Quantifying the relationship between human Lyme disease and *Borrelia burgdorferi* exposure in domestic dogs

Yan Liu,[1] Shila K. Nordone,[2] Michael J. Yabsley,[3,4] Robert B. Lund,[5] Christopher S. McMahan,[5] Jenna R. Gettings[3,5]

[1]School of Community Health Sciences, University of Nevada, Reno, NV; [2]Department of Molecular and Biomedical Sciences, Comparative Medicine Institute, North Carolina State University, College of Veterinary Medicine, Raleigh, NC; [3]Southeastern Cooperative Wildlife Disease Study, Department of Population Health, College of Veterinary Medicine, University of Georgia, Athens, GA; [4]Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA; [5]School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC, USA

Correspondence: Jenna R. Gettings, Southeastern Cooperative Wildlife Disease Study, Department of Population Health, College of Veterinary Medicine, University of Georgia, 589 D.W. Brooks Drive, Athens, GA 30602, USA.
Tel.: 706.542.1741 - Fax: 706.542.5865.
E-mail: Jenna.Gettings@uga.edu

Key words: Lyme disease; Canine sentinel; *Borrelia burgdorferi*; USA.

See online Appendix for a supplementary Figure.

Acknowledgements: IDEXX Laboratories, Inc. is thanked for their data contribution. The Companion Animal Parasite Council conceptualized the need for this analysis. Steve Meshnick is thanked for his advice during the conceptualization of the study.

Contributions: all authors made substantial contributions to the manuscript. YL, CSM, and RBL performed the statistical analysis and contributed to manuscript preparation. SKN, JRG, RBL, and MJY wrote parts of the manuscript. All authors have read and approve of the submission.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: YL and RBL thank National Science Foundation Grant DMS 1407480. CSM was partially supported by Grant R01 AI121351 from the National Institutes of Health. JRG is supported by The Boehringer Ingelheim Vetmedica-CAPC Infectious Disease Postdoctoral Fellowship.

Conference presentation: this work was partially presented at the 15th International Conference on Lyme Borreliosis and Other Tick-Borne Diseases, 2018 Sep 11-14, Atlanta, Georgia, USA.

Received for publication: 6 November 2018.
Revision received: 15 January 2019.
Accepted for publication: 20 January 2019.

©Copyright Y. Liu et al., 2019
Licensee PAGEPress, Italy
Geospatial Health 2019; 14:750
doi:10.4081/gh.2019.750

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Abstract

Lyme disease (LD) is the most common vector-borne disease in the United States. Early confirmatory diagnosis remains a challenge, while the disease can be debilitating if left untreated. Further, the decision to test is complicated by under-reporting, low positive predictive values of testing in non-endemic areas and travel, which together exacerbate the difficulty in identification of newly endemic areas or areas of emerging concern. Spatio-temporal analyses at the national scale are critical to establishing a baseline human LD risk assessment tool that would allow for the detection of changes in these areas. A well-established surrogate for human LD incidence is canine LD seroprevalence, making it a strong candidate covariate for use in such analyses. In this paper, Bayesian statistical methods were used to fit a spatio-temporal spline regression model to estimate the relationship between human LD incidence and canine seroprevalence, treating the latter as an explanatory covariate. A strong non-linear monotonically increasing association was found. That is, this analysis suggests that mean incidence in humans increases with canine seroprevalence until the seroprevalence in dogs reaches approximately 30%. This finding reinforces the use of canines as sentinels for human LD risk, especially with respect to identifying geographic areas of concern for potential human exposure.

Introduction

Lyme disease (LD), caused by the spirochete *Borrelia burgdorferi*, remains the most commonly reported vector-borne disease in the contiguous United States (US) (Mead, 2015; Rosenberg *et al.*, 2018). *Borrelia burgdorferi* is transmitted among vertebrate host species by ticks of the genus *Ixodes*, with both humans and dogs as incidental hosts (Lane *et al.*, 1991). Although neither humans nor dogs are significant in the maintenance of the pathogen and the subsequent transmission to other ticks (Mather *et al.*, 1994), both species can suffer clinical consequences as a result of infection. Early infection in humans often presents with erythema migrans (red rash) and flu-like symptoms. As the infection becomes disseminated in the body, clinical signs associated with the nervous system, heart and joints may manifest (Steere, 2001). With the exception of erythema migrans, the clinical presentation in dogs is similar involving fever, lethargy, arthritis, joint swelling and lymphadenopathy (Little *et al.*, 2010).

Cases affecting the heart, kidneys and the nervous system have been reported, but these manifestations are not commonly reported (Littman *et al.*, 2006; Little *et al.*, 2010; Detmer *et al.*, 2016). The potential severity and chronicity of LD, in humans motivates the need for an accurate understanding of its incidence and distribution, with the main objective being to facilitate early diagnosis and allocation of educational resources necessary for disease prevention.

The Centers for Disease Control and Prevention (CDC) in the US recognizes standard 2-tiered testing (STTT) as the gold standard for confirmatory diagnosis of LD recommending its use since 1995. The first-tier STTT utilizes an immunofluorescence assay or enzyme immunoassay that, when positive or equivocal, should be followed by a second-tier immunoblot (IgM and/or IgG) test (Schriefer, 2015). Current recommendations to initiate STTT are driven by pretest probability in the clinical setting, with exposure history being a critical factor in deciding to test (Moore *et al.*, 2016). Despite human LD being notifiable in the US, under-reporting of infection has been cited as a concern (Nelson *et al.*, 2015; Tseng *et al.*, 2015; Moore *et al.*, 2016; White *et al.*, 2016), most notably in paediatric patients (White *et al.*, 2016). Under-reporting is common in most passive infectious disease surveillance systems, particularly when only high-risk regions are studied, and data collections are not representative of the entire country. Surveillance of human LD is further complicated by the non-specific nature of some clinical symptoms and the pitfalls of STTT (Moore *et al.*, 2016). The end result has been a reported potential 3- to 12-fold discrepancy ratio between the number of CDC-confirmed cases and the number of infections actually occurring in the population (Hinckley *et al.*, 2014; Nelson *et al.*, 2015). Furthermore, this under-reporting may delay recognition of newly endemic regions.

Domestic dogs can serve as a proxy for the risk of exposure to infected ticks for humans. Although the risk of exposure may vary between humans and dogs, an infected dog has typically been exposed to all the necessary conditions for transmission of the pathogen to an incidental host, including humans. These factors include those associated with the vector such as density of infected nymphs (Pepin *et al.*, 2012; Eisen *et al.*, 2016b), presence of reservoir and maintenance hosts (Ley *et al.*, 1995; Schauber *et al.*, 2005; Kilpatrick *et al.*, 2014), climate (Schauber *et al.*, 2005; Moore *et al.*, 2014; Watson *et al.*, 2017), landscape and environment (Ley *et al.*, 1995; Pepin *et al.*, 2012; Tuite *et al.*, 2013; Seukep *et al.*, 2015; Walter *et al.*, 2016), demographics and socioeconomics (Tuite *et al.*, 2013; Seukep *et al.*, 2015) and geographical location (Tuite *et al*., 2013). The use of pathogen exposure in dogs as an indicator of human incidence has been shown to be sound in several small-scale studies (Lindenmayer *et al.*, 1991; Rand *et al.*, 1991; Guerra *et al.*, 2001; Johnson *et al.*, 2004; Duncan *et al.*, 2005; Stone *et al.*, 2005) and more recently at the national level (Mead *et al.*, 2011).

While the aforementioned studies are supportive of the concept that canine seroprevalence can be used to assess risk of human exposure to infected ticks, their design and methodology limit actionable decision-making by human healthcare providers. To the point, analyses in these prior studies focused on either correlation statistics or employed simple linear regression techniques to explore associations. Importantly, previous analyses fail to allow for a thorough examination of the relationship at low canine seroprevalence. In particular, the assumption of a linear association leads to inaccurate conclusions (*e.g.* over- and under-estimation) and fails to describe the dynamic nature of LD over the past decade in which human cases are rising in numbers and expanding in geographic distribution (Kugeler *et al.*, 2015), most notably in transitional regions bordering endemic areas of disease. Dynamic changes in LD have been associated with changes in the range of *I. scapularis* (Eisen *et al.*, 2016a); however, annual assessment of vector and disease distribution remain logistically and financially unfeasible. Dogs may be a more sensitive marker of the presence of *B. burgdorferi* at the leading edge of these changes (Mead *et al.*, 2011) and an understanding how their exposure relates to human cases is therefore of significance. In light of recent observations, this study updates, refines and hones an analysis that was published by Mead and colleagues performed at the state and county level using canine data from 2001-2006 (Mead *et al.*, 2011). Presented here are data from 2012 through 2016, a more contemporary observation of the current state of Lyme disease in the US. In addition, the canine test is used more routinely during this time period compared to the early 2000s, providing a broader sample population in terms of geography and number of dogs tested (over 16 million, compared to the approximately 1 million previously considered).

Data availability has also improved the scope for human LD diagnosis in the last 5 years, in part due to growing awareness of disease incidence and testing (Rosenberg *et al.,* 2018). Recent spatial analysis revealed clusters of human cases near the state borders in Maryland, Pennsylvania, Virginia and West Virginia, regions that border established endemic regions of LD (Hendricks *et al.*, 2017). These researchers highlight the need for exploratory surveillance approaches to mitigate the extent at which state level reporting affects accurate estimation of LD progression. We similarly reasoned that exploratory spatial data analyses are necessary for public health professionals on a national level to establish baseline risk assessment and initiate surveillance to monitor changes in LD occurrence. To that end, we estimated the relationship between canine *B. burgdorferi* seroprevalence (Watson *et al.*, 2017) and human LD incidence (Nelson *et al.*, 2014; NNDSS, 2017) using updated information and a spatio-temporal model that can quantify both linear and non-linear associations, thus avoiding erroneous assumptions regarding the form of the relationship. By demonstrating the strength of the association, particularly in areas of low canine seroprevalence, we can open the discussion on the utility of canine data as an explanatory covariate in spatial analyses that could be used to guide pretest probability and evidence-based testing decisions and identify areas of concern to facilitate public health awareness and educational campaigns.

Materials and Methods

Data structure

The observed canine data consist of test results from 16,595,623 *B. burgdorferi* modified enzyme-linked immunosorbent assays (ELISAs) performed from 2012 to 2016 in the contiguous US based on the following assays: SNAP®4Dx® and SNAP®4Dx® Plus (IDEXX Laboratories, Inc., Westbrook, ME, USA) available from the Companion Animal Parasite Council website (CAPC, 2017). SNAP®4Dx® Plus is a combination test that detects antibodies against *B. burgdorferi*, *Anaplasma* spp., *Ehrlichia* spp., and antigen from *Dirofilaria immitis*. The county in which the veterinary clinic is located is reported with the test result. These data are collated through the use of software that connects clinics to IDEXX, and all test results (negative or positive) captured by the network are added to the database on a monthly basis. The ELISA is a qualitative test that detects antibodies that are seroreactive to the C6 peptide which is based on a region of the outer membrane protein (VlsE) of *B. burgdorferi* (Stillman *et al.*, 2014) and has a reported sensitivity and specificity of 96.7% and 98.8%, respectively. Similar to human infection, detection of seroconversion to C6 antigen in dogs is indicative of disseminated infection (Bacon *et al.*, 2003; Nyman *et al.*, 2006; Levy *et al.*, 2008; Wagner *et al.*, 2012; Embers *et al.*, 2016). Seroconversion can occur as early as 3-4 weeks post-infection if the bacterial burden is high (Wagner *et al.*, 2012); however, in the data used in this study, the time since infection was not known and cannot be determined from the test results. After treatment, antibodies will decline over time (Levy *et al.*, 2008), but persist for months (or years) in untreated animals. It is also important to note that the C6 ELISA does not detect antibodies to current vaccines (Ting Liang *et al.*, 2000), so vaccinated dogs will not test positive with the exception of prior infections or vaccine failure and subsequent disseminated infection.

The predominate population represented by the data presented here was that of owned dogs brought to a veterinarian for care. The reason for testing was not provided and may include wellness screening, suspicion of disease (*e.g.*, heartworm disease, Lyme disease, anaplasmosis, ehrlichiosis), or monitoring response to therapy. Other demographic details, such as age, sex, or breed, were also not available. Travel and testing history were also not available, so location of exposure is not known. At the population level, it is reasonable to make the assumption that most dogs were exposed within the county and during the year of testing. Some observations may, however, have been misclassified by county or year. It is expected that there are few of these misclassifications with most representing animals that could have been exposed in a neighbouring county and/or previous year. Thus, given the strong spatio-temporal dependence, it is expected that these misclassifications will not unduly influence the analysis.

Overall, 1,062,524 tests were positive, yielding an empirical seroprevalence of 6.40%. For the purposes of fitting the model and matching the spatio-temporal granularity of the CDC data, the canine test data were aggregated by county and by year. Figure 1 depicts the overall empirical seroprevalence over the 2012-2016 time period for each county. The second data set consisted of confirmed and probable human cases of LD reported by the CDC from 2012 to 2016 at the county level (CDC, 2017). All tests during this period use the same case definition, which is available on the CDC website (NNDSS, 2017). The overall incidence per 100,000 humans by county for the 2012-2016 time period are depicted in Figure 2.

Statistical models

Our statistical analysis proceeded in two steps. First, a Bayesian spatio-temporal model was fitted to the canine test data to obtain seroprevalence estimates at each county throughout the contiguous US at all years of interest. The second step analyzed the human incidence data via a different model using the seroprevalence estimates from the first step as an explanatory factor. The overarching goal of this analysis was to accurately quantify the association between human incidence and canine exposure preva-

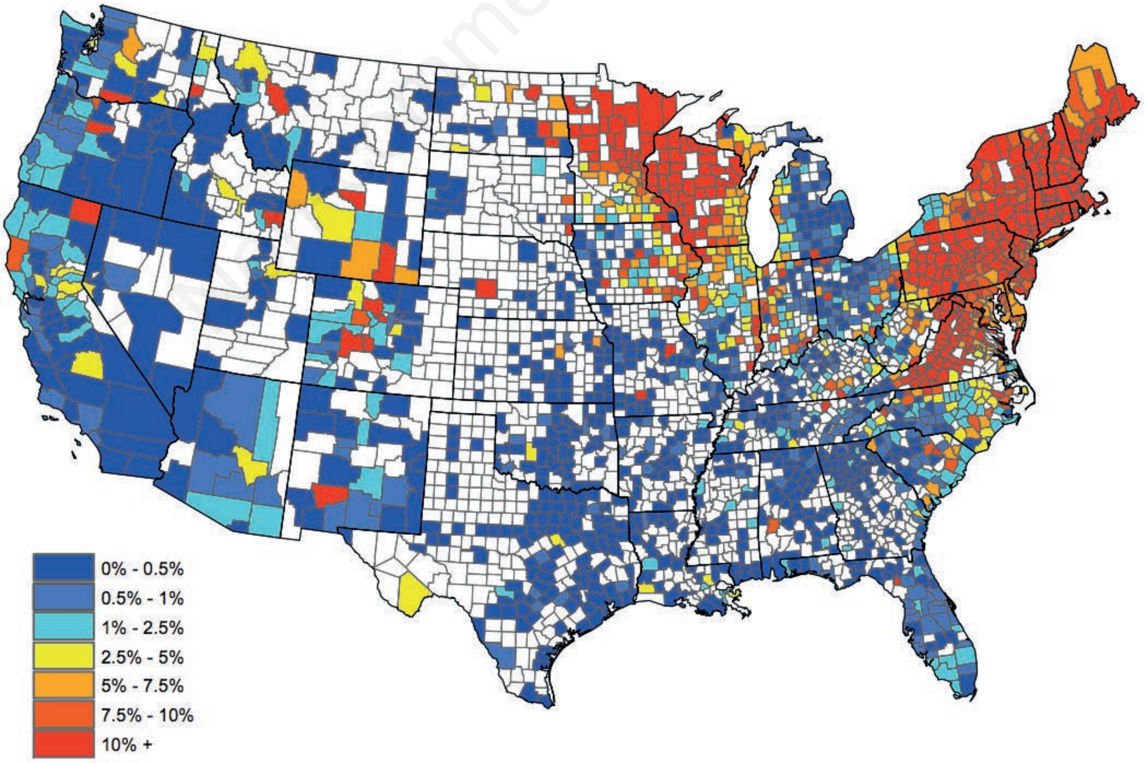

Figure 1. Average empirical canine *Borrelia burgdorferi* seroprevalence for 2012-2016. The overall seroprevalence for the five-year period is displayed for each county reporting data.

lence. In the first step, we used the Bayesian hierarchical spatio-temporal model, the salient features of which have been described by Banerjee *et al.* (2014) and Watson *et al.* (2017). Let Y_{st} and n_{st} denote the number of positive and total number of canine tests in county s during year t, respectively, for counties $s \in \{1, \dots, S\}$ and years $t \in \{1, \dots, T\}$. Our step one model was a space-time Poisson regression:

$$Y_{st} \mid n_{st}, p_{st} \sim \text{Poisson}\ \{n_{st} p_{st}\}, \qquad \text{Eq. 1}$$

$$\log\ \{p_{st}\} = \beta_0 + \xi_{st} \qquad \text{Eq. 2}$$

where β_0 is an intercept parameter and ξ_{st} a spatio-temporal random effect for county s at time t. Here p_{st} is the seroprevalence of *B. burgdorferi* antibodies in dogs in county s at time t. The spatio-temporal dependence in the data are described through a multivariate autoregressive model obeying:

$$\xi_1 = \phi_1; \qquad \text{Eq. 3}$$

$$\xi_t = \varphi \xi_{t-1} + \phi_t, \text{ for } t \in \{2, \dots, T\}; \qquad \text{Eq. 4}$$

$$\phi_t \sim \text{CAR}\ (\tau^2; \rho), \text{ for } t \in \{1, \dots, T\} \qquad \text{Eq. 5}$$

where $\xi_t = (\xi_{1t}, \dots, \xi_{St})'$ and $\phi_t = (\phi_{1t}, \dots, \phi_{St})'$ are random vectors (′ denotes matrix transpose). In this model, the correlation parameter φ lies between (–1,1) and controls the degree of temporal correlation in the model. Values of this parameter near 1(–1) indicate strong positive (negative) temporal correlation and values near zero indicate weaker temporal correlation. Similarly, the parameter ρ controls the spatial correlation and is constrained to lie between (0,1). Values of this parameter close to unity indicate strong positive spatial dependence, while values close to zero indicate weak spatial correlation. Eq. 5 specifies that ϕ_t are independent and identically distributed (iid) random vectors, whose distribution follows a conditional autoregressive (CAR) model. See Besag (1974), Banerjee *et al.* (2014) and Watson *et al.* (2017) for details and the statistical properties of CAR schemes. To complete the Bayesian model formulation, the following prior distributions were used:

$$\beta_0 \sim \text{N}\ (0{,}1000), \qquad \text{Eq. 6}$$

$$\varphi \sim \text{Uniform}\ (-1{,}1); \qquad \text{Eq. 7}$$

$$\rho \sim \text{Uniform}\ (0{,}1); \qquad \text{Eq. 8}$$

$$\tau^{-2} \sim \text{Gamma}\ (0.5{,}0.05). \qquad \text{Eq. 9}$$

These specifications put a diffuse prior on β_0; *i.e.* the prior exerts little influence on the posterior distribution and inferences are based primarily on the data. Uninformative (uniform) prior distributions were used for φ and ρ for the same reasons. The prior

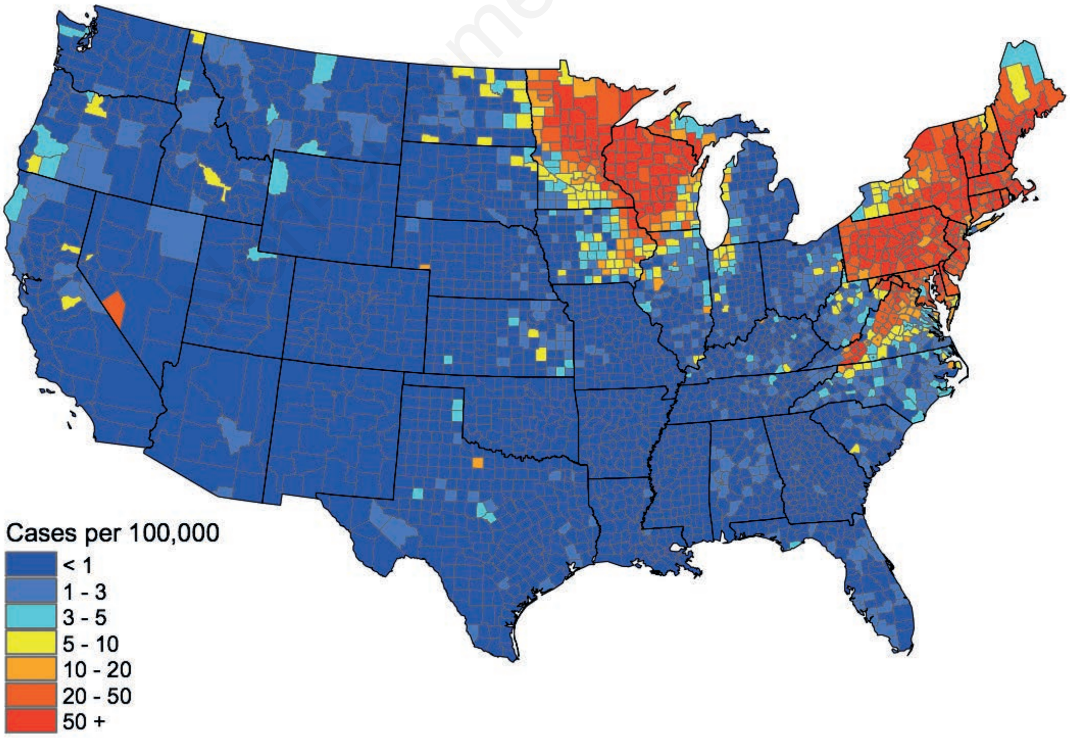

Figure 2. Average empirical human Lyme disease incidence per 100,000 population by county for 2012-2016. County-level data are publicly available from the Centers for Disease Control and Prevention. The overall incidence for the five-year period is displayed for each county reporting data.

for τ^{-2} was chosen to be uninformative with statistical conjugacy (Banerjee *et al.*, 2014) in mind, which facilitates computation.

The unknown parameters in the model were β_0, φ, ρ, τ^{-2}, and ξ_{st}, for $s \in \{1, \dots, S\}$ and $t \in \{1, \dots, T\}$. A posterior sampling algorithm was developed, in the usual statistical fashion, to sample all model parameters and random effects from their posterior distributions. This Markov chain Monte Carlo (MCMC) sampling algorithm uses a combination of Gibbs and Metropolis-Hastings steps. To complete model fitting, Y_{st} for counties not reporting test results were treated as latent variables and were sampled along with the other model parameters. The posterior sampling algorithm was implemented using R and C ++. For more information on Bayesian models and MCMC methods, see Gelman *et al.*, 2014. In what follows, an estimate, denoted by $\hat{p}_{st}$, of p_{st} was obtained as the sample mean from G posterior samples $p_{st}^{(g)}$, for $g \in \{1, \dots, G\}$, where $p_{st}^{(g)} = \exp\{\beta_0^{(g)} + \xi_{st}^{(g)}\}$ with $\beta_0^{(g)}$ and $\xi_{st}^{(g)}$ denoting the gth posterior of β_0 and ξ_{st}, respectively. The Bayesian spatio-temporal model for the second step of this analysis is similar to the first model, but has several key differences. Let X_{st} and m_{st} denote the number of CDC-confirmed human Lyme cases and the population size of county s during year t, respectively, for counties $s \in \{1, \dots, S\}$ and years $t \in \{1, \dots, T\}$. Our model for this step was another Poisson regression:

$$X_{st} \mid m_{st}, r_{st} \sim \text{Poisson}\ \{m_{st} r_{st}\}, \qquad \text{Eq. 10}$$

$$\log\{r_{st}\} = \gamma_0 + g(\hat{p}_{st}) + \delta_{st} \qquad \text{Eq. 11}$$

where γ_0 is an intercept parameter, $g(\cdot)$ an unknown function and δ_{st} a spatio-temporal random effect. In the above, r_{st}, is the risk of a human Lyme disease case in county s at time t. Specifications for the intercept and spatio-temporal random effects are as described for the step one model. The key quantity here was the function $g(\cdot)$, as this quantifies the association between canine and human risk. One may posit that $g(p)$ is non-decreasing in $p \in [0,1]$; *i.e.* the mean incidence in humans increases with canine seroprevalence. The unknown function in Eq. 11 represents an infinite dimensional parameter. To reduce the dimensions of the problem, our approach used regression splines to approximate $g(\cdot)$, a classic statistical practice (Ruppert *et al.*, 2003). In particular, the monotone splines of Ramsay (1988) were utilized:

$$g(p) = \sum_{\ell=1}^{L} \gamma_\ell I_\ell(p), \quad 0 \le p \le 1, \qquad \text{Eq. 12}$$

where the $I_\ell(\cdot)$ are integrated spline basis functions and the γ_ℓ are spline coefficients. Once a user specifies the degree of the polynomials and the knot set, the L basis functions are fully determined; see Ramsay (1988) for further details. The degree controls function smoothness, with adequate smoothness typically being attained with degree 2 or 3. The knot set controls the flexibility of the approximation, with more knots giving more flexibility to adapt to the data. This makes knot selection slightly more tenuous as too few (many) knots may lead to under- (over-) fitting problems. A common statistical approach used to circumvent this problem involves first specifying a large collection of knots and then regularizing the spline coefficients; *i.e.*, the non-useful spline coefficients are drawn toward zero during the estimation procedure. Proceeding in this fashion guards against both over- and under-fitting. To complete the model formulation, the priors in Eq. 6 need to be specified. To perform regularization, the generalized double Pareto shrinkage prior in Armagan *et al.* (2013) was chosen for the γ_ℓ's. The degree of the regression splines was set to be 3 and 5 equally spaced knots were specified between the minimum and maximum of the $\hat{p}_{st}$'s. Model fitting and parameter estimation was conducted in the same fashion as in the step one model. The model estimates are displayed as maps for ease of interpretation after Kriging (implemented in ArcGIS using the default parameters). Kriging is a spatial interpolation method, in which the values are modelled by a Gaussian process; for further details see Deutsch and Journel (1992). To demonstrate that the model and methodology were robust to spurious outcomes in the canine data (likely attributable to travel-related exposure) observed in non-endemic regions, the model was also fitted using a subset of the described data. This subset included only states classified as high incidence by the CDC (CDC, 2017) and those immediately surrounding: Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, Pennsylvania, Delaware, Maryland, Virginia, West Virginia, North Carolina, Tennessee, Kentucky, Ohio, Indiana, Michigan, Illinois, Iowa, Wisconsin, Minnesota, North Dakota, South Dakota and Missouri.

Results

By fitting the model in Eq. 1, two tasks were accomplished. First, canine seroprevalence estimates at every county and year can be obtained, even if a county did not report data during some years. Second, the effect of small sample sizes can be mitigated; *i.e.* patterns attributable to small sample sizes will not unduly influence results. For example, sporadic positive cases (likely attributable to travel) observed in non-endemic areas that report few test results. Figure 3 depicts these estimates after Kriging, averaged over the five study years.

Figure 4 presents the estimated association between canine seroprevalence and reported human incidence along with 95% point-wise credible intervals for all states (blue line) and selected high-incidence and neighbouring states (black line). The estimated association is highly non-linear. In particular, as canine seroprevalence increases from 0% to nearly 10%, human incidence rises quickly. In contrast to a prior study (Mead *et al.*, 2011), a strong association was observed below 5% seroprevalence, as exhibited by the narrow credible intervals. Between seroprevalence of 10% and 30%, the human incidence still rises with increasing seroprevalence, but at a slower rate. The association appears to flatten completely beyond 30%, with no further increase in human incidence. Perhaps most important is the similarity in the two curves, which demonstrates that the model and the methodology are robust to spurious outcomes in the canine data that might be present in

Table 1. Estimates of parameters.

Parameter	Estimate (posterior mean)	95% Credible Interval
β_0	-12.506	(-12.594, -12.411)
ϕ	0.877	(0.852, 0.904)
ρ	0.999	(0.999, 1.000)
τ^{-2}	0.814	(0.762, 0.872)

non-endemic regions. Table 1 provides point estimates and 95% credible intervals for the other model parameters.

County level estimates of the number of cases per 100,000 humans can be obtained from the model for each considered year. Figure 5 shows these estimates as the average of the number of cases per 100,000 humans after Kriging. From this map, the states that contain mostly high-incidence areas (10 cases or more per 100,000 humans) were: Minnesota, Wisconsin, Maine, New Hampshire, Vermont, Massachusetts, New York, Connecticut, Rhode Island, New Jersey, Pennsylvania, Maryland, Delaware and Virginia. Bordering states that contain some areas of high-incidence included the Upper Peninsula of Michigan, Iowa, Illinois, West Virginia and North Carolina, while North Dakota and Ohio had the leading edge of high-incidence on their eastern borders. The remainder of the country was found to have low incidence (<10 cases per 100,000 humans) but areas with slightly increased incidence (>1 cases per 100,000 humans) were seen on the leading edges of the endemic regions: the north-western and south-western corners of Indiana and Michigan, respectively, and along the eastern Lake Michigan shore, most of North Carolina and along the Atlantic coast into South Carolina, Northern California and Southern Oregon. Smaller foci were observed in Florida, Southern Kentucky and at the border of Kentucky and Indiana.

Discussion

This paper quantifies the association between canine seroprevalence and human LD incidence within the contiguous US at the county level. The relationship appears well described by a monotone increasing function (denoted by g previously). A spatio-temporal spline regression model was used to avoid making assumptions about the form of the association. That is, splines are very flexible functions which allow the model to fit the data at all levels of the explanatory covariate, capably illuminating any underlying relationship (Howe *et al.*, 2011). Importantly, regularization was used to prevent the spline model from overfitting, while spatio-temporal random effects were utilized to account for unexplained spatio-temporal autocorrelation. The estimated association displayed in Figure 4, is our primary result. This analysis demonstrates that mean human incidence monotonically increases as dog seroprevalence increases from 0% to 30%. Notable in this analysis is the narrow credible intervals at low seroprevalence, indicating low variability in the model estimates and a strong association between seroprevalence and human incidence. If dogs were unreliable sentinel for human LD, this interval would be wider, and the association would be interpreted as weaker. Above 30% canine seroprevalence, mean human incidence seems to plateau. The reason for this plateau cannot be elucidated from this study alone and, in part, may be caused by geospatial limitations: there are very few regions of the US where canine seroprevalence is >30%, and with limited data points it is difficult to fully describe the association with human incidence at higher seroprevalence. Beyond these limitations, differential behaviours in hyperendemic counties compared to other regions of the country exist that may impact data acquisition, including but not limited to preventative care practices in both humans and dogs.

Estimates derived from Eq. 11 are displayed in Figure 5,

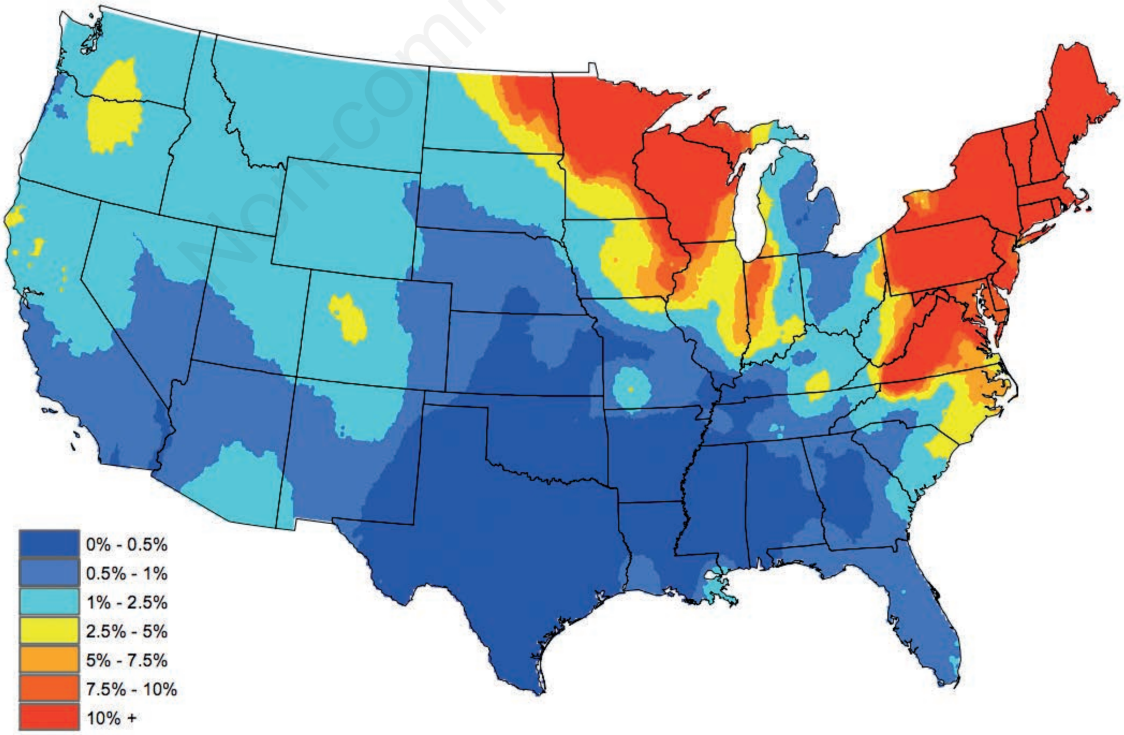

Figure 3. Model-estimated canine seroprevalence, averaged for 2012-2016.

which accurately describes incidence while adjusting for the spatial and temporal random effects unique to each county and demonstrates the ability of seroprevalence to be used as a covariate for spatial models of human incidence. The difference between the model-based estimates and the observed cases of human Lyme disease can be seen in Figure S1 (in the Appendix). The finding that the mean human incidence increases with canine seroprevalence and further that the association is strong at low seroprevalence levels suggest that future research could be aimed at using canine data to monitor areas of transition between endemic and non-endemic regions. This and a previous study by our group (Watson *et al.,* 2017) demonstrate that the widespread availability and timeliness of canine data provides a consistent and spatially uniform surveillance tool in comparison to human data, and therefore enhances the ability to detect areas where *B. burgdorferi* may be diagnosed (either due to travel-acquired infections or due to an area of new transmission of the pathogen). If canine seroprevalence continues to increase, these data would suggest a need to examine the ticks or human case data to determine if endemic transmission was occurring. Future research should focus on the use of a combination of canine data with human case reports and entomological data in defining emergent endemic counties and thus guide pretest probability. For this, a stronger understanding of which specific risk factors influence the association at the county level

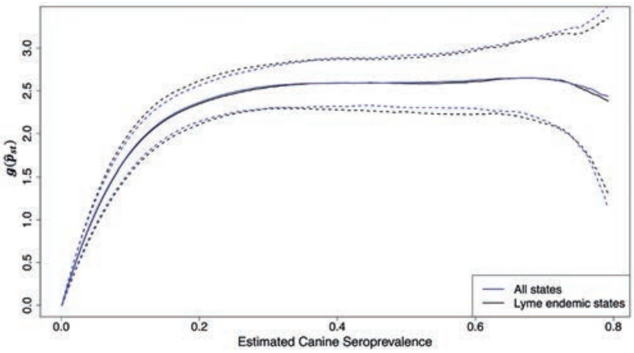

Figure 4. Estimated association between human Lyme disease incidence and canine *B. burgdorferi* seroprevalence. The solid curves depict the estimated $g(\cdot)$, with the dashed lines providing 95% point-wise credible intervals. The measure is unit-less, but represents the logarithm of the proportion of the county's risk (r_{st}) to the county's baseline risk ($r_{st0} = \gamma_0 + \delta_{st}$) at a given canine seroprevalence level, *i.e.* log (r_{st}/r_{st0}) = g($\hat{p}_{st}$). The blue line depicts all 48 contiguous states, while the black line depicts the following high-incidence and neighbouring states: Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, Pennsylvania, Delaware, Maryland, Virginia, West Virginia, North Carolina, Tennessee, Kentucky, Ohio, Indiana, Michigan, Illinois, Iowa, Wisconsin, Minnesota, North Dakota, South Dakota and Missouri.

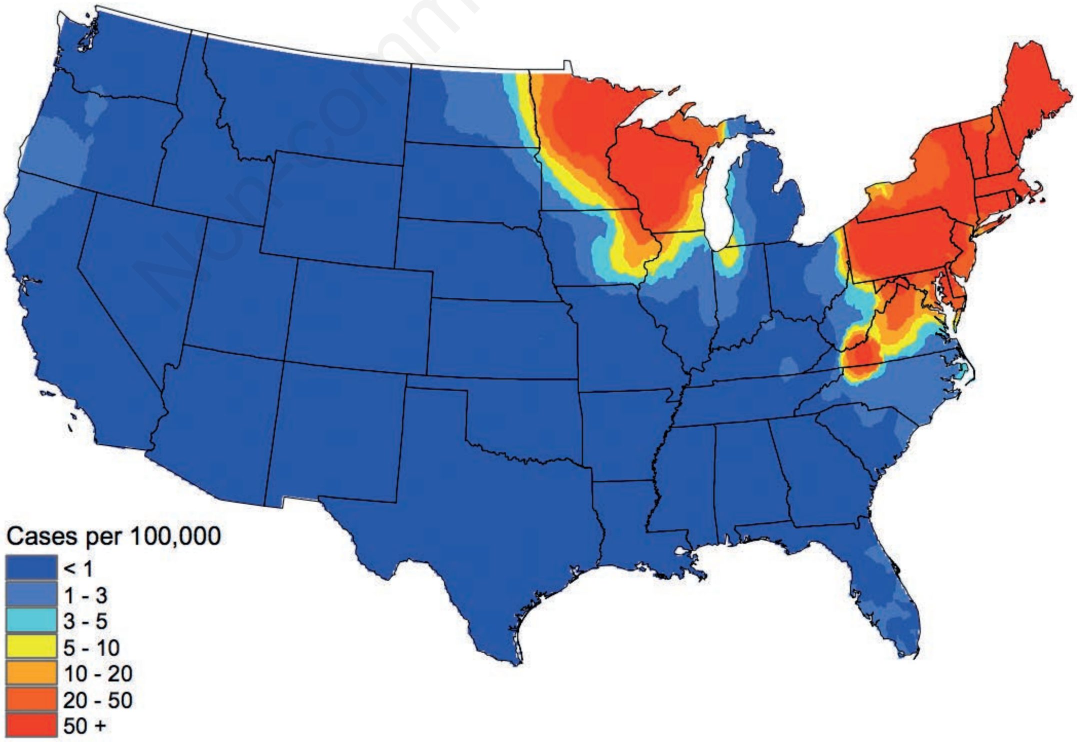

Figure 5. Smoothed map depicting the average estimated counts for human Lyme disease cases per 100,000 population based on canine *B. burgdorferi* seroprevalence.

is needed (Eisen *et al.,* 2003). Our model does not predict areas of temporal emergence and the results should not be interpreted as such. Time trends were not included in the model because of the short five-year study period. In light of this limitation, it is currently recommended that canine data be examined annually. Limitations do exist in the data that preclude simple explanations of the association. These include movement of dogs (Millen *et al.*, 2013) that does not follow that of humans (*i.e.* shelter and rescue dogs); non-human biting *B. burgdorferi* vectors such as *I. affinis* that may bite dogs; cross-reactivity with other borrelial organisms; and false positive test results. Importantly, our model mitigates these factors, as can be seen by comparing the estimated human incidence map (Figure 5) to the observed incidence (Figure 2). However, not all travel-related positive tests should be disregarded. For example, despite Florida being a non-endemic state, many positive tests are reported, and it is easy to conjecture that these are a result of travel between Florida and endemic regions. Since pets do not travel alone, if several canine cases are present within a county, it suggests that the population of that county has some behaviour (travel, camping, hiking) that exposes pets, and therefore likely their owners, to the risk of LD. Using rickettsial disease as an example of a vector-borne disease in which illness in dogs can precede that of their owners (Elchos and Goddard, 2003), monitoring of canine LD prevalence at the county level (CAPC, 2017) and enhanced communication between veterinarians and human health care providers would be optimal when LD is suspected.

The validation of the strength of the association between canine and human LD opens the door to focused studies in areas of concern where the presence of *B. burgdorferi* in ticks is difficult to ascertain and sampling humans is expensive and time consuming. Future considerations will include models that describe the emergence of risk in non-endemic areas. Enhanced sampling of dogs, both owned and non-owned, in combination with the emergence of novel human LD diagnostics (Wormser *et al.*, 2013; Magni *et al.*, 2015; Molins *et al.*, 2015; Embers *et al.*, 2016) that improve upon STTT, have the potential to drive further improvements of national-level risk assessment tools for public health promotion. Data could also be used to guide physicians toward evidence-based use of protective vaccines as they emerge in the marketplace (Izac *et al.*, 2017), with an emphasis on high-risk individuals (children 5-9 and adults 45-59 years of age (CDC, 2017)) and/or potential travel-related exposure. This would ultimately contribute to reducing the financial burden placed on the US health care system by LD (Adrion *et al.*, 2015).

Conclusions

Lyme disease can be difficult to diagnose, and chronic manifestations may be severe and debilitating, but it is preventable with appropriate awareness and protection. The model presented here quantifies the association between canine seroprevalence and human LD incidence. Given the availability and timeliness of canine seroprevalence data, we describe a link to human LD incidence that may ultimately serve as an annual early-warning system in transitional and non-endemic regions facilitating awareness and aid in early and accurate diagnosis.

References

Adrion ER, Aucott J, Lemke KW, Weiner JP, 2015. Health care costs, utilization and patterns of care following Lyme disease. PLoS One 10:e0116767.

Armagan A, Dunson DB, Lee J, 2013. Generalized double pareto shrinkage. Stat Sinica 23:119.

Bacon RM, Biggerstaff BJ, Schriefer ME, Gilmore Jr. RD, Philipp MT, Steere AC, Wormser GP, Marques AR, Johnson BJ, 2003. Serodiagnosis of Lyme disease by kinetic enzyme-linked immunosorbent assay using recombinant VlsE1 or peptide antigens of *Borrelia burgdorferi* compared with 2-tiered testing using whole-cell lysates. J Infect Dis 187:1187-99.

Banerjee S, Carlin B, Gelfand AE, 2014. Hierarchical modeling and analysis for spatial data. Chapman Hall/CRC , Boca Raton, FL, USA.

Besag J, 1974. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Series B Stat Methodol 36:192-236.

Centers for Disease Control and Prevention (CDC), 2017. Lyme disease, data and statistics. Available from: https://www.cdc.gov/lyme/stats/index.html Accessed: July 10, 2017.

Companion Animal Parasite Council (CAPC), 2017. Parasite prevalence maps, canine Lyme disease. Available from: https://capcvet.org/maps/#2016/all/lyme-disease/dog/united-states/ Accessed: July 10, 2017.

Detmer SE, Bouljihad M, Hayden DW, Schefers JM, Armien A, Wunschmann A, 2016. Fatal pyogranulomatous myocarditis in 10 boxer puppies. J Vet Diagn Invest 28:144-9.

Deutsch CV, Journel AG, 1992. Geostatistical software library and user's guide. Oxford University Press, New York, USA. pp 119-147.

Duncan AW, Correa MT, Levine JF, Breitschwerdt EB, 2005. The dog as a sentinel for human infection: prevalence of *Borrelia burgdorferi* C6 antibodies in dogs from southeastern and mid-Atlantic States. Vector Borne Zoonotic Dis 5:101-9.

Eisen RJ, Eisen L, Castro MB, Lane RS, 2003. Environmentally related variability in risk of exposure to Lyme disease spirochetes in northern California: effect of climatic conditions and habitat type. Environ Entomol 32:1010-8.

Eisen RJ, Eisen L, Beard CB, 2016a. County-scale distribution of *Ixodes scapularis* and *Ixodes pacificus* (*Acari: Ixodidae*) in the continental United States. J Med Entomol 53:349-86.

Eisen R, Eisen L, Ogden N, Beard C, 2016b. Linkages of weather and climate with *Ixodes scapularis* and *Ixodes pacificus* (*Acari*: *Ixodidae*), enzootic transmission of *Borrelia burgdorferi*, and Lyme disease in North America. J Med Entomol 53:250-61.

Elchos BN, Goddard J, 2003. Implications of presumptive fatal rocky mountain spotted fever in two dogs and their owner. J Am Vet Med Assoc 223:1450-2.

Embers ME, Hasenkampf NR, Barnes MB, Didier ES, Philipp MT, Tardo AC, 2016. Five-antigen fluorescent bead-based assay for diagnosis of lyme disease. Clin Vaccine Immunol 23:294-303.

Gelman A, Carlin JB, Stern HS, Rubin DB, 2014. Bayesian data analysis. Chapman & Hall/CRC, Boca Raton, FL, USA.

Guerra MA, Walker ED, Kitron U, 2001. Canine surveillance system for Lyme borreliosis in Wisconsin and northern Illinois: geographic distribution and risk factor analysis. Am J Trop Med Hyg 65:546-52.

Hendricks B, Mark-Carew M, 2017. Using exploratory data analysis to identify and predict patterns of human Lyme disease case clustering within a multistate region, 2010-2014. Spat Spatiotemporal Epidemiol 20:35-43.

Hinckley AF, Connally NP, Meek JI, Johnson BJ, Kemperman MM, Feldman KA, White JL, Mead PS, 2014. Lyme disease testing by large commercial laboratories in the United States. Clin Infect Dis 59:676-81.

Howe CJ, Cole SR, Westreich DJ, Greenland S, Napravnik S, Eron Jr JJ, 2011. Splines for trend analysis and continuous confounder control. Epidemiology 22:874.

Izac JR, Oliver LD, Earnhart CG, Marconi RT, 2017. Identification of a defined linear epitope in the ospa protein of the lyme disease spirochetes that elicits bactericidal antibody responses: Implications for vaccine development. Vaccine 35:3178-85.

Johnson L, Ginsberg HS, Zhioua E, Whitworth UG, Markowski D, Hyland KE, 2004. Passive tick surveillance, dog seropositivity, and incidence of human Lyme disease. Vector Borne Zoonotic Dis 4:137-42.

Kilpatrick HJ, Labonte AM, Stafford KC, 2014. The relationship between deer density, tick abundance, and human cases of Lyme disease in a residential community. J Med Entomol 51:777-84.

Kugeler KJ, Farley GM, Forrester JD, Mead PS, 2015. Geographic distribution and expansion of human Lyme disease, United States. Emerg Infect Dis 21:1455-7.

Lane RS, Piesman J, Burgdorfer W, 1991. Lyme borreliosis: relation of its causative agent to its vectors and hosts in North America and Europe. Annu Rev Entomol 36:587-609.

Levy SA, O'Connor TP, Hanscom JL, Shields P, Lorentzen L, Dimarco AA, 2008. Quantitative measurement of C6 antibody following antibiotic treatment of *Borrelia burgdorferi* antibody-positive nonclinical dogs. Clin Vaccine Immunol 15:115-9.

Ley C, Olshen EM, Reingold AL, 1995. Case-control study of risk factors for incident Lyme disease in California. Am J Epidemiol 142:S39-47.

Lindenmayer JM, Onderdonk B, Du W, Marshall A, 1991. Dogs as sentinels for Lyme disease in Massachusetts. Am J Public Health 81:1448-55.

Little SE, Heise SR, Blagburn BL, Callister SM, Mead PS, 2010. Lyme borreliosis in dogs and humans in the USA. Trends Parasitol 26:213-8.

Littman MP, Goldstein RE, Labato MA, Lappin MR, Moore GE, 2006. ACVIM small animal consensus statement on Lyme disease in dogs: diagnosis, treatment, and prevention. J Vet Intern Med 20:422-34.

Magni R, Espina BH, Shah K, Lepene B, Mayuga C, Douglas TA, Espina V, Rucker S, Dunlap R, Emanuel III F, 2015. Application of nanotrap technology for high sensitivity measurement of urinary outer surface protein a carboxyl-terminus domain in early stage Lyme borreliosis. J Transl Med 13:346.

Mather TN, Fish D, Coughlin RT, 1994. Competence of dogs as reservoirs for Lyme disease spirochetes (*Borrelia burgdorferi*). J Am Vet Med Assoc 205:186-8.

Mead PS, 2015. Epidemiology of Lyme disease. Infect Dis Clin North Am 29:187-210.

Mead P, Goel R, Kugeler K, 2011. Canine serology as adjunct to human Lyme disease surveillance. Emerg Infect Dis 17:1710-2.

Millen K, Kugeler KJ, Hinckley AF, Lawaczeck EW, Mead PS, 2013. Elevated Lyme disease seroprevalence among dogs in a nonendemic county: Harbinger or artifact? Vector Borne Zoonotic Dis 13:340-1.

Molins CR, Ashton LV, Wormser GP, Hess AM, Delorey MJ, Mahapatra S, Schriefer ME, Belisle JT, 2015. Development of a metabolic biosignature for detection of early Lyme disease. Clin Infect Disc 60:1767-75.

Moore SM, Eisen RJ, Monaghan A, Mead P, 2014. Meteorological influences on the seasonality of Lyme disease in the United States. Am J Trop Med Hyg 90:486-96.

Moore A, Nelson C, Molins C, Mead P, Schriefer M, 2016. Current guidelines, common clinical pitfalls, and future directions for laboratory diagnosis of Lyme disease, United States. Emerg Infect Dis 22:1169.

National Notifiable Diseases Surveillance System (NNDSS), 2017. Lyme disease (*Borrelia burgdorferi*) 2011 case definition. Available from: https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/2011/. Accessed: July 10, 2017.

Nelson C, Hojvat S, Johnson B, Petersen J, Schriefer M, Beard CB, Petersen L, Mead P, Centers for Disease Control and Prevention, 2014. Concerns regarding a new culture method for *Borrelia burgdorferi* not approved for the diagnosis of Lyme disease. MMWR Morb Mortal Wkly Rep 63:333.

Nelson CA, Saha S, Kugeler KJ, Delorey MJ, Shankar MB, Hinckley AF, Mead PS, 2015. Incidence of clinician-diagnosed Lyme disease, United States, 2005-2010. Emerg Infect Dis 21:1625-31.

Nyman D, Willen L, Jansson C, Carlsson SA, Granlund H, Wahlberg P, 2006. VlsE C6 peptide and IgG ELISA antibody analysis for clinical diagnosis of Lyme borreliosis in an endemic area. Clin Microbiol Infect 12:496-7.

Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, Barbour AG, Hamer S, Diuk-Wasser MA, 2012. Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking *Ixodes scapularis* nymphs in the eastern United States. Am J Trop Med Hyg 86:1062-71.

Ramsay JO, 1988. Monotone regression splines in action. Statist Sci 3:425-41.

Rand PW, Smith RP, Lacombe EH, 1991. Canine seroprevalence and the distribution of *Ixodes dammini* in an area of emerging Lyme disease. Am J Public Health 81:1331-4.

Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, Hooks HH, Partridge SK, Visser SN, Beard CB, Petersen LRP, 2018. Vital signs: Trends in reported vectorborne disease cases – United States and territories, 2004-2016. MMWR Morb Mortal Wkly Rep 67:496-501.

Ruppert D, Wand MP, Carroll RJ, 2003. Semiparametric regression. Cambridge University Press, New York, USA.

Schauber EM, Ostfeld RS, Evans AS, 2005. What is the best predictor of annual Lyme disease incidence: weather, mice, or acorns? Ecol Appl 15:575-86.

Schriefer ME, 2015. Lyme disease diagnosis: Serology. Clin Lab Med 35:797-814.

Seukep S, Kolivras K, Hong Y, Li J, Prisley S, Campbell J, Gaines D, Dymond R, 2015. An examination of the demographic and environmental variables correlated with Lyme disease emergence in Virginia. Ecohealth 12:634-44.

Steere AC, 2001. Lyme disease. N Engl J Med 345:115-25.

Stillman BA, Monn M, Liu J, Thatcher B, Foster P, Andrews B, Little S, Eberts M, Breitschwerdt EB, Beall MJ, Chandrashekar R, 2014. Performance of a commercially available in-clinic ELISA for detection of antibodies against *Anaplasma phagocytophilum*, *Anaplasma platys*, *Borrelia burgdorferi*, *Ehrlichia canis*, and *Ehrlichia ewingii* and *Dirofilaria immitis* antigen in dogs. J Am Vet Med Assoc 245:80-6.

Stone EG, Lacombe EH, Rand PW, 2005. Antibody testing and Lyme disease risk. Emerg Infect Dis 11:722-4.

Ting Liang F, Jacobson RH, Straubinger RK, Grooters A, Philipp MT, 2000. Characterization of a *Borrelia burgdorferi* VlsE invariable region useful in canine Lyme disease serodiagnosis by enzyme-linked immunosorbent assay. J Clin Microbiol 38:4160-6.

Tseng Y-J, Cami A, Goldmann DA, DeMaria Jr A, Mandl KD, 2015. Using nation-wide health insurance claims data to augment Lyme disease surveillance. Vector Borne Zoonotic Dis 15:591-6.

Tuite AR, Greer AL, Fisman DN, 2013. Effect of latitude on the rate of change in incidence of Lyme disease in the United States. CMAJ Open 1:E43-7.

Wagner B, Freer H, Rollins A, Garcia-Tapia D, Erb HN, Earnhart C, Marconi R, Meeus P, 2012. Antibodies to *Borrelia burgdorferi* OspA, OspC, OspF, and C6 antigens as markers for early and late infection in dogs. Clin Vaccine Immunol 19:527-35.

Walter KS, Pepin KM, Webb CT, Gaff HD, Krause PJ, Pitzer VE, Diuk-Wasser MA, 2016. Invasion of two tick-borne diseases across New England: harnessing human surveillance data to capture underlying ecological invasion processes. Proc Biol Sci 283.

Watson SC, Liu Y, Lund RB, Gettings JR, Nordone SK, McMahan CS, Yabsley MJ, 2017. A Bayesian spatio-temporal model for forecasting the prevalence of antibodies to *Borrelia burgdorferi*, causative agent of Lyme disease, in domestic dogs within the contiguous United States. PLoS One 5:e0174428.

White J, Noonan-Toly C, Lukacik G, Thomas N, Hinckley A, Hook S, Backenson P, 2016. Lyme disease surveillance in New York state: An assessment of case underreporting. Zoonoses Public Health 65:238-46.

Wormser GP, Schriefer M, Aguero-Rosenfeld ME, Levin A, Steere AC, Nadelman RB, Nowakowski J, Marques A, Johnson BJ, Dumler JS, 2013. Single-tier testing with the C6 peptide ELISA kit compared with two-tier testing for Lyme disease. Diagn Microbiol Infect Dis 75:9-15.