SPATIAL ANALYSIS OF THE DISTRIBUTION OF *EHRLICHI A CHAFFEENSIS*, CAUSATIVE AGENT OF HUMAN MONOCYTOTROPIC EHRLICHIOSIS, ACROSS A MULTI-STATE REGION

MICHAEL J. YABSLEY, MICHAEL C. WIMBERLY, DAVID E. STALLKNECHT, SUSAN E. LITTLE, AND WILLIAM R. DAVIDSON
Southeastern Cooperative Wildlife Disease Study, D. B. Warnell School of Forest Resources, and Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, Georgia

Abstract. *Ehrlichia chaffeensis*, which causes human monocytotropic ehrlichiosis (HME), is an important emerging tick-borne pathogen in the southeastern and southcentral United States. The endemcity probability of *E. chaffeensis* and, by implication, locations with risk for HME, was predicted by using two modeling methods. This is first large-scale study to use geospatial analyses to estimate the distribution of *E. chaffeensis*, and it was conducted using data from a prototypic surveillance system that used white-tailed deer as natural sentinels. Analyses included the *E. chaffeensis* serostatus for 563 counties from 18 states. Both kriging and logistic regression models provided very reliable portrayals of *E. chaffeensis* occurrence and predicted that *E. chaffeensis* distribution had good concordance with human case data. The integration of a deer surveillance system with geospatial analyses was useful in developing HME risk maps that will be useful for identifying high-risk areas for public health interventions such as prevention and control efforts.

INTRODUCTION

*Ehrlichia chaffeensis*, which causes human monocytotropic ehrlichiosis (HME), is an important emerging tick-borne pathogen in the southeastern and southcentral United States. It is maintained in a zoonotic cycle involving white-tailed deer (*Odocoileus virginianus*) as principal vertebrate reservoirs and lone star ticks (*Amblyomma americanum*) as biologic vectors. The lone star tick ranges from mid-Texas north to Iowa and east to the Atlantic coast and recently has been detected as far north as Maine. Evidence of *E. chaffeensis* has been detected in white-tailed deer and/or ticks throughout the range of the lone star tick; however, the distribution of *E. chaffeensis* in these states is not continuous. Some known distributional limits of *E. chaffeensis* in white-tailed deer include a western boundary extending across Kansas, Oklahoma, and Texas, a southern boundary across peninsular Florida, and a cluster of many seronegative populations centered along the Appalachian Mountains. Other studies of white-tailed deer have documented an apparent northern range limit for *E. chaffeensis* in southern portions of Iowa, Illinois, Indiana, and Ohio. Currently the only estimate of HME risk is based on human case data reported to the Centers for Disease Control and Prevention through the National Electronic Telecommunications System for Surveillance. Although the Centers for Disease Control and Prevention and the Council of State and Territorial Epidemiologists recommended that human ehrlichioses be nationally reportable in 1998, currently 11 states do not report ehrlichiosis cases. The incidence of HME in three of these states (Louisiana, Maryland, and Mississippi) is suspected to be high based on numbers of cases in neighboring states. Even for states that consider HME reportable, inconsistencies in diagnosis and reporting of HME limits the usefulness of these data for risk mapping.

Recently, an extensive relatively fine-scale (county) serologic surveillance system using white-tailed deer as natural sentinels was evaluated as a means of discerning the distribution of *E. chaffeensis*. This alternative surveillance system included 563 populations of white-tailed deer throughout 18 states in the southeast and southeastern United States. The goals of the present study were to evaluate alternative methodologies for interpolating between the sampled counties to generate continuous maps of predicted serologic status, and to determine whether data from these maps reflected HME risk by examining its spatial correlation with the distribution of reported HME cases. To accomplish this goal, we conducted two analyses: (1) geostatistical analyses (kriging) to estimate the spatial pattern of *E. chaffeensis* and (2) logistic regression modeling to identify ecologic predictor variables that could be useful in estimating distributional limits of locales suitable for *E. chaffeensis* transmission. The distributions of *E. chaffeensis* projected by kriging and logistic regression were then compared with Centers for Disease Control and Prevention HME case data extracted from McQuiston and others and Gardner and others. Because distribution data of diseases or disease agents are rarely available for all possible locations, geographic information systems (GISs) and geospatial analyses are being used with increasing frequency to estimate geographic distributions. Geostatistics provides a set of methods for explicitly incorporating the spatial pattern data into analysis (geospatial analysis). Spatial autocorrelation among the observations are characterized across multiple scales using semivariograms, and these spatial dependencies can be used to interpolate between the sampled locations by kriging. The predictions obtained using this technique are based on the assumptions that spatial variability in the data is related to the distance between observations, and that values in unsampled areas can be predicted as a weighted average of observations at nearby locations. Geostatistics have been widely used in the fields of ecology, physical geography, and natural resource management, but have only recently been applied to the problem of disease mapping.

An alternative method of spatial interpolation is based on ecologic relationships between the variable of interest (the presence of *E. chaffeensis*) and one or more independent variables. The variable of interest is typically available at only a limited number of spatial locations, whereas the independent variables are mapped across the entire landscape. Statistical methods such as logistic regression are used to develop predictive equations, and these equations are then applied to the unsampled areas to generate a predictive map. This ap-
proach differs from kriging in that predictions of *E. chaffeensis* serostatus are based on relationships with ecologic variables sampled at the same location, whereas kriging predictions are based on spatial patterns of *E. chaffeensis* serostatus across multiple locations.

**MATERIALS AND METHODS**

*Ehrlichia chaffeensis* database. The white-tailed deer *E. chaffeensis* serologic database used was derived from samples collected by Southeastern Cooperative Wildlife Disease Study researchers. The database comprised a total of 563 white-tailed deer populations tested for antibodies reactive to *E. chaffeensis* by the indirect immunofluorescent antibody test as previously described. Serologic data for each population were categorized by county or parish and linked to a map of county boundaries in a GIS. If one or more deer with antibodies reactive to *E. chaffeensis* was detected, that county or parish was classified as positive.

Geostatistical modeling (kriging). Because county serostatus was coded as an indicator variable (present = 1 and absent = 0), the spatial distribution of *E. chaffeensis* was interpolated using indicator kriging. The predictions obtained by this technique are based on the assumptions that 1) spatial variability in *E. chaffeensis* serostatus is related to the distance between counties and 2) the presence of *E. chaffeensis* in an unsampled county can be predicted based on the status of nearby sample locations. An empirical semivariogram was calculated over twelve 50-km lag intervals using data for the sampled counties (n = 563), with the spatial location of each county represented by its centroid. A spherical model was fitted to this semivariogram and used to generate a kriged map of the probability of *E. chaffeensis* presence or absence.

Kriging accuracy was assessed using subsets of data. The 563 data points were randomly divided into a training set (75%, 422 data points) and a test set (25%, 141 data points) using ArcGIS 8.3 (Environmental Systems Research Institute, Redlands, CA). The training set was used to develop the kriging model that was then used to predict the values of the 141 test data points. The subset analysis was independently conducted 10 times. Counties were classified as positive for *E. chaffeensis* if the predicted probability was > 0.5, and negative otherwise. Accuracy (percent of counties correctly classified), sensitivity (percent of positive counties correctly classified), and specificity (percent of negative counties correctly classified) were computed. Semivariogram analysis, kriging, and validation were conducted using ArcGIS 8.3.

Logistic regression modeling. Explanatory variables. Spatial predictor variables for the logistic regression model were obtained as GIS datasets from a variety of sources and a full description of variables is given in Table 1. Climate variables included average monthly and annual minimum temperature, maximum temperature, number of frost days, precipitation, and mean water vapor pressure summarized for 1980–1997. These data were obtained as 1-km Advanced Very High Resolution Radiometer data collected from April 1992 to March 1993. Elevation data was obtained as a 1-km digital elevation model. A soil drainage index was computed for each county using the State Soil Geographic (STATSGO) database (1:250,000). Low values characterized frequently saturated soils, whereas high values characterized well-drained soils. For each mapped soil polygon, a composite index for each polygon was computed based on the area-weighted average of drainage indices of the associated soil series. Deer density data from 1999 were obtained as a paper map from the Quality Deer Management Association (Watkinsville, GA). These data were mapped as polygons with five density levels: 1) deer absent, rare, or urban with known population; 2) < 15 deer/km²; 3) 15–30 deer/km²; 4) 31–45 deer/km²; and 5) > 45 deer/km². The states included in this study were digitized, georeferenced, and converted to a 1-km grid.

These data sets were overlaid on the county boundaries to compute summary values for each county. Climate variables and soil drainage were summarized as mean values for each county. Land cover was summarized as the percentage of each county occupied by various land cover classes. Elevation was summarized as mean value for each county, and a ruggedness index was computed by subtracting the lowest elevation from the highest elevation within each county. Deer density was summarized by assigning the density class that covered the majority of each county to represent the entire county.

Logistic regression models. Logistic regression analysis was carried out in SAS 8.2 (SAS Institute, Inc., Cary, NC) using the serostatus of white-tailed deer populations as the dependent variable. We used stepwise regression with a P < 0.05 cutoff to reduce the large set of ecologic predictor variables to a more parsimonious subset. Because we had a priori reasons (e.g., differing climates in western versus eastern states) to believe that different predictive variables were important in different portions of the 19-state study region, we analyzed the region as a whole and then analyzed three separate subregions containing 18 of the 19 original states. The western subregion included Kansas, Oklahoma, and Texas (n = 110 of 563 counties), the central subregion included Alabama, Arkansas, Louisiana, Mississippi, and Missouri (n = 179), and the eastern subregion included Georgia, Delaware, Kentucky, New Jersey, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia (n = 256). Florida was excluded from other subregions because of its semi-isolated geography and unique climate. Because of small sample size (n = 18), data for Florida could not be analyzed separately.

The logistic regression models were applied in ArcView 3.2 (Environmental Systems Research Institute) using spatial predictor variables from all of the counties in the 19-state region to create a map of the probability of *E. chaffeensis* occurrence (hereafter termed endemicity probability). The overall fit of the models was measured by comparing observed and predicted values for the sampled counties, with *E. chaffeensis* classified as present when the predicted probability of occurrence was greater than 0.5, and absent otherwise. Accuracy, sensitivity, and specificity were computed as described previously. We also calculated the area under the receiver operating characteristic curve (AUC ROC) and the pseudo-$R^2$ as overall indices of model fit.

Comparison of predicted *Ehrlichia chaffeensis* distribution with HME case data. Incidence rates of HME reported to the
Centers for Disease Control and Prevention were obtained by combining data from McQuiston and others and Gardner and others. In both studies, annual incidence rates (per million human population) had been calculated using the corresponding yearly census estimate. Incidence groups were low (< 2.5), low-medium (2.5 to < 5), medium-high (5 to < 15), and high (> 15) by McQuiston and others and low (< 3.5), low-medium (3.5 to < 8.3), medium-high, (8.3 to < 17), and high (> 17) by Gardner and others. Although category limits differed slightly in the two studies, the four levels of incidence (low to high) were combined to obtain unified sequential incidence estimates. If a county was positive in both studies but classified in different incidence categories, the higher of the two was used in this study. These incidence estimates per county were entered into ArcView 3.2 and compared with the endemicity probability maps generated using our geostatistical and logistic regression modeling.

**RESULTS**

**Geostatistical modeling (kriging).** The fitted semivariogram had a range of 463.3 km, a nugget of 0.13014, and a sill of 0.21846 (Figure 1). The range identified the maximum distance at which spatial autocorrelation was detected among the sampled counties. The nugget quantified the minimum variability at a lag distance of zero, whereas the sill quantified maximum variability among spatially independent samples. The high ratio of nugget to sill values (0.60) indicated that a considerable amount of fine-scale (< 50 km) spatial variability in *E. chaffeensis* distribution was not accounted for in the kriging model. With a cutoff value of 0.5, the mean accuracy of the model was 85.4%, the mean sensitivity was 96.2%, and the mean specificity was 60.4% (Table 2).

Large irregularly shaped clusters of both positive and negative counties were detected in the study region (Figure 2). Positive counties were clustered in Atlantic coastal states from Florida to New Jersey and in a broad area extending north from Texas to Missouri east to Kentucky and south to Mississippi. Two large negative clusters were apparent. One was along the western edge of the region through western portions of Kansas, Oklahoma, and Texas and a second was centered along the Appalachians in Virginia, Kentucky, West Virginia, Tennessee, and Georgia. Two smaller negative clus-
For the entire 19-state region—

$E. \text{CHAFFEENSIS}$—

$E. \text{chaffeensis}$

The highest correlation with correctly classified positive counties and nearly 100% of high incidence counties were classified as positive by our analyses (Figure 6).

**Logistic regression models.** For the entire 19-state region-wide model, five variables were significantly related to $E. \text{chaffeensis}$ serostatus in white-tailed deer (Table 3 and Figure 3). The $E. \text{chaffeensis}$ endemnicity probability decreased with elevation, increased with percent savannah, increased with percent mixed deciduous forest (combination of deciduous and mixed), decreased with percent wooded wetlands, and increased with summer maximum temperature. With a cutoff value of 0.5, this model accurately predicted the $E. \text{chaffeensis}$ serostatus of 421 (74.8%) of 563 counties (Table 2). Only 30 counties were misclassified as negative (sensitivity = 92.1%).

The predicted distribution of seropositive white-tailed deer populations in some areas (southern Louisiana, Mississippi, and along the Mississippi River alluvial basin) was overestimated using this region-wide model and the predicted distribution in eastern Kansas and northern Missouri was underestimated (Figure 3).

In the eastern subregion model, five variables were significantly related to $E. \text{chaffeensis}$ serostatus (Table 3 and Figure 4). Predicted distribution decreased with elevation, and increased with percent total forest, soil moisture content, annual minimum temperature, and summer relative humidity. Three explanatory variables were associated with serostatus of white-tailed deer populations in the central subregion model (Table 3). The predicted distribution decreased with percentage cropland and annual precipitation, and increased with the density of white-tailed deer (Table 3 and Figure 4). In the western subregion model, predicted positive counties were associated with increasing annual precipitation and increasing summer minimum temperature (Table 3 and Figure 4).

The overall fit of the three subregion models was better than the region-wide model. The accuracy of the eastern subregion model compared with the region-wide model increased by only 3.3%, but the specificity increased by 15.1% (Table 2). The accuracy of the central subregion model was greater than the overall model (79.3% versus 74.8%, Table 2). The sensitivity was slightly lower than the overall model, but the specificity was higher (47.7% versus 38.5%). The western subregion model correctly classified the serostatus of 84.5% counties, resulting in a 9.7% increase in predictive power over the overall model (Table 2). The sensitivity of the region-wide and western subregion models was the same, but the specificity of the western model was 29.1% higher than that of the region-wide model. All three subregion models had higher max-rescaled $r^2$ values and AUC ROC (Table 2) indicating that they better fit the data.

**Comparison with HME case data.** Within our 19 state study region, 271 counties in 15 states (Figure 5) had reported cases of HME during the years 1997–2002. The highest correlation with human case data was observed for the kriging analysis, with a total of 251 counties (92.6%) with human cases being classified as positive by our analysis (Figure 6). Overall, a high correlation was noted for the three subregion logistic models; 244 counties (90%) with reported HME cases were predicted to be positive. As HME incidence increased, there was higher correlation with correctly classified positive counties and nearly 100% of high incidence counties were classified as positive by our analyses (Figure 6).

**DISCUSSION**

We used two different modeling methods to predict the endemicity probability of $E. \text{chaffeensis}$ within the southeastern and southcentral United States and, by implication, identify locations with risk for HME. This is first large-scale study to use geospatial analyses to estimate the geographic distribution of $E. \text{chaffeensis}$, and it was conducted using data from a prototypic surveillance system that used white-tailed...
deer as natural sentinels. Data for this deer sentinel system was derived by serologically testing deer from 563 populations (counties) in 18 southcentral and southeastern states. Because geographic gaps existed since deer were not tested from every county, the present study used kriging and logistic regression modeling to predict the distribution of *E. chaffeensis* across all counties in these states. Logistic regression analyses also identified several climatic and landcover vari-

**Table 3**

Logistic regression models for predicting distribution of *Ehrlichia chaffeensis*

<table>
<thead>
<tr>
<th>Region</th>
<th>Variable</th>
<th>Coefficient</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>19-state region</td>
<td>Intercept</td>
<td>-11.51</td>
<td>14.73</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
<td>-0.09</td>
<td>35.45</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Savanna</td>
<td>3.78</td>
<td>4.67</td>
<td>0.0308</td>
</tr>
<tr>
<td></td>
<td>Deciduous forest mix</td>
<td>1.18</td>
<td>10.79</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>Wooded wetlands</td>
<td>-214.9</td>
<td>5.06</td>
<td>0.0244</td>
</tr>
<tr>
<td></td>
<td>Summer maximum temperature</td>
<td>0.41</td>
<td>20.62</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>-36.72</td>
<td>11.06</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
<td>-0.31</td>
<td>28.61</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Forest</td>
<td>1.88</td>
<td>8.36</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>Soils</td>
<td>0.08</td>
<td>14.30</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Annual minimum temperature</td>
<td>0.20</td>
<td>5.62</td>
<td>0.0178</td>
</tr>
<tr>
<td></td>
<td>Summer relative humidity</td>
<td>0.52</td>
<td>11.63</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>Annual precipitation</td>
<td>0.09</td>
<td>21.69</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Central subregion</td>
<td>Intercept</td>
<td>18.99</td>
<td>23.80</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Annual precipitation</td>
<td>-1.55</td>
<td>21.69</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Crops</td>
<td>-2.69</td>
<td>11.47</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>Deer density</td>
<td>0.59</td>
<td>7.65</td>
<td>0.0057</td>
</tr>
<tr>
<td>Western subregion</td>
<td>Intercept</td>
<td>-13.95</td>
<td>19.45</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Annual precipitation</td>
<td>0.85</td>
<td>16.66</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Summer minimum temperature</td>
<td>0.37</td>
<td>7.16</td>
<td>0.0074</td>
</tr>
</tbody>
</table>
ables useful as predictors of *E. chaffeensis* occurrence. To relate these geospatial analyses to the occurrence of HME, comparisons were made with the geographic distribution of laboratory-confirmed or suspected cases of HME for the period 1997–2002 compiled by public health agencies as previously published. These comparisons disclosed spatial concordance between HME cases and *E. chaffeensis* endemicity projected by both geospatial models.

The usefulness of modeling is dependent upon the reliability of input data, how well model predictions fit the actual status of variables being modeled, and in regard to geospatial modeling, especially the issues of scale and scope of sampling. Input data from the white-tailed deer surveillance sentinel system used in this study has previously been demonstrated to meet critical reliability criteria including diagnostic accuracy, adequacy of sample sizes and sampling intensity, key epidemiologic associations with the lone star tick vector, and ability to detect spread of *E. chaffeensis*. Data for the present study included more than one-third (33.8%) of counties in the 19-state study region, and measures of reliability for both kriging and logistic regression models disclosed high accuracy, sensitivity, and somewhat lower specificity (Table 2). Furthermore, analyses confirm that the logistic regression models fit the data well and that variables included in the models had biologic relevance based on previous studies.

Kriging accurately estimated the distribution of *E. chaffeensis* across the entire study region. This procedure investigates and projects the spatial pattern of *E. chaffeensis* based on spatial autocorrelation among sample populations, but does not consider relationships with secondary variables such as landcover and climate. This purely spatial analysis is dependent on a sufficient number of populations being sampled because an increase in distances between sampled populations leads to increased uncertainty. Large clusters of counties with similar serologic status were accurately detected; however, when variability in *E. chaffeensis* serostatus occurred among counties in close proximity, this model tended to over predict the presence of *E. chaffeensis*. Two prominent clusters of seronegative populations in the Mississippi River alluvial basin were misclassified as seropositive, in this case, small numbers of seronegative counties were encircled by seropositive counties. Three smaller negative clusters in central Alabama, southcentral Georgia, and southeastern Texas also were projected to have moderate risk (0.5 < *P* < 0.7). Similarly, two positive counties in western Kansas, one in southern Louisiana, and one in western North Carolina fell well within areas projected to have a low probability of *E. chaffeensis* occurrence. The failure to correctly classify individual or small clusters of counties reflects the high nugget value of our semivariogram, which indicated that some fine scale variability (< 50 km) was present but not being detected.

Logistic regression analyses also accurately classified the *E.
chaffeensis status within the study region and identified several ecologic variables associated with the distribution of E. chaffeensis. In contrast to kriging, logistic regression modeling relies on correlations with ecologic variables that are then used to estimate the distribution of E. chaffeensis. The region-wide model had good accuracy (75%) and high specificity (92%), but low sensitivity (39%). Distributions of E. chaffeensis were accurately predicted in many states, but the region-wide logistic regression model produced results that conflicted with observed distributions in northern portions of Kansas and Missouri, central South Carolina, New Jersey, and portions of the Mississippi River alluvial basin and Gulf Coast. Partitioning into three subregions improved the accuracy, sensitivity, and specificity of the logistic regression modeling method. The most significant changes were the increased ability of subregion models to accurately predict positive status in northern Kansas/Missouri and New Jersey and to predict negative status for areas in eastern Kentucky, within the Mississippi River alluvial basin, and along the central Gulf coast. Both region-wide and subregional logistic regression models projected limited positive areas in north central Kentucky, northeastern Alabama, northwestern Georgia, and central South Carolina that were outside the current distribution based on observed serostatus and kriging analysis.

As with most predictive disease mapping studies, we did not explicitly account for spatial autocorrelation in our logistic regression models. Thus, our comparisons between spatial modeling methods highlight the differences between a purely spatial approach (kriging) and a purely ecologic approach (logistic regression). Both kriging and subregion logistic regression models provided very reliable portrayals of E. chaffeensis occurrence, and the similarity of maps produced by these two modeling processes tends to substantiate the projected distributions. Because kriging was more accurate than logistic regression, models the observed distribution of serostatus, and does not require collection of additional ecologic variables for use in a GIS, kriging analysis probably provides the most accurate estimate of current E. chaffeensis distribution and would be more useful in monitoring spread to new locations. Conversely, because logistic regression analysis is based on association with ecologic variables, the projected extensions of E. chaffeensis into currently negative areas suggests these areas probably are suitable for maintenance of E. chaffeensis and that they are likely locations of future expansion. Alternative statistical methods such as kriging with varying local means, autologistic regression, and geographically weighted regression have the potential to integrate spatial and ecologic information into a unified modeling framework.16,30,31 These approaches hold promise for improving prediction accuracy and for elucidating stronger re-

![FIGURE 4. Maps of endemnicity probabilities for Ehrlichia chaffeensis based on separate logistic regression models for the eastern, central, and western subregions. Projections for Florida are from the original region-wide model. Solid circles represent positive E. chaffeensis serologic data and open circles represent negative serologic data obtained from a previous white-tailed deer surveillance study.9 This figure appears in color at www.ajtmh.org.](image-url)
relationships with ecologic predictor variables, and we plan to examine them in future studies.

The distribution of *E. chaffeensis* is highly dependent on the presence of white-tailed deer to serve as reservoirs, and the presence of the lone star tick to serve as a vector, and indirectly on any factors that limit the distribution of white-tailed deer or the lone star tick. Based on extensive biologic data relevant to the lone star tick, five critical variables, and interactions thereof, have been identified as being important to lone star tick population dynamics. These variables consist of measures of moisture, temperature, habitat type, day length, and host density (white-tailed deer and/or cattle). During warm months, temperature and moisture are positively correlated with survival of the lone star tick and activity, while day length is negatively correlated with lone star tick activity. Abundance and survival of the lone star tick are positively associated with wooded habitats, particularly young second-growth forests containing thick underbrush that provides proper microhabitat, but are negatively associated with grass or crop land habitats. The landcover variables in our models (Table 3) conformed in general terms with these previously identified critical variables important to lone star tick population dynamics. The presence of *E. chaffeensis* was correlated with higher soil moistures and higher proportions of forest/savannah, which are habitats that provide better microhabitat for the lone star tick, while wooded wetlands, cropland (Figure 7), and high elevations (probably to some extent reflecting temperature) were negatively correlated with *E. chaffeensis*.

Regarding climatic variables, our logistic regression models were in agreement with previous laboratory and field studies on the lone star tick in which high annual precipitation and high relative humidity prevent desiccation and higher average temperatures (Figure 7) help prolong questing, all of which conspire to increase lone star tick survival and densities. Of interest was that the presence of *E. chaffeensis* was negatively correlated with increasing annual precipitation in the central subregion model, possibly because high
Precipitation in this low-lying region may lead to flooding or excessive wetness. The abundance of another ixodid tick, *Ixodes scapularis*, is negatively impacted by excessively wet soils, and our negative correlation with wooded wetlands suggests that *A. americanum* also may not be well suited for habitats subject to standing ground water.

The presence of white-tailed deer is clearly important to maintenance of high densities of lone star ticks, however, interestingly, we found that the density of white-tailed deer was not a significant predictor variable in the region-wide model or two of the three subregion models. Although deer were important to modeling population dynamics of lone star ticks, there are critical differences between modeling lone star ticks and modeling the probability of *E. chaffeensis* infection in white-tailed deer. Because the force of transmission is so heavily focused on white-tailed deer, low numbers of lone star ticks can quickly introduce and maintain *E. chaffeensis* within white-tailed deer populations. Furthermore, based on almost 100% concordance between the presence of lone star ticks and the presence of *E. chaffeensis* in whitetailed deer, and detection of *E. chaffeensis* by a polymerase chain reaction in nearly all surveys of white-tailed deer, it is reasonable to conclude that a very high proportion of lone star tick populations are infected. White-tailed deer were present throughout our study area (Figure 7), frequently at high densities. Only the central subregion logistic regression model included deer density as a variable, and this appeared attributable to sizable highly agriculturalized areas within the Mississippi River alluvial basin where deer are rare, absent, or exist at low densities or in semi-isolated populations.

Comparison of the predicted distribution for *E. chaffeensis* with HME case data showed that > 90% of counties with HME cases were predicted positive for *E. chaffeensis* by both kriging and logistic regression. In fact, counties in the high HME incidence category were nearly always (> 98%) identified as positive for *E. chaffeensis* based on our analyses. Several counties with a low incidence of HME were predicted to be negative. This could be because our analyses misclassified the county status or that the HME cases were acquired in another locale but reported in the county of residence. This comparison provides evidence that use of white-tailed deer as sentinels is an efficient alternative to human surveillance for predicting *E. chaffeensis* distribution and HME risk.

The distribution of *E. chaffeensis* projected in the current study should be validated by testing additional white-tailed deer populations from unsampled counties, especially in areas where the prediction data were inconsistent between the two.
models. Although highly correlated with HME case data, a better means for validation of the sentinel deer surveillance system is needed because HME case data are known to contain a variety of biases and under-represent the distribution and magnitude of HME.\textsuperscript{1,2,13} An extensive cross-sectional serosurvey of humans for \textit{E. chaffeensis} antibodies, which would provide a less biased and more accurate estimate of exposure to \textit{E. chaffeensis} than does HME case data, should be conducted to validate the ability of the sentinel white-tailed deer surveillance system to accurately predict HME risk.

Received October 5, 2004. Accepted for publication December 14, 2004.

Acknowledgments: We thank Scott Dansk and Sarah Cross for excellent help with building the GIS databases used in this study. We also thank the numerous staff at federal, state, and private wildlife agencies and the Southeastern Cooperative Wildlife Disease Study for collecting samples used in this study.

Financial support: This work was supported primarily by the National Institutes of Allergy and Infectious Diseases (grant 5 R01 AI044235). Additional support was provided by the Federal Aid to Wildlife Restoration Act (50 Stat. 917) and through sponsorship from fish and wildlife agencies in Alabama, Arkansas, Florida, Georgia, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, Puerto Rico, South Carolina, Tennessee, Virginia, and West Virginia.

Authors’ addresses: Michael J. Yabsley, Southeastern Cooperative Wildlife Disease Study, College of Veterinary Medicine, University of Georgia, Athens, GA 30602, Michael C. Wimberly, D. B. Warnell School of Forest Resources, The University of Georgia, Athens, GA 30602. David E. Stallknecht and Susan E. Little, Department of Infectious Diseases, College of Veterinary Medicine, The University of Georgia, Athens, GA 30602. William R. Davidson, Southeastern Cooperative Wildlife Disease Study, College of Veterinary Medicine, University of Georgia, Athens, GA 30602 and D. B. Warnell School of Forest Resources, The University of Georgia, Athens, GA 30602.

Reprint requests: Michael J. Yabsley, Wildlife Health Building, Southeastern Cooperative Wildlife Disease Study, College of Veterinary Medicine; University of Georgia, Athens, GA 30602. Telephone: 706-542-1741, Fax: 706-542-5865, E-mail: myabsley@vet.uga.edu.

REFERENCES


