

## Geographic Variation in the Relationship between Human Lyme Disease Incidence and Density of Infected Host-Seeking *Ixodes scapularis* Nymphs in the Eastern United States

Kim M. Pepin,\* Rebecca J. Eisen, Paul S. Mead, Joseph Piesman, Durland Fish, Anne G. Hoen, Alan G. Barbour, Sarah Hamer, and Maria A. Diuk-Wasser

Fogarty International Center, National Institutes of Health, Bethesda, Maryland; Bacterial Diseases Branch, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado; Yale School of Public Health, New Haven, Connecticut; Department of Community and Family Medicine, Dartmouth Medical School, Lebanon, New Hampshire; Departments of Microbiology and Molecular Genetics and Medicine, University of California, Irvine, California; Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan; Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas

**Abstract.** Prevention and control of Lyme disease is difficult because of the complex biology of the pathogen's (*Borrelia burgdorferi*) vector (*Ixodes scapularis*) and multiple reservoir hosts with varying degrees of competence. Cost-effective implementation of tick- and host-targeted control methods requires an understanding of the relationship between pathogen prevalence in nymphs, nymph abundance, and incidence of human cases of Lyme disease. We quantified the relationship between estimated acarological risk and human incidence using county-level human case data and nymphal prevalence data from field-derived estimates in 36 eastern states. The estimated density of infected nymphs (mDIN) was significantly correlated with human incidence ( $r = 0.69$ ). The relationship was strongest in high-prevalence areas, but it varied by region and state, partly because of the distribution of *B. burgdorferi* genotypes. More information is needed in several high-prevalence states before DIN can be used for cost-effectiveness analyses.

### INTRODUCTION

Lyme disease is the most common vector-borne disease in the United States, with more than 25,000 confirmed cases in 2009.<sup>1</sup> Incidence of reported cases continues to rise. Between 2005 and 2009, mean incidence in the 13 states with highest incidence had increased from  $29.6 \pm 10.6$  per 100,000 in 2005 to  $49.6 \pm 15.5$  per 100,000 in 2009, whereas in 11 states with lower incidence, mean incidence has increased from  $1.3 \pm 0.7$  to  $2.3 \pm 1.7$  per 100,000.<sup>2</sup> Efforts to stop this emergence have been hampered by the complex ecology of the disease and the lack of effective and affordable interventions.<sup>3–7</sup>

In the Northeastern and Midwestern United States, Lyme disease is caused by the bacterium *Borrelia burgdorferi* sensu stricto and transmitted to humans by *Ixodes scapularis*, primarily nymphs.<sup>8</sup> Humans acquire the disease mainly during the months of May and July after contact with the vector during outdoor activities in tick habitat, which is characterized by closed-canopy deciduous and mixed forest. Three previous studies have shown a positive relationship between human cases and the density of *B. burgdorferi*-infected ticks, but these studies were limited in geographic scope to Connecticut and Rhode Island, which are both states with very high prevalence.<sup>9–11</sup> Another two studies showed a positive correlation between tick abundance and human cases, again focusing on specific regions (Wisconsin<sup>12</sup> and Westchester, New York<sup>13</sup>). Two more location-specific studies have examined the link between numbers of ticks sampled by passive surveillance and cases in Rhode Island and Maine.<sup>14,15</sup> To our knowledge, these studies are the only studies that have attempted to quantify and/or characterize the relationship between acarological risk and human cases in the eastern United States. However, an understanding of this relationship and how it varies geographically is impor-

tant for models used in identifying optimal strategies for implementing vector- and reservoir-targeted interventions.

In 2009, 95% of Lyme disease cases in the United States were reported from the Northeast and Midwest.<sup>2</sup> However, there is large variation in incidence among states where transmission of *B. burgdorferi* occurs, possibly because of differences in reservoir host ecology,<sup>16–18</sup> climate,<sup>19–23</sup> rates of human contact,<sup>13</sup> genetic variation among strains,<sup>23–26</sup> or differences between states in reporting practices. Thus, relationships between acarological risk and incidence may differ among geographic regions. Investigating the differences is an important stepping stone to a mechanistic understanding of the underlying drivers linking density of infected nymphs (DIN) and human incidence. Improved knowledge of the relative importance of individual drivers in different locations might aid in tailoring prevention and control recommendations to different geographical regions.

We have previously reported on a large-scale study to measure and develop models of the density of *I. scapularis* nymphs and *B. burgdorferi* prevalence in 36 states (2,411 counties) at an  $8 \times 8$ -km spatial scale.<sup>27–29</sup> Here, we used the raw data (rDIN and raw density of nymphs [rDON]) to examine the strength of the relationship between DON, DIN, and human incidence at the county level to evaluate whether the model-based predictions of acarological indices (mDON and mDIN) significantly correlated with reported human Lyme disease incidence and identify factors that modify the relationship between DIN and human cases. We then used the estimated indices to investigate geographical differences in the relationship between acarological risk and county-level incidence. To our knowledge, this study is the first study to examine the relationship over such a broad spatial extent and evaluate geographic differences in how DIN translates to human incidence.

### MATERIALS AND METHODS

**Human case data.** Human case data were obtained from case reports to the Centers for Disease Control and Prevention

\* Address correspondence to Kim M. Pepin, Fogarty International Center, National Institutes of Health, 31 Center Drive, MSC 2220, Bethesda, MD 20892-2220. E-mail: uzu2@cdc.gov

(CDC) by state and territorial health departments as part of the National Notifiable Diseases Surveillance System (NNDSS); detailed information on methods and case definitions has been published previously.<sup>30</sup> For each county in the United States, the number of cases between 2004 and 2006 was averaged to get the mean number of cases per county during the time period that tick data were collected (Figure 1C shows these data as incidence per 100,000). Note that cases are reported based on the county of residence, and thus, it is possible that transmission occurred outside the county of residence if a person had traveled recently.

**Density of nymphs.** Host-seeking *I. scapularis* nymphs were collected by drag sampling at 301 sites in the United States east of the 100th meridian between 2004 and 2006<sup>27,28</sup> (Figure 1A). All sites were closed-canopy deciduous forests in state parks, state forests, or other natural areas with public access. The spatial distribution of sites included 36 states (286 counties), which we classified based on the incidence of human cases as follows: > 10/100,000, DE, CT, ME, MD, MA, NH, NJ, NY, PA, RI, VT, WI, and MN (high); 0.5–5/100,000, VA, WV, OH, SC, NC, IL, IN, IA, MI, and ND (low); < 0.5/100,000, AL, AR, FL, GA, KS, KY, MS, MO, NE, OK, SD, TN, and TX (rare). Each site was dragged three to six times in summer between

late May and September, each time covering 1 km<sup>2</sup> per visit. Visits were spaced as uniformly as possible throughout this time frame in each location to prevent detection bias because of sampling time (Supplemental Figure 1) (note that there were several sites where this spacing was not possible, but they were not localized in a single area and thus, should not consistently bias the results). Daily nymphal density per 1,000 m<sup>2</sup> (i.e., DON) was calculated as the area under the nymphal frequency curve between the first and last sampling dates divided by the total number of days during this time period. Although nymphs tend to be most active in May and June in the northeast, we continued to sample through late summer (August and September), because (1) peak nymphal activity dates could differ across a large geographic area sampled, (2) human cases are most prevalent in late summer, and (3) humans tend to be most active in forested areas during summer holidays (May and September). In most counties, the same site was dragged each year except in three locations (Cumberland, ME; Suffolk, NY; and Chippewa, WI), where different sites were visited in different years. Because each site was considered to be representative of the county in which it was situated, daily nymphal densities from multiple years and/or all sites within the same county were averaged to get the nymphal

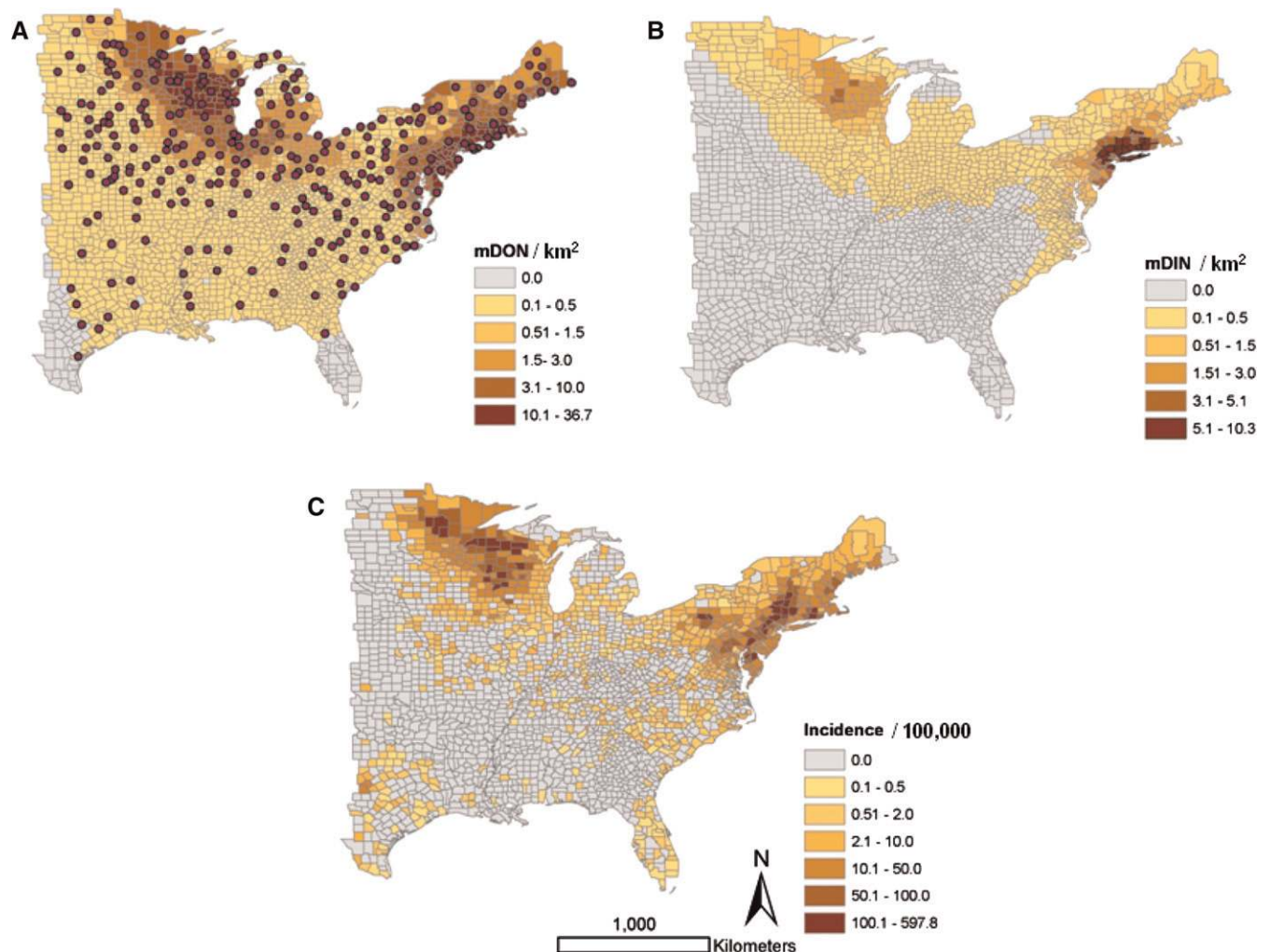


FIGURE 1. Map of study sites, estimated acarological risk, and human incidence. (A) Counties are color-coded by mDON/km<sup>2</sup>; unique study sites are shown as dark red dots. (B) Counties are color-coded by mDIN/km<sup>2</sup>. (C) Counties are color-coded by incidence/100,000 of human Lyme disease.

density per 1,000 m<sup>2</sup> during the entire study period. Density estimates were scaled to the county level by multiplying by the proportion of deciduous and mixed forest (i.e., *I. scapularis* habitat<sup>31</sup>) for the county where each site was located. This area was estimated using the ArcMap Zonal Statistics tool (ESRI, Redlands, CA) based on the US Geological Survey National Land Cover Database (<http://www.mrlc.gov/nlcd.php>). Hereafter, we refer to this estimate of DON as rDON. Note that accounting for land cover in the calculation of rDON was necessary to compare the tick data with human case data (which were available at the county level), because there is high variability between counties in forest coverage (amount of *I. scapularis* habitat).

A previous study<sup>28</sup> developed a regression model for predicting DON at an 8 × 8-km resolution throughout the eastern United States. This model assumed a zero-inflated negative binomial error structure and included climate, landscape, and spatial structure as covariates. Specifically, elevation, monthly vapor pressure deficit, annual amplitude of the maximum daily temperature, and annual phase of the minimum daily temperature were covariates in the zero-inflated component of the model, whereas the annual amplitude of the normalized vegetation index and spatial autocorrelation were covariates in the negative binomial component. For the current study, we calculated the average DON by county using the previously developed model. We refer to model-derived estimates of DON as mDON. Nine counties were excluded because of missing climate or landscape data that were used as covariates. This lack of data was because the climate/landscape data, obtained from National Aeronautics and Space Administration (NASA), consisted of 8-km grids that were clipped to minimize the number of pixels that contained both land and water. Some counties that were coastal and small did not either contain data or have a sufficient number of pixels to calculate an average value. These 10 counties included Dukes, Nantucket, Suffolk, and Barnstable in Massachusetts; Bronx, Kings, New York, and Richmond in New York; and Bristol and Newport in Rhode Island.

**Density of infected nymphs.** Detection and quantification of *B. burgdorferi* in nymphal tick specimens was performed by species-specific quantitative polymerase chain reaction (PCR) as previously described.<sup>23</sup> Individual ticks were screened for the presence of *B. burgdorferi*. Nymphal infection prevalence (NIP) was calculated as the proportion of nymphs that tested positive for *B. burgdorferi*. Using a binomial probability distribution, it was expected with 95% confidence that at least 1 of 14 ticks would be infected if the site harbored *B. burgdorferi*-infected nymphs with an infection prevalence of 0.20. Thus, the lower limit for pathogen detection was 14 ticks, and sites with fewer ticks were excluded from analyses with rDIN. In the high and low prevalence areas (for which we conducted the analyses using rDIN and mDIN), the average number of ticks collected per site was 77.6 ± 21.6 (2 standard errors of the mean [SEMs]), meaning that most sites had at least 55 ticks. Yearly prevalence per county was calculated as the mean prevalence for all sites within the same county during the study period. The density of infected nymphs (rDIN) was calculated as the product of rDON and prevalence.

Similar to mDON, model DIN (mDIN) was estimated at the county level using a model that was developed previously

for predicting DIN at a finer spatial scale<sup>29</sup> (Figure 1B). The model used for predicting DIN was similar to the DON model, except that maximum daily temperature was excluded from the zero-inflation component and the largest forest patch index (which is a surrogate for fragmentation describing the size of the largest forested patch in a county) was substituted for the normalized difference vegetation index in the negative binomial component. Predictions were made for all counties within the study region, except for the nine mentioned above for which there was no climate/landscape data.

**Additional covariates in the model of rDIN and human incidence.** We included *B. burgdorferi* genotype frequency, climate, and forest fragmentation variables in a model selection routine that included rDIN as a fixed component of the model (described in *Statistical analyses*) to examine whether these variables accounted for some of the variation in incidence left unexplained by rDIN. Our logic was that these variables could impact human behavior or *B. burgdorferi* transmission from vectors to humans. For example, the 16S-23S rRNA intergenic spacer restriction fragment length polymorphism sequence types (RSTs) have been previously linked to differential capacity for hematogenous dissemination in humans and severity of Lyme disease symptoms<sup>26,32</sup> as well as disease severity in animal models and transmission rates in reservoir mice.<sup>25,33</sup> Because the frequency of these types varies geographically,<sup>23</sup> we hypothesized that differences in pathogenicity and transmissibility between RST types may explain part of the geographic variation in Lyme disease incidence. As for the fragmentation and weather variables, we are unaware of work that has directly tested whether they could impact human contact with vectors or transmission rates, but weather has been hypothesized to impact transmission.<sup>34</sup> Furthermore, the direction of the relationship between fragmentation and DIN compared with fragmentation and human incidence is conflicting,<sup>17</sup> suggesting that fragmentation could act directly on incidence through some unexplained mechanism. Thus, we incorporated weather and fragmentation variables from a hypothesis-generating standpoint. These variables included the amplitude of county-level maximum temperature (TMAX<sub>mag</sub>) derived through temporal Fourier transformation, the mean area of forest patches (mean patch area [MPA]), and the mean area occupied by the largest forest patch (largest patch index [LPI]).

The proportion of each of the three RST types was estimated for 29 sites (thus, there was some missing data in this covariate, because the others were  $N = 40$ ) as described previously.<sup>35</sup> In brief, real-time PCR with probes specific for RST 1 or RST 2 strains was carried out for all individual *B. burgdorferi*-positive tick extracts. A nymph was characterized as containing an RST 3 strain if it was positive by species-specific PCR for *B. burgdorferi* but negative by PCR specific for RST 1 or RST 2. These data were previously reported in the work by Gatewood and others,<sup>23</sup> which found that the frequency of RST types was associated with variability in the synchronization of *I. scapularis* life stages.

**Statistical analyses.** All statistical analyses were conducted in R and Matlab. We evaluated the relationship between rDON and cases ( $N = 286$  counties in the full raw data) by negative binomial regression on different subsets of the data (based on human incidence). County population size was included as an offset, and thus, the model essentially predicts incidence (frequency of infection per county with weighted

effects from population size). Inspection of residuals showed that a negative binomial model was better for rDON, but a Poisson model with square-root transformed case data was equally good for rDIN ( $N = 40$ ; only sites with  $> 14$  ticks were included) analyses. Thus, we used a Poisson model for all DIN analyses; the negative binomial model tended to largely overestimate case numbers in high prevalence areas, and the Poisson model is simpler with one less parameter.

We evaluated performance of mDON and mDIN by fitting them to incidence in counties for which there were observed data ( $N = 40$ ) using a Poisson model. We compared these fits to those fits with rDON and rDIN by Akaike Information Criterion (AIC). Lower scores indicate better model fits (a two-point difference is significant). Because both mDON and mDIN fitted the incidence data better than rDON and rDIN, we used the mDIN data to investigate spatial differences in the relationship between DIN and human cases. Thus, we assumed that mDIN in the 40 counties where rDIN was measured was similarly representative of mDIN in the other 1,141 counties in high- and low-prevalence regions where rDIN was not measured. We fitted data from all counties in high- and low-incidence states to cases ( $N = 1,181$ ) using mDIN as a continuous covariate and state as a factor. We also fitted mDIN to cases for each state individually and compared predictions from the full model to those predictions from the state-specific models by using the likelihood function to calculate AIC for each set of model predictions in each state. Performance of the model predictions against observed cases was evaluated by Spearman's rank correlation  $r$ .

We investigated geographical effects on the relationship of mDIN and incidence by including region (east versus west) as a factor. East included states east of and including Ohio (classification of the 23 states included is indicated in *Results*). To identify factors that were responsible for modifying the relationship between DIN and incidence regionally, we conducted model selection with rDIN in the model to start and six covariates: RST types 1–3, LPI, MPA, and  $TMAX_{mag}$  (described above). Only main effects were examined. Variables were selected by forward selection and 10-fold cross-validation (where coefficients are estimated from 90% of the data and used to predict the other 10%) using the  $\chi^2$ -distributed difference in deviance for determining whether a variable is kept. This method of model selection decreases overfitting by requiring that candidate variables contribute to significantly accurate out of sample predictions for their incorporation into a model. Competing models generated by the selection procedure were compared by AIC. Last, to understand the hierarchical structure of the effects of region and other factors on the relationship between rDIN and incidence, we conducted a recursive partitioning analysis by conditional inference<sup>36</sup> implemented using the party package in R. This method is non-parametric, and it identifies covariates that explain the most significant amount of response variable (i.e., incidence) variation by repeatedly splitting the response variable into two groups. The resulting tree indicates the hierarchical nature of the covariate effects on rDIN.

## RESULTS

**Relationship of rDON and cases in different regions.** Overall, rDON was significantly correlated to reported incidence (Spearman's  $r = 0.66$ ,  $N = 286$ ). Next, we compared the

TABLE 1  
Fit of DON in different regions

Human incidence	$\Delta$ AIC	$P$ value	$n$
All	45	$< 0.0001$	286
Rare	-0.7	0.99	72
Low	80	0.03	125
High	13	$< 0.0001$	89

Cases at the county scale were modeled using a negative binomial model with a log link and a square root transformation of cases. The population size of the county was included as an offset. Fits were done on difference subsets of the data (rows 3–5) that correspond to rare- ( $< 0.5/10^4$ ), low- ( $0.5\text{--}5/10^4$ ), and high- ( $> 10/10^4$ ) incidence regions. Change in AIC is the difference between the model with DON and the null model (i.e., intercept only). Values above two indicate that rDON is a significantly better explanation of incidence than the null model. Columns 3 and 4 show the  $P$  value for rDON and the number of points in the dataset, respectively.

relationship of rDON with cases in regions where the incidence is rare, low, or high to investigate whether vector density could explain case variability in any or all of these areas (Figure 1A versus C). We found that, within rare areas, rDON did not explain variation in human Lyme disease incidence across counties ( $P = 0.99$ ; AIC change over the null model with only an intercept was -0.7, which is less than 2) (Table 1). However, rDON explained a significant amount of variation in human incidence in each of the low- and high-incidence areas ( $P < 0.03$  and  $P < 0.0001$ , respectively;  $\Delta$ AIC = 80 and 13, respectively) (Table 1).

**Addition of *B. burgdorferi* prevalence data.** Next, we examined whether the incorporation of *B. burgdorferi* prevalence data explained more of the variation in incidence relative to vector abundance alone (Figure 1B versus C and Table 2). For the raw data, all covariates explained a significant amount of the variation in incidence ( $\Delta$ AIC was greater than two above the null model) (Table 2), but much was left unexplained (Spearman's  $r = 0.57$  for rDON and 0.60 for rDIN). rDIN performed only slightly better than NIP, despite the apparently larger variation in rDIN by location (Supplemental Figure 2).

**Dissecting the roles of additional factors.** Because much of the variation in human incidence remained unexplained by the model with rDIN ( $r = 0.6$ ) (Table 2), we identified other factors that may be needed for interpreting human incidence in terms of rDIN. First, we examined whether accounting for geographical region (Northeast versus Midwest) improved the fit of rDIN and incidence. For this examination, we included region as a two-level factor and examined if either the intercept or slope of the relationship of rDIN and incidence differed by region. We found that there was no difference in the slope of the relationship between regions but that the relationship between rDIN and incidence had a significantly higher intercept in the Northeast relative to the Midwest ( $P < 0.0001$ ). Figure 2A and Table 3 show statistics

TABLE 2  
Comparison of raw data with model data covariates

	$\beta$	$\pm$ SE	$P$ value	$r^*$	Pseudo- $R^2$	$\Delta$ AIC	$n$
rDON	0.25	0.10	0.020	0.57	0.12	25	40
mDON	0.36	0.10	0.0016	0.58	0.23	47	40
rDIN	0.24	0.11	0.035	0.60	0.10	21	40
mDIN	0.42	0.09	$< 0.0001$	0.69	0.36	74	40
NIP	0.25	0.12	0.044	0.50	0.10	20	40

\*Spearman's rank correlation of predicted and observed cases ( $P < 0.001$  for all).

Cases at the county scale were modeled using a Poisson error structure with a log link and a square root transformation of the response variable.  $P$  values were adjusted with an estimated dispersion parameter. The population size of the county was included as an offset. Only points for which there was raw DON, prevalence, and DIN data are included. Columns 2–4 show the parameter estimates and their significance. Deviance-based pseudo- $R^2$  is a measure of model fit relative to the mean of the data. Change in AIC is the difference between the model with a covariate and the null model (i.e., mean of the data).

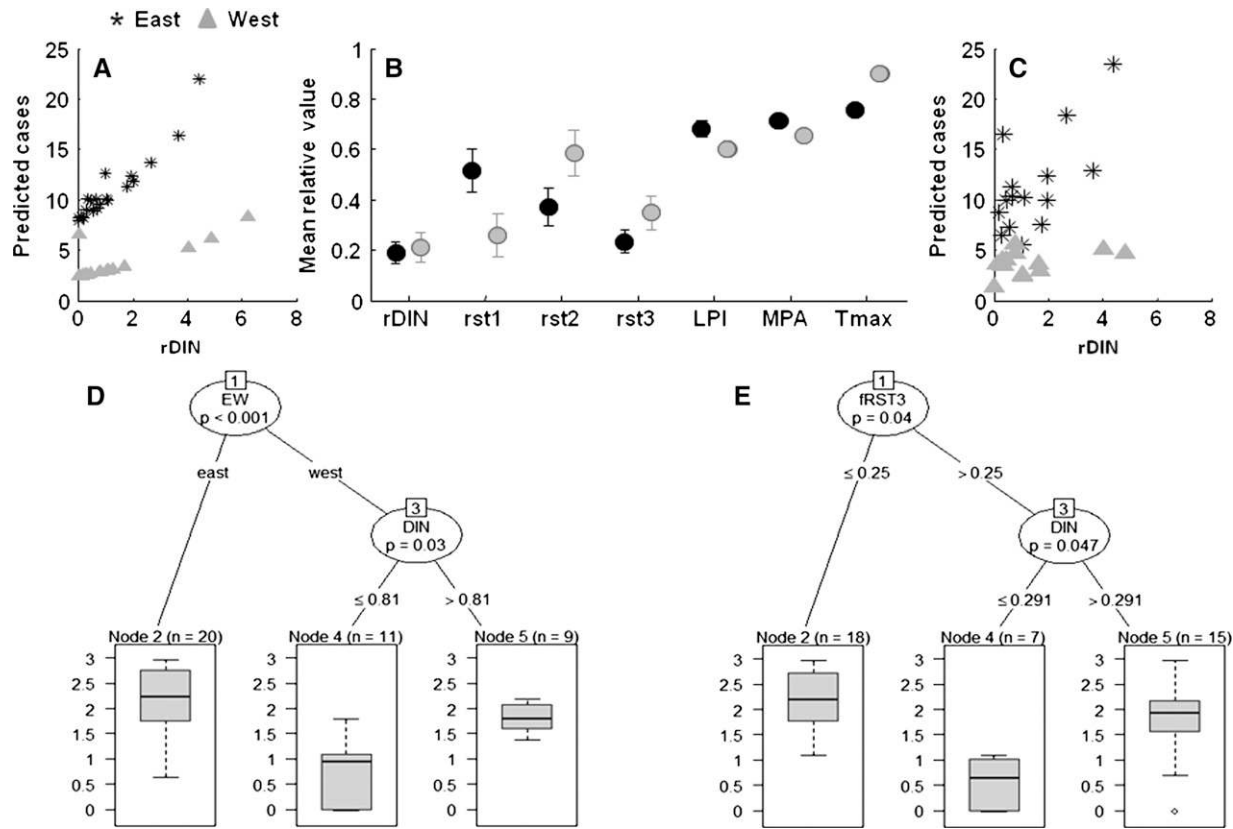


FIGURE 2. Factors influencing the relationship of DIN and cases. (A) Relationship of rDIN and incidence with geographic region included in the model (model details in the text). (A–C) Eastern counties are in black, and western counties are in grey. (B) Mean values of each candidate predictor of incidence in eastern versus western regions. Absolute values were rescaled to be between zero and one to compare them on the same plot. Error bars include 1 SEM. (C) Relationship of rDIN and incidence in the best model that was selected based on landscape, climate, and *B. burgdorferi* strain type variables. Model selection excluded the region factor and included rDIN in the model to start (*Materials and Methods*). Covariates in the best model included rDIN, the frequency of RST 3, and TMAXmag. (D and E) Regression trees for covariates in the models in A and C [log(sqrt[x]) – offset] for comparison between the two analytical methods. Box plots show the median value of the transformed cases for each terminal node on the tree. The stopping criterion for splitting was alpha = 0.05. EW = region; LPI = largest patch index; MPA = mean patch index; TMAXmag = magnitude of difference between summer and winter daily temperature maximums.

for the reduced model that excludes the non-significant interaction term between region and rDIN, and more of the variation was explained ( $r = 0.77$ ). Figure 2A and B highlights that similar levels of rDIN occurred in both regions, but mean levels of incidence are lower in the Midwest relative to the Northeast region (Figure 2A).

When region was excluded from the model, the best model of rDIN and incidence included the frequency of RST 3 and TMAX<sub>mag</sub> ( $r = 0.74$ ) (Figure 1C and Table 3). Both variables

were negatively related to incidence (Table 3), indicating that a higher frequency of RST3 strains and a larger difference between summer and winter temperatures are associated with lower incidence in humans. Although RST 1, RST 2, and the fragmentations indices (MPA and LPI) were not selected as the most significant modifiers of the rDIN–human incidence relationship, LPI and RST1 were significant in models that contained only rDIN and one other variable (LPI,  $P < 0.031$ ; RST 1,  $P < 0.0002$ ). The regression trees in Figure 2D and E

TABLE 3

Comparison of parameter estimates and model fits between a model that takes regional differences into account and the best model without region effects

Model	ΔAIC	Covariate	β	SE	P value	Partial pseudo-adjR <sup>2</sup>	Model pseudo-adjR <sup>2</sup>	P
Full model: Intercept + rDIN + region								
Intercept	0	Intercept	1.72	0.11	< 0.0001		0.49	2
Intercept + rDin	14.5	rDIN	0.28	0.08	0.002	0.11		
Intercept + rDin + Region	97.8	Region	0.59	0.11	< 0.0001	0.41		
Full model: Intercept + rDIN + rst3 + Tmax								
Intercept + rDin + rst3	84.7	Intercept	1.74	0.14	< 0.0001		0.61	3
Intercept + rDin + Tmax	78	rDIN	0.36	0.12	0.006	0.11		
		rst3	-0.37	0.14	0.02	0.25		
		Tmax	-0.43	0.12	0.002	0.20		

ΔAIC for the model with rDIN relative to the null model shows that rDIN is better than the null model. Likewise, models with region, RST 3, or TMAX are better than the model with only rDIN.

TABLE 4  
Model including all 23 states

Full model	<i>n</i> countries	Inc*/105	±SE	DIN	±SE	Pseudo-adjR <sup>2</sup>	AIC	β	Multi-state model of incidence	SE	<i>P</i> value	r†
Intercept	1,181	36	3	0.48	0.03	0.66	2,833	0.46		0.05	< 0.001	0.69
DIN								0.29		0.02	< 0.0001	
Factor: state				High-prevalence states all were significantly positive except for Vermont (not significant) Low-prevalence states all were significantly negative except for North Carolina (not available)								

\*Statistics from the national model; larger AIC and smaller pseudo-R<sup>2</sup> relative to the state-based model indicate that the national model performs worse than the state-based model.  
†Spearman's *r* for the correlation of predictions and observed data.

illustrate the hierarchical structure of the most significant combination of factors explaining the variation in incidence. The major factors that separate high from low incidence irrespective of rDIN are region (Figure 2D, top node) and RST 3 frequency (Figure 2E, top node). In the Midwest, where RST 3 frequencies are higher, incidence is lower, although rDIN occurs at similar levels as in the Northeast (Figure 2B). After variation because of region or RST 3 is segregated, in the Midwest, there is more variation in incidence that can be significantly classified into two groups (high versus low) according to levels of rDIN (second nodes). The second node does not occur for the Northeast, because the number of points with low rDIN was not high enough (minimum number for division was 10). Although TMAX<sub>mag</sub> was also significant, with a lower value in the Northeast relative to the Midwest (Figure 2B), it did not show significant partitioning in the regression tree analysis.

**Evaluation of model-estimated acarological indices: mDON and mDIN.** We compared the performance of previously estimated acarological indices, mDON and mDIN, with rDON and rDIN by fitting all four covariates individually to the

counties for which there were empirical observations. Both estimated indices performed significantly better than the raw data (Table 2) (pseudo-R<sup>2</sup> showed 11% and 26% improvements in fit over rDON and rDIN, respectively), and mDIN (*r* = 0.69) fit the incidence data better than mDON (*r* = 0.58). This finding shows that the estimated indices performed better at explaining variation in incidence, likely because they account for effects of other factors (i.e., landscape fragmentation, spatial autocorrelation, and weather) that are known to affect *I. scapularis* and *B. burgdorferi* distributions<sup>16–18,28</sup> and their interactions.<sup>2,3</sup>

**Spatial variation in the relationship between mDIN and human incidence at the state level.** Because mDIN fit the incidence data best and provided acarological risk data for all counties within the 23 states in high- and low-prevalence areas (listed in Tables 4 and 5), we used this index to examine whether differences in the relationship of mDIN and human incidence occurred at a finer spatial scale: the state level (rather than the east versus west). We investigated state-specific differences in the relationship between mDIN and incidence through two model structures: a national-level

TABLE 5  
Models for individual states

Region		<i>n</i> countries	Ranked state	Incidence*/10 <sup>5</sup>	±SE	DIN	±SE	Pseudo-R <sup>2</sup>		AIC		β	SE	<i>P</i> value	r‡
Incidence	Area							State	Full†	State	Full†				
H	E	8	CT	239	62	6.5	0.39	0.01	-1.09	16	71	0.02	0.11	0.84	
H	E	3	DE	183	61	1.0	0.06	0.99	0.22	4	48	0.28	0.03	0.063	
H	E	21	NJ	158	40	4.6	0.32	0.11	-0.24	66	133	0.14	0.09	0.13	
<b>H</b>	<b>W</b>	<b>72</b>	<b>WI</b>	<b>157</b>	<b>21</b>	<b>1.6</b>	<b>0.09</b>	<b>0.23</b>	<b>0.20</b>	<b>151</b>	<b>198</b>	<b>0.30</b>	<b>0.06</b>	<b>&lt; 0.0001</b>	<b>0.52</b>
<b>H</b>	<b>E</b>	<b>58</b>	<b>NY</b>	<b>125</b>	<b>40</b>	<b>1.3</b>	<b>0.30</b>	<b>0.50</b>	<b>0.50</b>	<b>387</b>	<b>429</b>	<b>0.60</b>	<b>0.08</b>	<b>&lt; 0.0001</b>	<b>0.86</b>
H	E	3	RI	114	75	4.4	0.29	0.13	-0.49	7	51	-0.14	0.36	0.77	
H	E	24	MD	105	17	0.7	0.06	0.01	0.01	78	120	0.05	0.12	0.66	
<b>H</b>	<b>E</b>	<b>10</b>	<b>MA</b>	<b>87</b>	<b>9</b>	<b>3.0</b>	<b>0.50</b>	<b>0.37</b>	<b>-3.42</b>	<b>16</b>	<b>128</b>	<b>-0.21</b>	<b>0.09</b>	<b>0.049</b>	<b>0.85</b>
<b>H</b>	<b>E</b>	<b>67</b>	<b>PA</b>	<b>79</b>	<b>14</b>	<b>0.6</b>	<b>0.10</b>	<b>0.42</b>	<b>0.27</b>	<b>299</b>	<b>415</b>	<b>0.47</b>	<b>0.07</b>	<b>&lt; 0.0001</b>	<b>0.62</b>
<b>H</b>	<b>W</b>	<b>87</b>	<b>MN</b>	<b>60</b>	<b>10</b>	<b>0.4</b>	<b>0.04</b>	<b>0.31</b>	<b>0.10</b>	<b>248</b>	<b>362</b>	<b>0.45</b>	<b>0.07</b>	<b>&lt; 0.0001</b>	<b>0.81</b>
H	E	10	NH	58	18	1.2	0.14	0.08	0.06	43	86	0.23	0.32	0.50	
<b>H</b>	<b>E</b>	<b>16</b>	<b>ME</b>	<b>47</b>	<b>12</b>	<b>0.6</b>	<b>0.04</b>	<b>0.31</b>	<b>0.05</b>	<b>42</b>	<b>99</b>	<b>0.47</b>	<b>0.18</b>	<b>0.019</b>	<b>0.47 (mns)</b>
H	E	14	VT	35	16	0.6	0.09	0.16	0.08	25	69	0.28	0.18	0.14	
<b>L</b>	<b>E</b>	<b>100</b>	<b>VA</b>	<b>9</b>	<b>2</b>	<b>0.1</b>	<b>0.01</b>	<b>0.18</b>	<b>0.03</b>	<b>210</b>	<b>289</b>	<b>0.44</b>	<b>0.10</b>	<b>&lt; 0.0001</b>	<b>0.48</b>
<b>L</b>	<b>W</b>	<b>99</b>	<b>IA</b>	<b>8</b>	<b>1</b>	<b>0.1</b>	<b>0.01</b>	<b>0.17</b>	<b>0.03</b>	<b>125</b>	<b>188</b>	<b>0.33</b>	<b>0.06</b>	<b>&lt; 0.0001</b>	<b>0.52</b>
<b>L</b>	<b>E</b>	<b>55</b>	<b>WV</b>	<b>5</b>	<b>2</b>	<b>0.05</b>	<b>0.01</b>	<b>0.14</b>	<b>0.01</b>	<b>83</b>	<b>137</b>	<b>0.42</b>	<b>0.14</b>	<b>0.004</b>	<b>2 (ns)</b>
<b>L</b>	<b>W</b>	<b>102</b>	<b>IL</b>	<b>4</b>	<b>2</b>	<b>0.1</b>	<b>0.02</b>	<b>0.47</b>	<b>0.05</b>	<b>105</b>	<b>226</b>	<b>0.81</b>	<b>0.08</b>	<b>&lt; 0.0001</b>	<b>0.48</b>
<b>L</b>	<b>W</b>	<b>83</b>	<b>MI</b>	<b>3</b>	<b>2</b>	<b>0.1</b>	<b>0.01</b>	<b>0.24</b>	<b>0.02</b>	<b>90</b>	<b>157</b>	<b>0.52</b>	<b>0.09</b>	<b>&lt; 0.0001</b>	<b>0.54</b>
L	W	23	ND	2	1	0.1	0.02	0.13	0.01	20	64	0.58	0.39	0.15	
L	E	100	NC	2	0	0.02	0.004	0.03	0.00	134	180	-0.28	0.18	0.13	
<b>L</b>	<b>W</b>	<b>92</b>	<b>IN</b>	<b>2</b>	<b>0</b>	<b>0.1</b>	<b>0.01</b>	<b>0.10</b>	<b>0.01</b>	<b>96</b>	<b>146</b>	<b>0.38</b>	<b>0.11</b>	<b>0.001</b>	<b>0.21</b>
L	E	46	SC	1	0	0.001	0.0005	0.01	0.00	49	92	0.12	0.13	0.38	
L	E	88	OH	1	0	0.1	0.01	0.03	0.00	105	151	-0.22	0.13	0.095	

\*Incidence/100,000 during the study period (2004–2006); basis for state rank.  
†Statistics from the national model; larger AIC and smaller pseudo-R<sup>2</sup> relative to the state-based model indicate that the national model performs worse than the state-based model.  
‡Spearman's *r* for the correlation of predictions and observed data.  
Significant relationships between DIN and cases are in bold; E = east; H = high prevalence; L = low prevalence; mns = marginally not significant; ns = not significant at *a* = 0.05; W = west.

model that included state as a factor versus separate models of mDIN for individual states. We found that there was a significant effect of state in the full model ( $P < 0.0001$ ) and that the state-level models performed better than the full model in all states (Tables 4 and 5, compare AIC for full with state). Lack of fit of the full model was not specific to any geographic location (Figure 3B–I) or levels of mDIN or incidence (Table 5, columns 6–9).

Of 23 states, 12 states showed a significant relationship between mDIN and human incidence, with an average correlation between model predictions and incidence of 0.55 (range = 0.2–0.86) (Tables 4 and 5). The other 11 states showed no significant relationship. Three of these states, Delaware, Rhode Island, and Connecticut (Table 4), had a very low number of counties, and thus, their lack of fit is likely because of small sample size. For states where the relationship was significant, the fit was good ( $r > 0.6$ ) in 4 of 12 states (New York, Pennsylvania, Massachusetts, and Minnesota), weak ( $r = 0.4$ – $0.6$ ) in 5 more states (Wisconsin, Virginia, Iowa, Illinois, and Michigan), and poor ( $r < 0.40$ ) in the other 3 states (Maine, West Virginia, and Indiana), despite the high significance of mDIN in the models (Tables 5, bold  $P$  values and Supplemental Figure 3). Furthermore, the slopes of the relationship were significantly different among the states where the relationship was significant, including an unexpected negative relationship for Massachusetts (Figure 3B–I and Table 5, bold  $b$ -values and SE), highlighting that there are quantitative differences between states, even where the relationship is highly significant.

## DISCUSSION

We quantified the relationship of acarological risk and human Lyme disease incidence for all of the known US range of *I. scapularis*-borne *B. burgdorferi*. Our study is the first to examine this association over such a large geographic range. We found that both rDON and rDIN explained a statistically significant amount of variation in human incidence in low- and high-incidence areas but that neither explained variation in incidence in rare areas. Restricting the analysis to low- and high-incidence areas, we found that the relationship of mDIN and incidence varied quantitatively by geographic location, that mDIN alone explained a statistically significant amount of variation in human incidence at the county level in only 12 of 20 states (including only states with more than eight counties), that the fit of the relationship was good in only four states (New York, Pennsylvania, Massachusetts, and Minnesota), and that fit was reversed in one of these states (Massachusetts). A subanalysis of 40 high incidence counties with NIP data identified some of the factors that contribute to geographic differences in the relationship of rDIN and incidence, most notably, genetic composition of *B. burgdorferi*.

One factor that has been proposed to modify the relationship between DIN and reported cases is variation in *B. burgdorferi* genotype frequency.<sup>26,32</sup> Differences in virulence<sup>37</sup> and transmission efficiency<sup>25,38</sup> may lead to differences in infection risk and/or reporting rates by humans infected with different strains. Thus, areas with low incidence but high rDIN could be explained by an overrepresentation

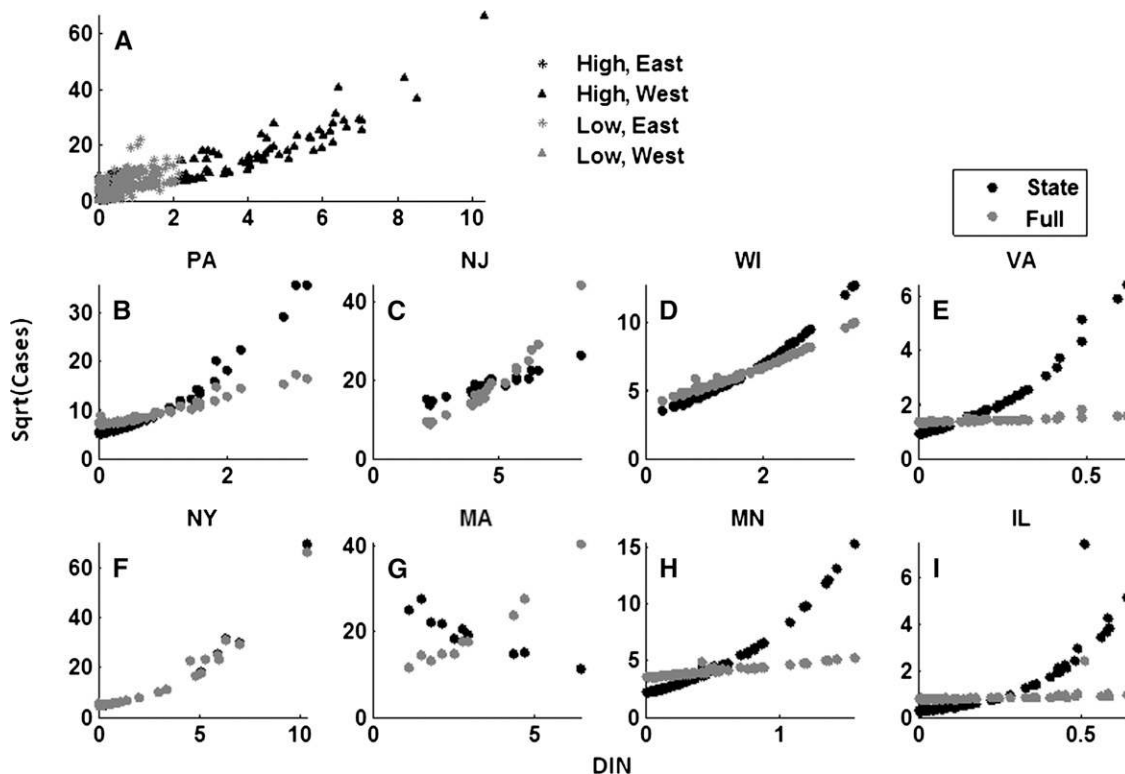


FIGURE 3. Relationship of DIN and cases by state. The full model (A) included DIN as a continuous predictor and state as a 23-level factor. Because state was highly significant, we also modeled the relationship of DIN and human cases individually for each state (B–I) and compared these relationships (black circles) to predictions from the full model (grey circles). Only some of the states from each region are shown to illustrate differences in performance between the national-level and state-based models. (B, C, F, and G) High prevalence, east. (D and H) High prevalence, west. (E and I) Low prevalence. Statistics are summarized in Table 4.

of strains with a low virulence in or transmission to humans. Although more studies are needed to confirm this hypothesis, our regression analyses support it for RST 3. We found that (1) incidence data could be partitioned similarly by region or RST 3, (2) RST 3 was more frequent in the region with lower incidence but similar rDIN, and (3) there was a negative relationship between RST 3 and incidence estimated by the Poisson regression model. Although the other two RST types each showed different regional distributions, neither of these two types alone caused deviations from the mean relationship of rDIN and reported incidence. Nevertheless, because higher levels of RST 3 (a genetically diverse group defined as non-RST 1 or 2) led to lower than expected incidence based on rDIN, by extrapolation, the sum of RST 1 and RST 2 frequencies was correlated with higher than expected incidence based on rDIN. Thus, our results suggest that the monophyletic RST 1 and 2 groups<sup>39</sup> cause higher than expected reported cases based on rDIN, whereas a high prevalence of other strains results in fewer reported cases. More generally, our results highlight that genotypic differences are important for predicting human incidence spatially through acarological risk indices. However, more analyses of the genetic determinants of virulence (and transmission) in humans and the spatial dynamics of *B. burgdorferi* genetic diversity are needed to understand the effects of *B. burgdorferi* genotype on reported incidence.

In addition to *B. burgdorferi* genotype, there are other factors not measured in our study that could contribute to geographical differences and weak associations in the relationship between mDIN and human incidence. One possibility is human movement. Although peridomestic exposure accounts for most cases of Lyme disease,<sup>12</sup> some cases are a consequence of travel to other counties for outdoor work or recreation. This movement could explain the negative relationship between mDIN and cases that was found for Massachusetts. For example, if Massachusetts residents tend to get infected in counties that are densely forested but live in counties (i.e., where cases would be reported) that are mainly urban (i.e., low estimates of mDIN), then such a negative relationship could be observed. More accurate spatial risk predictions might be obtained if case reports included probable place of exposure, although accurate estimates are very challenging to obtain.<sup>40</sup> Also, we found that spatial scale was crucial for interpreting the relationship between mDIN and incidence, because we showed that there are quantitative differences regionally (Midwest versus Northeast) as well as between states within these regions. It is possible that the county scale is appropriate in states with larger forested patches, such as New York and Pennsylvania (which showed relatively strong positive relationships between mDIN and incidence), whereas a finer scale is necessary in highly residential states with very dispersed forest cover, such as New Jersey and Maryland (which showed no significant relationship). For example, a study that measured DIN in forested areas in six Rhode Island towns found that DIN explained 97% of the variation in incidence among those towns.<sup>9</sup> However, most of our study sites were not located in residential areas, and the incidence data corresponded to a much broader range relative to the acarological data, which increases the variance in the relationship between mDIN and incidence. Last, geographical differences in Lyme disease control efforts at the residential or personal level could also

contribute to variation in the relationship between mDIN and human incidence.

There are also some experimental design caveats that should be noted when interpreting our results. First, our estimates of mDIN were based on data collected exclusively in public forested areas. Thus, our estimates of mDIN in counties that are mainly residential could be inaccurate. Second, sampling at different sites was done at different times between May and early September, and although most sites were sampled at least five times, some sites were only sampled three times. Although these sampling design caveats likely account for some of the unexplained variation in the relationship between DIN and human incidence, they were not spatially systematic (i.e., they occurred randomly with respect to state and region) and thus, should not introduce systematic bias in our results. Third, this analysis is based on original field efforts that were optimized for collections of *I. scapularis* nymphs based on phenology in the northeastern United States (emergence in late May); therefore, areas in which nymphs become active earlier would have inaccurate estimates of nymphal densities. For example, a previous study found that nymphs were observed as early as April in the south.<sup>41</sup> However, even with intensive sampling, only a few specimens were recovered at this time; thus, this difference should not introduce much bias in our results. There are also two other factors unaccounted for in our sampling design that could cause geographic variation in human incidence between northern and southern areas. First, because we only sampled *I. scapularis*, the existence of an alternative vector might uncouple the relationship between mDIN and incidence. However, this possibility would only occur if the alternative vector caused a significant proportion of human cases, and there is no evidence to support this case. Second, it has been reported that nymphs in southern areas very rarely feed on humans,<sup>41</sup> although it is still unclear whether this difference from northern areas is because of the low densities of *I. scapularis* nymphs or an actual feeding preference. Future environmental surveillance should aim to quantify DIN in residential areas as well as other habitats and potential vectors to better understand how DIN translates to incident human cases across varying ecological conditions.

## CONCLUSION

Reservoir- or vector-targeted control of human Lyme disease relies on understanding how DIN corresponds to human incidence geographically. Additionally, a mechanistic understanding of the link between DIN and incidence should be sought, because it is important for model frameworks that anticipate changes in the geographic range of human incidence and evaluate the cost efficacy of all types of prevention methods over time. Our multistate analysis showed large variation in the relationship of DIN and incidence between states, with only three states showing a reliable quantification of the relationship. To more accurately quantify how acarological risk translates to human Lyme disease risk spatially, longitudinal environmental surveillance of several factors at spatial scales finer than the county level would be needed. Unfortunately, such studies are not likely to be cost-effective if conducted using a random sampling design. Thus, future research to determine how acarological risk translates to human incidence spatially should first aim to



identify probable sites of transmission to humans over a broad spatial scale through epidemiological studies that monitor human movement (preferably to and from areas where acarological risk is quantified). Intensive environmental sampling in some of these sites would help in quantifying the relationship between DIN and human incidence more cheaply.

Received October 10, 2011. Accepted for publication February 26, 2012.

Note: Supplemental figures are available at [www.ajtmh.org](http://www.ajtmh.org).

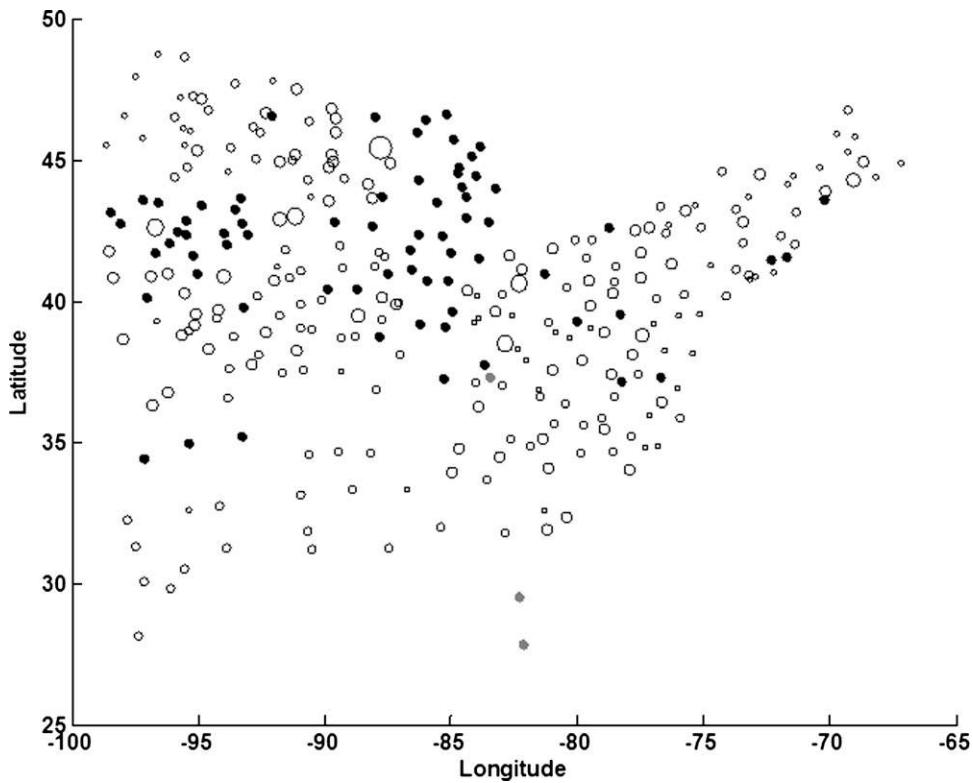
**Acknowledgments:** The authors thank Graham Hickling, Jean Tsao, Uriel Kitron, Edward Walker, Roberto Cortinas, Jonas Bunikis, and Michelle Rowland, who were key collaborators in the data collection and development of the acarological risk map. The authors thank the 80 field assistants who participated in the data collection; Tim Andreadis, Katherine Watkins, Laura Krueger, Jessica Payne, Elizabeth Racz, Kelly Liebman, Liza Lutzker, and David Boozer for tick identification and logistic support; Paul Cislo, Gwenaël Vourc'h, and Robert Brinkerhoff for statistical and data analyses support; Carlos Diuk for database support; Forrest Melton, Brad Lobitz, and Andrew Michaelis for technical assistance; and Dennis Grove, Lindsay Rollend, and Russell Barbour for arranging collection permits. This project was funded by Centers for Disease Control and Prevention—Division of Vector-Borne Infectious Diseases Cooperative Agreement no. C100171-01. K.M.P. was supported by the Research and Policy in Disease Dynamics (RAPIDD) program of the Science and Technology Directorate, the US Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. A.G.B. was supported by National Institute of Allergy and Infectious Disease Grant AI-065359.

**Authors' addresses:** Kim M. Pepin, Fogarty International Center, National Institutes of Health, Bethesda, MD; and Bacterial Diseases Branch, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, Fort Collins, CO, E-mail: [uzu2@cdc.gov](mailto:uzu2@cdc.gov). Rebecca J. Eisen, Paul S. Mead, and Joseph Piesman, Bacterial Diseases Branch, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, Fort Collins, CO, E-mails: [dyn2@cdc.gov](mailto:dyn2@cdc.gov), [pfm0@cdc.gov](mailto:pfm0@cdc.gov), and [jfp2@cdc.gov](mailto:jfp2@cdc.gov). Durland Fish and Maria A. Diuk-Wasser, Yale School of Public Health, New Haven, CT, E-mails: [durland.fish@yale.edu](mailto:durland.fish@yale.edu) and [maria.diuk@yale.edu](mailto:maria.diuk@yale.edu). Anne G. Hoen, Department of Community and Family Medicine, Dartmouth Medical School, Lebanon, NH, E-mail: [anne.g.hoen@dartmouth.edu](mailto:anne.g.hoen@dartmouth.edu). Alan G. Barbour, Departments of Microbiology and Molecular Genetics and Medicine, University of California, Irvine, CA, E-mail: [abarbour@uci.edu](mailto:abarbour@uci.edu). Sarah Hamer, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI; and Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, E-mail: [shamer@cvm.tamu.edu](mailto:shamer@cvm.tamu.edu).

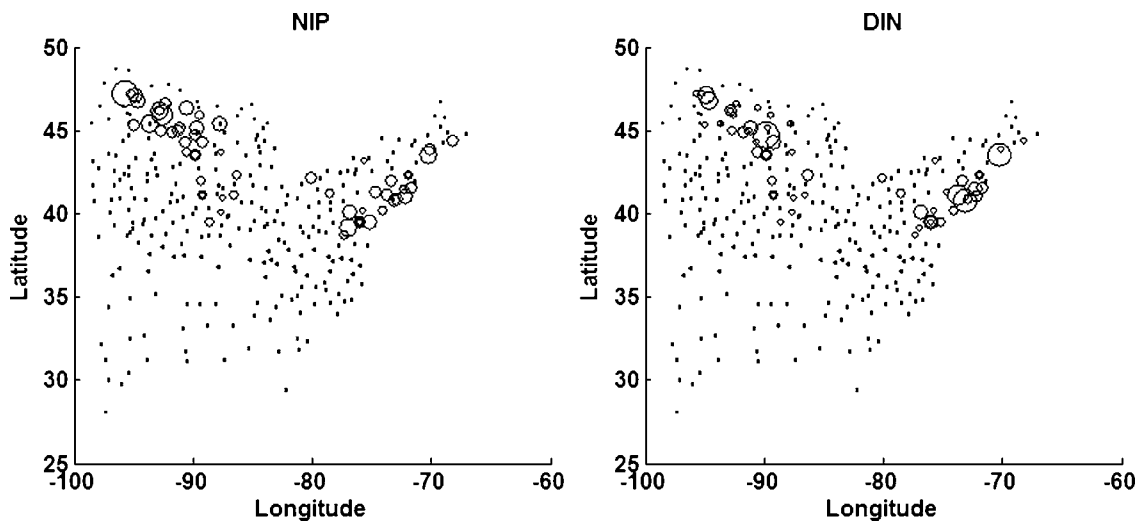
## REFERENCES

- CDC, 2011. *Reported Cases of Lyme Disease by Year, United States, 1996–2010*. Available at: <http://www.cdc.gov/lyme/stats/chartstables/casesbyyear.html>. Accessed January 2011.
- CDC, 2011. *Lyme Disease Incidence Rates by State, 2005–2010*. Available at: <http://www.cdc.gov/lyme/stats/chartstables/incidencebystate.html>. Accessed January 2011.
- Connally NP, Durante AJ, Yousey-Hindes KM, Meek JI, Nelson RS, Heimer R, 2009. Peridomestic lyme disease prevention results of a population-based case-control study. *Am J Prev Med* 37: 201–206.
- Gould LH, Nelson RS, Griffith KS, Hayes EB, Piesman J, Mead PS, Cartter ML, 2008. Knowledge, attitudes, and behaviors regarding lyme disease prevention among Connecticut residents, 1999–2004. *Vector Borne Zoonotic Dis* 8: 769–776.
- Hayes EB, Piesman J, 2003. Current concepts—how can we prevent Lyme disease? *N Engl J Med* 348: 2424–2430.
- Piesman J, Eisen L, 2008. Prevention of tick-borne diseases. *Annu Rev Entomol* 53: 323–343.
- Shen AK, Mead PS, Beard CB, 2011. The lyme disease vaccine—a public health perspective. *Clin Infect Dis* 52: S247–S252.
- Fish D, 1993. *Population ecology of Ixodes dammini*. Ginsberg H, ed. *Ecology and Environmental Management of Lyme Disease*. New Brunswick, NJ: Rutgers University Press, 25–42.
- Mather TN, Nicholson MC, Donnelly EF, Matyas BT, 1996. Entomologic index for human risk of Lyme disease. *Am J Epidemiol* 144: 1066–1069.
- Stafford KC, Cartter ML, Magnarelli LA, Ertel SH, Mshar PA, 1998. Temporal correlations between tick abundance and prevalence of ticks infected with *Borrelia burgdorferi* and increasing incidence of Lyme disease. *J Clin Microbiol* 36: 1240–1244.
- Connally NP, Ginsberg HS, Mather TN, 2006. Assessing peridomestic entomological factors as predictors for Lyme disease. *J Vector Ecol* 31: 364–370.
- Kitron U, Kazmierczak JJ, 1997. Spatial analysis of the distribution of Lyme disease in Wisconsin. *Am J Epidemiol* 145: 558–566.
- Falco RC, McKenna DF, Daniels TJ, Nadelman RB, Nowakowski J, Fish D, Wormser GP, 1999. Temporal relation between *Ixodes scapularis* abundance and risk for Lyme disease associated with erythema migrans. *Am J Epidemiol* 149: 771–776.
- Johnson JL, Ginsberg HS, Zhioua E, Whitworth UG, Markowski D, Hyland KE, Hu RJ, 2004. Passive tick surveillance, dog seropositivity, and incidence of human Lyme disease. *Vector Borne Zoonotic Dis* 4: 137–142.
- Rand PW, Lacombe EH, Dearborn R, Cahill B, Elias S, Lubelczyk CB, Beckett GA, Smith RP, 2007. Passive surveillance in Maine, an area emergent for tick-borne diseases. *J Med Entomol* 44: 1118–1129.
- Allan BF, Keesing F, Ostfeld RS, 2003. Effect of forest fragmentation on Lyme disease risk. *Conserv Biol* 17: 267–272.
- Brownstein JS, Skelly DK, Holford TR, Fish D, 2005. Forest fragmentation predicts local scale heterogeneity of Lyme disease risk. *Oecologia* 146: 469–475.
- Estrada-Pena A, 2009. Diluting the dilution effect: a spatial Lyme model provides evidence for the importance of habitat fragmentation with regard to the risk of infection. *Geospat Health* 3: 143–155.
- Ashley ST, Meentemeyer V, 2004. Climatic analysis of Lyme disease in the United States. *Clim Res* 27: 177–187.
- Brownstein JS, Holford TR, Fish D, 2003. A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. *Environ Health Perspect* 111: 1152–1157.
- Ogden NH, Maarouf A, Barker IK, Bigras-Poulin M, Lindsay LR, Morshed MG, O'Callaghan CJ, Ramay F, Waltner-Toews D, Charron DF, 2006. Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada. *Int J Parasitol* 36: 63–70.
- Ostfeld RS, Canham CD, Oggenfuss K, Winchcombe RJ, Keesing F, 2006. Climate, deer, rodents, and acorns as determinants of variation in Lyme-disease risk. *PLoS Biol* 4: 1058–1068.
- Gatewood AG, Liebman KA, Vourc'h G, Bunikis J, Hamer SA, Cortinas R, Melton F, Cislo P, Kitron U, Tsao J, Barbour AG, Fish D, Diuk-Wasser MA, 2009. Climate and tick seasonality are predictors of *Borrelia burgdorferi* genotype distribution. *Appl Environ Microbiol* 75: 2476–2483.
- Dykhuizen DE, Brisson D, Sandigursky S, Wormser GP, Nowakowski J, Nadelman RB, Schwartz I, 2008. Short report: the propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. *Am J Trop Med Hyg* 78: 806–810.
- Hanincova K, Ogden NH, Diuk-Wasser M, Pappas CJ, Iyer R, Fish D, Schwartz I, Kurtenbach K, 2008. Fitness variation of *Borrelia burgdorferi* sensu stricto strains in mice. *Appl Environ Microbiol* 74: 153–157.
- Wormser GP, Brisson D, Liveris D, Hanincova K, Sandigursky S, Nowakowski J, Nadelman RB, Ludin S, Schwartz I, 2008. *Borrelia burgdorferi* genotype predicts the capacity for hematogenous dissemination during early Lyme disease. *J Infect Dis* 198: 1358–1364.
- Diuk-Wasser MA, Gatewood AG, Cortinas MR, Yaremych-Hamer S, Tsao J, Kitron U, Hickling G, Brownstein JS, Walker E, Piesman J, Fish D, 2006. Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. *J Med Entomol* 43: 166–176.

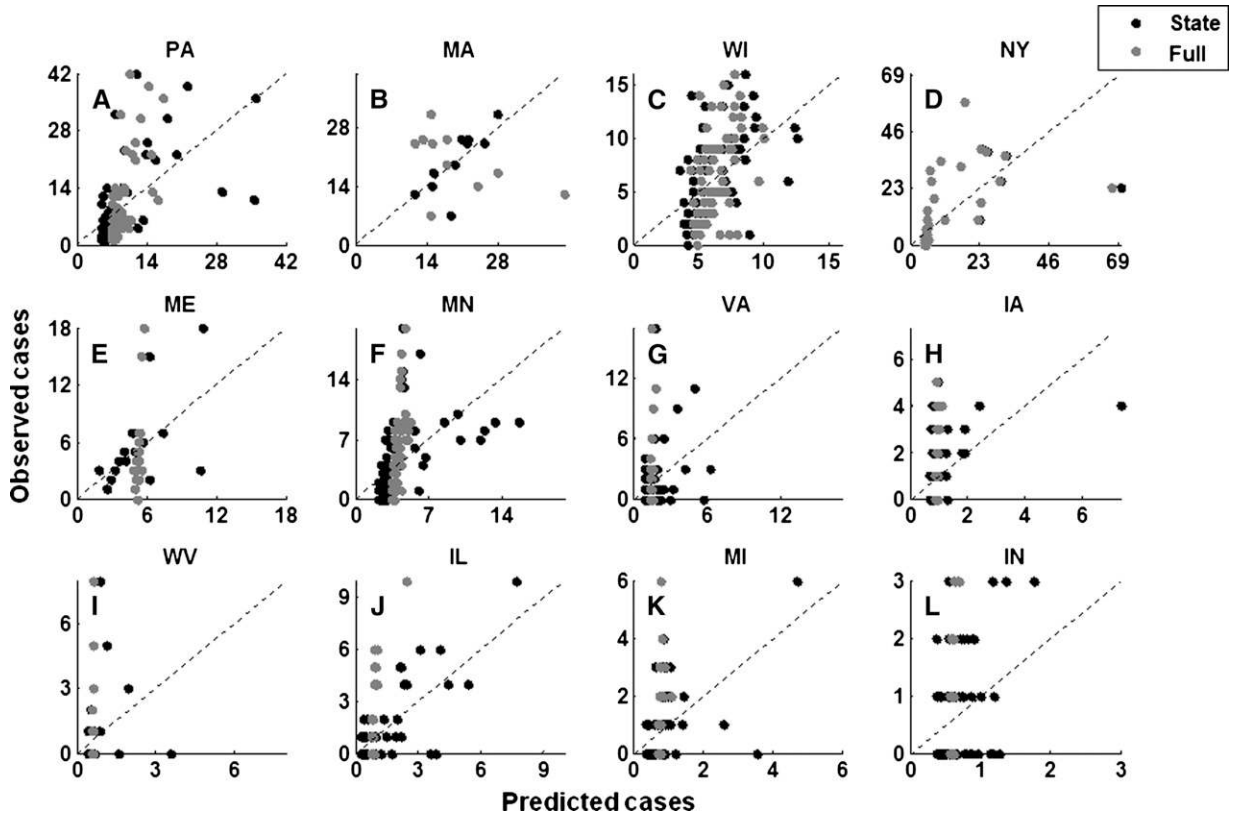
28. Diuk-Wasser MA, Vourc'h G, Cislo P, Hoen AG, Melton F, Hamer SA, Rowland M, Cortinas R, Hickling GJ, Tsao JI, Barbour AG, Kitron U, Piesman J, Fish D, 2010. Field and climate-based model for predicting the density of host-seeking nymphal *Ixodes scapularis*, an important vector of tick-borne disease agents in the eastern United States. *Glob Ecol Biogeogr* 19: 504–514.
29. Diuk-Wasser M, Gatewood A, Cislo P, Brinkerhoff R, Hamer S, Rowland M, Cortinas R, Melton F, Hickling G, Tsao J, Bunikis J, Barbour A, Kitron U, Piesman J, Fish D, 2012. Human risk of infection with *Borrelia burgdorferi*, the Lyme disease agent, in eastern United States. *Am J Trop Med Hyg*, 86: 320–327.
30. Bacon RM, Kugeler KJ, Mead PS, 2008. Surveillance for Lyme disease—United States, 1992–2006. *MMWR Surveill Summ* 57: 1–9.
31. Guerra M, Walker E, Jones C, Paskewitz S, Cortinas MR, Stancil A, Beck L, Bobo M, Kitron U, 2002. Predicting the risk of Lyme disease: habitat suitability for *Ixodes scapularis* in the north central United States. *Emerg Infect Dis* 8: 289–297.
32. Wormser GP, Liveris D, Nowakowski J, Nadelman RB, Cavaliere LF, McKenna D, Holmgren D, Schwartz I, 1999. Association of specific subtypes of *Borrelia burgdorferi* with hematogenous dissemination in early Lyme disease. *J Infect Dis* 180: 720–725.
33. Wang G, Ojaimi C, Iyer R, Saksenberg V, McClain SA, Wormser GP, Schwartz I, 2001. Impact of genotypic variation of *Borrelia burgdorferi sensu stricto* on kinetics of dissemination and severity of disease in C3H/HeJ mice. *Infect Immun* 69: 4303–4312.
34. Estrada-Pena A, 2009. Tick-borne pathogens, transmission rates and climate change. *Front Biosci* 14: U2674–U2780.
35. Tsao JI, Wootton JT, Bunikis J, Luna MG, Fish D, Barbour AG, 2004. An ecological approach to preventing human infection: vaccinating wild mouse reservoirs intervenes in the Lyme disease cycle. *Proc Natl Acad Sci USA* 101: 18159–18164.
36. Hothorn T, Hornik K, Zeileis A, 2006. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Statist* 15: 651–674.
37. Jones KL, Glickstein LJ, Damle N, Sikand VK, McHugh G, Steere AC, 2006. *Borrelia burgdorferi* genetic markers and disseminated disease in patients with early Lyme disease. *J Clin Microbiol* 44: 4407–4413.
38. Derdakova M, Dudioak V, Brei B, Brownstein JS, Schwartz I, Fish D, 2004. Interaction and transmission of two *Borrelia burgdorferi sensu stricto* strains in a tick-rodent maintenance system. *Appl Environ Microbiol* 70: 6783–6788.
39. Bunikis J, Garpmo U, Tsao J, Berglund J, Fish D, Barbour AG, 2004. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology* 150: 1741–1755.
40. Eisen L, Eisen RJ, 2007. Need for improved methods to collect and present spatial epidemiologic data for vectorborne diseases. *Emerg Infect Dis* 13: 1816–1820.
41. Goddard J, Piesman J, 2006. New records of immature *Ixodes scapularis* from Mississippi. *J Vector Ecol* 31: 421–422.



SUPPLEMENTAL FIGURE 1. Spatial variation in sampling design. The proportion of drags that were completed in early (May and July) versus late (August and September) summer by location. Open circles are scaled relative to the proportion of drags (early/late summer); larger circles indicate stronger sampling intensity during early summer. Filled black circles indicate locations where no samples were collected in late summer; filled grey circles indicate locations where no samples were collected in early summer. Size of filled circles is scaled relative to the mean proportion at all other sites.



SUPPLEMENTAL FIGURE 2. Spatial variation in prevalence (NIP) and DIN. NIP and DIN are plotted for each site sampled. Dots indicate sites where there were too few ticks to estimate NIP (and hence, DIN). Size of the circles is scaled to the levels of NIP or DIN.



SUPPLEMENTAL FIGURE 3. Accuracy of model predictions. Observed cases are plotted against predicted cases from the state-based model (black) and the national-level model (grey). Cases were square-root transformed. Points that fall along the dotted line indicate a perfect fit. Only states for which the state-based model was significant are shown.