complex relationship between land cover and LD incidence rates, with the most critical being a segmented response to NDVI, a proxy of landscape vegetation greenness, showing a strong decrease of rates with increasing average NDVI (up to NDVI ~0.6), then a strong positive relation with rates increasing with increasing average NDVI. We suspect this complex association could be capturing aspects of the structure of landscapes, in particular fragmentation and hence the quality of habitats. Furthermore, our study presented a simple framework to fuse remote sensing time series data and census data and the incidence of LD that have not been carried out previously to elucidate how the distribution and trends in landscapes may be playing a role in moderating reservoirs for ticks and spread of disease.

Ethical approval

Not applicable.

CRediT authorship contribution statement

Sydney Westra: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Mark S. Goldberg:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Kamel Didan:** Writing – review & editing, Visualization, Resources, Formal analysis.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data are published and accessible as described in the methodology. Data can be found from the CDC (https://www.cdc.gov/lyme/stats/survfaq.html), from the USA Census (https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html), and from The Vegetation Index and Phenology Lab (https://lpdaac.usgs.gov/products/mod13c2v006/; https://lpdaac.usgs.gov/products/vip30v004/).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crpvbd.2023.100132.

References

Allan, B.F., Keesing, F., Ostfeld, R.S., 2003. Effect of forest fragmentation on Lyme disease risk. Conserv. Biol. 17, 267–272.

Alonso-Carné, J., García-Martín, A., Estrada-Peña, A., 2015. Assessing the statistical relationships among water-derived climate variables, rainfall, and remotely sensed features of vegetation: Implications for evaluating the habitat of ticks. Exp. Appl. Acarol. 65, 107–124.

Bawa, K., Rose, J., Ganeshaiah, K.N., Barve, N., Kiran, M.C., Umashaanker, R., 2002. Assessing biodiversity from space: an example from the Western Ghats, India. Conserv. Ecol. 6, 7.

Brownstein, J.S., Skelly, D.K., Holford, T.R., Fish, D., 2005. Forest fragmentation predicts local scale heterogeneity of Lyme disease risk. Oecologia 146, 469–475.

Cameron, A.C., Trivedi, P.K., 2013. Regression Analysis of Count Data, vol. 53. Cambridge University Press, New York, NY, USA.

Cao, J., Valois, M.F., Goldberg, M.S., 2006. An S-Plus function to calculate relative risks and adjusted means for regression models using natural splines. Comput. Methods Progr. Biomed. 84, 58–62.

CDC, 2019. Lyme disease: Data and surveillance. Centers for Disease Control and Prevention, Atlanta, USA. https://www.cdc.gov/lyme/stats/survfaq.html.

Connally, N.P., Ginsberg, H.S., Mather, T.N., 2006. Assessing peridomestic entomological factors as predictors for Lyme disease. J. Vector Ecol. 31, 364–370.

Connally, N.P., Durante, A.J., Yousey-Hindes, K.M., Meek, J.I., Nelson, R.S., Heimer, R., 2009. Peridomestic Lyme disease prevention: Results of a population-based case–control study. Am. J. Preventive Med. 37, 201–206.

DeLong, D.C., 1996. Defining biodiversity. Wildl. Soc. Bull. 24, 738–749.

Didan, K., Barreto, A., 2016a. NASA MEaSUREs vegetation index and phenology (VIP) vegetation indices 15 Days global 0.05 Deg CMG [Data set]. NASA EOSDIS Land Processes DAAC 4. https://lpdaac.usgs.gov/products/vip15v004/. (Accessed 8 November 2019).

Didan, K., Barreto, A., 2016b. NASA MEaSUREs vegetation index and phenology (VIP) vegetation indices monthly global 0.05 Deg CMG. NASA EOSDIS Land Process. DAAC 4. https://lpdaac.usgs.gov/products/vip30v004/. (Accessed 8 November 2019).

Dister, S.W., Fish, D., Bros, S., Frank, D.H., Wood, B.L., 1997. Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. Am. J. Trop. Med. Hyg. 57, 687–692.

Diuk-Wasser, M.A., VanAcker, M.C., Fernandez, M.P., 2021. Impact of land use changes and habitat fragmentation on the eco-epidemiology of tick-borne diseases. J. Med. Entomol. 58, 1546–1564.

Falco, R.C., Fish, D., 1988. Prevalence of *Ixodes dammini* near the homes of Lyme disease patients in Westchester County, New York. Am. J. Epidemiol. 127, 826–830.

Gabriele-Rivet, V., Koffi, J.K., Pelcat, Y., Arsenault, J., Cheng, A., Lindsay, L.R., et al., 2017. A risk model for the Lyme disease vector *Ixodes scapularis* (Acari: Ixodidae) in the prairie provinces of Canada. J. Med. Entomol. 54, 862–868.

Glass, G.E., Schwartz, B.S., Morgan III, J.M., Johnson, D.T., Noy, P.M., Israel, E., 1995. Environmental risk factors for Lyme disease identified with geographic information systems. Am. J. Publ. Health 85, 944–948.

Grantham, H.S., Moilanen, A., Wilson, K.A., Pressey, R.L., Rebelo, T.G., Possingham, H.P., 2008. Diminishing return on investment for biodiversity data in conservation planning. Conserv. Lett. 1, 190–198.

Hamer, S.A., Goldberg, T.L., Kitron, U.D., Brawn, J.D., Anderson, T.K., Loss, S.R., et al., 2012. Wild birds and urban ecology of ticks and tick-borne pathogens, Chicago, Illinois, USA, 2005–2010. Emerg. Infect. Dis. 18, 1589–1595.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Rem. Sens. Environ. 83, 195–213.

Huete, A., Didan, K., van Leeuwen, W., Miura, T., Glenn, E., 2010. MODIS vegetation indices. In: Land Remote Sensing and Global Environmental Change. Springer, New York, pp. 579–602.

Jackson, L., Levine, J., Hilborn, E., 2006a. A comparison of analysis units for associating Lyme disease with forest-edge habitat. Community Ecol. 7, 189–197.

Jackson, L.E., Hilborn, E.D., Thomas, J.C., 2006b. Towards landscape design guidelines for reducing Lyme disease risk. Int. J. Epidemiol. 35, 315–322.

Jamison, A., Tuttle, E., Jensen, R., Bierly, G., Gonser, R., 2015. Spatial ecology, landscapes, and the geography of vector-borne disease: A multi-disciplinary review. Appl. Geogr. 63, 418–426.

Jobe, D.A., Nelson, J.A., Adam, M.D., Martin Jr., S.A., 2007. Lyme disease in urban areas, Chicago. Emerg. Infect. Dis. 13, 1799–1800.

Justice, C., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., Saleous, N., et al., 2002. An overview of MODIS Land data processing and product status. Rem. Sens. Environ. 83, 3–15.

Justice, C.O., Vermote, E., Townshend, J.R.G., Defries, R., Roy, D.P., Hall, D.K., et al., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. IEEE Trans. Geosci. Rem. Sens. 36, 1228–1249.

Killilea, M.E., Swei, A., Lane, R.S., Briggs, C.J., Ostfeld, R.S., 2008. Spatial dynamics of Lyme disease: A review. EcoHealth 5, 167–195.

Kilpatrick, A.M., Dobson, A.D.M., Levi, T., Salkeld, D.J., Swei, A., Ginsberg, H.S., et al., 2017. Lyme disease ecology in a changing world: Consensus, uncertainty and critical gaps for improving control. Philos. Trans. R. Soc. B: Biol. Sci. 372, 20160117.

Kitron, U., Kazmierczak, J.J., 1997. Spatial analysis of the distribution of Lyme disease in Wisconsin. Am. J. Epidemiol. 145, 558–566.

Kotchi, S.O., Bouchard, C., Brazeau, S., Ogden, N.H., 2021. Earth observation-informed risk maps of the Lyme disease vector *Ixodes scapularis* in Central and Eastern Canada. Rem. Sens. 13, 524.

Krishnaswamy, J., Bawa, K.S., Ganeshaiah, K.N., Kiran, M.C., 2009. Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. Rem. Sens. Environ. 113, 857–867.

Kugeler, K.J., Farley, Grace M., Forrester, J.D., Mead, P.S., 2015. Geographic distribution and expansion of human Lyme disease, United States. Emerg. Infect. Dis. 21, 1455–1457.

LoGiudice, K., Ostfeld, R.S., Schmidt, K.A., Keesing, F., 2003. The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk. Proc. Natl. Acad. Sci. USA 100, 567–571.

Lüdecke, D., 2018. ggeffects: Tidy data frames of marginal effects from regression models. J. Open Source Softw. 3, 772.

Lukacic, G., White, J., Noonan-Toly, C., DiDonato, C., Backenson, P.B., 2018. Lyme disease surveillance using sampling estimation: Evaluation of an alternative methodology in New York State. Zoonoses Public Health 65, 260–265.

Mather, T.N., 1993. The dynamics of spirochete transmission between ticks and vertebrates. In: Ecology and Environmental Management of Lyme Disease. Rutgers University Press, New Brunswick, NJ.

Maupin, G.O., Fish, D., Zultowsky, J., Campos, E.G., Piesman, J., 1991. Landscape ecology of Lyme disease in a residential area of Westchester County, New York. Am. J. Epidemiol. 133, 1105–1113.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall, London, UK.

Mead, P.S., 2015. Epidemiology of Lyme disease. Inf. Dis. Clinics. 29, 187–210.

Murphree Bacon, R., Kugeler, K.J., Mead, P.S., 2008. Surveillance for Lyme Disease - United States, 1992–2006.

Myneni, R., Williams, D., 1994. On the relationship between FAPAR and NDVI. Rem. Sens. Environ. 49, 200–211.

Myneni, R.B., Hall, F.G., Sellers, P.J., Marshak, A.L., 1995. The interpretation of spectral vegetation indexes. IEEE Trans. Geosci. Rem. Sens. 33, 481–486.

Orioski, K.A., Campbell, G.L., Genese, C.A., Beckley, J.W., Schriefer, M.E., Spitalny, K.C., et al., 1998. Emergence of Lyme disease in Hunterdon County, New Jersey, 1993: A case-control study of risk factors and evaluation of reporting patterns. Am. J. Epidemiol. 147, 391–397.

Ostfeld, R.S., Keesing, F., 2000. Biodiversity series: The function of biodiversity in the ecology of vector-borne zoonotic diseases. Can. J. Zool. 78, 2061–2078.

Rutz, H.J., Wee, S.B., Feldman, K.A., 2018. Characterizing Lyme disease surveillance in an endemic state. Zoonoses Public Health 65, 247–253.

Schiffman, E.K., McLaughlin, C., Ray, J.A.E., Kemperman, M.M., Hinckley, A.F., Friedlander, H.G., et al., 2018. Underreporting of Lyme and other tick-borne diseases in residents of a high-incidence county, Minnesota, 2009. Zoonoses Public Health 65, 230–237.

Schwartz, A.M., Hinckley, A.F., Mead, P.S., Hook, S.A., Kugeler, K.J., 2017. Surveillance for Lyme disease - United States, 2008–2015. MMWR Surveill. Summ. 66, 1–12.

Seukep, S.E., Kolivras, K.N., Hong, Y., Li, J., Prisley, S.P., Campbell, J.B., et al., 2015. An examination of the demographic and environmental variables correlated with Lyme disease emergence in Virginia. EcoHealth 12, 634–644.

Shu, S., Wu, H., Jennifer, Y.G., Zeid, R., Harris, I.S., Jovanović, B., et al., 2020. Synthetic lethal and resistance interactions with BET bromodomain inhibitors in triple-negative breast cancer. Mol. Cell 78, 1096–1113.

Smith, G., Wileyto, E.P., Hopkins, R.B., Cherry, B.R., Maher, J.P., 2001. Risk factors for Lyme disease in chester county, Pennsylvania. Publ. Health Rep. 116 (Suppl. 1), 146–156.

Tourre, Y., Jarlan, L., Lacaux, J.P., Rotela, C.H., Lafaye, M., 2008. Spatio-temporal variability of NDVI-precipitation over southernmost South America: Possible linkages between climate signals and epidemics. Environ. Res. Lett. 3, 044008.

Tran, P.M., Waller, L., 2013. Effects of landscape fragmentation and climate on Lyme disease incidence in the northeastern United States. EcoHealth 10, 394–404.

Trischchenko, A.P., Cihlar, J., Li, Z., 2002. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors. Rem. Sens. Environ. 81, 1–18.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Rem. Sens. Environ. 8, 127–150.

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., et al., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. Int. J. Rem. Sens. 26, 4485–4498.

USA Census Bureau, 2020. Methodology for the United States population estimates: Vintage 2019. USA Census Bureau. https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html.

VanAcker, M.C., Little, E.A.H., Molaei, G., Bajwa, W.I., Diuk-Wasser, M.A., 2019. Enhancement of risk for Lyme disease by landscape connectivity, New York, New York, USA. Emerg. Infect. Dis. 25, 1136.

Weier, J., Herring, D., 2000. Measuring Vegetation (NDVI & EVI). NASA. https://earthobservatory.nasa.gov/features/MeasuringVegetation.

White, J., Noonan-Toly, C., Lukacik, G., Thomas, N., Hinckley, A., Hook, S., et al., 2018. Lyme disease surveillance in New York State: An assessment of case underreporting. Zoonoses Public Health 65, 238–246.

Wood, C.L., Lafferty, K.D., 2013. Biodiversity and disease: A synthesis of ecological perspectives on Lyme disease transmission. Trends Ecol. Evol. 28, 239–247.