Meteorological Influences on the Seasonality of Lyme Disease in the United States

Sean M. Moore,* Rebecca J. Eisen, Andrew Monaghan, and Paul Mead

Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado; Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado

Abstract. Lyme disease (Borrelia burgdorferi infection) is the most common vector-transmitted disease in the United States. The majority of human Lyme disease (LD) cases occur in the summer months, but the timing of the peak occurrence varies geographically and from year to year. We calculated the beginning, peak, end, and duration of the main LD season in 12 highly endemic states from 1992 to 2007 and then examined the association between the timing of these seasonal variables and several meteorological variables. An earlier beginning to the LD season was positively associated with higher cumulative growing degree days through Week 20, lower cumulative precipitation, a lower saturation deficit, and proximity to the Atlantic coast. The timing of the peak and duration of the LD season were also associated with cumulative growing degree days, saturation deficit, and cumulative precipitation, but no meteorological predictors adequately explained the timing of the end of the LD season.

INTRODUCTION

Lyme disease is a tick-borne zoonotic disease caused by the bacterial spirochete Borrelia burgdorferi.1 Since its initial description in 1977,2 Lyme disease has been increasing in both incidence and geographic distribution, and it is now the most common vector-transmitted infectious disease in the United States.3 The occurrence of Lyme disease in the United States is concentrated in the northeast from Maine south to Virginia and in the Midwestern states of Wisconsin and Minnesota. Like many vector-transmitted diseases in temperate regions, the occurrence of Lyme disease is highly seasonal. Approximately two-thirds of the cases from 1992 to 2006 had a reported onset date in June, July, or August.3 The seasonality of case occurrence varies geographically with the beginning of the main transmission season (and the peak) occurring earlier in southern endemic states (Virginia and Maryland) and later in the northern endemic states (Maine, New Hampshire, and Minnesota). Seasonality might also vary from year to year, with the timing of the beginning and peak of the period when the majority of infections occur differing depending on environmental factors.

The seasonality of Lyme disease case occurrence is driven in part by the life cycle of Ixodes scapularis, the main vector of Lyme disease spirochetes in the eastern United States. Ixodid ticks have four life stages (egg, larva, nymph, and adult), and only one blood meal per host-seeking life stage.5 The life cycle of I. scapularis typically takes 2 years to complete in the northern United States latitudes where Lyme disease occurrence is most common. Nymphal ticks typically emerge and quest (seek hosts) in the late spring and early summer, although the majority of adults seek a blood meal in late summer and early autumn.4 The majority of human Lyme disease cases are believed to be transmitted by nymphal ticks because of their small size, which prevents timely detection, and the strong overlap in the timing of nymphal activity and Lyme disease occurrence.5 The timing of peak nymphal activity depends on both temperature and moisture. Ixodid tick development rates increase with temperature,6–9 and temperature also influences nymphal questing activity.10–12 Nymphal questing activity may also be reduced when relative humidity is low or saturation deficit is high,11–13 or when precipitation is heavy.

Although influenced by meteorological conditions, tick activity is not the sole determinant of human illness. Human infection requires contact between infectious ticks and susceptible humans, and patterns of human activity can potentially augment or decrease the risk of disease transmission. Late spring through early fall is also the time period when humans are most active outdoors, which increases the contact rates between questing ticks and humans. Nevertheless, it is essential to recognize that human behavior may be independently influenced by temperature and precipitation in ways that also affect disease.

Strategies for preventing Lyme disease include reducing the abundance of ticks in the environment and modifying human behavior to reduce tick exposure and pathogen transmission.14–16 One common control method in the northeast is the area-wide application of acaricides in the peridomestic setting to reduce the number of host-seeking nymphal ticks.17,18 A single springtime application of synthetic acaricide timed to coincide with nymphal emergence can significantly reduce the number of questing ticks for several months.19 Natural, plant-based acaricides, which may be more acceptable to some homeowners,20,21 can also reduce the abundance of I. scapularis nymphs; however, their efficacy begins to wane after several weeks.19,22,23 Knowledge regarding the time period when humans have a high risk of acquiring infection could therefore be used to target the timing of application of conventional or natural acaricides. Public awareness campaigns emphasize the avoidance of high-risk habitat for tick exposure, the use of clothing and repellents to avoid tick bites, and tick checks and bathing after outdoor activity to prevent pathogen transmission.24 A better understanding of the seasonality of Lyme disease could be used to target when these awareness campaigns would be most effective, and also provide the public with a better sense of when risk levels are highest. The timely transfer of information regarding the level of risk to the medical community is also important because early diagnosis and treatment with antibiotics reduces complications and improves clinical outcome.25,26 The goal of our study was to identify the meteorological factors associated with the timing of the primary Lyme disease season, with the goal of using this knowledge to improve the timing of control and prevention efforts.
The first objective of this study was to develop a measure of the seasonality of Lyme disease occurrence in humans by identifying the period or season when the majority of Lyme disease cases occur. This was accomplished using novel methods to calculate the beginning, peak, end, and duration of the season when the majority of cases occur. Next, we identified which meteorological variables are associated with seasonal patterns of Lyme disease, particularly the beginning of the period when the majority of cases occur, across all states where Lyme disease is highly endemic. Other studies of vector-borne diseases have linked seasonal variation in incidence or occurrence to climate and meteorology in general, but have not tried to predict or explain temporal variation in the timing of the seasonal onset and peak, or the length of the season when the majority of cases occur. In addition to examining the association between meteorological factors and seasonality across a broad region of 12 highly endemic states, we also tested whether these associations held for 4, multi-state sub-regions. The meteorological variables connected to the variability in the timing of the Lyme disease season both geographically and annually in different regions of the United States could then be used to help forecast the annual timing of the Lyme disease season in highly endemic areas.

MATERIALS AND METHODS

**Epidemiological data.** Human Lyme disease cases have been voluntarily reported to the Centers for Disease Control and Prevention (CDC) by state and territorial health departments as part of the National Notifiable Disease Surveillance System (NNDSS) since 1991. A total of 275,518 human cases of Lyme disease were reported between 1992 and 2007. During this period, a case of Lyme disease was defined as either 1) a physician-diagnosed erythema migrans rash of ≥ 5 cm in diameter or 2) at least one objective late manifestation (i.e., musculoskeletal, cardiovascular, or neurologic) with laboratory evidence of infection with *B. burgdorferi.* The national surveillance case definition was revised in 2008 to include probable cases, so we limited our analysis to 1992 to 2007. However, we did use the confirmed case data from 2008 to include probable cases, so we limited our analysis to 1992 to 2007. However, we did use the confirmed case data from 2008 to 2007 to evaluate our model results. We analyzed every year with ≥ 100 cases in a state (Table 1; total number of cases = 195,765). Years with fewer than 100 cases in a particular state were excluded because we could not accurately calculate seasonal statistics for that year.

**Seasonal response variables.** To examine seasonal variability in Lyme disease occurrence we calculated the beginning week of the primary Lyme disease season, the peak week of occurrence, and the week ending the primary Lyme disease season. In addition, we calculated the percentage of cases that occurred before, within, and after the primary season. The beginning of the primary Lyme disease season was defined as the week when the number of Lyme disease cases increased (accelerated) at its most rapid pace (i.e., not the week with the largest increase in the number of cases, which occurs a few weeks later in the season, but the week with the maximum percent increase in number of cases over the previous week). This variable is calculated by maximizing the second derivative of the weekly case curve (or the third derivative of the cumulative weekly case curve). The number of cases recorded in a given week is subject to stochastic factors independent of any meteorological drivers, which can inflate the change in the number of cases from one week to the next and lead to errors in calculating the beginning week of the Lyme disease season. To minimize the influence of this stochasticity on our assessment of seasonal variables, a fifth-order polynomial regression function was fit to the cumulative weekly case curve for each year for each state. A fifth-order polynomial was necessary to represent the non-normality in

<table>
<thead>
<tr>
<th>Region</th>
<th>Years</th>
<th>Start (SD)</th>
<th>Peak (SD)</th>
<th>End (SD)</th>
<th>Duration (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>21</td>
<td>22.1 (1.1)</td>
<td>27.6 (0.8)</td>
<td>35.5 (1.1)</td>
<td>13.4 (1.5)</td>
</tr>
<tr>
<td>Maine</td>
<td>6</td>
<td>22.0 (1.8)</td>
<td>27.8 (1.2)</td>
<td>36.5 (1.0)</td>
<td>14.5 (2.1)</td>
</tr>
<tr>
<td>Mass</td>
<td>13</td>
<td>22.2 (0.7)</td>
<td>27.5 (0.7)</td>
<td>35.1 (1.0)</td>
<td>12.9 (1.0)</td>
</tr>
<tr>
<td>NH</td>
<td>2</td>
<td>22.0 (1.4)</td>
<td>27.5 (0.7)</td>
<td>35.5 (0.7)</td>
<td>13.5 (2.1)</td>
</tr>
<tr>
<td>East</td>
<td>66</td>
<td>21.3 (1.2)</td>
<td>27.1 (0.7)</td>
<td>35.1 (0.9)</td>
<td>13.8 (1.7)</td>
</tr>
<tr>
<td>CT</td>
<td>16</td>
<td>21.9 (1.0)</td>
<td>27.2 (0.4)</td>
<td>34.6 (0.7)</td>
<td>12.7 (1.4)</td>
</tr>
<tr>
<td>RI</td>
<td>14</td>
<td>22.1 (1.1)</td>
<td>27.6 (0.9)</td>
<td>35.1 (0.8)</td>
<td>13.1 (1.2)</td>
</tr>
<tr>
<td>NJ</td>
<td>7</td>
<td>20.7 (1.4)</td>
<td>27.0 (0.6)</td>
<td>35.9 (1.3)</td>
<td>15.1 (2.6)</td>
</tr>
<tr>
<td>NY</td>
<td>15</td>
<td>21.1 (0.9)</td>
<td>27.2 (0.4)</td>
<td>35.5 (0.7)</td>
<td>14.3 (1.4)</td>
</tr>
<tr>
<td>PA</td>
<td>14</td>
<td>20.2 (0.9)</td>
<td>26.4 (0.5)</td>
<td>34.9 (0.9)</td>
<td>14.6 (1.3)</td>
</tr>
<tr>
<td>South</td>
<td>26</td>
<td>18.6 (1.3)</td>
<td>25.4 (0.6)</td>
<td>35.5 (2.9)</td>
<td>16.9 (3.4)</td>
</tr>
<tr>
<td>MD</td>
<td>16</td>
<td>19.2 (0.9)</td>
<td>25.6 (0.5)</td>
<td>34.9 (1.3)</td>
<td>15.8 (2.4)</td>
</tr>
<tr>
<td>VA</td>
<td>10</td>
<td>17.7 (1.3)</td>
<td>25.1 (0.7)</td>
<td>36.4 (4.0)</td>
<td>18.7 (4.0)</td>
</tr>
<tr>
<td>Midwest</td>
<td>32</td>
<td>22.3 (1.6)</td>
<td>27.6 (1.5)</td>
<td>34.8 (1.8)</td>
<td>12.4 (2.0)</td>
</tr>
<tr>
<td>MN</td>
<td>16</td>
<td>23.1 (1.5)</td>
<td>27.7 (1.2)</td>
<td>34.3 (1.2)</td>
<td>11.2 (1.6)</td>
</tr>
<tr>
<td>WI</td>
<td>16</td>
<td>21.6 (1.4)</td>
<td>27.4 (1.7)</td>
<td>35.3 (2.1)</td>
<td>13.7 (1.4)</td>
</tr>
<tr>
<td>Mean</td>
<td>145</td>
<td>21.2 (1.8)</td>
<td>27.0 (1.2)</td>
<td>35.2 (1.7)</td>
<td>14.0 (2.6)</td>
</tr>
</tbody>
</table>

*Years are the number of years between 1992 and 2007 where > 100 Lyme disease cases occurred in a state and included in analysis. Start, peak, and end are the week when the Lyme disease season starts, reaches its peak, and ends, respectively; duration in the length of the Lyme disease season in weeks. Numbers in parentheses are standard deviations (in weeks) for each variable.

Over 95% of Lyme disease cases in the United States occurred in 13 states during our study period, primarily in the Northeast and Upper Midwest (Connecticut, Delaware, Maine, Maryland, Massachusetts, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Virginia, and Wisconsin). During the time period of our study, reports from Delaware did not include an illness onset date and were therefore excluded from our analysis. In the remaining 12 states, Lyme disease cases for each state were aggregated on a weekly timescale for each year based on the reported date of illness onset. We analyzed every year with ≥ 100 cases in a state (Table 1; total number of cases = 195,765). Years with fewer than 100 cases in a particular state were excluded because we could not accurately calculate seasonal statistics for that year.
both tails of the annual case curve ($R^2 > 0.994$ for each year; Figure 1). The beginning week was then calculated by maximizing the second-order weekly difference in case counts, which approximates the second-order derivative, from this fifth-order polynomial function (Figure 1C).

The week with the most cases (peak) was also calculated from the fifth-order polynomial function. The end week of the primary Lyme disease season was calculated using the same method as the beginning week calculation, except we identified the week when the maximum deceleration in cases occurred following the annual peak. The seasonal duration (in weeks) was calculated by subtracting the beginning week from the end week.

**Meteorological variables.** All meteorological variables were obtained from primary forcing data for Phase 2 of the North American Land Data Assimilation System (NLDAS-2). The primary forcing data were originally developed to provide a high-quality, long-term dataset to drive a suite of research-grade land-surface models, but have subsequently been used for a variety of climate research applications over central North America. The non-precipitation NLDAS-2 forcing data were interpolated to 1/8th degree spatial resolution and disaggregated to hourly temporal resolution from the 32-km, 3-hourly North American Regional Reanalysis (NARR), which has been shown to represent near-surface climatic fields with high fidelity. Precipitation fields in the NLDAS-2 forcing data are not from NARR, but rather directly from gauge data augmented with satellite- and radar-derived precipitation estimates. Hourly meteorological fields from the NLDAS-2 forcing data were used to calculate the weekly minimum, mean, and maximum temperatures, saturation deficit, relative humidity, weekly precipitation, and cumulative precipitation from the beginning of the year for each grid cell (Table 2). Temperature data were used to calculate the weekly growing degree days (GDD) above a particular temperature threshold, and the cumulative growing degree days above that temperature threshold from the beginning of the year. Because several aspects of tick biology and human behavior may respond to different temperature thresholds we examined potential GDD thresholds of 5, 6, 10, 14, and 18°C by using the model selection techniques described below to determine which threshold value provided the best estimate of Lyme disease seasonality. The beginning of the Lyme disease season was strongly correlated with GDD at each threshold value, but model fit was maximized using a GDD threshold of 10°C. A threshold of 10°C has also been used in
of the seasonal statistics among states or among years. Indi-
determine whether there was a significant difference in each
version 2.9 in R.40 was used to perform an exhaustive search
and duration of the Lyme disease season. The leaps package
tify meteorological predictors of the beginning, peak, end,
Linear multiple regression models were constructed to iden-
Bonferroni correction to adjust for multiple comparisons.
line were calculated in a similar fashion for each state by
(GFigure 2). The latitude and distance from the Atlantic coast-
ables. In addition to weekly meteorological variables, we
coastline were also included as potential explanatory vari-
Therefore, we considered weekly values of saturation deficit,
were included in the same cluster. An optimum grouping
by the percent of annual observations from each state that
explained by the cluster groupings. Precision was determined
had the best combination of precision and percent of variance
number of clusters was varied from three to six, with the
occur varied from Week 16 to Week 26 (mean
s several studies of ixodid tick populations38,39; therefore, we used a
GDD threshold of 10°C for all of our analyses. A county-level
mean value for each weekly meteorological variable was calcu-
lated using the Zonal Statistics tool in ArcGIS 9.3 (Environ-
mental Systems Resource Institute, 2010; ESRI, Redlands, CA).
A statewide average was then derived by calculating a
weighted average of the county-level values within each state.
Each county-level statistic was weighted by the percentage of
1992–2007 Lyme diseases cases from a state occurring in that
county, so that meteorological variables from the counties
with the most cases within a state had the highest weights
(Figure 2). The latitude and distance from the Atlantic coast-
line were calculated in a similar fashion for each state by
weighting the mean latitude and coastline distance of each
county by its percentage of statewide cases.

Statistical analysis. Analysis of variance was conducted to
determine whether there was a significant difference in each
of the seasonal statistics among states or among years. Indi-
vidual states were compared using pairwise t tests with a
Bonferroni correction to adjust for multiple comparisons.
Linear multiple regression models were constructed to iden-
tify meteorological predictors of the beginning, peak, end,
and duration of the Lyme disease season. The leaps package
version 2.9 in R.40 was used to perform an exhaustive search
of the best fit models using Akaike’s information criterion
(AIC) as the model selection criterion. The mean beginning
of the Lyme disease season across all states was 21.2 weeks.
Therefore, we considered weekly values of saturation deficit,
relative humidity, GDD, and precipitation from Weeks 1 to
20 as potential explanatory variables for the beginning of
the Lyme disease season. Cumulative precipitation and cumula-
tive GDD through Weeks 1 to 20, latitude, and distance to
coastline were also included as potential explanatory vari-
ables. In addition to weekly meteorological variables, we
calculated the saturation deficit, relative humidity, GDD, pre-
cipitation, cumulative precipitation, and cumulative precipita-
tion after Week 8 (because the amount of precipitation that
occurs as rain versus snow in the first few months of the year
varies significantly) in the 5 weeks before the beginning of
the Lyme disease season. We also calculated the week at which a
cumulative threshold of 100 to 500 GDD was reached (in incre-
ments of 50). For the analysis of the duration and end of
the Lyme disease season, we also included all weekly meteoroe-
logical variables through Week 40 as potential explanatory
variables. Several of the meteorological variables were highly
correlated, so the predictors in the model were tested for
multicollinearity by comparing variance inflation factors
(VIF). All models with a VIF > 5 were excluded from analysis.
Because of the relatively high correlation between relative
humidity and saturation deficit, models including both variables
from the same week were also excluded. All statistical analyses
were performed in R 2.11 (R Development Core Team 2010).

Regional analysis. The 12 states were grouped into different
regions based on the similarity of their climate using k-means
cluster analysis. K-means cluster analysis partitions observa-
tions from each state into k clusters, where k is a pre-
determined value, by placing each observation in the cluster
with the nearest mean. The means for each cluster were calcu-
lated using the annual GDD, annual precipitation, and
mean weekly saturation deficit for each observation. Because
several states had a few years with > 100 Lyme disease cases
between 1992 and 2007, we included every year from 1992 to
2007 from each state for clustering purposes (N = 192). The
number of clusters was varied from three to six, with the
appropriate number of clusters determined by which grouping
had the best combination of precision and percent of variance
explained by the cluster groupings. Precision was determined
by the percent of annual observations from each state that
were included in the same cluster. An optimum grouping
of four clusters (regions) was chosen because only k = 3 and
k = 4 had at least 50% of observations from each state (except
MA) in one cluster, and the percentage of variance explained
(ratio of between-group variance to total variance) was higher
for k = 4 than k = 3 (75.8% versus 69.0%). The four regions
roughly correspond to the upper midwest (MN and WI),
northern New England (MA, NH, and ME), southern New
England and the northern mid-Atlantic (RI, CT, NY, NJ, and
PA), and the southern mid-Atlantic (MD and VA) (Table 3).

The beginning week of the Lyme disease season was ana-
yzed using a generalized linear mixed-effects regression
model (GLMM) with region, state, and year as nested random
effects to provide an estimate of the variance associated with
each factor after accounting for variance at the other hierar-
chical levels (variance components analysis) using the lme4
package in R.41 The best fitting linear regression model(s) for
all states were then fit separately for each region to determine
whether the overall relationship between Lyme disease sea-
onality and climate held within the different regions. An
exhaustive search of the best fit models of the beginning of
the Lyme disease season for each region was also conducted
using the same approach outlined for the overall model.

RESULTS

Seasonal statistics. Across all states and all years, the begin-
nning of the season when a majority of Lyme disease cases
occur varied from Week 16 to Week 26 (mean ± 21.2 weeks,
SD = 1.8 weeks; Figure 3 ; Table 1). The mean duration of the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly minimum temperature</td>
<td>Minimum temperature during the week</td>
</tr>
<tr>
<td>Weekly maximum temperature</td>
<td>Maximum temperature during the week</td>
</tr>
<tr>
<td>Weekly mean temperature</td>
<td>Mean temperature for the week</td>
</tr>
<tr>
<td>Weekly precipitation</td>
<td>Total precipitation for the week</td>
</tr>
<tr>
<td>Cumulative precipitation</td>
<td>Total cumulative precipitation from the beginning of the year</td>
</tr>
<tr>
<td>Growing degree days (GDD)</td>
<td>Total degrees above 10°C for the week calculated using daily mean temperatures</td>
</tr>
<tr>
<td>Cumulative growing degree days (GDD)</td>
<td>Total degrees above 10°C since the beginning of the year calculated using daily mean temperatures</td>
</tr>
<tr>
<td>Saturation deficit</td>
<td>Index of humidity that measures how far the amount of water vapor is below saturation for a given temperature and pressure</td>
</tr>
<tr>
<td>Relative humidity</td>
<td>Weekly mean of the ratio of the measured water vapor pressure to the saturation vapor pressure</td>
</tr>
</tbody>
</table>
season was 14.0 weeks (SD = 2.6 weeks). An average of 71.6% of cases occurred within the main Lyme disease season (12% occurred before, 16.4% after). The mean seasonal peak occurred at week 27.0 (SD = 1.2 weeks). The average timing of the beginning of the Lyme disease season varied significantly across states (F11,132 = 19.8; P < 0.0001). The beginning of the Lyme disease season occurred significantly earlier in Virginia (mean = 17.7 weeks) than in any other state (expect Maryland which had a mean of 19.2 weeks; P = 0.12). Minnesota had the latest beginning to the Lyme disease season (mean = 23.1 weeks). The peak of the Lyme disease season also varied significantly among states (F11,132 = 11.3; P < 0.0001), as did the duration (F11,132 = 12.0; P < 0.0001). The end of the Lyme disease season also varied among states (F11,132 = 1.8; P = 0.052), but differences among the states only explained 13.1% of the variance in the end week. The duration of the Lyme disease season in Virginia was significantly longer than in all of the other states (mean = 18.7 weeks), whereas Minnesota had the shortest season at 11.2 weeks. No state had a significantly earlier or later end week compared with the mean of week 35.2.

**Beginning of the Lyme disease season.** The strongest univariate predictor of the beginning of the Lyme disease season was the cumulative GDD through Week 20 (r² = 0.54), with higher cumulative GDD associated with an earlier start week (Figure 4). Cumulative GDD through Weeks 10 to 19 were
Results of k-means cluster analysis with the number of clusters set to \(k = 4^*\)

<table>
<thead>
<tr>
<th>Region</th>
<th>CT</th>
<th>MA</th>
<th>MD</th>
<th>ME</th>
<th>MN</th>
<th>NH</th>
<th>NJ</th>
<th>NY</th>
<th>PA</th>
<th>RI</th>
<th>VA</th>
<th>WI</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>East</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>South</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Midwest</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

*Each column represents the number of years from 1992 to 2007 for each state that were clustered within a particular region based on the annual growing degree days (GDD), annual precipitation, and mean weekly saturation deficit for that year. Bolded values indicate the region into which each state was grouped.

Model validation was performed using climate and case data from 2008 to 2010 from each of the 12 states (\(N = 35\); Rhode Island, 2008 was excluded from model validation because 40% of cases had a reported onset date of 3/22). Our best-fit model performed well for the 2008–2010 period (adj-R\(^2\) = 0.70), with the cumulative GDD through Week 20 and mean saturation deficit variables remaining highly statistically significant (\(P < 0.001\)), but the cumulative precipitation since Week 8 and distance to the Atlantic coastline variables were not significant at a \(P\) value of 0.05 (\(P = 0.08\) and \(P = 0.12\)).

**Seasonal peak and duration.** The timing of the peak and duration of the Lyme disease season were explained by similar variables as the beginning of the season. The strongest univariate predictors of the peak and duration of the Lyme disease season were the cumulative GDD through Week 22 and Week 21, respectively (R\(^2\) = 0.50 and 0.31). The best model for the peak week included cumulative GDD through Week 22, the saturation deficit 1 week before the beginning of the season, the cumulative precipitation after Week 8, and the distance to the Atlantic coastline (R\(^2\) = 0.59). The best model of the duration of the Lyme disease season included cumulative GDD through Week 20, the mean saturation deficit before the beginning of the season, the cumulative precipitation up to 1 week before the beginning of the season, and the distance to the Atlantic coastline (R\(^2\) = 0.54). None of the meteorological variables were strong predictors of the end of the Lyme disease season. No single predictor explained more than 5% of the variance in the end of the season, and no multi-variable model had an adjusted R\(^2\) > 0.15.

**Regional analysis.** The majority of the variation in the beginning of the Lyme disease season was determined by factors at the regional scale (59.8%), with differences among states within a region accounting for less of the variation (11.2%). The remainder of the variation in the beginning of the season was a result of differences among years within each state (29.0%). The beginning of the Lyme disease season started significantly earlier in the southern mid-Atlantic (mean = 18.6 weeks) than in the other regions, and earlier in the east (mean = 21.3 weeks) than in the north.
The best-fit overall model (model 1 in Table 4) performed relatively well in the southern mid-Atlantic ($r^2 = 0.784$), east ($r^2 = 0.614$), and midwest ($r^2 = 0.614$) regions, but not in the north ($r^2 = 0.375$). All four variables were significant for the southern mid-Atlantic region model, distance to coastline was the only variable that was not statistically significant in the east and midwest regions, and only cumulative precipitation after Week 8 was significant in the north region model. The best-fit model for the north region included cumulative precipitation before the beginning of the Lyme disease season, precipitation 4 weeks before the Lyme disease season, Week 5 saturation deficit, and Week 3 precipitation (Table 5). All of the other regional best-fit models included variables representing cumulative GDD, saturation deficit before the beginning of the Lyme disease season, and cumulative precipitation as three of the four model parameters (Table 5).

**DISCUSSION**

Previous research has linked the development and seasonal activity patterns of *Ixodes* ticks to meteorological variables, specifically temperature and humidity. However, because non-entomologic factors can also influence the occurrence of tick-borne disease in humans, the degree of coupling between meteorological factors and human Lyme disease has been open to question. Here, we showed that the geographical and annual variation in the timing of human Lyme disease can also be largely explained by meteorological factors. The beginning, peak, and duration of the Lyme disease season, and cumulative precipitation as three of the four model parameters (Table 5).

![Figure 4. Beginning week of Lyme disease season across all states and years as a function of the cumulative growing degree days (GDD) above 10°C through Week 20 for that state and year ($r^2 = 0.54$).](image-url)

**Table 4**

Best fit models with the beginning week of the Lyme disease season as the response variable

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of parameters</th>
<th>Adj. R²</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>Model parameters</th>
<th>Parameter estimates</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0.785</td>
<td>368.4</td>
<td>0</td>
<td>Week 20 cumul. GDD</td>
<td>−0.014</td>
<td>−0.016 to −0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean SD before onset</td>
<td>0.945</td>
<td>0.696–1.194</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cumul. precip. after Week 8</td>
<td>0.009</td>
<td>0.007–0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Distance to coastline</td>
<td>0.093</td>
<td>0.055–0.131</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.784</td>
<td>369.4</td>
<td>1.0</td>
<td>Week 20 cumul. GDD</td>
<td>−0.014</td>
<td>−0.016 to −0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean SD before onset</td>
<td>0.932</td>
<td>0.683–1.181</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cumul. precip. after Week 8</td>
<td>0.009</td>
<td>0.008–0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Longitude</td>
<td>−0.052</td>
<td>−0.090 to −0.014</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.773</td>
<td>376.7</td>
<td>8.3</td>
<td>Weeks to 150 GDD</td>
<td>0.530</td>
<td>0.445–0.614</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean SD before onset</td>
<td>1.062</td>
<td>0.801–1.322</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cumul. precip. after Week 8</td>
<td>0.010</td>
<td>0.008–0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Distance to coastline</td>
<td>0.098</td>
<td>0.059–0.137</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.772</td>
<td>377.3</td>
<td>8.9</td>
<td>Weeks to 150 GDD</td>
<td>0.568</td>
<td>0.487–0.648</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean SD before onset</td>
<td>1.055</td>
<td>0.792–1.318</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cumul. precip. after Week 8</td>
<td>0.010</td>
<td>0.008–0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Longitude</td>
<td>−0.056</td>
<td>−0.078 to −0.033</td>
</tr>
</tbody>
</table>

*Number of model parameters, model adjusted $r^2$, AIC, and ΔAIC values and parameter estimates with 95% confidence intervals for all models with ΔAIC < 10. ΔAIC represents the difference between a model’s AIC value and the AIC value of the best fit overall model. Italicized parameters are not statistically different from 0 at $\alpha = 0.05$ confidence level.
Lyme disease season were associated with several meteorological variables, including cumulative GDD, cumulative precipitation, and saturation deficit at both the national and regional levels. The strongest predictor at the national level was the cumulative GDD above 10°C through the first 20 weeks of the year, with higher cumulative GDD associated with an earlier beginning to the Lyme disease season. An early beginning to the Lyme disease season was also associated with less cumulative precipitation and a lower mean saturation deficit before the start of the season.

Tick population dynamics and phenology are highly sensitive to temperature at multiple life stages. The host-seeking behavior of the nymphal stage is most relevant for the transmission of *B. burgdorferi* to humans because nymphs are believed to be responsible for most human infections. Interstitial development rates in ixodid ticks typically increase with increasing temperature, and the timing of nymphal host seeking is correlated with ambient temperatures. The negative relationship between cumulative GDD through Week 20 and the week of the year when the Lyme disease season begins, indicating that higher cumulative GDD at Week 20 are associated with an earlier season (or the positive relationship between the number of weeks to reach 300 GDD and the week the Lyme disease season begins), is likely due in part to this positive effect of temperature on the activity of newly emerged nymphs in the spring. Cumulative GDD were a better predictor of seasonal timing than weekly minimum, mean, or maximum temperatures. Nymphal questing activity may track not just immediate environmental conditions, but temperatures earlier in the year as well. Newly emerged nymphs of some *Ixodes* species may undergo a behavioral diapause between emergence and host-seeking activity (questing) that is dependent on climatic conditions such as temperature. In addition, larvae that feed later in the summer or fall may undergo developmental (morphogenetic) diapause and overwinter before molting. In this case larval development and nymphal emergence would likely be subject to cumulative spring temperatures. Tick development, survival, and activity levels are all potentially sensitive to moisture levels in addition to temperature. In the overall best-fit model (and three of the four regional models) a higher mean saturation deficit in the 5 weeks before the beginning of the Lyme disease season was associated with a delayed start to the season. Nymphal *I. scapularis* activity increases with ambient relative humidity, and the activity and density of other ixodid ticks is also negatively correlated with saturation deficit. *Ixodes scapularis* nymphs also reach higher questing heights at higher relative humidity levels, which may increase their likelihood of encountering humans. Higher saturation deficits may delay the beginning of the Lyme disease season because it decreases questing activity and decreases the likelihood of questing nymphs encountering humans. The cumulative precipitation from Week 8 to the beginning of the Lyme disease season was associated with a later start to the season. Although spring or summer rainfall is positively correlated with ixodid tick densities in California, this association is generally measured during the time of peak nymphal density rather than earlier in the year when nymphal activity begins and the first human Lyme disease cases are likely to occur. Heavy rainfall events may impede nymphal questing activity, which could delay the beginning of the Lyme disease season in humans.

In addition to meteorological effects on pathogen and vector biology, seasonal patterns in disease occurrence can also arise because of human activities that lead to an increased exposure to pathogens and their vectors. It is likely that people will spend more time outdoors in areas where they may encounter questing ticks in warmer and drier weather. Therefore, the association between cumulative GDD, cumulative precipitation, and the beginning of the Lyme disease season may be partially driven by the influence of weather on human behavior in a way that affects the exposure of the human population.

<table>
<thead>
<tr>
<th>Region</th>
<th>Model</th>
<th>Model parameters</th>
<th>Adj. $R^2$</th>
<th>AIC</th>
<th>ΔAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>1</td>
<td>(+) Cumulative precip. before LD season</td>
<td>0.832</td>
<td>34.9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-) Precip. 4 weeks before LD season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Week 5 saturation deficit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-) Week 3 precip.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(+) Cumulative precip. before LD season</td>
<td>0.827</td>
<td>35.5</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-) Precip. 4 weeks before LD season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Week 7 relative humidity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-) Week 5 precip.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>1</td>
<td>(-) Week 20 cumulative GDD</td>
<td>0.723</td>
<td>137.2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Mean SD before onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Cumulative precip. after Week 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-) Week 9 precipitation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>1</td>
<td>(-) Week 17 cumulative GDD</td>
<td>0.852</td>
<td>42.9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Mean SD before onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Cumulative precip. before LD season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Week 17 relative humidity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>1</td>
<td>(-) Cumulative GDD from Week 16–21</td>
<td>0.819</td>
<td>72.7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Mean SD before onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Cumulative precip. after Week 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Week 8 saturation deficit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(-) Cumulative GDD from Week 16–21</td>
<td>0.809</td>
<td>74.4</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Mean SD before onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+) Cumulative precip. before LD season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-) Week 8 saturation deficit</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Model adjusted $R^2$, AIC, and ΔAIC values for all models with ΔAIC < 2. The (+/−) signs indicate whether parameter coefficients are positive or negative. ΔAIC represents the difference between a model's AIC value and the AIC value of the best fit overall model.
to infected *I. scapularis* nymphs. However, the strength of the relationship between cumulative GDD and human behavior is likely to be weaker than the relationship between cumulative GDD and tick development rates because human activity levels are mostly associated with the current temperature rather than cumulative temperatures since the beginning of the year. Likewise, human activity is probably more strongly associated with the amount of current precipitation rather than cumulative precipitation from Week 8 through the beginning of the Lyme disease season. However, in the northern areas considered in our study where snowpack in some years can last until springtime the cumulative GDD and precipitation before the beginning of the Lyme disease season may have a more lasting effect on the amount of time people spend outdoors in areas where they are likely to encounter questing nymphs. In addition, cumulative temperatures and precipitation could influence other environmental factors such as seasonal flooding and plant phenology that may affect human activity levels.

The best-fit overall model performed relatively well in three of the four climatically defined regions examined in this study. All three meteorological parameters (cumulative GDD, cumulative precipitation, and mean saturation deficit) were statistically significant in the east, southern mid-Atlantic, and midwest regions, but only cumulative precipitation was statistically significant in the north region. The majority of the overall variation in the beginning of the Lyme disease season was caused by geographical differences between regions rather than interannual differences within a state or region. However, the continued significance of these meteorological variables at the within-region level suggests that there is an association between weather and the interannual variation in seasonality as well. With the exception of the north, cumulative GDD was consistently associated with an earlier start to the Lyme disease season and higher cumulative precipitation was consistently associated with a later beginning to the season. In the northern region (MA, ME, and NH) cumulative precipitation was still positively associated with the start of the Lyme disease season, but there was no relationship between temperature and the timing of seasonality. This lack of a relationship between temperature and the beginning of the Lyme disease season is somewhat surprising because Maine is the current northern limit of Lyme disease’s geographic range in the United States, and the distribution of *I. scapularis* is believed to be limited by temperature. Therefore, we might expect *I. scapularis* activity to be sensitive to climate variability in a region where mean temperatures are close to their lower temperature limit for survival and development. However, the interannual variation in cumulative GDD through Week 20 was lower in the north region than it was in the other three regions, therefore temperature might not have varied enough from year to year to influence the timing of the Lyme disease season in these northern states. The timing of the Lyme disease season in the north might also be associated with weather-independent human activity patterns.

The best-fit overall model, and all of the best-fit regional models, included several parameters (cumulative precipitation from Week 8 to the beginning of the Lyme disease season and the mean saturation deficit before the Lyme disease season) that can only be calculated once the Lyme disease season begins. Because of this constraint these models cannot be used to precisely forecast when the Lyme disease season will begin in advance. Despite this limitation, the strong correlation between cumulative GDD and the beginning of the Lyme disease season suggests that GDD calculations could be used to predict whether the Lyme disease season is likely to start earlier or later than usual. The strongest correlation was with the cumulative GDD through Week 20, which does not provide much lead time for forecasting, particularly because infection typically occurs a week or two before symptoms first appear, but the cumulative GDD as early as Week 10 was still correlated with the beginning of the season \((R = 0.64)\) and could provide several weeks advance notice. Such information could be used to better time prevention efforts such as public awareness campaigns and acaricidal treatments, particularly since the peak of the Lyme season was also strongly associated with cumulative GDD. The correlation between cumulative GDD and the beginning of the season increased steadily from Week 10 \((R = 0.64)\) to Week 20 \((R = 0.74)\), providing a tradeoff between lead time and accuracy.

By analyzing the relationship between meteorological factors and the seasonality of Lyme disease occurrence in humans we have identified several variables associated with variations in when the Lyme disease season begins and peaks, as well as how long the season lasts, but not when the season ends. By identifying inflection points in the annual case curves from 1992 to 2007 we found that the beginning of the Lyme disease season varied by up to 10 weeks among the states where the disease is highly endemic and by up to 6 weeks within a state over this 16-year period. These methods could be used to calculate statistics for other infectious diseases with clear seasonal patterns. An analysis of the association between the variation in seasonality and meteorological variables, and other environmental and social factors, can be used to understand what factors drive disease occurrence as we have done here. Links between environmental factors and the timing of seasonal transmission can also be used to develop forecasting tools that inform disease prevention efforts. For infectious diseases that can cause large seasonal epidemics, such as influenza or dengue, these methods could also be used to identify factors that influence the timing and severity of peak disease occurrence, and the duration of the epidemic.

Received April 5, 2013. Accepted for publication August 5, 2013.

Published online January 27, 2014.

Financial support: The National Center for Atmospheric Research is funded in part by the National Science Foundation.

Disclosure: The NLDAS-2 primary forcing data used in this study were acquired as part of the mission of NASA’s Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

Authors’ addresses: Sean M. Moore and Andrew Monaghan, National Center for Atmospheric Research, Boulder, CO, E-mails: smoore@ucar.edu and monaghan@ucar.edu. Rebecca J. Eisen and Paul Mead, Centers for Disease Control and Prevention, Division of Vector-Borne Diseases, Fort Collins, CO, E-mails: Dyn2@cdc.gov and Pfm0@cdc.gov.

REFERENCES


