Title: LYME DISEASE IN NEW YORK STATE: SPATIAL PATTERN AT A REGIONAL

SCALE

Date of revision: January 30, 2001

Corresponding Author: Stephan Glavanakov, Department of Biological Sciences, University at

Albany, Albany, NY 12222; e-mail: sg8873@albany.edu

phone: 518-442-4343; FAX: 518-442-4767

6 Figures, 2 Tables, 32 References, and 22 Pages

Manuscript prepared in Microsoft Word 2000 (9.0.2831 SR-1)

LYME DISEASE IN NEW YORK STATE:

SPATIAL PATTERN AT A REGIONAL SCALE

STEPHAN GLAVANAKOV, DENNIS J. WHITE, THOMAS CARACO

ANDREI LAPENIS, GEORGE ROBINSON, BOLESLAW K. SZYMANSKI

AND WILLIAM A. MANIATTY

Departments of Biological Sciences, Geography and Planning, and Computer Science, University at Albany, Albany, New York; New York State Department of Health, Slingerlands, New York; Department of Computer Science, Rensselaer Polytechnic Institute, Rensselaer, New York

Abstract. Lyme disease occurs commonly in New York State, but its geographic distribution is heterogeneous. For each of nine consecutive years, incidence rates from 57 New York State counties were subjected to spatial autocorrelation analysis. Although the epidemic advanced during the study period, the analyses reveal a consistent pattern of spatial dependence. The correlation distance, the distance over which incidence rates covary positively, remained near 120 km over the nine years. A local spatial analysis around Westchester County, a major disease focus, indicated that the global correlation distance matched the extent of the most intense local clustering; statistically weaker clustering extended to 200 km from Westchester. Analyzing the spatial character of the epidemic may reveal the epizootic processes underlying patterns in human infection, and may help identify a spatial scale for regional control of disease.

2

Lyme disease remains the most frequently reported vector-borne disease in the Northern Hemisphere, and the world's most common tick-borne infection.[1--3] In the US more than 103,000 cases have been recorded since 1982; the estimated cost for early, acute disease has risen to $60 million annually.[4] The geographic range of Lyme disease has increased steadily, particularly across the northeastern US.[5] Because Lyme disease is a zoonosis, regional epidemiological patterns depend on the spatial and temporal distribution of infection in both the tick-vector and its reservoir hosts.[6--8] So, analyzing the spatial character of the epidemic should promote our understanding of the epizootic processes underlying patterns of human infection. Furthermore, a geostatistical analysis of disease incidence may indicate important spatial scales for regional control.[9]

We begin by summarizing the epizootiology of Lyme disease. Next, we employ spatial autocorrelation methods[10] to estimate a regional correlation distance from disease incidence reported by New York State (NYS) counties. We treat each year from 1988 to 1996 separately, and find surprising consistency. We proceed to map the cumulative incidence rates and the case numbers for the epidemic in NYS. We suggest that a diffusion-like wave of disease has proceeded from a major focus of infection in NYS, and describe its local correlation distance. Finally, we interpret the combined patterns in terms of epizootic dispersal processes.

ECOLOGICAL BACKGROUND

Several authors present details of the cycle of infection responsible for the epidemic of Lyme disease,[2,11--13] so we provide only a brief summary. The black-legged tick (*Ixodes scapularis*) is the most common vector of Lyme disease in the northeastern and north central United States. Larval ticks hatch during summer and quest for their first blood meal. They may acquire the

pathogen (the spirochetal bacterium *Borrelia burgdorferi*) by feeding on an infectious host, most often a white-footed mouse (*Peromyscus leucopus*).[11,12] After advancing to the nymph stage and overwintering, infected ticks pass the pathogen to the next generation of mice; transplacental transmission of the spirochete is extremely rare in mice.[14] But an infectious tick, most often a nymph, can inadvertently transmit the pathogen to a human, who may develop Lyme disease.[15]

Most nymphs that acquire a blood meal from a small mammal mature as adults the same year. Adults ordinarily take the third and last blood meal from white-tailed deer (*Odocoileus virginianus*). When replete, adult female ticks mate on the deer, overwinter in leaf litter and then oviposit. When larvae hatch, the two-year life cycle of the tick is complete.

Since Lyme disease is a zoonosis, spatial patterns in the tick-mouse cycle of infection should govern pattern in human disease. So, we conducted a series of spatial analyses of Lyme disease in NYS, and interpret the patterns in terms of host-vector ecology.[16] Pattern at local ecological scales refers to differences between adjacent locations, such as patches of forest or field. Local patterns, and the processes driving them, combine to produce regional pattern at extended spatial scales.[13,17] Our data on the incidence of Lyme disease let us characterize heterogeneity among NYS counties, and we refer to this as regional pattern. Our data analyses use similar terms with slightly different meanings, following standard convention. A local spatial analysis describes the size of a disease cluster about a specific location, and a cluster may extend beyond the local ecological scale. A global spatial analysis addresses spatial autocorrelation of disease across the entire set of locations sampled simultaneously.[18--21]

Pathogen dispersal. At a local ecological scale dispersal by deer influences the spatial pattern of hatching tick larvae, since an adult female tick may spend several weeks on a deer.[6,12] Adult ticks die after oviposition, and the spirochete is rarely transmitted transovarially in the

4

Northeast,[22] so essentially all larvae hatch susceptible. The pathogen cannot survive outside its hosts,[1] and the deer lack reservoir competence. Therefore, deer affect the spatial dynamics of tick populations, but do not disperse the pathogen directly or indirectly. The spatial advance of *B. burgdorferi* infection must be driven primarily by dispersal of white-footed mice and other mammalian hosts of juvenile ticks.[8,23] Local dispersal of infectious mice can expose new populations of tick larvae to the pathogen. In addition, a mouse may transport infected tick larvae and disperse the parasites while they feed.

At a regional scale longer-distance dispersal of juvenile ticks by birds, especially if they exhibit reservoir competence, may play a role in the spread of Lyme disease.[24-26] Regional pattern in human infection likely reflects a combination of consistent, local advance of the tick-mouse cycle of infection and occasional, longer-distance advance via dispersal of ticks and/or *B. burgdorferi* by birds. Local advance of infection, perhaps a diffusive spread, should produce clustering of incidence rates, case numbers, or both. Spatial clustering of infection will induce positive correlations between incidence rates at nearby locations. Counties with high rates of disease will occur near each other, situated among counties where infection is rare. Longer-distance dispersal via birds could establish new disease foci, initiating a new spatial cluster of infection.

Given this general depiction for the advance of Lyme disease, regional heterogeneity will result from the pattern of spatial dependence. Clustering implies that the level of positive spatial autocorrelation between locations should, at least initially, decline as the distance between locations increases. A key measure of the presumed spatial dependence is correlation distance, an indicator of average cluster size. We estimated global correlation distance as the length at which spatial autocorrelation reaches zero. Locations within a cluster exhibit positive spatial autocorrelation (spatial dependence), while spatial independence implies zero autocorrelation.[18] For any particular

disease focus, we can also employ a local spatial analysis to describe the size of the surrounding cluster, and help explain the global autocorrelation.[20,21] Larger-scale control procedures and public-health education efforts will more likely succeed if they encompass most locations within the correlation distance of disease foci. That is, a rigorous quantification of disease clustering might guide coordination of regional responses to the Lyme disease epidemic.

METHODS

The NYS Department of Health maintains a Human Registry listing the date and location of each case of Lyme disease in the state, confirmed by physician diagnosis or two-test serodiagnosis.[5] Reporting by the state's 57 counties outside of New York City began in 1986. More than 29,000 cases were confirmed within the first decade. Our analysis addresses incidence rates and case numbers reported by these 57 counties for the 11 years 1986--1996. Data can be obtained from the NYS Department of Health, Bureau of Communicable Disease, Albany, New York 12223 USA. The data are expressed as incidence rates: cases/10^5 individuals/yr. In general we might expect the variance of incidence-rate estimates to vary inversely with population size.[27] However, population sizes of NYS counties are quite large; the mean population size (in 1995) of counties outside of New York City is 189,800.[28] Therefore, the standard error of an incidence rate estimate for any NYS county will be small. So, following Ord and Getis,[29] our analyses (specifically, the null distribution in statistical tests) treat counties as equivalent sampling units. That is, we do not need precision-weighted analyses.

Lyme disease: correlation distance. To estimate a regionally scaled correlation distance, we calculated the autocorrelation of incidence rates as a function of distance between county centroids. We chose Moran's statistic *I* as a spatial autocorrelation coefficient.[10,19] To assure that the data reflected uniform reporting procedures among counties, we restricted this

6

analysis to the period 1988 to 1996. Incidence rates in each year t ($t = 1$-- 9) were positively skewed, so we transformed the rates logarithmically for these analyses.

We included all 57 NYS counties outside of New York City. We grouped distances by 40-km intervals between county centroids. This distance assured that at least 18 county pairs fell into each of the 13 distance categories; other choices left less than 8 pairs in either the first or last category. The mean number of county pairs per distance category was 122.8.

We estimated spatial autocorrelation as a function of distance. $I(d)$ is Moran's statistic, where d is the distance between the centroids of paired counties. For each year we generated a spatial correlogram, a plot of $I(d)$ against distance. As d increases, the value of the spatial autocorrelation coefficient I should decrease, since increasing distance (at least initially) diminishes correlation. We take the regional correlation distance as the maximal distance d before $I(d)$ first fails to differ significantly from zero (see Appendix 1 for significance test). Moran's I is a global measure;[21] its evaluation simultaneously considers all locations a distance d apart. So, the correlogram combines effects of both disease foci and areas of low incidence on the pattern of spatial dependence.

Let $y_i(t)$ represent the ln(incidence rate + 0.5) for Lyme disease in county i during year t; addition of the constant lets us include any zeros in the data. Let $m(t)$ represent the sample mean ln(incidence rate + 0.5) for the 57 counties in year t. Define $\xi_i(t)$ as the deviation about the mean for county i: $\xi_i(t) = y_i(t) - m(t)$. As a function of the distance d between county centroids, Moran's spatial autocorrelation coefficient is calculated as:

$$I(d) = \frac{N \sum_{ij} w_{ij}(d) \xi_i \xi_j}{\left(\sum_{ij} w_{ij}(d) \right) \left(\sum_i \xi_i^2 \right)} \qquad i \neq j$$

N is the total number of counties. The weight $w_{ij}(d)$ given to the county pair $(i, j; i \neq j)$ is 1 if their centroids are exactly d distance units apart, and is 0 otherwise.

Appendix 1 explains the statistical procedures used to analyze global spatial autocorrelation and to estimate correlation distance. Moran's I has a normal distribution asymptotically as N increases.[30] Hence individual estimates $I(d)$ can be tested for significance via evaluation of standard normal variates.[10,19]

Westchester County: local clustering. Moran's I and the resulting correlograms address global spatial autocorrelation. That is, they summarize disease incidence over the entire set of counties. A local spatial statistic serves to describe spatial pattern about a single location of interest.[16,19,20] Westchester County (1995 pop. 891,044) was the first NYS county to report more than 3000, 4000, and 5000 cases of Lyme disease (in 1991, 1992, and 1993, respectively). So, we conducted a local analysis with Westchester as a preselected focal location. Given the geography of Lyme disease in NYS, we would expect the size of the disease cluster around Westchester to exceed the global correlation distance estimated from regionally averaged spatial correlograms.

A local spatial analysis helps clarify global autocorrelation by describing cluster size. For our local analysis we chose the Getis-Ord statistic G_i^*, where the subscript implies that the procedure refers to pattern about a specific location.[29] We calculated G_i^* for Westchester county and the set of counties with centroids $\leq d$ distance units ($d = 1$--10) from Westchester. The distance d' where G_i^* reaches a maximum indicates the scale of the most intense clustering[20] around and

including Westchester County. Significant G_i^* values at distances $> d'$ suggest clustering, but at lower intensity.

Let d again represent distance between county centroids. Calculating $G_i^*(d)$ requires that the ln(incidence $+ 0.5$) at the focal location be summed together with the j values at all locations no farther than d distance units from the focal location. So, let $w_{ii} = 1$, let $w_{ij} = 1$ if the distance between county j and the focal location $\leq d$, and let $w_{ij} = 0$ otherwise. For convenience, let:

$$W_i^* = \sum_j w_{ij}(d) \qquad j = 1,\ 2,...,\ i,...,\ N$$

Then $G_i^*(d)$, in standardized form, is calculated as:[20]

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)\,y_j - m(t)W_i^*}{s\left\{\left(NW_i^* - \left[W_i^*\right]^2\right)\middle/(N-1)\right\}^{1/2}} \qquad all\ j$$

where y_j is ln(incidence $+ 0.5$) in county j during year t, $m(t)$ is the sample mean for the N counties in year t, and s is the sample standard deviation of the N values of y in year t.

For each of the nine years we calculated $G_i^*(d)$ for 10 distance categories. To increase the number of observations in the first category to six, we grouped data for counties' centroids not more than 80 km from Westchester. Each category thereafter increased the distance by 40 km.

When $G_i^*(d)$ is calculated as above, it approximates a standard normal variate. So, its expected value in the absence of clustering is 0. Positive values suggest disease clustering at levels exceeding the state-wide mean for that year. Significance of the $G_i^*(d)$ about a preselected focus (Westchester County), especially when global spatial autocorrelation is significant, is not always clear-cut.[29] So, we conservatively use the local statistic as an exploratory tool.

RESULTS

Figure 1 shows state-wide incidence rates for each year, and shows the cumulative number of cases in NYS over the 11 years included in our data. As the epidemic has progressed, Lyme disease has been reported in all 57 counties analyzed, but spatial clustering of higher incidence persists.

Correlation distance. Table 1 lists the 117 Moran's I values, with statistical significance indicated. Figure 2 shows $I(d)$ plotted as a function of distance for years $t = 1988$--1996. Together, the 9 correlograms reveal that, despite the growth in reported cases of Lyme disease over 9 years, the regional pattern retains a similar spatial dependence. That is, disease clusters and areas of low disease incidence have produced a consistent pattern of spatial heterogeneity from 1988 through 1996.

Spatial autocorrelation is significantly positive at $d = 1$ in 8 of the 9 years. In all 9 years, spatial autocorrelation is significantly positive at $d = 2$ and 3. Since the incidence data are grouped by 40-km intervals, the regional correlation distance has an average approaching 120 km.

For $d = 7$--12, 24 of 45 estimates are significantly negative. So, we observe positive spatial autocorrelation at smaller distances, and negative spatial autocorrelation at greater distances. Overall, the pattern of spatial dependence indicates a cline[10] in the incidence of Lyme disease. The qualitative pattern is not surprising; the correlation distance provides a detailed quantitative characterization.

We mapped cumulative (*i.e.*, summed from the beginning of reporting) incidence and cases for the NYS region using a commercially available software package (Surfer for Windows, Version 7.00, Golden Software Corp., Golden, CO). Figure 3 shows a contour map generated by kriging the cumulative incidence rates (scaled logarithmically) for each county; each contour line connects

locations with the same estimated incidence. Figure 3 also shows a similar contour map generated from the cumulative case numbers (scaled logarithmically) for each county in NYS. Both maps use the county-centroid data to estimate a smoothed surface for cumulative incidence or cases. Coincidence of both high incidence and high case number marks important disease clusters. The maps indicate three such clusters (four, if we consider Long Island separately from the rest of NYS). Within a cluster, local dispersal of mice likely plays a major role in advancing the infection. Longer-distance, avian dispersal may have established one or more of the separate foci. The largest cluster occurs, of course, in the southeast corner of the state, and we took Westchester County as the center of the most important disease focus in our data.

Westchester County. Figure 4 shows the number of cases in each of the 9 years, and the yearly running total number of cases, for Westchester County. Figure 5 shows the incidence rate at Westchester and the average incidence at different spatial lags ($d = 1$--12) away from Westchester, for years $t = 1, 3, 5, 7$, and 9. Prior to 1994 (*i.e., $t < 7$*), Westchester's incidence rate exceeds rates in both adjacent and more distant counties. However, in 1994 the mean incidence at lag $d = 1$ exceeds the rate at Westchester. During the last year shown (1996), the average incidence at both lags $d = 1$ and 2 (*i.e.*, at distances less than 80 km) surpasses the rate at Westchester. Local diffusion of the tick-mouse cycle of infection results in not only an overall increase in the incidence of Lyme disease, but in a diffusive-like dispersal of Lyme disease moving North and West (within NYS) away from Westchester County.

Table 2 lists the 90 $G_i^*(d)$ values. Figure 6 shows the plot of the $G_i^*(d)$, for $d = 1$--10, for each of the nine years. The various curves indicate surprising among-year consistency, given the advance of the epidemic during the years analyzed. Seven of the 9 plots reach a maximum at the third distance category, where the distance between county centroids is 120 to 160 km. This implies

that the most intense clustering around Westchester has a cluster size slightly larger than the overall correlation distance suggested by the correlograms. However, after a Bonferroni-type adjustment,[29] the local statistic retains significance through the fifth distance category in six of nine years. This implies some clustering through 200--240 km. The strongest clustering around and including Westchester matches the global pattern of spatial dependence. But the advance of Lyme disease has produced a spatial dependence around a focus that exceeds the strength of the overall regional pattern; see Figure 3.

DISCUSSION

Correlation distance for Lyme disease indicates the average distance over which incidence rates remain positively associated. Hence correlation distance identifies a characteristic spatial scale for the pattern of infection. Similarly, distance-dependent plots of a local statistic can quantify the extent and intensity of clustering about disease foci. Each method offers insight concerning spatial pattern and, plausibly, the processes generating the pattern.

The correlograms for Lyme disease indicate a consistent (among years) correlation distance close to 120 km. So, incidence rates for counties less than 120 km apart covary positively, so that regional heterogeneity in Lyme disease exhibits a strong spatial dependence. Data collected at a finer spatial scale, or grouped at a more coarse scale, might not yield such a consistent pattern.

The generally monotonic decline of Moran's *I* with distance indicates clinal variation. Within any particular year, the incidence of Lyme disease tends to decline to both the north and west of the major focus in the southeastern corner of NYS. However, the cumulative incidence and case maps (Figure 3, estimated from 9-yr sums) suggest that slightly greater complexity underlies the correlogram averaging. The contour maps together indicate at least three foci of

12

Lyme disease in NYS; the westernmost focus may be an extension of a disease cluster in northern Pennsylvania.[31] The most significant focus in NYS, around Westchester County, appears to be a point about which a diffusion-like wave of Lyme disease has advanced to nearby counties of NYS.

Our results concern human Lyme disease, but they ma help clarify the epizootological processes driving the epidemic. Local, diffusive advance of the tick-mouse cycle of infection, perhaps augmented by local avian dispersal, establishes and maintains disease clusters around infection foci. These processes contribute significantly to the average correlation distance, and the consistency of that distance, defining the regional pattern of spatial dependence in human disease incidence. Longer-distance dispersal via birds may occasionally move *B. burgdorferi* or infectious juvenile ticks far enough to establish new disease foci, after which local processes can generate a new cluster.

Our analyses treat Lyme disease incidence as if each case occurs exactly at the centroid of the county where it was reported. This is a necessary constraint imposed by data reported according to administrative districts. If the data were arrayed on a regular lattice, rather than at irregularly located centroids, more powerful geostatistical analyses could be applied to investigation of the spatio-temporal dynamics of the epidemic process.[9] Furthermore, a more comprehensive spatial analysis would include incidence data from states or counties adjacent to NYS. In any case, we have been able to estimate the regional cluster size associated with the spatial pattern of Lyme disease in NYS.

Cluster size offers insight concerning the ecological processes underlying an epidemic, but to understand fully the ecology of the epizootic processes, direct study is required.[8] Cluster

size does, however, suggest a scale for control measures and, perhaps, for public education efforts.

For a given distance d (d = 1--13) between county centroids, and for a given yr t (t = 1--9), ξ_i is the deviation of the ln(incidence rate + 0.5) for county i about the sample mean for that year. Adding a constant (0.5) permits inclusion of observed zeroes in the data. The mean is calculated over all N counties. Moran's spatial autocorrelation coefficient invokes weights $w_{ij}(d)$,where $w_{ij}(d)$ = 1 if counties i and j are d distance units apart; otherwise w_{ij} = 0. Given our weights, $\Sigma_{ij}\, w_{ij}(d)$ is twice the number of county pairs distance d apart.

We tested $I(d)$ against the null distribution based on all possible permutations of the observed ξ_i over localities. This null model is preferred for systematic analyses like ours; see discussion of an alternative in Sokal and Oden.[10] $I(d)$ has expected value $E_d[I] = -[N-1]^{-1}$ in the absence of spatial dependence.[19] To calculate the variance of $I(d)$ under the null model, we need $S_1 = 2\Sigma_{ij}w_{ij}(d)$ and $S_2 = 4\Sigma_i[w_i(d)]^2$, where $i \neq$ j, and $w_i(d)$ is the number of pairs involving county i for distance d. We also require an index of kurtosis (fourth moment about the mean) b_2 for the ξ_i values:[10]

$$b_2 = \frac{N \sum_i \xi_i^4}{\left(\sum_i \xi_i^2\right)^2}$$

Let W represent $(\Sigma_{ij}\, w_{ij})$, for convenience. Given S_1, S_2 and b_2, the variance of $I(d)$, $V_d[I]$, is:[10]

$$V_d[I] = \frac{N\left[S_1\left(N^2 - 3N + 3\right) - S_2 N + 3W^2\right] - b_2\left[S_1\left(N^2 - N\right) - 2S_2 N + 6W^2\right]}{(N-1)(N-2)(N-3)W^2} - \frac{1}{(N-1)^2}$$

We test the significance of $I(d)$ via a z-score: $z = (I(d) - \mathrm{E}_d[I])/(\mathrm{V}_d[I])^{1/2}$. Since the incidence rates for county i may be used in the calculation of several $I(d)$, we effectively conduct a simultaneous test of the entire correlogram's significance.[32] Given that d takes 13 values, we set $\alpha = 0.05/13 = 0.004$ for each $I(d)$, so that the combined probability of a Type I error was 0.05.

Authors' addresses: Stephan Glavanakov and Thomas Caraco, Department of Biological Sciences, University at Albany, Albany, NY 12222. Dennis J. White, New York State Department of Health, Griffin Laboratory, Slingerlands, NY 12159. Andrei Lapenis, Department of Geography and Planning, University at Albany, Albany, NY 12222. Boleslaw K. Szymanski, Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180. William A. Maniatty, Department of Computer Science, University at Albany, Albany, NY 12222.

REFERENCES

1. Barbour AG, Fish D, 1993. The biological and social phenomenon of Lyme disease. *Science 260:* 1610--1616.

2. Bennett CE, 1995. Ticks and Lyme disease. *Adv Parasitol 36:* 343--405.

3. White DJ, 1998. Lyme disease. Palmer SR, Lord Soulsby, Simpsin DIH, eds. *Zoonoses*. New York, NY: Oxford University Press, 141--153.

4. Centers for Disease Control, 1997. Lyme disease: United States, 1996. *J Am Med Assoc 278:* 12.

5. 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.

6. Wilson ML, Ducey AM, Litwin TS, Gavin TA, Spielman A, 1990. Microgeographic distribution of immature *Ixodes dammini* correlated with that of deer. *Med Vet Entomol 4:* 151--159.

7. Ginsberg HS, 1993. Geographic spread of *Ixodes dammini* and *Borrelia burgdorferi*. Ginsberg H, ed. *Ecology and Environmental Management of Lyme Disease*. New Brunswick, NJ: Rutgers University Press, 63--82.

8. Van Buskirk J, Ostfeld RS, 1998. Habitat heterogeneity, dispersal, and local risk of exposure to Lyme disease. *Ecol Appl 8:* 365--378.

9. Cliff AD, Haggett P, 1988. *Atlas of Disease Distributions*. Oxford, United Kingdom: Blackwell.

10. Sokal RR, Oden NL, 1978. Spatial autocorrelation in biology 1. Methodology. *Biol J Linnean Soc 10:* 199--228.

11. Ostfeld RS, Cepeda OM, Hazler KR, Miller MC, 1995. Ecology of Lyme disease: habitat associations of ticks (*Ixodes scapularis*) in a rural landscape. *Ecol Appl 5:* 353--361.

12. Spielman A, Wilson ML, Levine JF, Piesman J, 1985. Ecology of *Ixodes*-borne human babesiosis and Lyme disease. *Ann Rev Entomol 30:* 439--460.

13. Caraco T, Gardner G, Maniatty W, Deelman E, Szymanski BK, 1998. Lyme disease: self-regulation and pathogen invasion. *J theor Biol 193:* 561--575.

14. Burgess EC, Wachal MD, Cleven TD, 1993. *Borrelia burgdorferi* infection of dairy cows, rodents, and birds from four Wisconsin dairy farms. *Vet Microbiol 35:* 61--77.

15. Mather TN, 1993. The dynamics of spirochete transmission between ticks and vertebrates. Ginsberg H, ed. *Ecology and Environmental Management of Lyme Disease*. New Brunswick, NJ: Rutgers University Press, 43--60.

16. Kitron U, Michael J, Swanson J, Haramis L, 1997. Spatial analysis of the distribution of Lacrosse encephalitis in Illinois, using a geographic information system and local and global spatial statistics. *Am J Trop Med Hyg 57:* 469--475.

17. Levin SA, Grenfell B, Hastings A, Perelson AS, 1997. Mathematical and computational challenges in population biology and ecosystems science. *Science 275:* 334--343.

18. Cliff AD, Ord JK, 1973. *Spatial Autocorrelation*. London, UK: Pion Press.

19. Cliff AD, Ord JK, 1981. *Spatial Processes: Models and Applications.* London, UK: Pion Press.

20. Getis A, Ord JK, 1996. Local spatial statistics: an overview. Longley P, Batty M, eds. *Spatial Analysis: Modelling in a GIS Environment*. Cambridge, UK: GeoInformation International, 261--277.

21. Sokal RR, Oden NL, Thomson BA, 1998. Local spatial autocorrelation in a biological model. *Geogr Analysis 30:* 331--354.

22. Piesman J, 1988. Transmission of Lyme disease spirochetes. *Exper Appl Acarol 7:* 71--80.

23. Jones CG, Ostfeld RS, Richard MP, Schauber EM, Wolff JO, 1998. Chain reactions linking acorns to gypsy moth outbreaks and Lyme disease risk. *Science 279:* 1023--1026.

24. Anderson JF, Johnson RC, Magnarelli LA, Hyde FW, 1986. Involvement of birds in the epidemiology of the Lyme disease agent *Borrelia burgdorferi*. *Infect Immun 51:* 394--396.

25. Anderson JF, 1988. Mammalian and avian reservoirs for *Borrelia burgdorferi*. *An NY Acad Sci 539:* 180--191.

26. Battaly GR, Fish D, 1993. Relative importance of bird species as hosts for immature *Ixodes dammini* (Acari: Ixodidae) in a suburban residential landscape of southern New York State. *J Med Entomol 30:* 740--747.

27. Besag J, Newell J, 1991. The detection of clusters in rare diseases. *J Roy Statist Soc A 154:* 143--155.

28. NYS Department of Commerce, 1998. New York population: counties and msa's, 1995. NYS Bureau of Economic Analysis.

29. Ord JK, Getis A, 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr Analysis 27:* 286--306.

30. Walter SD, 1993. Assessing spatial patterns in disease rates. *Stat Medicine 12:* 1885--1894.

31. Centers for Disease Control, 1999. Methods used for creating a national Lyme disease risk map. *Morbid Mort Wkly Rpt 48:* 21--25.

32. Oden NL, 1984. Assessing the significance of a spatial correlogram. *Geogr Analysis 16:* 1--16.

Figure Legends

FIGURE 1. Lyme disease epidemic in New York State. Histogram shows annual incidence (cases/10^5 individuals) in 57 counties of New York State outside of New York City; for 1988 through 1996. Ordinate plotted on left of figure. Filled diamonds show cumulative number of Lyme disease cases among the same population; ordinate plotted on right of figure. Since incidence rates tended to increase during the study period, cumulative case number apparently increases faster than linear.

FIGURE 2. Spatial autocorrelation depends on distance between county centroids. Ordinate is Moran's *I*, the spatial autocorrelation coefficient. Abscissa is distance between county centroids, grouped into 40-km categories. Estimates are plotted for each of nine years, 1988 through 1996. Despite increase in Lyme disease (see Figure 1), spatial dependence of incidence remains consistent. Spatial autocorrelation exceeds zero significantly at lags 1, 2 and 3 in most years. Hence the correlation distance for these data is approximately 120 km.

FIGURE 3. New York State Lyme disease maps. (a) Cumulative case numbers, scaled logarithmically. (b) Cumulative incidence rates, scaled logarithmically. Maps were obtained by kriging data associated with county centroids. Kriging uses data at discrete points to estimate a continuous surface. Disease foci indicated by combination of high case number and high incidence.

FIGURE 4. Lyme disease in Westchester County. Histogram shows annual case number for Westchester, 1988 through 1996. Ordinate plotted on left of figure. Filled diamonds show cumulative number of Lyme disease cases in Westchester; ordinate plotted on right of figure.

FIGURE 5. Local incidence rates around Westchester County. Lag 0 is Westchester County. Lag d (d = 1--12) shows overall incidence for counties at distance d from Westchester. Data plotted for alternate years; see insert. For 1988, 1990 and 1992, incidence at Westchester exceeds incidence in surrounding counties. Incidence at Westchester increases in 1994, but counties at distance lag 1 have still greater incidence. In 1996, incidence at Westchester declines, and counties at lags 1 and 2 have incidences exceeding the rate at Westchester. Pattern suggests locally increasing infection at Westchester producing a wave that advances spatially, after which annual incidence at Westchester declines.

FIGURE 6. Local spatial clustering, Westchester County. The Getis-Ord statistic indicates clustering of incidence rates above the state-wide mean in and around Westchester County. Lag distance 1 includes county centroids no more than 80 km from Westchester; distances thereafter increase by 40 km. Consistency among years is marked. Most intense clustering (largest values of local statistic) occurs at third category. Hence the strongest clustering includes Westchester and the counties no further than 120 to 160 km from Westchester.

Table 1

Values of Moran's *I* statistic, 1988--1996

	t = 88	89	90	91	92	93	94	95	96
d = 1	**0.75**	**0.51**	0.18	**0.89**	**0.69**	**0.70**	**0.85**	**0.91**	**0.65**
2	**0.32**	**0.33**	**0.29**	**0.32**	**0.37**	**0.27**	**0.42**	**0.48**	**0.42**
3	**0.24**	**0.28**	**0.22**	**0.19**	**0.27**	**0.18**	**0.35**	**0.38**	**0.31**
4	0.07	0.14	0.06	0.07	0.05	0.10	**0.20**	**0.21**	0.15
5	-0.03	-0.01	0.07	-0.01	-0.02	0.01	0.10	-0.02	0.07
6	-0.09	-0.14	-0.11	0.01	-0.21	-0.08	-0.04	-0.12	-0.11
7	**-0.27**	**-0.25**	-0.16	-0.15	**-0.26**	-0.17	**-0.29**	**-0.26**	-0.19
8	**-0.29**	-0.15	-0.19	**-0.27**	-0.22	-0.10	**-0.33**	**-0.37**	-0.16
9	-0.15	-0.28	**-0.31**	-0.23	-0.15	-0.15	**-0.41**	**-0.35**	**-0.38**
10	-0.22	-0.30	-0.24	**-0.33**	-0.21	**-0.34**	**-0.50**	**-0.36**	**-0.50**
11	-0.23	-0.35	-0.18	**-0.42**	-0.03	**-0.46**	**-0.56**	**-0.51**	**-0.41**
12	0.00	-0.14	-0.21	-0.34	-0.10	-0.23	-0.31	-0.32	**-0.56**
13	-0.24	-0.22	-0.21	-0.45	-0.27	-0.46	-0.56	-0.38	-0.51

Bold indicates significance; *P* ≤ 0.004 after Bonferroni type correction.

Table 2

Values of Getis-Ord local statistic, 1988--1996

	$t = 88$	89	90	91	92	93	94	95	96
$d = 1$	**4.23**	**4.24**	3.52	**3.96**	**3.99**	**4.14**	**4.74**	**4.86**	**4.56**
2	**5.08**	**4.86**	**4.35**	**4.90**	**5.03**	**5.04**	**5.32**	**5.72**	**5.22**
3	**5.30**	**4.92**	**4.74**	**5.13**	**4.70**	**4.45**	**5.41**	**5.96**	**5.49**
4	**3.97**	**4.59**	**3.59**	**3.86**	**3.93**	**4.07**	**5.00**	**5.68**	**4.83**
5	3.00	**3.93**	3.21	3.35	**3.59**	**3.60**	**4.59**	**4.65**	**4.29**
6	1.77	2.32	2.99	2.68	2.20	1.83	**3.96**	3.43	3.44
7	1.65	1.48	2.15	2.64	0.91	2.18	**3.65**	2.92	2.76
8	0.26	1.15	0.76	1.91	0.73	2.05	2.04	2.25	2.33
9	0.23	0.49	0.47	1.41	0.62	1.23	1.42	0.96	1.31
10	0.04	0.23	0.66	1.05	0.07	1.03	0.43	0.21	1.18

Bold indicates significance, $P \leq 0.01$; critical values from Table 3 of Ord and Getis[29]

a)

b)