How does Climate Change impact the spread of Lyme Disease?

Phillip Kim

Abstract

Lyme disease is a common vector-borne illness that has experienced a significant rise in cases lately. The frequency of Lyme Disease is highly dependent on the climate as ticks, the primary source of Lyme Disease, change activity based on seasonal weather changes. Therefore, it is safe to assume that climate change will have some influence on the occurrence of Lyme disease. Other research papers have conducted similar investigations, and even though these models only focus on New York, this does not mean the effects described are limited to New York. This model is not intended to simulate the consequences solely in New York, but to serve as a potential microcosm of how climate change can affect Lyme Disease in other areas. The linear regression models (a model created to predict future data based on current data) I used aimed to test this relationship and tested it against a model created by Nicholas H. Ogden. Due to the unpredictability of the spread of disease, reliable models can be challenging to create. However, based on my results, a positive relationship between climate change and the incidence of Lyme Disease is almost certainly not due to random chance and is worth further investigation.

Author Summary

Other research papers have conducted similar investigations. However, this one primarily aims to observe the relationship between Lyme Disease and climate change in New York, a state located in the Northeast and has had a particularly significant issue with Lyme Disease. Overall, I found that while indirect, the relationship between climate change and Lyme Disease is significant, and is worth investigating further in environmental and/or public health decisions.

Introduction

Lyme Disease is a common vector-borne illness that has experienced a significant rise in recent years. It is common in deer and rodents. Even though these animals do not directly infect humans, a tick can draw blood from an infected host, typically a rodent. It can then make contact with a human host and transmit the disease. Once infected, the disease starts with an itchy bump, a rash, and a fever (Mayo Foundation for Medical Education and Research, 2023). However, it may eventually transition to more severe symptoms, such as arthritis (Mayo Foundation for Medical Education and Research, 2023). As stated previously, Lyme Disease has experienced a significant increase in recent years. According to the CDC, “In 2022, reported case counts were 1.7 times the annual U.S. average during 2017–2019” (Kugeler, 2024).

Lyme Disease is highly dependent on the climate as Ticks, the primary vector, change activity based on seasonal weather changes. In addition, Lyme Disease can be affected by other factors. Therefore, climate change may significantly impact ticks and Lyme disease. However, it is largely unknown to what extent climate change influences the spread of Lyme Disease. However, considering that the CDC conducted research into Lyme Disease in

Massachusetts, a comparable area, with conclusive results (Cocoros et al., 2023) this relationship is likely to be significant. In this paper, I set out to answer three key research questions: How has Lyme disease changed in recent decades across New York (NY)? To what extent are observed degree days and precipitation associated with Lyme disease? How will climate change impact the spread of Lyme disease?

Materials and Methods

Data Collection and Preparation

The Lyme Disease data was collected from the New York Department of Health (Communicable Disease Annual Reports and Related Information, n.d.). The data was separated by county. New York City's data was entered as one county. Even though the data on case counts was collected from all of New York City, most Lyme Disease cases in New York City come from Staten Island. Therefore, Staten Island's temperature was used for all of New York City's data. There are two measurements for Lyme Disease cases, per 100,000 and overall cases. Both metrics were over the course of 1 year. Throughout this analysis, cases per 100,000 were used to account for potential changes in population over time and between counties. This dataset began including Lyme disease data on a county level in 1994, which is why this analysis began in 1994.

Historical temperature and precipitation climate data was collected from the Physical Sciences Laboratory of NOAA (Home: NOAA Physical Sciences Laboratory, n.d.), using temperature data from the climate section from 1994 to 2019 inclusive (PSL TDS Catalog, n.d.).

Precipitation data was collected from the same source and during the same years. Both are Netcdf files (this type of file is used to store information about the earth and its atmosphere in one area), organizing information based on the coordinates from which they were collected. The data extracted was organized into 57 different counties of New York aside from New York City.

To model the relationships between temperature, precipitation, and Lyme disease, I fit a 2-variable linear regression model (a model created to predict future data based on current data) similar to that used in the LOW model (Moore et al., 2014). The two explanatory variables used in this paper are degree days (days above freezing temperature) and the average cumulative precipitation each month in a year. All the county-based data was aggregated into one larger dataset for New York State.

The years 2020 and beyond were excluded from the dataset due to the COVID-19 pandemic and the lack of normality in individual behaviors during 2020-2022, and 2024 gives too short of a time frame for any significant additional information.

I extracted data from temperature and precipitation Netcdf files through the Xarray module in Python. The temperature dataset contained data at various latitudes and longitudes. These coordinates were the locations of weather stations. Using the center point of each county in New York, I matched it to the closest available coordinate in the temperature dataset. (Micheal, J/County table, 2020). I used a for loop to organize the information by year, then I found the data for every county and calculated the number of degree days.

From there, the correlation coefficient (A number from -1 to 1 used to test the strength of a relationship between two variables) was obtained between precipitation vs cases and degree days vs cases. Regressions were not fitted on the county-by-county basis as I found that this led to overfitting, because the p-values (the chance of a relationship between two variables being purely due to chance, and not actually due to one causing a change in the other) when comparing were high in county-based regressions even though their correlation coefficients were higher in general. The p-value threshold for significance used was the 0.05 threshold.

Results

Observed Lyme Disease Trends Lyme disease trends have generally increased in recent decades. For example Albany county has experienced an increase in average case counts per 100,000 from 4.9 in 1994 to 139.2 in 2019 (an increase of 28-fold). Furthermore, by county the average correlation coefficient of case counts per 100,000 individuals vs years is 0.621 with a 25 percentile of 0.311, a median of 0.687 and a 75th percentile of 0.763. In addition, the average p-value (the chance of a relationship between two variables being purely due to chance, and not actually due to one causing a change in the other) for each county is 1.16 * 10^-26 with a 25th percentile of 3.7*10^-99, a median of 2.00 * 10^-69 and a 75th percentile of 3.94 * 10^-69. In other words, Lyme Disease has increased significantly in recent decades with a near zero chance of this trend being due to random chance.

Analysis

I also investigated the relationship between degree days precipitation and Lyme disease by fitting a 2-variable linear regression across all of NY state. The resulting linear regression fit to the data is in Equation 1.

Case_counts = 25.8 + 0.498 * degree_days + -229400 * precipitation (Eq. 1).

The coefficient for degree days is 0.498. This means that for every additional degree_day observed, the model predicts an increase in case_counts of 0.498. For every additional precipitation of kg/m2/s observed, we predict a decrease in case_counts of 229400. This may seem unusual; however, the level of precipitation is usually very small. The p-value of temperature vs cases was less than 10 * 10^-99, and the p-value of precipitation vs cases was 4.84 * 10^-71. Both of these values are extremely small. This result implies that the coefficients for degree_days and precipitation are statistically significant at the .05 level.

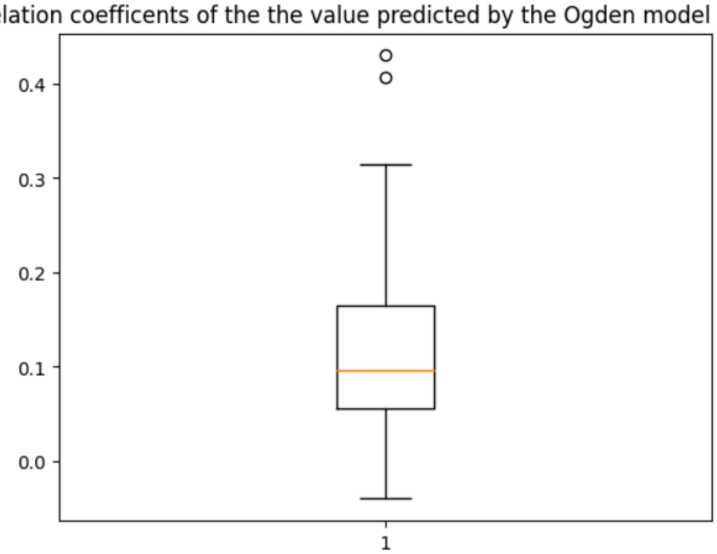
On the other hand, the model had a low correlation coefficient of 0.178, but this regression had a very low p-value of 6.89 * 10^-14. This value is a generally low correlation coefficient; however, a very low p-value suggests that this relationship is still significant and is likely not due to chance.

I also benchmarked the state-level regression model performance by comparing its predictions against those from a Lyme disease model from the literature, finding that it outperforms the Ogden et al. model. The Ogden et al. model:

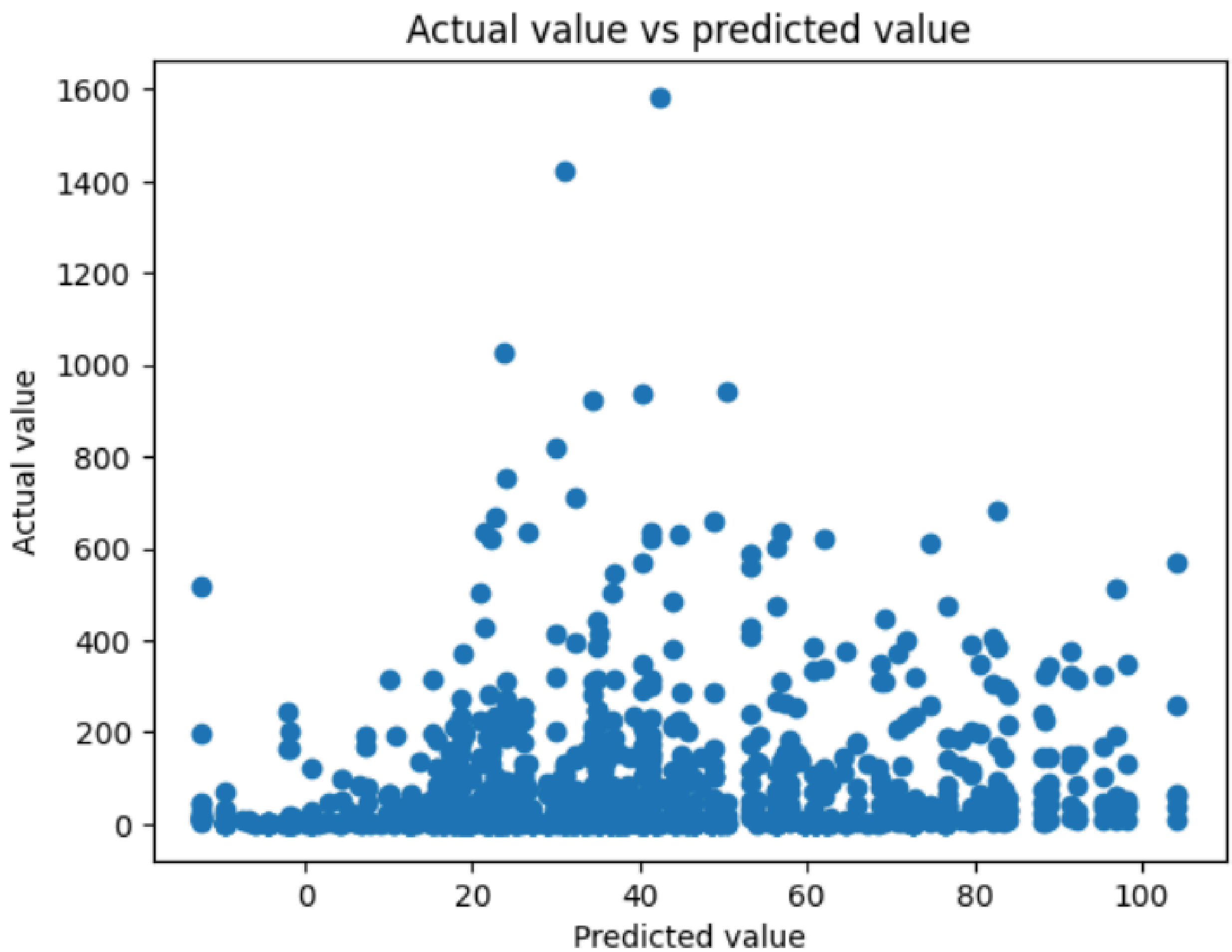
R0 = 5.556 + 1.072 × 10–6 DD –4.658 × 10–3 DD2 (Eq. 2)

Is also based on degree days but not on precipitation, where DD is the number of days per year exceeding 0°C and R0 is a variable reflecting how easily the disease spreads. I used their fitted model to make predictions of case counts rather than R0 since I assume they are likely to be correlated across time for New York and compared the predictions. When using the Ogden et al. model (Eq. 2) to predict case counts and correlating it with the actual amount of cases, this model produced an average correlation coefficient of 0.122 between predicted and observed with some variations with an average p-value of 0.0797. When fitted to the entire state the correlation coefficient was -0.144 with a p-value of 4.57*10^-62. A negative correlation implies that as predicted case counts decrease, the amount of cases increases, suggesting poor model performance. On the other hand, the multivariable regression (Eq. 1) for the entire state produced a correlation coefficient of 0.178, higher in magnitude than the previous two. Since it is higher in magnitude, this model (Eq. 1) outperformed the Ogden et al. model. Additionally, it had a p-value of 6.89 * 10^-14, a value significantly lower than the 1st model and insignificantly larger than the 2nd model.

Variance in the correlation coefficents of the the value predicted by the Ogden model vs the actual case counts



There was a pattern with predicted vs actual cases as they tend to cluster in a similar spot showing that the regression is at least somewhat accurate.



Discussion

This data may not be entirely accurate due to variability in location and data collection. The center of each county may not be where much of the population is centered. In addition, the New York Department of Health reports the incidence of disease in the amount of emergency service responses in these areas leaving room for underreporting by hospitals or individuals reported as an additional case in a county separate from where they caught Lyme disease.

The relationship between the number of degree days, the level of precipitation, and the incidence of Lyme Disease is very variable yet significant. Due to variable differences in how disease spreads, and how often people go outside in a single year, the results were generally unpredictable. A correlation coefficient of 0.178 and a p-value of 6.89 * 10^-14 suggests that this relationship is weak but still worth investigating due to the unpredictability of climate and diseases. This information would imply that a similar conclusion would apply to the extent to which Lyme Disease affects.

References

1. Mayo Foundation for Medical Education and Research. (2023, February 10). *Lyme disease*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/lyme-disease/symptoms-causes/syc-20374651
2. Kugeler, K. J. (2024). Surveillance for Lyme Disease After Implementation of a Revised Case Definition — United States, 2022. MMWR. Morbidity and Mortality Weekly Report, 73. https://doi.org/10.15585/mmwr.mm7306a1
3. Cocoros, N. M., Kluberg, S. A., Willis, S. J., Forrow, S., Gessner, B. D., Nutt, C. T., Cane, A., Petrou, N., Sury, M., Rhee, C., Jodar, L., Mendelsohn, A., Hoffman, E. R., Jin, R., Aucott, J., Pugh, S. J., & Stark, J. H. (2023). Validation of Claims-Based Algorithm for Lyme Disease, Massachusetts, USA. Emerging Infectious Diseases, 29(9). https://doi.org/10.3201/eid2909.221931
4. Communicable Disease Annual Reports and Related Information. (n.d.). Health.ny.gov. Retrieved July 16, 2024, from https://health.ny.gov/statistics/diseases/communicable/
5. Home: NOAA Physical Sciences Laboratory. (n.d.). Psl.noaa.gov. https://psl.noaa.gov/
6. PSL TDS Catalog. (n.d.). Psl.noaa.gov. Retrieved July 16, 2024, from https://psl.noaa.gov/thredds/catalog/Datasets/ncep.reanalysis2/Dailies/gaussian_grid/catalog.html
7. Moore, S. M., Eisen, R. J., Monaghan, A., & Mead, P. (2014). Meteorological Influences on the Seasonality of Lyme Disease in the United States. The American Journal of Tropical Medicine and Hygiene, 90(3), 486–496. https://doi.org/10.4269/ajtmh.13-0180
8. User:Michael J/County table. (2020, July 7). Wikipedia. https://en.wikipedia.org/wiki/User:Michael_J/County_table
9. Ogden, N. H., Radojevic´, M., Wu, X., Duvvuri, V. R., Leighton, P. A., & Wu, J. (2014). Estimated Effects of Projected Climate Change on the Basic Reproductive Number of the

Lyme Disease Vector Ixodes scapularis. Environmental Health Perspectives, 122(6), 631–638. https://doi.org/10.1289/ehp.1307799

10. Disease Reporting | EpiQuery. (n.d.). A816-Health.nyc.gov. Retrieved July 16, 2024, from https://a816-health.nyc.gov/hdi/epiquery/disease-reporting#:~:text=The%20Syndromic%20Surveillance%20Unit%20in