Environ Ecol Stat (2007) 14:83–100
DOI 10.1007/s10651-006-0002-z

ORIGINAL ARTICLE

Spatio-temporal patterns in county-level incidence and reporting of Lyme disease in the northeastern United States, 1990–2000

Lance A. Waller · Brett J. Goodwin · Mark L. Wilson · Richard S. Ostfeld · Stacie L. Marshall · Edward B. Hayes

Received: 16 September 2004 / Revised: 28 February 2005 /
Published online: 20 January 2007

Abstract We present an exploratory analysis of reported county-specific incidence of Lyme disease in the northeastern United States for the years 1990–2000. We briefly review the disease ecology of Lyme disease and the use of risk maps to describe local incidence as estimates of local risk of disease. We place the relevant elements of local environmental and ecological variables, local disease incidence, and (importantly) local disease reporting in a conceptual context to frame our analysis. We then apply hierarchical linear models of increasing complexity to summarize observed patterns in reported incidence, borrowing information across counties to improve local precision. We find areas of increasing incidence in the central northeastern Atlantic coast counties, increasing incidence branching to the north and west, and an area of fairly stable and slightly decreasing reported incidence in western New York.

L. A. Waller (✉)
Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA
e-mail: lwaller@sph.emory.edu

B. J. Goodwin
Department of Biology, University of North Dakota, Grand Forks, ND 58202, USA

M. L. Wilson
Departments of Biology and Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA

R. S. Ostfeld
Institute of Ecosystem Studies, Millbrook, NY 12545, USA

S. L. Marshall
US Centers for Disease Control and Prevention, Emory University, Atlanta, GA 30322, USA

E. B. Hayes
US Centers for Disease Control and Prevention, Colorado State University, Fort Collins, CO 80523, USA

Keywords Hierarchical linear model · Risk map · Local risk · Conceptual context · Local precision

1 Introduction

Lyme disease is a zoonosis caused by the spirochete *Borrelia burgdorferi*, which cycles between tick vectors and dozens of species of vertebrate hosts. The first medical description of Lyme disease appears in Steere et al. (1977), although earlier reports describe incident cases of erythema migrans, the localized rash serving as a primary symptom of Lyme disease (Scrimenti 1970). The US Centers for Disease Control and Prevention (CDC) began surveillance for the disease in 1982 (Schmid et al. 1985), working with state and local health departments and health care providers to encourage reporting of cases. In 1990, the Council of State and Territorial Epidemiologists approved a standardized case definition for reporting and agreed to implement state requirements to report Lyme disease; these policies were implemented nationally in 1991 (Orloski et al. 2000). Currently, Lyme disease is one of the most common vector-borne diseases in the United States, with almost 90,000 cases reported between 1992 and 1998 (Orloski et al. 2000). Although cases have been reported from 49 states and the District of Colombia (Orloski et al. 2000), the bulk of reported cases has occurred in the northeastern United States (Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont, and the District of Columbia), with a second concentration in the midwestern United States (Michigan, Minnesota, and Wisconsin).

The ecology of Lyme disease depends on the life cycles and interactions between the causative agent, the vector, and the various hosts. Larval and nymphal tick vectors (the black-legged tick, *Ixodes scapularis*, in the eastern United States) feed primarily on smaller vertebrates (Lane et al. 1991; Barbour and Fish 1993; Mather 1993). White-footed mice (*Peromyscus leucopus*) play a crucial role as both a preferred immature tick host and principal reservoir of *B. burgdorferi* (Levine et al. 1985; Mather et al. 1989; Anderson and Magnarelli 1993; Fish 1993; Mather and Ginsberg 1994; Keirans et al. 1996; Donahue et al. 1997). Adult ticks feed primarily on white-tailed deer (*Odocoileus virginianus*) (Piesman and Spielman 1979; Wilson et al. 1990). Each generation of ticks must become infected via these vertebrate hosts, because there is no transovarial transmission between egg-laying females and their offspring (Piesman et al. 1986; Patrican 1997). Once infected, ticks can infect a human host during a blood meal (Barbour and Fish 1993).

There can be a great deal of spatial variability in Lyme disease incidence. The tick vectors are patchily distributed both regionally and locally, regardless of infection status (Kitron and Kazmierczak 1997; Wilson 1998). In particular, the presence of *I. scapularis* is positively associated with sandy soils, woody or shrubby vegetation, and the presence of deer (Ginsberg and Ewing 1989; Kitron et al. 1991, 1992; Duffy et al. 1994; Glass et al. 1994, 1995). Countering this spatial variability in vector presence is the tendency of tick loads on mice to remain relatively constant, even in the face of substantial variation in the densities of both mice and questing ticks (Goodwin et al. 2001). Host abundance and community composition also can vary dramatically in both space and time (van Buskirk and Ostfeld 1995; Giardina et al. 2000).

Spatial variability of Lyme disease hosts and vectors suggests that construction of a *risk map*, i.e., a map of the potential risk of (human) infection could be

useful. Such a map would indicate areas where tick control, public education, or other interventions might be most beneficial (Kitron 2000; Orloski et al. 2000). Maps of vector abundance (perhaps as functions of environmental variables), vector infection rates, human-vector interactions, and reported human cases all address different components of the spatial pattern of risk of Lyme disease. Each map requires different data and reveals different elements of the interacting processes determining Lyme transmission. The literature contains examples of risk maps for Lyme disease for particular communities (Glass et al. 1995; Dister et al. 1997), for particular states (Kitron and Kazmierczak 1997; Frank et al. 2002), and for the entire nation (Dennis et al. 1998; Estrada-Pena 1998; US CDC 2001). Risk maps for Lyme disease have been constructed from incidence data (Kitron and Kazmierczak 1997; Frank et al. 2002), tick distributions (Dennis et al. 1998), and environmental risk factors (Glass et al. 1995; Dister et al. 1997; Estrada-Pena 1998; US CDC 2001). Risk maps tend to be constructed either from a snapshot of the disease at a particular point in time (Glass et al. 1995; Dister et al. 1997), or by pooling or averaging disease incidence across the entire period of surveillance (Kitron and Kazmierczak 1997; Estrada-Pena 1998; US CDC 2001; Frank et al. 2002).

Risk maps are potentially useful tools for public health practitioners, but Lyme disease presents particular challenges in their creation, analysis, and interpretation. Surveillance data for Lyme disease can be problematic for a number of reasons: recognition by health-care providers of the signature symptoms over the study period, the associated potential lack of accurate diagnoses, imprecise serologic results, uneven case detection, reporting biases, and difficulty relating location of report to location of exposure (Kitron and Kazmierczak 1997). Thus, Lyme disease may be substantially under-reported, with the probability of reporting varying with stage of disease, age of the patient, and the interest and collaboration of healthcare providers (Orloski et al. 1998; Naleway et al. 2002). Despite these concerns, surveillance data show that both the incidence and the geographic range of human cases has increased steadily, particularly in the northeastern United States (White et al. 1991; Orloski et al. 2000), spreading outward from the initial diagnoses near Lyme, Connecticut. Kitron (2000) provides a brief but thorough discussion of risk maps for vector-borne diseases, which we expand on in the discussions that follow.

In this study, we explore spatio-temporal patterns in temporal changes in *reported* incidence for human cases for the years 1990–2000. Our aim is primarily exploratory, as any observed patterns and changes in reported incidence reflect a combination of patterns in true incidence and local diagnosis and reporting rates. The descriptive goal is important, as it provides insight into the evolving geographic coverage of the disease, as well as into evolving coverage of the surveillance system monitoring the disease. Although we cannot entirely separate the two components, we can draw conclusions on particular aspects of each.

In particular, we report county-level trends in reported Lyme disease incidence (rates per 100,000 persons) in the northeastern United States for the years 1990–2000, using data from the CDC's national surveillance system. The goal of this study is a descriptive analysis of county-specific *changes* in the annual rate of reported cases, i.e., assessment of the spatial pattern of rate changes in an attempt to identify areas experiencing fast observed growth. It is important to note that observed growth (or decline) could be a result of true increases (decreases) in disease incidence, or of changes in reporting practices, surveillance effort, and accuracy of diagnosis. All of these issues are likely to be of interest here, and motivate our discussion of risk maps

in Sect. 3. We also examine geographic patterns of missing county-level reports and their relationships with observed patterns in the incidence of disease. Previous investigations by the CDC concentrate on the years 1994–2000, because of varying data availability prior to 1994. We include available data from 1990–1994 and explore its compatibility with the proposed models.

2 Lyme disease surveillance data

Since 1992, the CDC has compiled reported cases of Lyme disease based on the standardized case definition approved by the Council of State and Territorial Epidemiologists (Orloski et al. 2000). Using this data base, we determined the annual number of cases of Lyme disease in each county in the northeastern United States (Pennsylvania, Maryland, Delaware, Massachusetts, New Jersey, New Hampshire, Vermont, New York, Connecticut, Rhode Island, and Maine, and the District of Columbia) from 1990 to 2000. To standardize for differences in county population size, we converted raw case numbers into incidence proportions per 100,000 people at risk, using the 1990 US Census population sizes. For simplicity, we used the 1990 Census population sizes as denominators, noting that adjustments for population growth would change results only slightly.

Before examining the data, some review of the nature of public health surveillance data (Teutsch and Churchill 1994; Brookmeyer and Stroup 2004) is in order to provide a context for the analyses that follow. In particular, Lyme disease surveillance data comprise a mix of *passive*, *active* and *laboratory* surveillance (Orloski et al. 2000). *Passive surveillance* relies on health-care providers to report any new diagnoses to either local or state public health departments. The health department in turn assesses the report with respect to the standardized case definition (perhaps including confirmatory laboratory tests), then electronically submits those cases meeting the standard to the CDC via the National Electronic Telecommunication System for Surveillance (NETSS). *Active surveillance* involves proactive contact of health-care providers by local or state health departments requesting information regarding any incident diagnoses of Lyme disease, followed by case assessment and reporting to the CDC. Finally, *laboratory surveillance* requires diagnostic laboratories to report all positive Lyme disease test results directly to the state or local health department. Because laboratory results typically include only limited patient information, the health department often must follow up with health care providers to verify that a case meets the standardized case definition before reporting the case to the CDC. Not all states include active and laboratory surveillance efforts, and those in the study area that do (Connecticut, Maryland, Massachusetts, New Jersey, New York, and Rhode Island) include them for only part of the study period (Orloski et al. 2000).

For any particular year, a number of counties do not provide reports regarding Lyme disease. We term such occurrences as “no reports” (NRs). The NRs may represent an absence of cases, an absence of diagnoses, or an absence of reports. Table 1 gives the number of counties with NRs by year. In each year and in each state, some reported cases were attributed to the state, but not to a specific county, and we ignore these cases in the following analysis. In a typical year, we find approximately 45–50 NRs from the 245 counties in the study area. The higher numbers of NRs associated with 1992 (99 counties) and 1993 (106 counties) are due primarily to Pennsylvania (67 counties) reporting cases only at the state level for these 2 years.

Table 1 Numbers of reported cases of Lyme disease in the northeastern United States, 1990–2000

Year	Total number of cases reported	Number of counties with (non-zero) reports	Number of counties with "no report" (NR)	Cases identified to state level, but not county
1990	6,108	189	56	130
1991	7,355	195	49	455
1992	8,041	146	99	1,333
1993	6,829	139	106	1,235
1994	11,454	200	45	81
1995	10,380	197	48	237
1996	15,023	201	44	296
1997	11,280	199	46	184
1998	15,111	201	44	114
1999	14,480	192	53	55
2000	15,621	189	56	156

Figure 1 provides maps of the total county-specific number of reported cases per 100,000 population (left map) and the county-specific number of years with NRs. We shade each map by quintile so roughly one-fifth of the counties appear in each grayscale. For visual comparison, darker grayscales indicate *higher* quintiles of cases and *lower* quintiles of NRs. The visual similarity in pattern roughly suggests a pattern of reporting mirroring that of (reported) disease incidence.

In general, as might be expected, we note higher numbers of reported cases per 100,000 persons at risk during the study period in counties along the central Atlantic coast, with reduced numbers radiating out from the locations of the initial documented cases in Connecticut. We note a pocket of reduced numbers of reports in the burroughs of New York City, and in western Massachusetts.

Figure 1 also appears to reveal some isolated pockets of higher reporting coverage that appear ahead of the generally perceived wavefront of incidence, as indicated by higher numbers of areas reporting cases. These areas generally correspond to counties with higher population densities than their surrounding neighbors (e.g., counties near Ithaca, Rochester, and Buffalo, New York). This could be due to the increased number of individuals at risk, but it might also reflect communication patterns among physicians and subsequent raised awareness (making cities "closer" in a communication sense than they appear in a mileage sense).

3 Elements of a risk map

Any spatial summary (map or analysis) of disease incidence requires a proper context for interpretation. As mentioned in the introduction, multiple spatial processes contribute to the observed pattern of Lyme disease incidence (in humans). Figure 2 shows these components and focuses the purpose of this report.

Beginning on the left-hand side of Fig. 2, we observe the environmental composition providing a background setting for the disease process. These factors (e.g., type of soil, land use, forest) typically are available for mapping by site surveys or remote sensing. The factors describe the habitats for hosts (deer mice, white tailed deer, and humans), as well as the vector (*I. scapularis*), and define areas

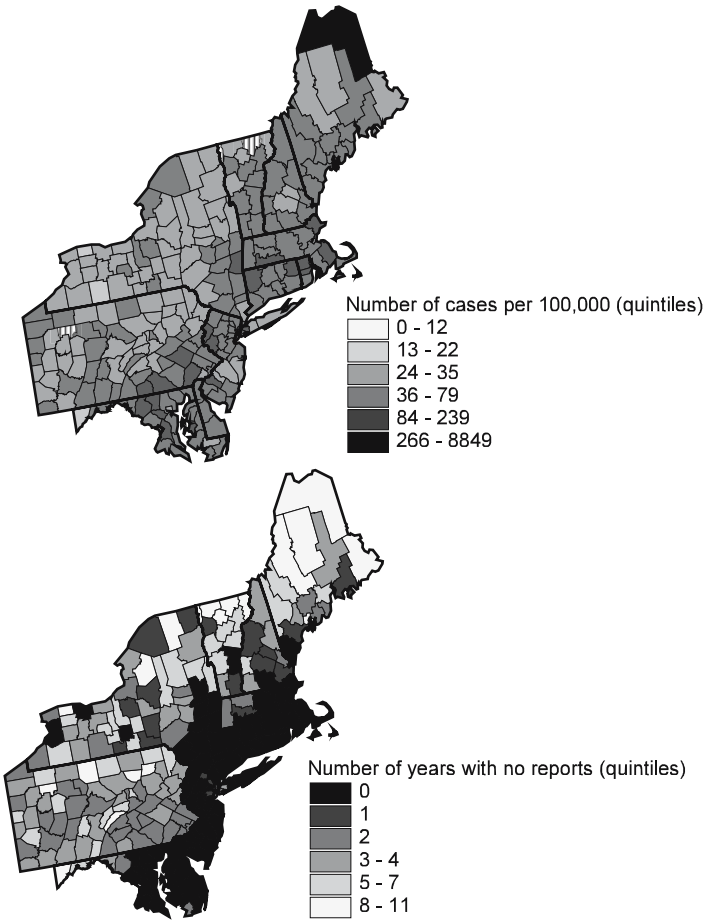

Fig. 1 Maps of numbers of Lyme disease cases per 100,000 population (1990 US Census) reported to the US Centers for Disease Control (*left*) and number of years with no (county-specific) reports (*right*)

where the hosts and (infected) vectors are likely to interact, allowing transmission. People, mice, deer, and ticks moving through this environment interact, at times allowing transmission of infection to humans at points indicated in the middle map.

The (unobserved) middle map in Fig. 2 represents the location of infection for all true cases. Coupled with information regarding the number of people at any location at a given time, this map would provide information regarding the true risk of infection as a function of location and time. However, we often do not observe the true incidence map, but rather a "filtered version" represented by the two schematic maps on the right-hand side of Fig. 2. The shaded map represents the probability of a given case being diagnosed and reported as a function of county; darker shading represents counties with higher probabilities of accurate diagnosis and reporting. Diagnosis and reporting are two separate processes, each with its own potential for geographic variation, but for simplicity, we aggregate them as a single

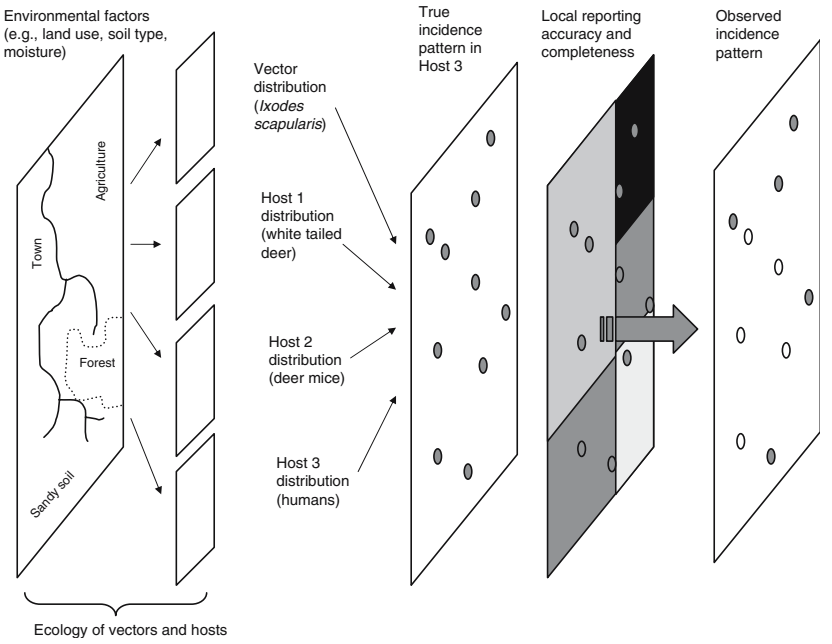

Fig. 2 Conceptual elements of a Lyme disease risk map. Moving from left to right, we find the geographic habitat structure for both hosts and vector, next the geographic distributions of various hosts and the vector, then the unobserved true incidence pattern which is "filtered" by county-specific reporting rates, yielding the observed pattern of reported locations (**gray circles**) and the hidden pattern of unreported locations (**open circles**)

surveillance reporting filter in Fig. 2. The result of this "filter" is the observed incidence pattern on the right, where we observe only the filled circles (we also show the "lost" cases via empty circles in Fig. 2, but these would not be observed). We note that the surveillance filter can add or delete cases (e.g., due to false positive or false negative diagnoses, respectively), an issue of particular importance for diseases and syndromes with high public awareness and relatively modest diagnostic tools (e.g., Lyme disease). Finally, we note that location uncertainty also has an impact on reporting, because the reported residential locations might not correspond to infection locations, as mentioned in the introduction, but we ignore such errors here for clarity.

Figure 2 is helpful, not only in providing a view of the complexity of the situation, but also in focusing attention and aiding interpretation. An accurate risk map (for human infection) seeks an unbiased estimate of the center map, but descriptions of the components of this risk map are also of interest. Some previous risk maps of Lyme disease focus only on mapping the habitat and location of the vector (e.g., Glass et al. 1995); others address only the reported outcomes (e.g., US CDC 2001). The former limits attention to the left-hand maps in Fig. 2, the latter the map on the far right.

Analysts typically view the (shaded) "filter" map as a nuisance, a source of bias between the observed and true incidence maps, but we suggest this map merits attention in its own right as a summary of spatial pattern and variation in the disease

surveillance system itself. That is, estimates of this shaded map provide valuable information regarding the surveillance process, identifying areas with adequate reporting and others needing improvement. The filter map can be further deconstructed into additional maps of local reporting practices, local diagnostic accuracy, and even local health-seeking behavior. Each map in Fig. 2 merits public health interest in its own right, but also contributes to the overall map of observed disease incidence and risk.

Closer examination of the maps of incidence counts and no reports (Fig. 1) provides insight into patterns in the typically unobserved "reporting map" conceptually defined in Fig. 2. More specifically, consider the five burroughs of New York City and the counties in western Massachusetts that reflect concentrations of NRs within the general "core" area of reporting. In both, reduced reporting also appears to play a role in defining the pattern of reported cases.

For this report, our approach is to view temporal changes in the spatial pattern on the right-hand side of Fig. 2 as a means of gaining partial insight into both the reported incidence map and the shaded reporting map. Although we cannot completely separate the two (which would provide clearer insight into the true risk map in the center of Fig. 2), we do find interesting patterns suggesting evolution of both incidence and reporting over the study period. These patterns provide an important step toward understanding both the spread of infection and local aspects of the surveillance system itself.

4 Statistical models to describe patterns

To describe the evolution of Lyme disease incidence and reporting in the northeastern United States, we consider local linear regressions of the natural logarithm of the annual crude incidence ratios for each county. For simplicity (and supported by general diagnostics), we assume that log-incidence ratios follow normal (Gaussian) distributions, and explore temporal trends in the reported ratio of cases to the 1990 county-level population size for each county.

We ignore NRs in our local regressions of log incidence rates, but then incorporate them as missing observations in a broader Bayesian context, allowing posterior prediction of the missing values rather than simply ignoring them. Some NRs, particularly those near the periphery of the study area, could reflect zero counts that were not reported, while likely values of NRs within prevalent areas are more difficult to ascertain. In any case, assumptions of "missing at random" or "missing completely at random" appear overly simplistic. Finally, recall that some NR patterns, such as those in Pennsylvania, are direct results of evolving reporting practices and requirements.

To begin, let Y_{it} and n_i denote the number of reported cases and the 1990 population size, respectively, for county i in year t (we omit a time subscript on n_i since we use only the 1990 population sizes). For each county, we fit (via least squares) the local linear regression model

$$\log(Y_{it}/n_i) = \beta_0 + \beta_1 t + \varepsilon_{it},$$

where the error term, ε_{it}, follows a normal (Gaussian) distribution with variance σ_i^2. For the local models, we assume a constant variance within (but not necessarily between) counties. For comparison, we also fit a single regression to the county-specific data within each state (now assuming constant variance across each state).

Not surprisingly, there is considerable variation in county-specific reported Lyme disease incidence. We next consider random effects models, wherein both the slope and the intercept vary by county. We use a hierarchical Bayesian formulation, allowing either spatially unstructured (exchangeable) or spatially structured slopes or intercepts. First, for comparison with the local (maximum likelihood) regression models, we consider a Bayesian version of the foregoing overall regression model, in which we assign diffuse prior distributions to both β_0 and β_1, as a comparison to the least squares (maximum likelihood) results for the model.

Next, we generalize the model to

$$\log(Y_{it}/n_i)|u_{0i}, u_{1i} \sim \mathrm{N}(\mu_{it}, \sigma^2),$$
$$\mu_{it} = \beta_0 + \beta_1 t + u_{0i} + u_{1i} t,$$

where μ_{it} represents the expected local log incidence rate (given the values of the local random effects, u_{0i} and u_{1i}). We now assume a constant variance across the entire study area in order to link regression models across counties. Retaining diffuse priors for β_0 and β_1, we assign mutually independent, exchangeable priors

$$u_{0i} \sim \mathrm{N}(0, {\sigma_{u0}}^2) \text{and}$$
$$u_{1i} \sim \mathrm{N}(0, {\sigma_{u1}}^2)$$

to the county-specific random effects that serve as local adjustments to the overall intercept and slope. We complete the model by assigning vague conjugate inverse-gamma hyper-priors to ${\sigma_{u0}}^2$ and ${\sigma_{u1}}^2$—specifically inv-gamma(0.5, 0.0005) priors as suggested by Kelsall and Wakefield (1999). The result is a model with county-specific adjustments centered around an overall slope and intercept. The impact of the local random effects is to yield posterior estimates of the local intercept, $(\beta_0 + u_{0i})$, and slope, $(\beta_1 + u_{1i})$, representing compromises between the overall values across all counties and the values suggested by each set of county-specific data, i.e., the model "borrows strength" across all counties to improve estimates of the county-specific intercepts and slopes.

Since the elements of a Lyme disease risk map outlined in Fig. 2 involve spatial patterns, we also consider borrowing strength locally rather than globally by replacing one or both of the exchangeable priors with a spatially correlated prior, effectively yielding either intercepts or slopes that are compromises between values supported by the local data and those supported by spatially neighboring counties. A popular family of such spatial prior distributions is the set of conditionally autoregressive (CAR) priors (Besag et al. 1991; Wakefield et al. 2000; Waller and Gotway 2004, Sect. 9.5), and we consider those defined by

$$u_{1i}|u_{1j}, j \neq i \sim \mathrm{N}\left[\left(\sum_j w_{ij}u_{1j} \Big/ \sum_j w_{ij}\right), 1 \Big/ \left({\sigma_{u1}}^2 \sum_j w_{ij}\right)\right] \textit{and}$$
$$u_{0i}|u_{0j}, j \neq i \sim \mathrm{N}\left[\left(\sum_j w_{ij}u_{0j} \Big/ \sum_j w_{ij}\right), 1 \Big/ \left({\sigma_{u0}}^2 \sum_j w_{ij}\right)\right],$$

where $w_{ij} = 1$, if regions i and j share a common border, and 0 otherwise. By convention, $w_{ii} = 0$. Besag (1974) shows that this set of conditional distributions defines a valid joint normal distribution, although the use of adjacency weights defines a singular precision matrix yielding an improper prior distribution (Besag and Kooperberg 1995;

Besag et al. 1995; Waller and Gotway 2004, Sect. 9.5). However, Besag et al. (1995) show that the impropriety is caused by the joint distribution being defined for pairwise contrasts between local random effect values; the addition of a "sum to zero" constraint on the random effects allows proper posterior inference for all model parameters. We complete the model by again defining diffuse conjugate inverse-gamma(0.5, 0.0005) hyper-prior distributions for ${\sigma_{u0}}^2$ and ${\sigma_{u1}}^2$ (Kelsall and Wakefield 1999). Such models are very common in spatial smoothing of regional disease rates (disease mapping), but most models use only random intercepts with some notable exceptions (Knorr-Held and Besag 1998; Schootman and Sun 2004). Our interests here involve changes over time, motivating consideration of the CAR prior for the slope parameters.

5 Results

Figure 3 reveals the crude county-specific 1990–2000 time trends in $\log(Y_{it}/n_i)$ by state, with state-specific trends indicated by dashed gray lines. The plots suggest the adequacy of a linear model of log incidence and indicate a fair amount of relatively symmetric variation around the general trend observed in each state, with possible exceptions in New York (NY), Pennsylvania (PA), and one county in Vermont (VT). Gaps in reporting are also apparent.

For comparison with results from more complicated models, Fig. 4 provides a map of the ordinary least squares (OLS) estimates of county-specific time trends (β_{1i}) in the log incidence proportion, ignoring NRs. The general pattern is one of higher

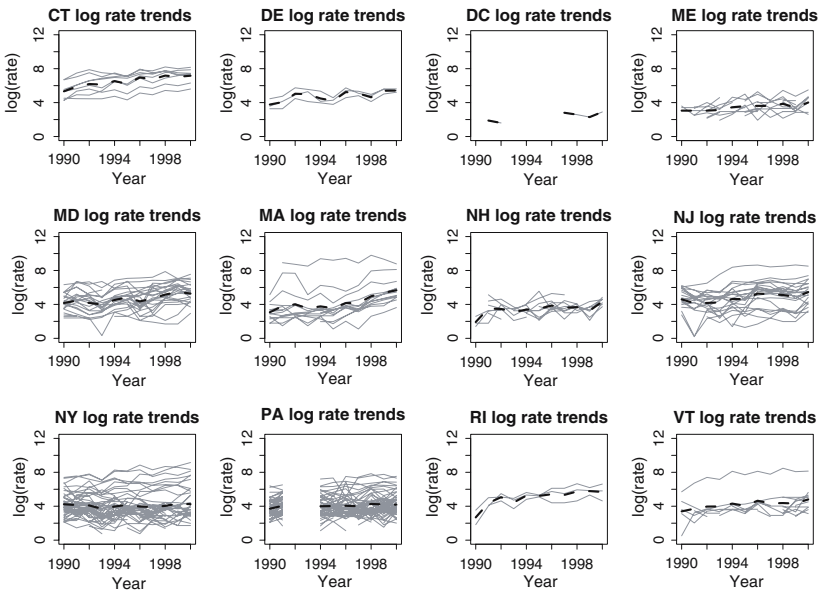

Fig. 3 County-specific log incidence rates by state (1990–2000). *Thin lines* connect county-specific annual log incidence rates with gaps representing "no reports". *Thick segments* connect annual statewide log incidence rates

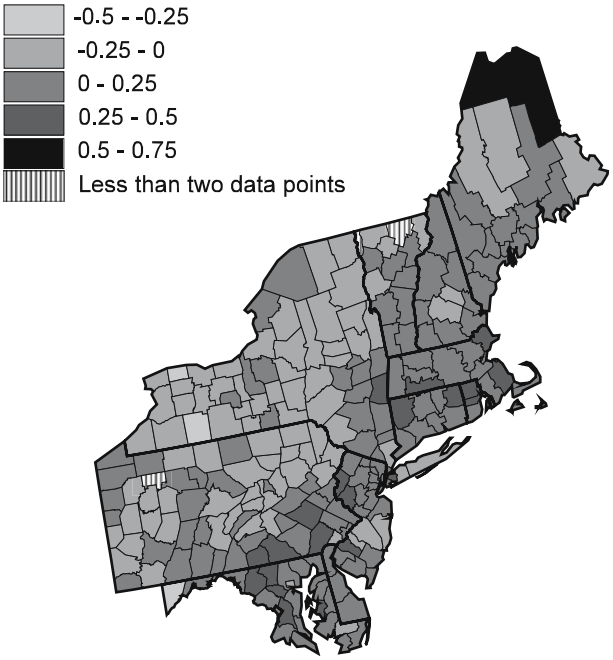

Fig. 4 Ordinary least squares estimates of county-specific time trends in log incidence proportions, 1990–2000. Three counties (indicated by *stripes*) do not have enough reports to enable estimation of a time trend

Table 2 Ordinary least squares estimates of the global intercept and slope

	Estimate	SE
β_0	3.931	0.063
β_1	0.071	0.010

slopes (faster increases) in counties with the highest incidence rates, and small but negative slopes roughly in a band across upstate New York. We also note that three counties (two in northern New York and one in western Pennsylvania) have zero or one reported values, making local regression estimation impossible without additional assumptions regarding the NRs.

The results shown in Fig. 4 build estimates based on local data alone. In moving toward compromises between local, global, and neighboring data, we next consider global OLS estimates based on all counties in the study area. Table 2 provides summary statistics indicating a small, positive increase in the log incidence proportion. This small positive slope is offset by considerable variation among counties, as shown in Fig. 3, resulting in an R^2 of only 0.02, a situation we hope to improve through the use of county-specific parameters.

Moving to Bayesian formulations, Table 3 presents posterior inference for the same model, with diffuse priors on the (still global) intercept and slope, and an inverse gamma (0.5,0.005) hyper-prior on the overall variance ${\sigma_Y}^2$. We see very close agreement with the OLS (equivalent to maximum likelihood) estimates in Table 2. Again

Table 3 Posterior inference for global intercept and slope

	2.5 Percentile	Median	97.5 Percentile
$[\beta_0 \mid \boldsymbol{Y}]$	3.807	3.931	4.058
$[\beta_1 \mid \boldsymbol{Y}]$	0.050	0.072	0.092
$[1/\sigma^2 \mid \boldsymbol{Y}]$	0.414	0.440	0.467

the slight positive global increase in incidence is offset by the large variation across counties.

We next add random effects as described in the preceding section. Fig. 5 provides the 95% interval estimates (confidence intervals for OLS estimates and posterior credible intervals for the Bayes' estimates) for each model considered, namely: OLS, the global "general Bayes" model summarized in Table 3, the random effects model with exchangeable (spatially unstructured) priors on both intercepts and slopes ("Exchangeable"), the random effects model with exchangeable intercepts and spatially structured slopes ("Spatial slope"), and the random effects model with spatially structured intercepts and slopes ("Both spatial"). For the random effects models, the intervals in Fig. 5 represent the posterior variation associated with the global values β_0 and β_1, and do not incorporate the additional variation associated with the random effects. While the mean values of the estimated global intercept and global slope drop with the addition of random effects, the overall slope estimates remains "significantly" positive (the 95% interval estimate excludes zero). We also note that the model with the exchangeable intercept random effects and spatially structured slope random effect (the "spatial slope" model) increases the precision of the overall

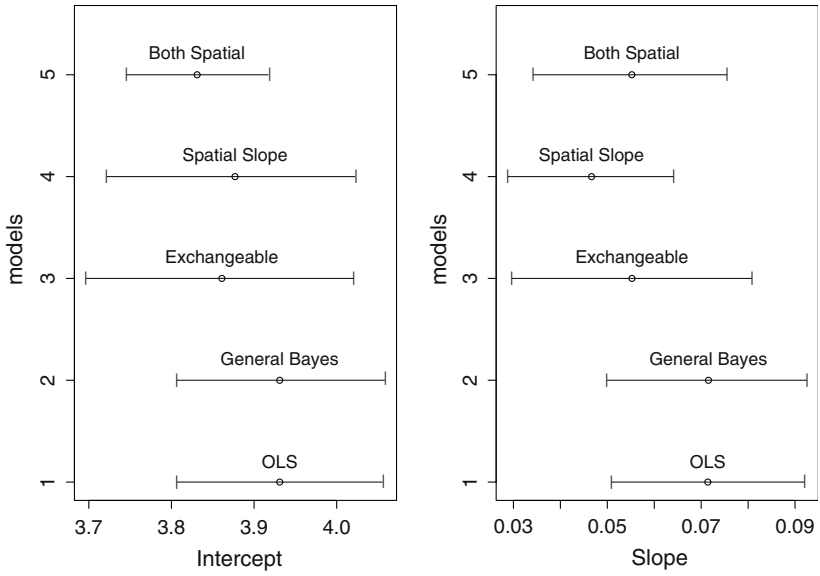

Fig. 5 Ninety-five percent estimation intervals (confidence intervals for ordinary least squares and credible sets for the others) for each of five models fit to the 1990–2000 county-specific log incidence proportions (see text for details)

Fig. 6 Posterior median estimates of the slope for models using spatially unstructured intercept and slope random effects, spatially unstructured intercept and spatially structured slope random effects, and spatially structured intercept and slope random effects, using same categories as the local OLS slopes in Fig. 4)

estimate of the time trend as evidenced by the narrower credible interval in Fig. 5, suggesting some benefit to borrowing strength from neighboring counties. The model with spatial priors on random effects (the "both spatial" model) associated with both the intercept and the slope suggests even more precision in the estimation of the intercept parameter, but these results are coupled with some unusual behavior, as the following detailed discussion shows.

Figure 6 provides maps of the posterior median local slope estimates, $(\beta_1 + u_{1i})$, $i = 1, \ldots, 245$, for each of the three random effects models, using the same shading categories we used for the local OLS estimates in Fig. 4. We note that no estimates

occur in the lowest or highest categories, reflecting the influence of data from other counties and an overall "shrinkage" of slopes and intercepts toward global or local means. This is perhaps most notable for Aroostook County, Maine (the northernmost county on the map), where the local OLS estimate (based on two data points) is highly positive in Fig. 1, but the compromise estimate between either all (exchangeable) or neighboring (spatial) data results in the small negative slope estimates shown in Fig. 6. All models mirror the initial pattern exhibited in Fig. 4, i.e., the greatest increases in the (log) incidence proportions occur in counties near the focus of the epidemic, not near the edges, as we might expect if we were capturing a strong "wavefront" signal. The model with exchangeable intercept and spatial slope random effects exhibits the greatest amount of spatial smoothing among the posterior slope estimates and suggests a general spatial pattern with arms of increasing local incidence surrounding a broad area of slightly decreasing incidence in upstate New York for the time period considered.

Curiously, posterior estimates of the local slope ($\beta_1 + u_{1i}$) shown in Fig. 6 for the model with spatial intercept and spatial slope random effects do not differ markedly from those in the exchangeable (non-spatial) model. To see why this may be, Fig. 7 shown the local log incidence rates and the fitted linear models under OLS and each of the random effects models for NY. Here we see the impact of adding random effects reflected in the "pinching" of the intercept estimates, and the adjustment to local slopes. In the case of the model with spatially structured intercept and slope random effects, the intercept "pinch" is severe, increasing precision on the intercept estimate, but resulting in slopes quite removed from the local data values. In contrast, the model with exchangeable intercept and spatial slope random effects incorporates

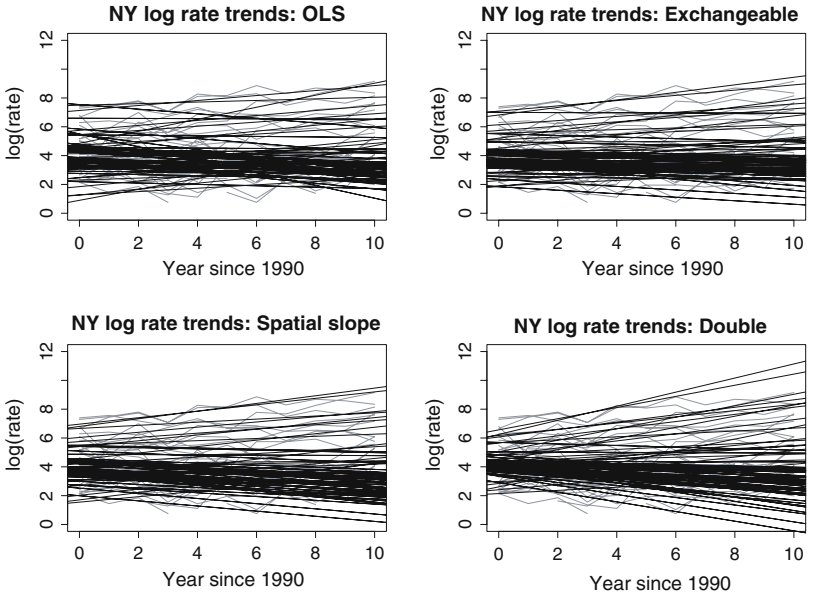

Fig. 7 County-specific fitted values for New York. *Gray lines* connect reported log incidence proportions within a county, *black lines* illustrate fitted relationships based on posterior median estimates of the associated slope and intercept

the spatial smoothing of slope estimates observed in Fig. 6, yet maintains a fair amount of fidelity to the original data. Based on this assessment, the model with exchangeable intercepts and spatial slopes seems to provide the best qualitative fit to the data.

6 Discussion

The foregoing concepts and analyses contribute several topics to the growing literature of risk maps for Lyme disease. First, the elements of a Lyme disease risk map, outlined in Fig. 2, synthesize several concepts raised, in part, by previous risk maps and provide a conceptual mechanism for comparing existing maps and proposing new analytic models. Second, regarding patterns of local reported incidence rates over the period 1990–2000, we see fairly linear changes in the local (log) incidence rates for most counties, with random variation of county-specific slopes and intercepts around central statewide rates. Local patterns include higher reported rates in the central northeastern Atlantic counties, near the area of the original identification of the disease, with continued increases over time along two "arms," one to the north and one to the south, skirting the Adirondack Mountains in eastern New York.

The collection of small-magnitude negative slopes in western New York raises several issues. Had a "wavefront" of reported incidence swept through the area during the study period, we might well expect a series of zero reported counts (or NRs), followed by a sudden increase generating steeper slopes in counties just behind the wavefront. No such pattern appears; in fact, the estimated local slopes in western New York suggest that any initial spread of disease through these counties occurred prior to the study period, as fairly stable and slightly decreasing (log) reported incidence rates appear in this area. However, while a general pattern of a wavefront does not appear over the time period and locations under investigation, the smoothing properties of the approach could conceivably hide small local anomalies consistent with more focused spread and outbreaks meriting additional investigation.

The relationship between incidence and NRs shown in Fig. 1 suggests a connection between the spread of the disease and the spread of reporting practices. We do not explicitly include this relationship in the models presented here, but the conceptual framework of Fig. 2, coupled with the hierarchical structure of the models considered, suggests modeling mechanisms for including spatially heterogeneous and perhaps spatially correlated reporting filters, provided we obtain data on reporting accuracy. In addition, the framework of Fig. 2 also suggests possibilities for simulation studies addressing the impact of variations in reporting practices and accuracy in modeled risk maps. Such surveillance of disease surveillance is an area of growing interest (Buenconsejo 2004) and could be linked into the model in future work.

The foregoing results represent our first steps toward more thorough analyses of the data, moving from summaries of observed patterns to prediction of patterns as a function of local covariates. Future work will involve environmental and ecological covariates, e.g., percent forest cover, soil types, forest fragmentation, deer density, and small mammal diversity, building on elements of the underlying disease ecology represented in the interaction between the maps on the left side of Fig. 2.

More formal modeling will require more formal model choice and fit assessments. Because of the exploratory nature of the present results, the preceding simple linear models provide insight into fairly consistent descriptions of incidence patterns across models, with the hierarchical structure allowing ready prediction of missing

reports and inclusion of spatial correlation through the prior distributions for the random effects. The somewhat unusual behavior of the model with spatially correlated intercepts and slopes merits closer scrutiny to determine whether its origin is related to some quirk of the specific data considered or lies deeper within the structure of the model.

Acknowledgements This research was supported in part by the National Center for Ecological Analysis and Synthesis (a center funded by NSF grant DEB-94-21535, University of California, Santa Barbara, the California Resources Agency, and the California Environmental Protection Agency) and by National Institute of Environmental Health Sciences grant R01-ES07750 (LAW). The views and opinions expressed reflect those of the authors and do not necessarily represent those of NCEAS, NSF, NIH, NIEHS, or CDC. The authors thank the many state and local health departments and the CDC for their collection and maintenance of the surveillance data and Les Real for many helpful comments and suggestions.

References

Anderson JF, Magnarelli LA (1993) Natural history of Borrelia burgdorferi in vectors and vertebrate hosts, in Ecology and Environmental Management of Lyme Disease, Ginsberg HS (ed) Rutgers University Press, New Brunswick, NJ, pp 11–24

Barbour AG, Fish D (1993) The biological and social phenomenon of Lyme disease. Science 260:1610–1616

Besag JE (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). J R Stat Soc Ser B 36:741–752

Besag J, Kooperberg C (1995) On conditional and intrinsic autoregressions. Biometrika, 82:733–746

Besag J, York J, Mollié A (1991) Bayesian image restoration with two applications to spatial statistics (with discussion). Ann Ins Stat Math 43:1–59

Besag J, Green P, Higdon D, Mengersen K (1995) Bayesian computation and stochastic systems (with discussion). Stat Sci 10:3–66

Brookmeyer R, Stroup D (2004) Monitoring the health of populations: statistical principles and methods for public health surveillance. Oxford University Press, New York

Buenconsejo J (2004) A Bayesian hierarchical model for estimation of disease incidence using two surveillance datasets. Unpublished Ph.D. Dissertation, Department of Biostatistics and Epidemiology, Yale University

Dennis DT, Nekomoto TS, Victor JC, Paul WS, Piesman J (1998) Reported distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the United States. J Med Entomol 35:629–638

Dister SW, Fish D, Bros SM, Frank DH, Wood BL (1997) Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. Am J Trop Med Hyg 57:687–692

Donahue JG, Piesman J, Speilman J (1997) Reservoir competence of white-footed mice for Lyme disease spirochetes. Am J Trop Med Hyg 36:92–96

Duffy DC, Campbell SR, Clark D, Dimotta C, Gurney S (1994) *Ixodes scapularis* (Acari: Ixodidae) deer tick mesoscale populations in natural areas: effects of deer, area, and location. J Med Entomol 31:152–158

Estrada-Pena A (1998) Geostatistics and remote sensing as predictive tools of tick distribution: a cokriging system to estimate *Ixodes scapularis* (Acari: Ixodidae) habitat suitability in the United States and Canada from advanced very high resolution radiometer satellite imagery. J Med Entomol 35:989–995

Fish D (1993) Population ecology of *Ixodes dammini*. In: Ginsberg H.S. (ed) Ecology and environmental management of Lyme disease. Rutgers University Press, New Brunswick, NJ, pp 25–42

Frank C, Fix AD, Pena CA, Strickland GT (2002) Mapping Lyme disease incidence for diagnostic and preventive decisions, Maryland. Emerg Infect Dis 8:427–429

Giardina AR, Schmidt KA, Schauber EM, Ostfeld RS (2000) Modeling the role of songbirds and rodents in the ecology of Lyme disease. Can J Zool 78:2184–2197

Ginsberg HS, Ewing CP (1989) Habitat distribution of *Ixodes dammini* (Acari: Ixodidae) and Lyme disease spirochetes on Fire Island, NY. J Med Entomol 26:183–189

Glass GE, Amerasinghe FP, Morgan JM, Scott TW (1994) Predicting *Ixodes scapularis* abundance on white-tailed deer using geographic information systems. Am J Trop Med Hyg 51:538–544

Glass GE, Schwartz BS, Morgan JM, Johnson DT, Noy PM, Israel E (1995) Environmental risk-factors for Lyme disease identified with geographic information systems. Am J Public Health 85:944–948
Goodwin BJ, Ostfeld RS, Schauber EM (2001) Spatiotemporal variation in a Lyme disease host and vector: blacklegged ticks on white-footed mice. Vector Borne Zoonotic Dis 1:129–138
Kelsall J, Wakefield J (1999) Discussion of "Bayesian models for spatially correlated disease and exposure data." In: Best NG Arnold RA, Thomas A, Conlon E, Waller LA in Bayesian Statistics 6, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Oxford University Press, Oxford, p 151
Keirans JE, Hutcheson HJ, Durden LA, Klompen JSH (1996) *Ixodes* (Ixodes) *scapularis* (Acari: Ixodidae): redescription of all active stages, distribution, hosts, geographical variation, and medical and veterinary importance. J Med Entomol 33:297–318
Kitron U (2000) Risk maps: transmission and burden of vector-borne diseases. Parasitol Today 16: 324–325
Kitron U, Bouseman JK, Jones CJ (1991) Use of the ARC/INFO GIS to study the distribution of Lyme disease in an Illinois county. Prev Vet Med 11:243–248
Kitron U, Jones CJ, Bouseman JK, Nelson JA, Baumgartner DL (1992) Spatial analysis of the distribution of *Ixodes dammini* (Acari: Ixodidae) on white tailed deer in Ogle County, IL. J Med Entomol 29:259–266
Kitron U, Kazmierczak JJ (1997) Spatial analysis of the distribution of Lyme disease in Wisconsin. Am J Epidemiol 145:558–566
Knorr-Held L, Besag J (1998) Modelling the risk of a disease in time and space. Stat Med 17:2045–2060
Lane RS, Piesman J, Burgdorfer W (1991) Lyme borreliosis: relation of its causative agent to its vectors and hosts in North America and Europe. Ann Rev Entomol 36:587–609
Levine JF, Wilson ML, Speilman A (1985) Mice as reservoirs of the Lyme disease spirochete. Am J Trop Med Hyg 34:355–360
Mather TN (1993) The dynamics of spirochete transmission between ticks and vertebrates. In: Ginsberg HS (ed) Ecology and environmental management of Lyme disease. Rutgers University Press, New Brunswick, NJ, pp 43–62
Mather TN, Ginsberg HS (1994) Vector-host-pathogen relationships: transmission dynamics of tick-borne infections. In: Sonenshine DE, Mather TN (eds) Ecological dynamics of tick-borne zoonoses. Oxford University Press, NY, pp 68–90
Mather TN, Wilson ML, Moore SI, Ribeiro JMC, Spielman A (1989) Comparing the relative potential of rodents as reservoirs of the Lyme disease spirochete (*Borrelia burgdorferi*). Am J Epidemiol 130:143–150
Naleway AL, Belongia EA, Kazmierczak JJ, Greenlee RT, Davis JP (2002) Lyme disease incidence in Wisconsin: a comparison of state-reported rates and rates from a population-based cohort. Am J Epidemiol 155:1120–1127
Orloski KA, Campbell GL, Genese CA, Beckley JW, Schriefer ME, Spitalny KC, Dennis DT (1998) Emergence of Lyme disease in Hunterdon County, NJ, 1993: a case-control study of risk factors and evaluation of reporting patterns. Am J Epidemiol 147:391–397
Orloski KA, Hayes EB, Campbell GL, Dennis DT (2000) Surveillance for Lyme disease—United States, 1992–1998. Morb Mortal Wkly Rep (Surveillance Summaries) 49:1–11
Patrican LA (1997) Absence of Lyme disease spirochetes in larval progeny of naturally infected *Ixodes scapularis* (Acari: Ixodidae) fed on dogs. J Med Entomol 34:52–55
Piesman J, Spielman A (1979) Host-associations and seasonal abundance of immature *Ixodes dammini* in southeastern Massachusetts. Ann Entomol Soc Am 72:829–832
Piesman J, Donahue JG, Mather TN, Spielman A (1986) Transovarially acquired Lyme disease spirochetes (*Borrelia burgdorferi*) in field-collected larval *Ixodes dammini* (Acari: Ixodidae). J Med Entomol 23:219
Schmid GP, Horsley R, Steere AC, Hanrahan JP, Davis JP, Bowen GS, Osterholm MT, Weisfeld JS, Hightower AW, Broome CV (1985) Surveillance of Lyme disease in the United States, 1982. J Infect Dis 151:144–1149
Schootman M, Sun D (2004) Small-area incidence trends in breast cancer. Epidemiology 15:300–307
Scrimenti RJ (1970) Erythema chronicum migrans. Arch Dermatol 102:104–105
Steere AC, Malawista SE, Snydman DR, Shope RE, Andiman WA, Ross MR, Steele FM (1977) Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three connecticut communities. Arthritis Rheum 20:7–17
Teutsch SM, Churchill RE (1994) Principles and practice of public health surveillance. Oxford University Press, New York
US CDC (2001). Lyme Disease—United States, 1999. Morb Mortal Wkly Rep 50:181–185

van Buskirk J, Ostfeld RS (1995) Controlling Lyme disease by modifying the density and species composition of tick hosts. Ecol Appl 5:1133–1140
Wakefield JC, Best NG, Waller L (2000) Bayesian approaches to disease mapping. In: Elliott P, Wakefield JC, Best NG, Briggs DG (eds) Spatial epidemiology: methods and applications. Oxford University Press, Oxford, pp 104–127
Waller LA, Gotway CA (2004) Applied spatial statistics for public health data. Wiley, New York
White DJ, Chang H-G, Benach JL, Bosler EM, Meldrum SC, Means RG, Debbie JG, Birkhead GS, Morse DL (1991) The geographic spread and temporal increase of the Lyme disease epidemic. J Am Med Assoc 266:1230–1236
Wilson ML (1998) Distribution and abundance of *Ixodes scapularis* (Acari: Ixodidae) in North America: ecological processes and spatial analysis. J Med Entomol 35:446–457
Wilson ML, Ducey AM, Litwin TS, Gavin TA, Spielman A (1990) Microgeographic distribution of immature *Ixodes dammini* ticks correlated with that of deer. Med Vet Entomol 4:151–159

Biographical sketches

Dr. Lance Waller received his Ph.D. in Operations Research from Cornell University in 1992. His research interests involve statistical analysis of spatial patterns in public health data. Past investigations include development of statistical tests of spatial clustering in disease incidence data, and implementation of spatial and space-time Markov random field models for maps of disease rates. He is currently investigating statistical methods to analyze environmental exposure, demographic, and disease incidence data linked through geographic information systems (GISs).

Dr. Brett Goodwin received his Ph.D. in landscape ecology/ecological modeling at Ottawa-Carleton Institute of Biology, Carleton University, Ottawa, ON. He received his B.S. (summa cum laude) in biology at McMaster University, in Hamilton, ON. His interests include landscape connectivity and modeling disease dynamics on dynamic landscapes.

Dr. Mark Wilson is currently Professor of Epidemiology and of Biology at the University of Michigan, MI, where his research and teaching cover the broad area of ecology and epidemiology of infectious diseases. After earning his doctoral degree from Harvard University in 1985, he worked at the Pasteur Institute in Dakar Senegal (1986-1990), was on the faculty at the Yale University School of Medicine (1991-1996), and then joined the University of Michigan, MI. He was a member of the National Academy of Sciences panel on "Climate, Ecosystems, Infectious Diseases and Human Health."

Dr. Richard Ostfeld is Associate Scientist and Animal Ecologist at the Institute of Ecosystem Studies. He is also adjunct Associate Professor of Ecology and Evolutionary Biology at the University of Connecticut and associate member of the graduate faculty in ecology at Rutgers University. He earned his Ph.D. in zoology from the University of California at Berkeley in 1985. His primary research interests are in understanding how the population dynamics of mammals, especially rodents, affect forest dynamics and disease risk.

Stacie Marshall is an epidemiologist at the US Centers for Disease Control and Prevention (CDC), in Atlanta, GA, where she regularly serves as a presenter and leads seminars on a variety of infectious disease topics, including Lyme disease.

Edward Hayes, M.D., is a doctor at the CDC in Fort Collins, CO, where he provides a pediatric perspective on infectious diseases.