
Collin J. O'Connor

University at Albany, State University of New York, collinoconnor3@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Biostatistics Commons, Ecology and Evolutionary Biology Commons, and the Epidemiology Commons

Recommended Citation


https://scholarsarchive.library.albany.edu/legacy-etd/2542

This Master's Thesis is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
A Comparative Spatial and Climate Analysis of Human Granulocytic Anaplasmosis and Human Babesiosis in New York State (2013-2018)

By

Collin J. O'Connor

A Thesis

Submitted to the University at Albany, State University of New York

in Partial Fulfillment of

the Requirements for the Degree of

Master of Science

School of Public Health

Department of Epidemiology and Biostatistics

May 2020
Abstract:

Human granulocytic anaplasmosis (HGA) and human babesiosis are tick-borne diseases spread by *Ixodes scapularis* (the blacklegged or deer tick) and are the result of infection with *Anaplasma phagocytophilum* and *Babesia microti*, respectively. In New York State (NYS), incidence rates of these diseases increased concordantly until around 2013, when rates of HGA began to increase more rapidly than human babesiosis, and the spatial extent of the diseases diverged. Surveillance data of tick-borne pathogens (2007 to 2018) and reported human cases of HGA (n=4,297) and human babesiosis (n=2,986) (2013 to 2018) from the New York State Department of Health (NYSDOH) showed a positive association between the presence/temporal emergence of each pathogen and rates of disease in surrounding areas. Comparing incidence rates among demographic groups showed higher rates of HGA among all age and sex groups, in addition to white (p<0.0001), and non-Hispanic/non-Latino individuals (p<0.0001). Human babesiosis exhibited higher rates among Black (p=0.0001), Asian (p=0.0106), Other (p<0.0001) and Hispanic/Latino (p<0.0001) individuals. Open-source climate, weather and landscape data were used to build a spatially weighted zero-inflated negative binomial (ZINB) model to examine and compare associations between the environment and rates of HGA and human babesiosis. HGA and human babesiosis ZINB models indicated similar associations with forest cover, forest land cover change, and winter minimum temperature, and differing associations with elevation, urban land cover change and winter precipitation. These results indicate that tick-borne disease ecology varies between pathogens spread by *Ixodes scapularis*. 
Introduction

Human granulocytic anaplasmosis (HGA) and human babesiosis are tick-borne diseases caused by the pathogens *Anaplasma phagocytophilum*, and *Babesia microti* respectively. HGA and human babesiosis exhibit similar non-specific flu-like symptoms including fatigue, fever, chills and headache. Both diseases may result in death, however, most patients recover with appropriate antibiotic and antiparasitic treatments, respectively.

In New York State (NYS), rates of HGA and human babesiosis had been increasing steadily and at similar rates since the early 2000s. However, rates of HGA have rapidly increased in recent years, surpassing those of human babesiosis by a factor of 1.67, 2.13, 1.83 and 1.72 in 2015, 2016, 2017 and 2018, respectively. The recent surge of HGA cases is of interest when examined in the context of human babesiosis, as both HGA and human babesiosis are spread by the same vector species, the blacklegged tick or deer tick (*Ixodes scapularis*) and have the same primary reservoir in nature, the white-footed mouse (*Peromyscus leucopus*). A potential explanation for diverging disease rates, despite similarities in the transmission pathway, is each pathogen’s unique relationship to the environment. Assessing how the environment affects tick-borne disease risk is an important research topic, as it has allowed researchers to spatially forecast risk of Lyme disease in response to the projections of a changing climate. Studies have examined the direct effects of environmental characteristics on tick behavior, which relates to the risk of all tick-borne diseases. Recent studies have shown that environmental conditions can affect the microbiome composition of *I. scapularis* and *Dermacentor andersoni* (the Rocky Mountain wood tick), providing evidence that environmental conditions affect tick-borne pathogens differently. Further, Neelakanta *et al.*
(2010) showed that *I. scapularis* infected with *A. phagocytophilum* express an antifreeze glycoprotein, called IAFGP, which gives *I. scapularis* an increased resistance to colder temperatures. Despite evidence for pathogen-specific interactions with the environment, large-scale spatial tick ecology studies have generally only examined the effect of the environment and landscape characteristics on the spatial density of *I. scapularis* and the prevalence of *Borrelia burgdorferi* (the causative agent of Lyme disease) in *I. scapularis* populations. The relationship between environmental conditions and the prevalence of *A. phagocytophilum* or *B. microti* and their respective human illnesses remains largely unstudied.

Large variation of climate and elevation in NYS may help to elucidate the potential effect of the environment on increasing rates of HGA and human babesiosis. The highest incidence rates of HGA and human babesiosis are localized to different regions of NYS; the upper and mid-Hudson Valley regions have exhibited the highest rates of HGA, while higher rates of human babesiosis have occurred in the lower Hudson Valley and the coastal region of Long Island.

This study aims to use a large-scale spatially weighted regression to comparatively examine the association between the environment of NYS and the rates of HGA and human babesiosis. Further, this study examines tick pathogen surveillance data from the New York State Department of Health (NYSDOH) spatiotemporally, to describe changing pathogen prevalence of *A. phagocytophilum* and *B. microti* among *I. scapularis* populations across NYS.

**Methods**

**Human granulocytic anaplasmosis and human babesiosis case criteria**
Cases of HGA and human babesiosis occurring between 2013 and 2018 in NYS (excluding NYC) from the NYSDOH Communicable Disease Electronic Surveillance System (CDESS) were analyzed using SAS v. 9.2 (SAS Institute, https://www.sas.com). Suspected or confirmed cases of HGA and human babesiosis are reportable to the NYSDOH under New York State Sanitary Code (10NYCRR 2.10), and are assigned a case status according to the prevailing CDC case definitions at the time of reporting.35,36 Only cases recorded in CDESS with a case status of either “confirmed” or “probable” were included in this study. CDESS records contain multiple clinical and demographic fields used for disease surveillance, and variables included in this study were the United States Postal Service (USPS) zip code of patient residence, clinical outcome, year the case occurred, age, sex, race and ethnicity.

Case demographics and hypothesis testing

Total population at risk and corresponding demographic distribution data were gathered from the American Community Survey (ACS) 2013 5-year estimates, aggregated to the spatial boundaries of zip code tabulation areas (ZCTAs).37 ZCTAs are used by the United States Census Bureau to differentiate from zip codes used by the USPS, and are used to spatially aggregate population data.38 In order to calculate incidence rates for HGA and human babesiosis, the USPS zip code of each case’s home address was manually converted to its corresponding ZCTA. Incidence rates of HGA and human babesiosis were then calculated for the 6-year study period (2013-2018), and were weighted by age, sex, race and ethnicity of each case and underlying population demographic distribution. Incidence rates were compared between HGA and human babesiosis for each demographic category using the exact rate ratio test ($\alpha=0.05$) in the R package ‘rateratio.test’39 in RStudio v. 1.2.133540. All rate-ratio tests were adjusted for multiple comparisons using the Bonferroni-Holm adjustment.41
Tick sampling and pathogen testing

Host-seeking adult and nymphal tick sampling was conducted on publicly accessible lands from the period of 2007 to 2018 as described elsewhere. Adult life-stage *I. scapularis* were sampled via active tick flagging during the months of April and May for spring-questing adults and between October and December for fall-questing adults. Nymphal life-stage *I. scapularis* were sampled via active tick dragging between the months of May and September. During all initial site visits, GPS coordinates of the site were recorded to be used for spatial analysis, and environmental data such as ambient air temperature (°C), relative humidity, estimated wind speed (mph), and general weather observations were noted at the time of sampling. *I. scapularis* were identified under a dissecting microscope (Model SMZ1000, Nikon, Tokyo, Japan) to species utilizing dichotomous keys and stored in 99.5% ethanol at -20°C until DNA extraction.

*I. scapularis* samples collected between 2007 and 2012 underwent total genomic DNA extraction via spin column method (DNeasy Blood and Tissue Kit, Qiagen USA, Germantown, MD), and were tested for the presence of (target gene): *A. phagocytophilum* (*msp3*), *B. microti* (*16S-like rRNA*), and *B. burgdorferi* (*ospA*) using a triplex Polymerase Chain Reaction (PCR) assay as described elsewhere. *I. scapularis* samples collected between 2013 and 2018 underwent automated total genomic DNA extraction via Qiagen QIAcube HT using QIAamp 96 kit (Qiagen USA, Germantown, MD) according to manufacturer protocols. Extracted DNA was then tested for the presence of (target gene) *A. phagocytophilum* (*msp2*), *B. microti* (*18S rDNA*), *B. burgdorferi* (*16S rDNA*), and *Borrelia miyamotoi* (*16S rDNA*) using a quadplex real-time PCR assay as previously described. Concordance between these two assays for the detection of *B.
*Burgdorferi, B. microti* and *A. phagocytophilum* was shown to be 100%, as previously established by testing a subset of samples using both assays during the validation processes.\textsuperscript{32}

Results for presence of *A. phagocytophilum* and *B. microti* in *I. scapularis* ticks were linked to the corresponding collection site data. Collection sites were then characterized based on the first temporal occurrence for a PCR positive result for either *A. phagocytophilum* or *B. microti* separately among *I. scapularis*. Study sites were grouped as either: sites where no PCR positive results for a pathogen had been found, sites where one or more PCR positive results for a pathogen were found during or after 2013 (the human case study period), or sites where one or more PCR positive results for a pathogen were found prior to 2013 or on the first visit to a collection site after 2013.

All adult *I. scapularis* collected during the spring season were assigned to the previous year. *I. scapularis* collected during the spring months of April and May did not successfully obtain a bloodmeal during the previous fall, thus, they overwintered and continued host-seeking into the spring. Spring collected *I. scapularis* have had the same number of bloodmeals and the same number of opportunities to acquire *A. phagocytophilum* and *B. microti* as *I. scapularis* adults collected during the previous fall.\textsuperscript{45}

**Tick pathogen relation to human case incidence**

Incidence rates of HGA and human babesiosis during the period of 2013-2018 for each NYS ZCTA were merged to a TIGER/Line shapefile of NYS ZCTAs\textsuperscript{46} in ArcMap v. 10.6.1.\textsuperscript{47} Tick collection sites categorized according to the temporal emergence of *A. phagocytophilum* and *B. microti* in *I. scapularis* populations were mapped and intersected with the NYS ZCTA shapefile to determine which collection sites were contained within any NYS ZCTAs. ZCTAs
were then assigned the same category of temporal emergence of *A. phagocytophilum* and *B. microti* as collection sites located within them. In the event of a ZCTA containing more than one collection site, all observations from those collection sites were combined and reassessed as a single collection site, and the resulting category was assigned to the ZCTA. Incidence rates among the assigned categories of pathogen emergence were then analyzed using the non-parametric Kruskal-Wallis test (α=0.05) to determine if mean incidence of HGA and human babesiosis differed between any of the three groups.48 Stepwise comparisons between temporal pathogen emergence groups were then assessed using the pairwise Wilcoxon rank-sum test with Bonferroni-Holm correction (α=0.05) to determine if temporal pathogen emergence groups had differing means of human incidence.49

**Climate/weather and landscape data collection**

Open-source climate, weather and landscape variables were selected to represent variables significantly associated with tick density in the current literature. All climate and weather variables were gathered from Parameter-elevation Regressions on Independent Slopes Model (PRISM)50 as raster files (PRISM Climate Group, Accessed 2/2020). Variables gathered from PRISM included: precipitation,51 mean temperature,52 minimum temperature, maximum temperature, mean dewpoint temperature,24 minimum vapor pressure deficit,53 and maximum vapor pressure deficit. Annual yearly averages and monthly averages for each PRISM variable were gathered and spatially averaged over each NYS ZCTA using the ‘zonal statistics as table’ tool in ArcMap during the study period.54 After a spatial average was calculated for each climate variable during its corresponding unit of time, values were averaged over the entire study to achieve single values for use in regression analysis. Winter averages were calculated as the
average of each monthly value from November through March for each year from 2012 through 2018.

Landscape covariates were selected from significant tick density predictors in the published literature, including elevation, deciduous and mixed forest cover and change in land use. Elevation data at 7.5 arc-second spatial resolution was gathered from the Global Multi-resolution Terrain Elevation Dataset (GMTED) 2010 in raster format. ‘Zonal statistics as table’ was used to find the mean elevation for each NYS ZCTA. Land cover and land cover change data were gathered from the National Land Cover Dataset (NLCD) 2016 Land Cover and 2016 Land Cover Change Index (LCCI), respectively. The NLCD Land Cover dataset was used to assess the percent of a ZCTA’s total area covered with deciduous or mixed forest. Deciduous and mixed forest layers were cut from the NYS ZCTA layer via the ‘Erase’ tool, leaving the total area for each ZCTA not covered with either deciduous or mixed forest. Remaining areas within each ZCTA were then used to calculate forest coverage percentages. This process was repeated to assess land cover change via the forest change and urban change layers in the NLCD LCCI. Both the forest and urban change layers indicate whether a spatial area has changed to or from its respective land cover class at any point between the period of 2001 and 2016.

**Spatial regression model development**

Human case counts of HGA and human babesiosis were predicted via environmental covariates using zero-inflated negative binomial (ZINB) regression model with ACS 2013 5-year population estimates as an offset. The goal of the model building procedure was to assess the association of environmental factors with cases of HGA and human babesiosis. Traditionally, a Poisson regression model is used to assess count data, but in the event of overdispersion,
negative binomial regression model is used. Further, when the count data exhibits an excess of zero count observations in addition to overdispersion, the ZINB model may be used. The ZINB model assumes that variation in number of successes greater than zero and the process that determines if an observation is either zero or non-zero are independent. To deal with the separate processes and excess zero counts, the ZINB model simultaneously assesses a negative binomial model for count values greater than zero, and a logistic model to measure the likelihood of being in the “not at risk” group.

ZINB models for HGA and human babesiosis were built using a three-step process. First, a correlation matrix was created with all variables in question using the ‘ggpairs’ function from the ‘GGally’ package in RStudio. Variables were then examined and adjusted for collinearity by removing the minimum amount of variables possible while leaving no covariates correlated at a Pearson’s correlation coefficient >0.70. Second, all remaining variables were used to build a backwards-stepwise logistic regression model to determine which variables would be used to build the zero-inflated portion of the ZINB model. Finally, the ZINB model was built using the backwards stepwise method using the ‘zeroinfl’ function from the ‘pscl’ package in RStudio. All variables were input into the negative binomial portion of the ZINB model and the significant variables from the previous logistic regression were input into the zero-inflated portion of the ZINB model. Variables were removed one by one when not statistically significant (α=0.05), starting with the variable with the highest p-value. After only statistically significant variables remained, the model was examined for best fit by comparing the AIC of all possible models using the ‘dredge’ function from the ‘MuMIn’ package in RStudio. The resulting model with the lowest AIC for each HGA and human babesiosis was deemed as the final model.
After final models for HGA and human babesiosis were determined, their residuals examined for spatial heterogeneity via Moran’s $I$, using inverse distance weighting of centroids of ZCTAs from the R ‘spdep’ package ($\alpha=0.05$). If Moran’s $I$ revealed spatial heterogeneity in the model, a spatial autocovariate term was fitted to the final models to improve fit and control for distance between sites. Models with and without an autocovariate term were compared using AIC, where the best fit occurred if the spatially controlled model had an AIC value of two or more points less than the non-spatially controlled model. Residuals were spatially analyzed by conducting a Kriging interpolation for both spatially adjusted and unadjusted HGA and human babesiosis models. Kriging interpolations allowed the spatial extent and magnitude of model residuals to be examined before and after spatial weighting. Further model diagnostics were performed by examining rootograms, in addition to comparing predicted 2019 case counts and distribution to preliminary 2019 case counts across NYS (unpublished preliminary data).

**Results**

**Human granulocytic anaplasmosis and human babesiosis case surveillance**

A total of 4,297 confirmed or probable cases of HGA and 2,986 confirmed or probable cases of human babesiosis occurred between 2013 and 2018 in NYS (excluding NYC). Incidence of both HGA and human babesiosis increased over the study period, however, the increase of HGA incidence was of a higher magnitude and extended further northward than human babesiosis (Figure 1). Of the 4,297 cases of HGA, 278 had known values for patient health outcome; 263 cases were known to have survived, while 15 cases died. Of the 2,986 cases of human babesiosis, 375 had known values for patient clinical outcome; 363 cases were known to have survived, while 12 cases died. Resulting fatality rates for HGA and human babesiosis
among cases with reported health outcomes were 0.35% and 0.40% respectively. Due to the high number of cases with missing clinical outcome values, comparative statistical analysis of health outcome data was not conducted between HGA and human babesiosis.

Both HGA and human babesiosis had a disproportionate number of white and male cases during the study period (Table 1). HGA surveillance data reported 61.39% of cases were male while human babesiosis surveillance data reported 65.74% of cases were male. Rates of HGA were significantly higher than human babesiosis among both male (exact rate ratio test, p<0.0001) and female (exact rate ratio test, p<0.0001) sexes.

Distributions of race categories showed 75.49% of HGA cases and 59.87% of human babesiosis cases were of white race. Surveillance data also showed that ethnicity responses had high percentages of answers of “unknown” for HGA and human babesiosis, at 35.63% and 49.73% respectively. Among cases where ethnicity responses that were known for HGA (n=2766) and human babesiosis (n=1501), 96.13% and 84.41% responded as non-Hispanic/non-Latino, respectively. All race and ethnicity categories were significantly different between HGA and human babesiosis (exact rate ratio test, α=0.05), apart from American Indian/Alaska Native (p=0.4375) and Pacific Islander (p=0.4375).

The age distribution of HGA and human babesiosis cases were similar, and the number of cases tended to increase with an increase in age category (Table 1). The mean age was slightly higher among human babesiosis (mean = 61.00 years old) cases than HGA (mean = 58.49 years old) (Table 1). Rates of HGA were significantly higher than human babesiosis among all age categories (exact rate ratio test, α=0.05).
**Ixodes scapularis testing results and temporal pathogen emergence**

During the period of 2007 through 2018, 34,076 adult and 18,539 nymphal *I. scapularis* ticks were tested for *A. phagocytophilum* and *B. microti*. Among adult *I. scapularis* tested, 2,190 samples (6.43%) were positive for *A. phagocytophilum* and 1,209 samples (3.55%) were positive for *B. microti*. Among nymphal *I. scapularis* tested, 799 samples (4.31%) were positive for *A. phagocytophilum* and 674 samples (3.64%) were positive for *B. microti*.

PCR results for *A. phagocytophilum* and *B. microti* aggregated to 240 field collection sites to show temporal emergence of each pathogen are exhibited (Figure 2). Results indicated 94 sites where no *A. phagocytophilum*-infected *I. scapularis* had been collected and 150 sites where no *B. microti*-infected *I. scapularis* had been collected (Figure 2). The number of sites with specific pathogen infections among *I. scapularis* prior to 2013 or on initial site visits were 87 for *A. phagocytophilum* and 38 for *B. microti*. The number of sites with specific pathogen infections occurring on or after 2013 were 59 for *A. phagocytophilum* and 52 for *B. microti*.

Violin plots showing the distribution of incidence for each disease among the three site categories and their respective Kruskal-Wallis and pairwise Wilcoxon rank sum tests (α=0.05) are shown (Figure 3 and Table 2). Results of violin plots indicate HGA and human babesiosis incidence at ZCTA level generally increase due to the presence of *A. phagocytophilum* and *B. microti* in *I. scapularis* from surrounding environments. Further, the violin plots show that the earlier *A. phagocytophilum* and *B. microti* are found in *I. scapularis* in the nearby environment, the higher the incidence of both HGA and human babesiosis are. The results of the Kruskal Wallis test (Table 2) indicate that means of incidence of disease differ between at least two of the site categories for *A. phagocytophilum* (p<0.0001) and *B. microti* (p<0.0001). Results from the pairwise Wilcoxon rank sum tests indicate that all site categories have differing values for mean
incidence of human babesiosis. Discordantly, the results for *A. phagocytophilum* sites indicate that mean incidence of HGA did not differ based on *A. phagocytophilum* emerging before or after 2013 (Table 2).

**Zero-inflated negative binomial regression of environmental factors on cases of human granulocytic anaplasmosis and human babesiosis**

Mean dewpoint temperature, mean temperature, maximum vapor pressure deficit, minimum vapor pressure deficit, and maximum temperature were excluded from regression analysis due to concerns of collinearity from high correlation coefficients between each other and the variables left in the analysis. Removing the above variables allowed all remaining correlation coefficients to be under 0.70 while still addressing the research questions. Following exclusions, variables examined via ZINB modeling were mean elevation, percent forest cover, percent forest land cover change, percent urban land cover change, winter precipitation, and winter minimum temperature.

Following initial ZINB model building procedures, Moran’s *I* of the residuals from the non-spatially controlled models provided evidence for spatial heterogeneity in both the HGA model (*p*<0.0001) and human babesiosis model (*p*<0.0001). In both the HGA and human babesiosis models, the autocovariate was significant in both the negative binomial and zero-inflated portions of the models. When comparing the AIC between the spatially adjusted and unadjusted HGA and human babesiosis models, the AICs were 607.61 and 223.92 points lower, respectively. Due to the decreases in AIC when accounting for spatial heterogeneity, the autocovariate terms were included in both the count and zero-inflated portions of the HGA and human babesiosis final models. Model diagnostics were assessed using rootograms and kriging.
interpolations of residuals to assess model fit and compare spatially adjusted and unadjusted models. Results from diagnostics are shown in Figures 4 and 5.

Results for final spatially weighted ZINB regression models for HGA and human babesiosis are shown (Table 3). Significant variables with similar magnitudes and direction of association in the count portions of both the HGA and human babesiosis models included: elevation, percent forest land cover change and percent mixed/deciduous forest. The count portion of the human babesiosis model additionally included percent urban land cover change, whereas the HGA model did not. Winter precipitation was significant in the count portions of both models but was negatively associated with HGA case counts and positively associated with human babesiosis case counts. Significant variables with similar magnitudes and directions of association in the zero-inflated portions of both HGA and human babesiosis models included: minimum winter temperature, percent mixed/deciduous forest cover, and percent urban land cover change. Additionally, the zero-inflated portion of the HGA model included elevation as a predictor, whereas the human babesiosis model did not.

Discussion

This study describes variations in epidemiological and spatial risk of HGA and human babesiosis, two tick-borne diseases spread by the same tick vector and maintained in the environment in the same major reservoir, across NYS. This study also establishes the varying influence that climate, weather and landscape characteristics have on the rates of HGA and human babesiosis. Although this study did not use time-series data to examine the change of climate across NYS, the spatial variation of climate, weather and landscape characteristics across NYS geography can be used to determine areas in NYS that are more susceptible to the encroachment of disease and overall increase of disease.
Epidemiological analysis of demographic distributions for age, sex, race and ethnicity showed similar trends between HGA and human babesiosis. Both diseases mostly impacted individuals above the age of 50 years old, individuals who were male, individuals who were white and individuals who were non-Hispanic and non-Latino, which is consistent with the current literature on tick-borne diseases. Individuals in the race categories of Black, Asian, other, and unknown had higher rates of human babesiosis than HGA. Human babesiosis surveillance data also showed a higher percentage of cases of unknown ethnicity than HGA. The slight differences in race and ethnicity between HGA and human babesiosis could be a result of the geographic difference in the spread of each disease. Figure 2 shows that human babesiosis incidence is higher in the more racially diverse region of Long Island, while HGA incidence is higher in the less racially diverse upper Hudson Valley region.

Cases of HGA and human babesiosis in NYS are localized to Long Island and the Hudson Valley region. Additionally, the magnitude of case counts exhibited high variability among ZCTAs with cases. These phenomena resulted in the spatial distributions of both diseases exhibiting overdispersion and an excess of zero case counts. To address each of these issues, the zero-inflated negative binomial (ZINB) regression method was used. ZINB regression is widely used in the literature, as it addresses overdispersion by switching from traditional Poisson regression to negative binomial regression, and addresses the excess zero case counts via a separate but simultaneous logit modeling process. The separate logit modeling process is justified by the assertion that the process generating zero occurrence of an outcome is different than the process determining the magnitude of positive counts. Recent published studies in tick ecology use ZINB regression to spatially examine the magnitude of infected tick density.
wherein the process generating a zero occurrence is related with the host’s ability to live in a specific environment.

When examining rates of a vector borne disease in humans using ZINB regression, the process causing zero occurrence of disease can be logically linked to the presence or absence of the disease-causing pathogen in the environment (e.g. presence of *I. scapularis* infected with *A. phagocytophilum* or *B. microti*). Arab (2015) used zero-inflated models to predict cases of Lyme disease in Illinois, but claimed that due to low Lyme disease prevalence in Illinois, zero occurrence of Lyme disease is not from a separate process or detectability issue which would justify use of a zero-inflated model. This study benefits by examining two tick-borne diseases that have been increasing in incidence and are localized to specific regions in NYS. The separate processes problem is addressed via spatially contrasting the excess zero disease counts appearing in the western and central regions of NYS and disease counts of varying magnitude on Long Island and in the Hudson Valley region. Further, this study is also able to address the separate processes problem by spatiotemporally examining whether the disease-causing pathogen was present in the environment, and if so, the time period in which the pathogen emerged. Other studies in the literature examining human incidence of tick-borne disease do not examine pathogen presence in the environment before ZINB modeling, and although it is not required, it provides justification for the use of the ZINB model.

The violin plots in Figure 3 show that the bimodal distribution of human incidence of both diseases shift away from the zero peak when a pathogen is present in the environment and shift further away from the zero peak the earlier the pathogen emerged. Interestingly, the results of the pairwise Wilcoxon rank sum tests (Table 2) indicate that only human babesiosis showed a significant difference in mean incidence between groups characterized by the temporal
emergence of their corresponding pathogen. The lack of a statistical difference in mean incidence values of HGA between the two different temporal pathogen emergence groups could be due to the inability of the PCR assays utilized to differentiate between two predominant pathogenic and nonpathogenic genetic variants of *A. phagocytophilum*. The variant Ap-v1 does not cause disease in humans and has a primary reservoir of white-tailed deer (*Odocoileus virginianus*), while the variant Ap-ha is pathogenic to humans and has reservoirs of white-footed mice (*Peromyscus leucopus*) and Eastern chipmunks (*Tamias striatus*). Studies in Ontario, Canada show a shift in the proportion of *A. phagocytophilum* infections in *I. scapularis* toward the Ap-ha variant and away from the Ap-v1 variant when comparing the populations before and after 2010. As Ontario neighbors NYS, it is plausible that populations of *I. scapularis* in NYS are exhibiting a similar shift in population dynamics. This phenomenon would result in the mischaracterization of collection sites, and in turn, bias the result of the Wilcoxon rank sum test towards the null hypothesis. Figure 2 illustrates this effect, as it shows collection sites throughout NYS that have had positive PCR results for *A. phagocytophilum* in *I. scapularis*, while also exhibiting low to zero incidence of HGA in their respective ZCTAs. Collection sites with *B. microti*-infected *I. scapularis* are less widespread in NYS compared to HGA, and the presence/absence of cases of human babesiosis at the ZCTA level seem to follow the spatial trend of pathogen presence/absence.

Finally, ZINB regression models provided the means to compare how HGA and human babesiosis are affected by climate and landscape characteristics. Variables from individual ZINB models can be broken down into “consonant”, “dissonant” and “neutral” trends due to the simultaneous interpretation of the count and zero-inflated models. The count model is interpreted as a standard negative binomial model, while the zero-inflated model is assessing the
odds of being a “structural zero”. A structural zero is an observation which can only be a zero count. In this way, the count and zero-inflated portions of the ZINB model have opposite in directionality when interpreting the effects of the variables on the outcome. Consonant trends are variables whose coefficients have opposite signs between the count and zero-inflated models. Dissonant trends are variables whose coefficients have the same signs in the count and zero-inflated models. Neutral variables are variables that are only significant in either the count or zero-inflated model. Variables with consonant trends in the HGA model included elevation and percent mixed/deciduous forest cover, while the human babesiosis model’s only consonant variable was percent mixed/deciduous forest cover. Variables with dissonant trends in the HGA model included percent urban land cover change and minimum winter temperature, while the human babesiosis model’s only dissonant variable was minimum winter temperature. All remaining variables in both models were considered neutral.

Elevation was negatively associated with magnitude of cases in the HGA [IRR=0.52, 95% CI = 0.50-0.54] and human babesiosis models [IRR=0.37, 95% CI = 0.34-0.40], which is in agreement with current literature regarding Lyme disease and tick density. The negative association between elevation and both HGA and human babesiosis in NYS is likely due to a combination of environmental and ecological factors. Generally, mean temperatures are lower at higher elevations (Figure 6), decreasing I. scapularis questing and development rates. Further, white-tailed deer tend to congregate at lower elevations, depriving adult-stage I. scapularis at higher elevations of its primary reproductive host. Elevation was positively associated with being a structural zero in the zero-inflated portion of the HGA model [OR=2.04, 95% CI = 1.74-2.40], which agrees with the findings of Diuk Wasser et al (2010). Conversely, elevation was not significant in the zero-inflated portion of the human babesiosis model. The null result for human
babesiosis is likely due to fewer ZCTAs with non-zero incidence (Figures 2 and 6), resulting in a broader range of elevation at which risk of human babesiosis does not exist.

The percent of a ZCTA covered by mixed/deciduous forest was positively associated with HGA [IRR=1.49, 95% CI = 1.45-1.52] and human babesiosis [IRR=1.56, CI = 1.52-1.60] in the count portions of each model. The percent mixed/deciduous forest variable also exhibited a consonant effect, as it was negatively associated with being a structural zero case of HGA [OR=0.55, 95% CI = 0.49-0.61] and human babesiosis [OR=0.57, 95% CI = 0.49-0.66]. Deciduous forests are the preferred habitat for *I. scapularis* and their hosts, and thus, ZCTAs that are more covered by these forest types provide an increased risk for tick encounters.

Land cover change may have varying effects on the risk for tick-borne disease due to fluctuations in *I. scapularis* and *P. leucopus* populations, as well as the creation of forest edge. The results of the ZINB models showed significant associations with both the forest and urban land cover change variables, but the interpretability of the results is limited due to the structure of the NLCD 2016 Land Cover Change Index data. The NLCD 2016 Land Cover Change Index combines the designation of areas that changed either to or from specific land cover classes during the period of 2001 to 2016, which presents a significant barrier in determining the exact effect that specific land cover changes have. The count portions of the final models showed an increase in the percent of a ZCTA that changed to or from forest land cover increased the risk for HGA [IRR=10.11, 95% CI = 7.38-13.85] and human babesiosis [IRR=6.13, 95% CI = 4.15-9.07], which agrees with current literature. A higher percentage of land area changing to or from forest land cover type would result in either more suitable *I. scapularis* habitat or new forest edge, respectively. A change to or from urban land cover may have a similar effect, but the HGA and human babesiosis models are inconclusive. The HGA
model shows a dissonant effect for percent change in urban land cover variable, while the human babesiosis model has a neutral effect. Percent urban land cover change has a negative association with magnitude of cases of HGA in the count portion of the model [IRR=0.38, 95% CI = 0.32-0.44], while the odds of being a structural zero of HGA are decreased with an increase in percent of urban land cover change in the zero-inflated portion of the model [OR=0.23, 95% CI = 0.08-0.66]. Similarly, the odds of being a structural zero case of human babesiosis decrease with an increase in the percent urban land cover change [OR = 0.24, 95% CI = 0.08-0.66]. It is likely that the inability to differentiate between a change to or from the urban land cover classes is driving the dissonant effect in the HGA model. The neutral effect in the human babesiosis model could be due to the different geography of the disease, as the southeastern portions of NYS have higher urbanicity and higher incidence of babesiosis (Figure 2). New data that separates the land cover class designations by direction of change may help elucidate the dissonant and neutral effects of both the HGA and human babesiosis models.

Minimum temperature was used to estimate the effect that low temperatures have on *I. scapularis* populations infected with both *A. phagocytophilum* and *B. microti*. Neelankata et al. (2010) showed that *I. scapularis* infected with *A. phagocytophilum* express the iafgp gene, which encodes an antifreeze glycoprotein that allows *I. scapularis* to survive better in colder temperatures. However, the effect of iafgp expression on *I. scapularis* overwintering survival in nature has not been demonstrated in the scientific literature. The count portions of both final models exhibited a negative association with minimum winter temperature and case counts of HGA [IRR = 0.54, 95% CI = 0.51–0.57] and human babesiosis [IRR = 0.77, 95% CI = 0.71-0.82]. The IRRs of both count models showed that lower minimum winter temperatures resulted in a larger increase of HGA cases than human babesiosis cases. This relationship may partly be
explained by the expression of the *iafgp* gene, allowing HGA incidence to increase in more northern latitudes. The zero-inflated portions of both the HGA [OR = 0.75, 95% CI = 0.65-0.87] and human babesiosis [OR = 0.39, 95% CI = 0.32–0.48] models exhibited a dissonant effect, likely because the coldest temperatures in the Adirondack mountain region exhibit zero cases of both diseases (Figures 2 and 6).

Winter precipitation was only significant in the count portions of both models, where it was negatively associated with HGA case counts [IRR = 0.97, 95% CI = 0.96-0.98] and positively associated with human babesiosis case counts [IRR = 1.12, 95% CI = 1.09-1.14]. The opposite signs in each count model may be due to the higher winter precipitation on Long Island, a location with generally low snowfall. PRISM does not have snowfall data available, thus, mean precipitation from November to March was used to estimate mean winter snowfall, likely including rainfall in areas where snow is less common. In areas with colder temperatures, *I. scapularis* remain dormant under the winter snow and use it to insulate themselves from the colder weather and adverse fluctuations in humidity. Due to the insulating effects of snow, it is likely that snowfall would have some effect modification with minimum winter temperatures. That is, in areas where minimum winter temperatures are low, an increase in snowfall would likely result in an increase in *I. scapularis* over-winter survival, while in areas where minimum winter temperatures and humidity are higher, such as the maritime climate of Long Island, the effect of snow on tick survivability may be negligible. Future studies should aim to examine only snowfall, and potential effect modification with minimum winter temperatures.

Model diagnostics using rootograms revealed that the use of the ZINB model was appropriate for the HGA and human babesiosis data (Figure 4). However, when comparing the fit of the spatially adjusted and spatially unadjusted models using rootograms, the differences are
negligible. When visually examining the results of kriging interpolations of residuals from the spatially adjusted and unadjusted models, the spatially adjusted models revealed that the value of residuals were decreased in some areas of NYS for both models (Figure 5). Figure 5 shows that the spatially adjusted HGA model reduces the relative magnitude of residuals in the upper Hudson valley region of NYS, where HGA incidence is the highest. Additionally, the spatially adjusted babesiosis model reduces the relative magnitude of residuals in the lower Hudson valley region of NYS, as well as areas along the United States-Canadian border and the eastern portion of Lake Ontario. These model diagnostics further justified the use of the autocovariate weights in the modeling process.

The limitations of this study are evident in the general interpretability of the results. Some variables examined are not specific enough to make inference on the cause-effect relationship between environment and disease, while others are clouded by the focality of higher rates of disease when compared to statewide environmental conditions. Additionally, some variables were excluded from the start of the analysis due to high collinearity, which is expected in a study examining the magnitude and direction of specific environmental variables on rates of disease. The association that NYS climate, weather and landscape features have with HGA and human babesiosis is complex and understudied. Scientific studies on tick-borne diseases generally examine the relationship of the environment with rates of Lyme disease or tick population density. It is important to note that not all tick-borne diseases have the same scope, and the pathogens that infect *I. scapularis* have different ecological dynamics and spatial extents. As tick-borne disease continues to rise, the scientific community should attempt to elucidate the relationship the environment has on each tick-borne disease individually, to better guide disease-specific prevention policies and public education efforts.
Tables and Figures:

Table 1: Demographic distributions, weighted incidence rates per 100,000 population and weighted exact rate ratio tests for human granulocytic anaplasmosis and human babesiosis.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Human Granulocytic Anaplasmosis (n=4297)</th>
<th>Incidence per 100,000</th>
<th>Human Babesiosis (n=2986)</th>
<th>Incidence per 100,000</th>
<th>Rate Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>80 (1.86%)</td>
<td>1.02</td>
<td>31 (1.04%)</td>
<td>0.39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>10-19</td>
<td>102 (2.37%)</td>
<td>1.09</td>
<td>38 (1.27%)</td>
<td>0.41</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>20-29</td>
<td>176 (4.09%)</td>
<td>2.04</td>
<td>77 (2.58%)</td>
<td>0.89</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>30-39</td>
<td>281 (6.54%)</td>
<td>3.67</td>
<td>168 (5.63%)</td>
<td>2.20</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>40-49</td>
<td>458 (10.66%)</td>
<td>4.68</td>
<td>326 (10.92%)</td>
<td>3.33</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>50-59</td>
<td>873 (20.32%)</td>
<td>8.75</td>
<td>602 (20.16%)</td>
<td>6.03</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>60-69</td>
<td>1084 (25.23%)</td>
<td>15.43</td>
<td>732 (24.51%)</td>
<td>10.42</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>70-79</td>
<td>808 (18.80%)</td>
<td>20.24</td>
<td>648 (21.70%)</td>
<td>16.24</td>
<td>0.0002</td>
</tr>
<tr>
<td>80+</td>
<td>433 (10.08%)</td>
<td>14.24</td>
<td>362 (12.12%)</td>
<td>11.91</td>
<td>0.0390</td>
</tr>
<tr>
<td>Missing</td>
<td>2 (0.05%)</td>
<td>-</td>
<td>2 (0.07%)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Mean (SD)</th>
<th>Median [Min, Max.]</th>
<th>Sex</th>
<th>Male</th>
<th>61.39%</th>
<th>2638 (74.21%)</th>
<th>7.98</th>
<th>1963 (57.97%)</th>
<th>5.94</th>
<th>&lt;0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1659 (38.61%)</td>
<td>4.84</td>
<td>1659 (38.61%)</td>
<td>4.84</td>
<td>1023 (34.26%)</td>
<td>2.99</td>
<td>1023 (34.26%)</td>
<td>2.99</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>White</th>
<th>3244 (75.49%)</th>
<th>5.93</th>
<th>1786 (59.81%)</th>
<th>3.26</th>
<th>&lt;0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>34 (0.79%)</td>
<td>0.58</td>
<td>82 (2.75%)</td>
<td>1.39</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>1 (0.02%)</td>
<td>0.41</td>
<td>5 (0.17%)</td>
<td>2.07</td>
<td>0.4375</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>35 (0.82%)</td>
<td>1.45</td>
<td>66 (2.21%)</td>
<td>1.74</td>
<td>0.00106</td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>1 (0.02%)</td>
<td>5.63</td>
<td>4 (0.13%)</td>
<td>22.51</td>
<td>0.4375</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>81 (1.89%)</td>
<td>2.02</td>
<td>177 (5.93%)</td>
<td>4.42</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>901 (20.97%)</td>
<td>-</td>
<td>866 (29.00%)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Hispanic or Latino</th>
<th>107 (2.48%)</th>
<th>0.18</th>
<th>234 (7.84%)</th>
<th>0.39</th>
<th>&lt;0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hispanic/Non-Latino</td>
<td>2659 (61.88%)</td>
<td>39.68</td>
<td>1267 (42.31%)</td>
<td>18.91</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1531 (35.63%)</td>
<td>-</td>
<td>1485 (49.73%)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Alive</th>
<th>3189 (74.21%)</th>
<th>-</th>
<th>1731 (57.97%)</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dead</td>
<td>15 (0.35%)</td>
<td>-</td>
<td>12 (0.40%)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>215 (5.00%)</td>
<td>-</td>
<td>249 (8.34%)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>878 (20.43%)</td>
<td>-</td>
<td>994 (33.29%)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

*Shaded cells indicate statistical significance at α = 0.05*
Table 2: Results of Kruskal-Wallis test and pairwise Wilcoxon rank sum tests on mean incidence of human granulocytic anaplasmosis and human babesiosis (2013-2018) for categories of respective pathogen emergence.

<table>
<thead>
<tr>
<th>Site Category</th>
<th>Kruskal-Wallis test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. No PCR positives</td>
<td>X</td>
</tr>
<tr>
<td>2. PCR positives found before 2013</td>
<td>0.0009 X</td>
</tr>
<tr>
<td>3. PCR positives found after 2013</td>
<td>&lt;0.0001 0.05734 X</td>
</tr>
</tbody>
</table>

Table 3: Final zero-inflated negative binomial regression models of environmental covariates and case counts of human granulocytic anaplasmosis and human babesiosis.

| Negative Binomial | OR 95% CI | Estimate | Std error | Pr >|t| |
|-------------------|----------|----------|-----------|-----|
| Intercept | 0.00 | 0.00-0.00 | 0.00 | 0.00 |
| Mean elevation (m)a | 0.52 | 0.50-0.54 | 0.65 | 0.02 | <0.0001 |
| Percent forest coverb | 1.49 | 1.45-1.52 | 0.40 | 0.01 | <0.0001 |
| Percent forest land cover changeb | 10.11 | 7.38-13.65 | 2.31 | 0.16 | <0.0001 |
| Percent urban land cover changeb | 0.38 | 0.32-0.44 | 0.08 | 0.00 | <0.0001 |
| Mean winter precipitation (cm)b | 0.97 | 0.96-0.98 | 0.03 | 0.01 | <0.0001 |
| Mean winter minimum temperature (°C) | 0.54 | 0.51-0.57 | 0.62 | 0.02 | <0.0001 |
| Autocovariate | 1.00 | 1.00-1.00 | 0.00 | 0.00 | <0.0001 |
| Logit | OR 95% CI | Estimate | Std error | Pr >|t| |
| Intercept | 0.01 | 0.00-0.11 | 0.25 | 0.50 |
| Mean elevation (m)a | 2.04 | 1.74-2.40 | 0.71 | 0.08 | <0.0001 |
| Percent forest coverb | 0.55 | 0.49-0.61 | 0.60 | 0.06 | <0.0001 |
| Percent urban land cover changeb | 0.23 | 0.18-0.28 | 1.48 | 0.54 | 0.0003 |
| Mean winter minimum temperature (°C) | 0.75 | 0.65-0.87 | 0.85 | 0.77 | 0.0002 |
| Autocovariate | 1.00 | 1.00-1.00 | 0.00 | 0.00 | <0.0001 |

| Logit | OR 95% CI | Estimate | Std error | Pr >|t| |
| Intercept | 0.01 | 0.00-0.11 | 0.25 | 0.50 |
| Mean elevation (m)a | 2.04 | 1.74-2.40 | 0.71 | 0.08 | <0.0001 |
| Percent forest coverb | 0.55 | 0.49-0.61 | 0.60 | 0.06 | <0.0001 |
| Mean winter minimum temperature (°C) | 0.75 | 0.65-0.87 | 0.85 | 0.77 | 0.0002 |

a variable divided by 100 for interpretability, one unit in table is equal to 100 units
b variable divided by 10 for interpretability, one unit in table is equal to 10 units
Figure 1: Incidence of human granulocytic anaplasmosis and human babesiosis per 100,000 individuals aggregated to zip code tabulation area in NYS excluding New York City (2013, 2015 and 2017).
Figure 2: Map of NYS with incidence of human granulocytic anaplasmosis and human babesiosis per 100,000 population by zip code tabulation area (2013-2018), overlaid with tick collection site locations categorized by temporal pathogen emergence of *Anaplasma phagocytophilum* and *Babesia microti*.
Figure 3: Violin plots comparing temporal emergence of disease-causing pathogen at collection sites and respective incidence of human disease per 100,000 population at zip code tabulation area level (2013-2018).

*Red circles correspond to mean incidence of human granulocytic anaplasmosis and human babesiosis at zip code tabulation area level*
Figure 4: Rootograms of spatially adjusted and unadjusted human granulocytic anaplasmosis and human babesiosis final models.
Figure 5: Kriging interpolations of residuals from spatially adjusted and unadjusted human granulocytic and human babesiosis final models.
Figure 6: Environmental covariates with statistically significant associations to either human granulocytic anaplasmosis or human babesiosis

A – Mean winter minimum temperature
B – Mean elevation
C – Percent mixed/deciduous forest cover
D – Percent urban land cover change
E – Percent forest land cover change
F – Mean winter precipitation

New York State shapefile gathered from NY’s GIS Clearinghouse
References


45. Centers for Disease Control and Prevention. Surveillance for Ixodes Scapularis and Pathogens Found in This Tick Species in the United States.


47. ESRI. ArcMap. Redlands, California. USA: ESRI; 2018.


54. ESRI. *Spatial Analyst Tools*. Redlands, California. USA: ESRI


