

# Social-ecological factors determine spatial variation in human incidence of tick-borne ehrlichiosis

B. R. BAYLES<sup>1\*</sup> AND B. F. ALLAN<sup>2</sup>

<sup>1</sup>School of Public Health, St Louis University, St Louis, MO, USA

<sup>2</sup>Department of Entomology and Integrative Biology, University of Illinois Urbana–Champaign, Urbana, IL, USA

Received 19 June 2013; Final revision 13 October 2013; Accepted 30 October 2013;  
first published online 26 November 2013

## SUMMARY

The spatial distribution of human cases of tick-borne diseases is probably determined by a combination of biological and socioeconomic factors. A zoonotic tick-borne pathogen, *Ehrlichia chaffeensis*, is increasing in human incidence in the USA. In this study, the spatial patterns of probable and confirmed *E. chaffeensis*-associated cases of ehrlichiosis from 2000 to 2011 were investigated at the zip-code level in Missouri. We applied spatial statistics, including global and local regression models, to investigate the biological and socioeconomic factors associated with human incidence. Our analysis confirms that the distribution of ehrlichiosis in Missouri is non-random, with numerous clusters of high incidence. Furthermore, we identified significant, but spatially variable, associations between incidence and both biological and socioeconomic factors, including a positive association with reservoir host density and a negative association with human population density. Improved understanding of local variation in these spatial factors may facilitate targeted interventions by public health authorities.

**Key words:** Emerging infections, infectious disease epidemiology, ticks, vectors, zoonoses.

## INTRODUCTION

Zoonotic pathogens transmitted by the lone star tick (*Amblyomma americanum*) are rapidly emerging throughout its range in the USA [1]. Foremost of these is *Ehrlichia chaffeensis*, causative agent of *Ehrlichiosis chaffeensis* infection in humans (formerly human monocytic ehrlichiosis and hereafter referred to as ‘ehrlichiosis’) [2]. In recent years, there have been significant increases in the number of cases reported from central regions of the country, with

rates in Missouri among the highest of all states [3]. Although advances in diagnostic techniques have facilitated the recognition of *E. chaffeensis* infection in humans, the rising incidence is thought to be partially attributable to anthropogenic influences on the increasing geographical range, abundance, and infection rates of *A. americanum* [4].

Transmission of tick-borne, zoonotic pathogens in susceptible human populations requires the relative proximity of vertebrate reservoir hosts, tick vectors, and humans. Thus the mechanisms that drive tick-borne disease risk and incidence are intrinsically spatial [5]. Further, the spatial distribution of entomological risk (often measured as the density of infected ticks for a given area) is dependent upon environmental factors occurring at multiple spatial scales [6]. For example, *A. americanum* is strongly influenced

\* Author for correspondence: B. R. Bayles, PhD, MPH, Institute on the Environment, University of Minnesota–Twin Cities, Learning and Environmental Sciences 325, 1954 Buford Ave, St Paul, MN 55108.  
(Email: brett.r.bayles@gmail.com)

by local-scale factors that determine its survival (e.g. temperature and humidity patterns) [7]. At larger spatial scales, tick distributions and pathogen infection rates are a factor of the presence and abundance of vertebrate host species, which are in turn influenced by the spatial arrangement of important habitats required for forage and reproduction [8]. As white-tailed deer (*Odocoileus virginianus*) are an important host for *A. americanum* and zoonotic reservoir for *E. chaffeensis* [4, 9], factors that influence the spatial distribution of white-tailed deer may also influence disease risk [10]. Finally, the presence/absence of pathogens, ticks, and hosts is determined at even larger spatial scales by phenomena associated with biogeography, including habitat characteristics and annual variation in temperature and precipitation [11].

Transmission of *E. chaffeensis* into human populations is probably further mediated by human socioeconomic and demographic patterns at multiple spatial scales. Factors that are associated with risk of exposure to *E. chaffeensis* include occupation (e.g. workers in resource extraction industries such as logging), age, sex and racial group (e.g. incidence is highest in white males aged >40 years), and even golf score [12–14]. Further, patterns in individual behaviours that influence risk of exposure to tick-borne diseases (hereafter ‘behavioural risk’) may also vary across spatial gradients [15]. However, these socioeconomic factors can only influence disease incidence when ecological parameters are suitable for transmission to occur.

An emerging discipline, termed ‘spatial epidemiology’, has been proposed as a means of elucidating these transmission dynamics by examining the spatial arrangement of social and ecological factors that contribute to patterns of disease risk and incidence [16]. Despite the importance of social-ecological factors in determining the entomological risk of exposure to *E. chaffeensis* or human incidence of ehrlichiosis [17, 18], little is known about the relative role these factors may play in determining the spatial variation in the distribution of ehrlichiosis incidence. Using a spatial epidemiological framework, we sought to better understand the manifestations of *E. chaffeensis* exposure within a highly endemic state to better inform medical surveillance efforts and guide public health interventions. Specifically, our objectives were to: (1) describe the spatial distribution of ehrlichiosis incidence in Missouri, (2) identify areas of unusually high numbers of clustered cases, and (3) quantify the large-scale effects of epidemiologically meaningful

socioeconomic, ecological, and biogeographical factors on the spatial variability of disease incidence.

## MATERIALS AND METHODS

### Epidemiological data

Human ehrlichiosis was first made a nationally notifiable infectious disease in 1998 [19], with subsequent case definition revisions made in 2000 and 2008 [20, 21]. We obtained a database of 1135 probable and confirmed cases of ehrlichiosis associated with *E. chaffeensis* in Missouri during 2000–2011 from the Missouri Department of Health and Senior Services (MDHSS). Probable and confirmed cases are each characterized by clinically compatible illness with evidence of elevated serological antibody titre, with case confirmation made with evidence of: a  $\geq$ fourfold increase in antibody titre, polymerase chain reaction assay, or pathogen isolation in cellular culture [2]. Of the cases reported during the study period, 1062 (93.6%) contained patient address information at the zip-code level, with  $\sim$ 50% of cases classified as confirmed (MDHSS IRB exemption no. 71612). To obtain robust estimates of disease distribution, probable and confirmed cases for each zip code were combined to calculate the standardized cumulative incidence/100 000 persons using 2000 U.S. decennial census population estimates [22].

### Socioeconomic, environmental and biogeographical variables

To evaluate the potential associations between human population characteristics and the spatial distribution of ehrlichiosis, we created socioeconomic variables with data obtained from the 2000 U.S. census [22]. Variables were chosen to reflect either known risk factors for ehrlichiosis or potential determinants of disease transmission [8, 14]. These included: gender (proportion male), age (proportion aged >64 years), race (proportion white), occupation (proportion working in agricultural, forestry, mining, fishing, or hunting industries), education (proportion with high-school-level diploma), and poverty (proportion under federal poverty line). In addition, we included human population density/km<sup>2</sup> and the proportion of vacant housing units to assess the contribution of land-use and housing characteristics to incidence in residentially developed regions.

We included several environmental variables that have been strongly associated with the presence of *E. chaffeensis*-infected *A. americanum*: forest cover and the density of white-tailed deer [4, 9, 18]. To characterize the distribution of forest cover, we calculated the percentage of each zip code covered by deciduous forest using data obtained from the 30 × 30 m resolution 2006 National Land Cover Database [23]. White-tailed deer population estimates from 2000 to 2009 were obtained at the county level using population projections calculated by the Missouri Department of Conservation from hunter harvest success rates. To rescale the average number of deer to the zip-code level, we divided the number of deer in each county by the weighted area (km<sup>2</sup>) of each zip code that fell within a particular county.

To understand the importance of environmental factors that vary at biogeographical scales, we included level III ecoregions derived by the Environmental Protection Agency (EPA). These ecoregions reflect a unique set of ecological communities defined by particular climatic and geological characteristics [24], which include (1) location, (2) climate, (3) vegetation, (4) hydrology, (5) terrain, (6) wildlife, and (7) human land-use. Level III ecoregions are derived from data compiled from many different sources with the intention that these offer concise and ecologically meaningful depictions of landscape-scale variation in biological and environmental variables. In the state of Missouri there are five major ecoregions: Central Irregular Plains, Interior River Valleys and Hills, Ozark Highlands, Mississippi Alluvial Plain, and Western Corn Belt Plains (Fig. 1a, Supplementary Table S1). We aggregated the analysis by ecoregion to identify how the effects of social-ecological factors on disease incidence may differ across large-scale environmental gradients and to control for the presumptive effects of local-scale environmental effects (e.g. microclimate) on disease distribution.

### Statistical analysis

We implemented an empirical Bayesian smoothing (EBS) on the cumulative incidence data to adjust for the random variance associated with the distribution of small numbers of cases in relatively small populations at risk across zip codes. EBS allows for better visualization of disease gradients, which may not otherwise be apparent with raw case numbers [25]. Using EBS incidence, we determined if the distribution of ehrlichiosis across the state exhibited spatial

dependence by calculating Moran's *I* measure of spatial autocorrelation [26]. To better delineate the location of significantly high-incidence clusters of ehrlichiosis, we computed a locally derived measure of Moran's *I* statistic known as the local indicator of spatial autocorrelation (LISA) statistic [27].

Comparisons were made between high-incidence LISA clusters and all other zip codes to determine which social-ecological factors were associated with elevated incidence. Values of each variable were compared using the Wilcoxon rank-sum test. To determine which factors predicted membership in a high-incidence LISA cluster, we used binary logistic regression with outcome of 1 for cluster vs. 0 for no cluster. A bivariate analysis was conducted first, with each variable included separately in the model. Significant variables were then entered into a multivariate logistic regression model with backward stepwise elimination used to fit a final set of parameters. To measure the importance of dividing the state into ecologically unique areas, we compared incidence, socioeconomic, and environmental variables across ecoregions using the Kruskal–Wallis one-way analysis of variance. Again, we adopted bivariate and multivariate binary logistic regression to model a dichotomized outcome variable based on the median incidence across all zip codes (1 for incidence  $\geq 2.31 \times 10^{-4}$  vs. 0 for incidence  $< 2.31 \times 10^{-4}$ ) to model factors associated with elevated incidence within each ecoregion. We selected the median as a cut-off value to represent elevated vs. non-elevated incidence in place of an *a priori* justification for other values and to retain sufficient statistical power. Analyses were performed using SAS v. 9.3 (SAS Institute Inc., USA) and GeoDa [28].

### Modelling spatial variation

When spatial structure exists in data, the use of global ordinary least squares (OLS) regression can suppress local variation in the relationships between model parameters [29]. Therefore, we implemented geographically weighted regression (GWR) to explore the important influence of local spatial variation on the relationships between social-ecological factors with smoothed ehrlichiosis incidence. GWR has been developed as a means of accounting for the effects of spatial non-stationarity by generating local models and parameter estimates for every point in space [30]. We used this technique to determine the extent of local-scale geographical variability in social-ecological

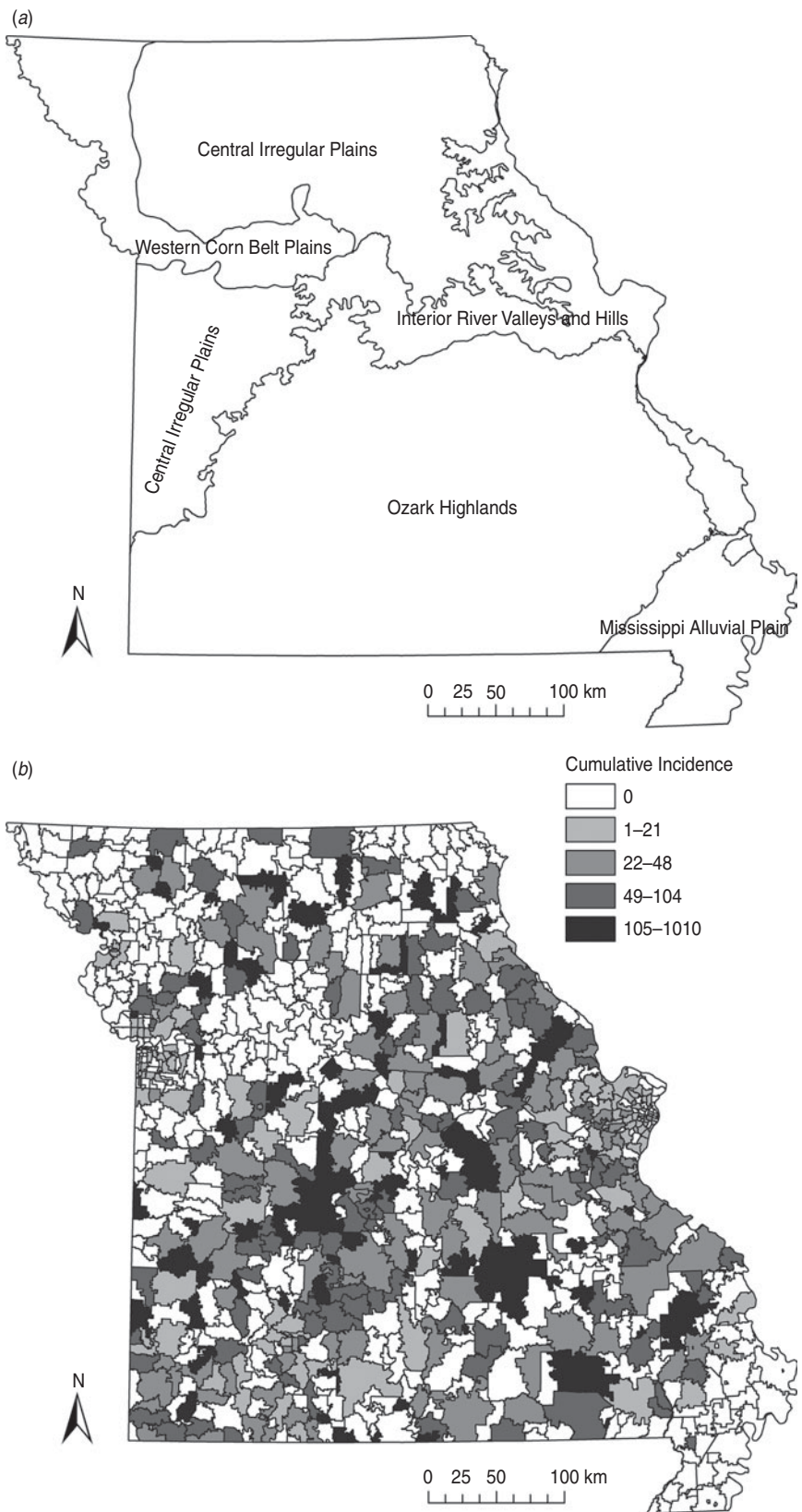
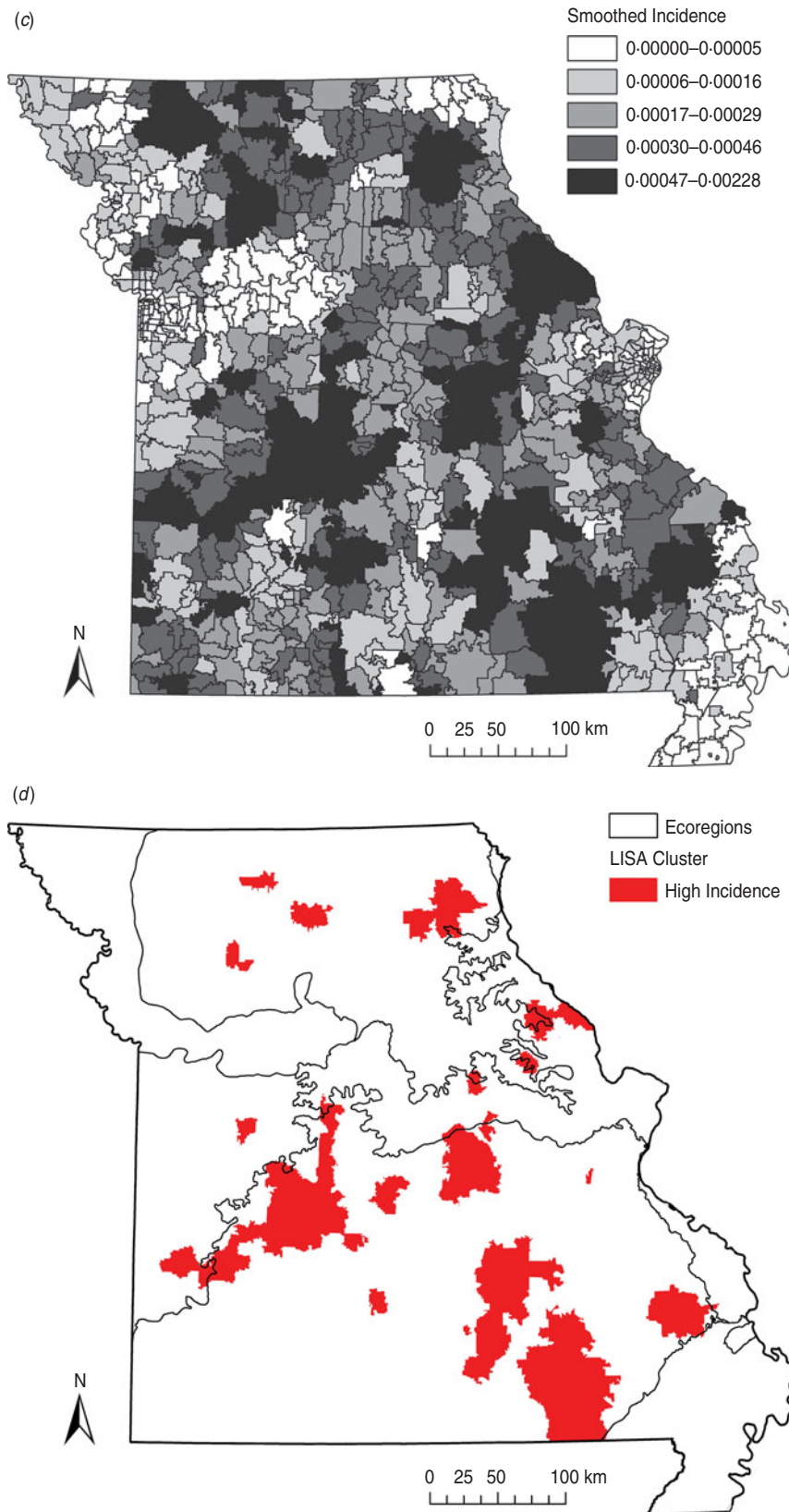


Fig. 1. For legend see next page





**Fig. 1.** Spatial distribution of: (a) Missouri level III ecoregions, (b) cumulative incidence of probable and confirmed *E. chaffeensis* infection/100 000 persons, (c) empirical Bayesian smoothed (EBS) incidence, and (d) high incidence local indicators of spatial autocorrelation (LISA) clusters.

variables and disease incidence for the centroid of each zip code, which may otherwise be obscured in an ecoregion-scale analysis.

We first built a global OLS regression to fit a set of social-ecological variables. Combinations of significant covariates were retained in the final OLS model by measuring the reduction of Akaike's Information Criterion (AIC), a technique that favours parsimony over complexity [31] (Supplementary Table S4). The model was screened for evidence of collinearity in the variables by ensuring all variance inflation factors remained low ( $<3.0$ ). Screening the OLS residuals for spatial heterogeneity using Moran's *I* statistic indicated that a non-random ( $P < 0.05$ ) spatial structure was present in the variables, justifying the inclusion of spatial weighting techniques. Therefore, we entered variables from the global OLS model into a GWR framework to explore the local variability in significant model parameters and incidence. Adaptive bandwidth selection was used to minimize AIC and ensure a sufficient number of zip codes were selected for each local model [32]. We measured improvement in model fit between the OLS and GWR models by comparing the corrected AIC (AICc) and adjusted  $R^2$  values (Table 3, Supplementary Fig. S1B). In addition, we tested for spatial autocorrelation in the GWR residuals (Supplementary Fig. S1C) using Moran's *I* statistic as a means of model validation and to ensure non-biased parameter estimates. To visualize the significance of each local parameter estimate for every zip code in the state, we plotted absolute *t* values by dividing each coefficient by its standard error and compared it against the magnitude of the coefficient [33]. Both OLS and GWR models were produced using ArcGIS v. 10.0 (ESRI, USA).

## RESULTS

### Spatial distribution of incidence

The average cumulative incidence of ehrlichiosis in Missouri was 37/100 000 persons, with incidence ranging from 0/100 000 to 1010/100 000 for the 11-year time period (Fig. 1b). At least one human case was reported in 423/1025 (41.3%) zip codes during this time. The EBS analysis revealed a clearer representation of areas with elevated incidence, particularly the Ozark Highlands ecoregion (Fig. 1c). Moran's *I* was 0.586 ( $P < 0.001$ ), indicating that neighbouring zip codes have statistically more similar incidence

than would be found at random. The LISA analysis revealed that multiple ( $n = 80$ ) zip codes had significantly similar high incidence (Fig. 1d). Within these clusters, the average cumulative incidence was 130/100 000 compared to 29/100 000 for the rest of the state.

### Determinants of elevated incidence

We found significant differences in the social-ecological characteristics between zip codes within and outside of high-incidence clusters. In the bivariate analysis, low population density, increased proportion aged  $\geq 65$  years, increased proportion white race, increased proportion of vacant housing units, increased proportion working in the agriculture, forestry, fishing, or hunting occupational sector, increased proportion with at least a high-school-level education, increased proportion living in poverty, increased proportion of deciduous forest cover, and high deer density were all significantly predictive of membership in a high-incidence cluster (Table 1, Supplementary Table S3). Several variables remained significant after inclusion in a multivariate model. The odds of membership in a high-incidence zip code decreased for every one unit increase in human population density [adjusted odds ratio (aOR) 0.98, 95% confidence interval (CI) 0.97–0.99], increased for every one unit increase in deer density (aOR 1.27, 95% CI 1.08–1.50), and increased with every 1% increase in vacant housing (aOR 1.04, 95% CI 1.02–1.06) and deciduous forest cover (aOR 1.03, 95% CI 1.02–1.04) (Table 1).

Disease incidence, socioeconomic and environmental indicators all differed significantly across ecoregions (Table 2). The Ozark Highlands contained the largest number of high-incidence zip codes ( $n = 55$ ) and the highest cumulative incidence (52.5/100 000 persons). Both the Western Corn Belt and Mississippi Alluvial Plain had no high-incidence clusters and the lowest cumulative incidence (10.6 and 9.4/100 000 persons, respectively). The Ozark Highlands had the lowest population density (60.26/km<sup>2</sup>), while the Western Corn Belt and Interior River had the highest (244.55/km<sup>2</sup> and 659.24/km<sup>2</sup>, respectively). The Central Irregular Region had the highest proportion of residents aged  $>64$  years (15.75%), largest proportion of residents with at least a high-school education (42.95%), highest proportion of outdoor occupations (9.92%) and the greatest density of deer (1.20/km<sup>2</sup>).

Table 1. Socioeconomic and environmental factors associated with elevated incidence clusters

Variable*	LISA cluster		P value	OR (95% CI)	aOR (95% CI)
	No cluster	Cluster			
Population density/km <sup>2</sup>	197.17	10.77	<0.001	0.24 (0.14–0.40)	0.98 (0.97–0.99)
Gender, male	49.76	50.27	0.06	1.04 (0.98–1.10)	—
Age, >64 years	14.87	17.36	0.004	1.07 (1.02–1.11)	—
Race, white	92.57	97.28	<0.001	1.58 (1.05–1.27)	—
Vacant housing	13.08	21.85	<0.001	1.06 (1.04–1.08)	1.04 (1.02–1.06)
Occupation†	7.17	10.08	0.001	1.04 (1.02–1.07)	—
Education‡	39.87	43.10	0.08	1.04 (1.01–1.06)	—
Poverty§	13.93	16.18	0.001	1.03 (1.00–1.05)	—
Deciduous forest	27.32	46.95	<0.001	1.04 (1.03–1.05)	1.03 (1.02–1.04)
Deer density/km <sup>2</sup>	0.87	1.47	<0.001	1.40 (1.20–1.63)	1.27 (1.08–1.50)

LISA, Local indicator of spatial autocorrelation; aOR, adjusted odds ratio; CI, confidence interval.

\* Values are presented as proportions unless referring to density/km<sup>2</sup>.

† Proportion working in agriculture, forestry, mining, fishing, or hunting industries.

‡ Proportion aged >25 years with high school or equivalent-level education.

§ Proportion living under the Federal poverty line.

Table 2. EPA level III ecoregion comparison

Variable	EPA level III ecoregion					P value
	Ozark Highlands	Central Irregular	Western Corn Belt	Interior River	Mississippi Plains	
<b>Incidence</b>						
LISA clusters*	55	21	0	4	0	<0.001
EBS incidence†	4.0 × 10 <sup>-3</sup>	3.0 × 10 <sup>-3</sup>	0.1 × 10 <sup>-3</sup>	0.2 × 10 <sup>-3</sup>	0.01 × 10 <sup>-3</sup>	<0.001
Cumulative incidence‡	52.6	38.3	10.6	22.8	9.4	<0.001
<b>Socioeconomic</b>						
Population density/km <sup>2</sup>	60.26	115.81	244.55	659.24	69.15	<0.001
Age	15.17	15.75	14.22	13.89	15.42	<0.001
Race	96.36	94.00	92.65	83.20	90.26	<0.001
Vacant housing	16.28	13.37	10.01	12.31	10.18	<0.001
Occupation	7.13	9.92	6.58	3.31	8.24	<0.001
Poverty	15.20	13.29	10.68	12.05	22.02	<0.001
Education	39.91	42.95	41.54	34.97	37.63	<0.001
Gender	49.99	49.80	50.41	49.19	48.74	0.001
<b>Environmental</b>						
Deciduous forest	49.49	17.24	10.70	22.57	3.38	<0.001
Deer density/km <sup>2</sup>	0.91	1.20	0.55	0.94	0.23	<0.001

EPA, Environmental Protection Agency; LISA, local indicator of spatial autocorrelation; EBS, empirical Bayesian smoothing.

\* Frequencies compared using the Fisher's exact test.

† Empirical Bayesian smoothed incidence.

‡ Standardized cases/100 000 people.

Clear regional differences in the associations between incidence and social-ecological variables were further explained with the ecoregion-specific multivariate models (Table 3). In the Ozark Highlands, nine variables were significantly associated with

ehrlichiosis incidence in the bivariate analysis. After entering these in the multivariate model, two remained significant predictors. Ehrlichiosis incidence was negatively associated with population density (aOR 0.60, 95% CI 0.50–0.71) and positively

Table 3. *Ecoregion-specific logistic regression analysis*

Ecoregion model	Variable	OR (95% CI)	<i>P</i> value*	aOR (95% CI)	<i>P</i> value*
Ozarks Highlands	Population density/km <sup>2</sup>	0.27 (0.19–0.40)	<0.001	0.60 (0.50–0.71)	<0.001
	Gender	1.13 (1.02–1.26)	0.02	—	—
	Age	1.07 (1.02–1.12)	0.004	—	—
	Race	1.10 (1.03–1.18)	0.004	—	—
	Vacant housing	1.05 (1.03–1.08)	0.002	1.03 (1.00–1.05)	0.04
	Occupation	1.04 (1.01–1.08)	0.02	—	—
	Poverty	1.04 (1.01–1.06)	0.01	—	—
	Deciduous forest	1.03 (1.02–1.04)	<0.001	—	—
	Deer density/km <sup>2</sup>	1.49 (1.14–1.95)	0.004	—	—
Central Irregular	Population density/km <sup>2</sup>	0.27 (0.17–0.41)	<0.001	—	—
	Age	1.12 (1.10–1.18)	<0.001	—	—
	Race	1.10 (1.04–1.16)	0.001	1.06 (1.02–1.11)	0.004
	Vacant housing	1.15 (1.10–1.20)	<0.001	1.10 (1.05–1.16)	<0.001
	Occupation	1.11 (1.07–1.15)	<0.001	—	—
	Education	1.06 (1.04–1.09)	<0.001	—	—
	Poverty	1.06 (1.02–1.10)	0.002	1.06 (1.02–1.11)	0.008
	Deciduous forest	1.04 (1.02–1.07)	0.001	1.03 (1.00–1.06)	0.04
	Deer density/km <sup>2</sup>	1.40 (1.12–1.76)	0.004	—	—
Western Corn Belt	Race	1.33 (1.02–1.73)	0.04	1.44 (1.05–1.98)	0.02
	Poverty	0.81 (0.69–0.95)	0.01	0.77 (0.64–0.93)	0.005
Interior River Valley	Population density/km <sup>2</sup>	0.24 (0.15–0.38)	<0.001	—	—
	Gender	1.26 (1.08–1.47)	0.003	—	—
	Race	1.07 (1.03–1.12)	0.002	—	—
	Vacant housing	1.04 (1.01–1.08)	0.03	—	—
	Occupation	1.37 (1.21–1.55)	<0.001	1.25 (1.10–1.43)	0.009
	Education	1.11 (1.06–1.16)	<0.001	1.06 (1.03–1.10)	<0.001
	Deciduous forest	1.08 (1.05–1.11)	<0.001	2.13 (1.23–3.69)	0.007
	Deer density/km <sup>2</sup>	4.64 (2.51–8.57)	<0.001	—	—
Mississippi	Poverty	1.06 (1.00–1.14)	0.04	1.06 (1.00–1.14)	0.04

aOR, Adjusted odds ratio; CI, confidence interval.

\* Wald  $\chi^2$  statistic.

associated with the proportion of vacant housing units (aOR 1.03, 95% CI 1.00–1.05). In the Central Irregular Plains, there were 10 variables significantly associated with incidence in the bivariate analysis. When entered into a multivariate model, four variables remained significant. Elevated incidence was positively associated with white race (aOR 1.06, 95% CI 1.02–1.11), proportion of vacant housing units (aOR 1.10, 95% CI 1.05–1.16), proportion living in poverty (aOR 1.06, 95% CI 1.02–1.11), and proportion of deciduous forest cover (aOR 1.03, 95% CI 1.00–1.06).

### Evidence of spatial non-stationarity

The global OLS model revealed that incidence was negatively associated with human population density, and positively associated with vacant housing,

deciduous forest cover, and deer density (Table 4). Using these variables in a local GWR model resulted in improved overall model fit ( $\Delta$ AICc=8.4) and explained a larger proportion of the variance (36.1%) compared to the global OLS model (25.4%). While the local GWR model supported the associations between incidence and the independent variables, the association between these covariates showed considerable differences across the state, suggesting spatial non-stationarity in some of the most significant predictors of elevated ehrlichiosis incidence (Supplementary Fig. S1A).

The relationship between incidence and human population density was generally negative across the state; however, there were several strong positive associations in the northeast and parts of the Kansas City and St Louis metropolitan areas (Fig. 2a). The distribution of vacant housing units had a strong



Table 4. Global OLS and local GWR model summary

Variable	OLS model			GWR model†		
	$\beta$ (S.E.)*	<i>t</i> statistic	<i>P</i> value	$\beta_{\min}$	$\beta_{\text{median}}$	$\beta_{\max}$
Intercept	$1.7 \times 10^{-4}$ ( $0.27 \times 10^{-4}$ )	6.41	<0.001	$-1.68 \times 10^{-4}$	$2.14 \times 10^{-4}$	$6.88 \times 10^{-4}$
Population density/km <sup>2</sup>	$-0.3 \times 10^{-4}$ ( $0.05 \times 10^{-4}$ )	-6.06	<0.001	$-2.14 \times 10^{-4}$	$0.02 \times 10^{-4}$	$0.25 \times 10^{-4}$
Vacant	$0.06 \times 10^{-4}$ ( $0.01 \times 10^{-4}$ )	6.96	<0.001	$-0.08 \times 10^{-4}$	$0.02 \times 10^{-4}$	$0.25 \times 10^{-4}$
Deciduous forest	$0.04 \times 10^{-4}$ ( $0.00 \times 10^{-4}$ )	9.22	<0.001	$-0.11 \times 10^{-4}$	$0.02 \times 10^{-4}$	$0.15 \times 10^{-4}$
Deer density/km <sup>2</sup>	$0.2 \times 10^{-4}$ ( $0.08 \times 10^{-4}$ )	2.61	0.009	$-1.01 \times 10^{-4}$	$-0.01 \times 10^{-4}$	$0.92 \times 10^{-4}$
Model fit						
AICc	-1399.1				-1407.5	
Adjusted <i>R</i> <sup>2</sup>	0.254			0.002	0.361	0.607

OLS, Ordinary least squares; GWR, geographically weighted regression.

\* Beta coefficient (standard error).

† Beta coefficients presented as the range of values across all zip codes, with median value derived from the interquartile range.

positive relationship in the northwestern and northeastern portions of the state, as well as many parts of the Ozark Highlands (Fig. 2*b*). The proportion of deciduous forest cover was significantly positively associated with increased incidence in the Mississippi Alluvial Plains, the St Louis metropolitan region, and portions of the Ozark Highlands and Central Irregular Plains (Fig. 2*c*). Deer density was generally positively associated with incidence; however, there were multiple regions where this relationship was significantly inversed, including the south-central portion of the state and the Western Corn Belt (Fig. 2*d*).

## DISCUSSION

Understanding spatial heterogeneity in environmentally transmitted infectious diseases, particularly zoonoses, is inherently challenging due to the numerous spatial scales at which interactions between hosts, vectors, pathogens and their environment occur. Further complicating these efforts is that human socioeconomic factors and behaviours relevant to disease risk may also vary across gradients in human land-use [15, 34], requiring a synthesis of both ecological and social factors to fully understand the causes of disease. Here, we demonstrate how a spatial epidemiological approach can both refine estimates of disease distribution and offer an improved understanding of the factors associated with transmission dynamics of an emerging vector-borne disease within a highly endemic state. To our knowledge, this study is the first to demonstrate a subcounty-level analysis of the distribution of human ehrlichiosis in Missouri. Unlike

metrics of risk based on entomological modelling, the use of incidence data decidedly demonstrates human contact with pathogen-infected vectors. This has allowed for the exploration of associations of epidemiologically important social-ecological factors that constrain human tick exposure.

Spatially explicit statistical analysis revealed that there were non-random patterns in the distribution of cases, including significantly high clustering of cases in the Ozark Highlands and Central Irregular Plains ecoregions. Identification of high-incidence clusters could prove invaluable for public health officials tasked with implementing disease surveillance and control programmes. We also identified social-ecological variables associated with incidence, including a generally positive association with an increased proportion of forest cover, increased proportion of vacant housing units, an increased density of white-tailed deer, and a generally negative association with human population density. This information may be used by public health agencies to develop integrated vector-borne disease mitigation programmes to include both environmental (e.g. deer population management) and social (e.g. awareness campaigns in rural communities) interventions.

Incidence significantly differed across ecoregions, suggesting that biogeographical-scale ecological factors may determine where humans are at greatest risk of exposure to ehrlichiosis. There was a significantly high clustering of cases in the Ozark Highlands and Central Irregular Plains, and environmental features of these two ecoregions that may facilitate high entomological risk include forest type (primarily oak-hickory woodlands, compared to bottomland

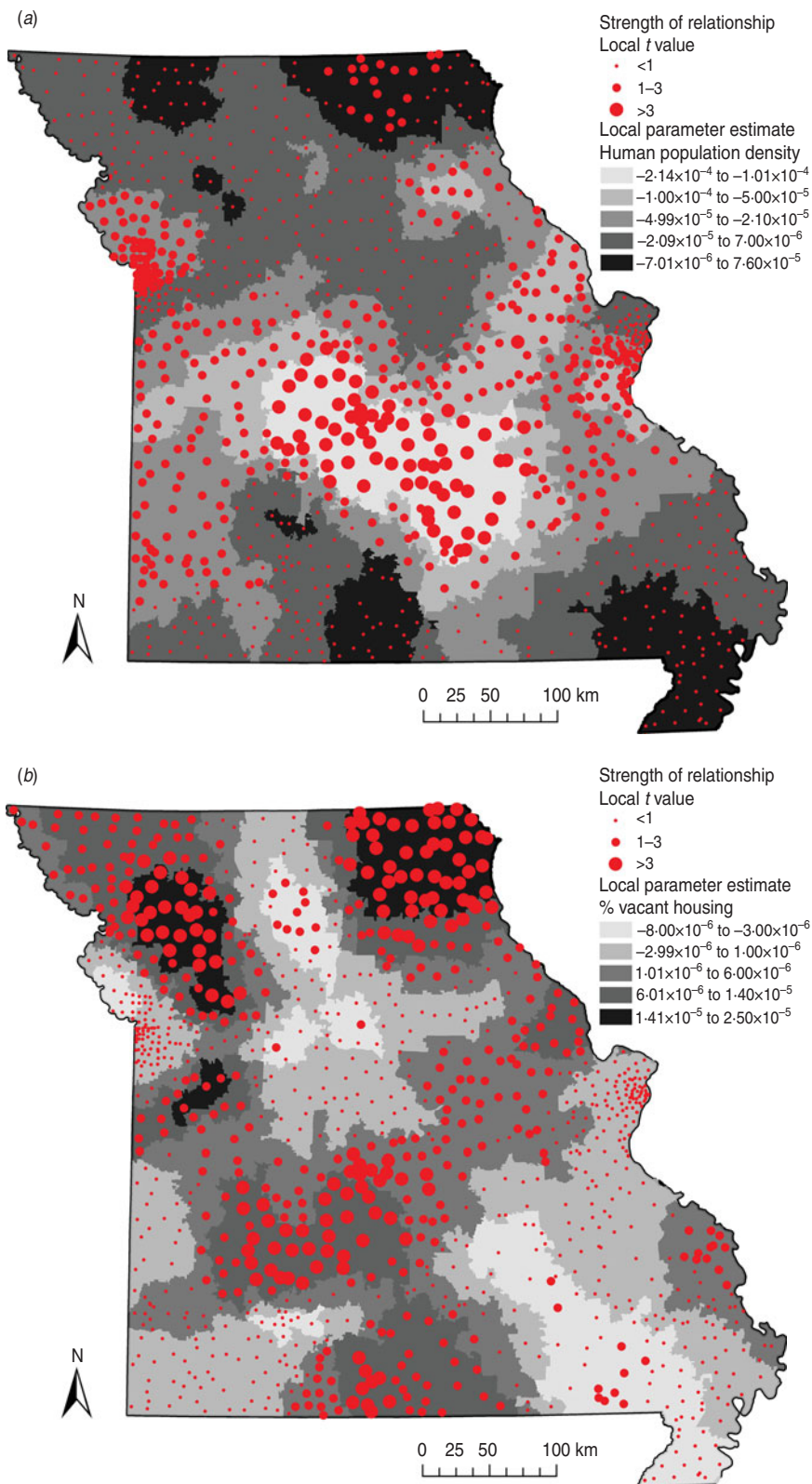
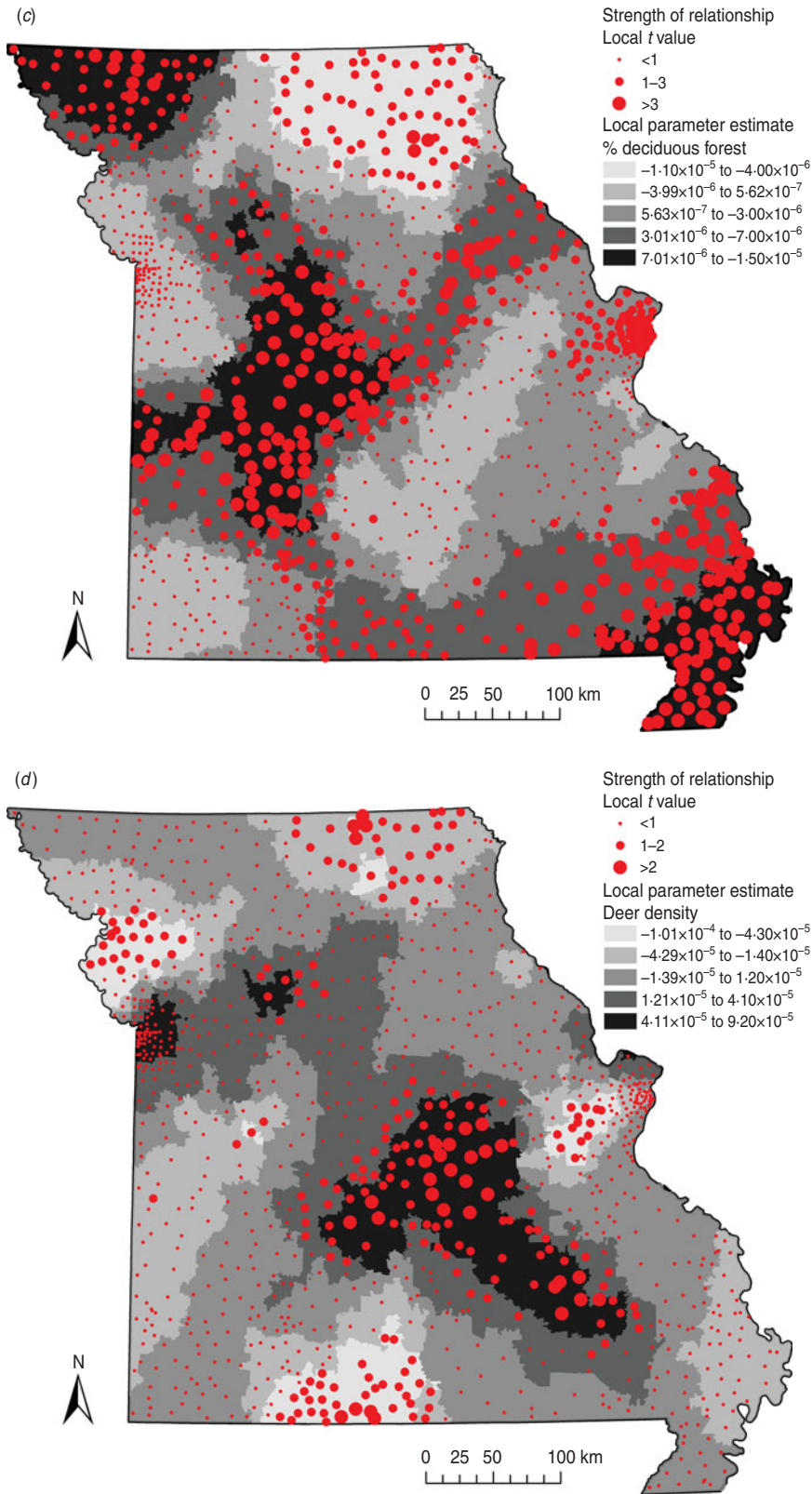


Fig. 2. For legend see next page



**Fig. 2.** [colour online]. Geographically weighted regression models predicting incidence of ehrlichiosis for the four variables selected from the global ordinary least squares analysis. Panels depict the local parameter beta coefficients (darker shades indicate a positive relationship while lighter shades indicate a negative relationship) and absolute value of the local  $t$  value (larger dots indicate a stronger statistical relationship) for: (a) human population density, (b) proportion of vacant housing units, (c) proportion of deciduous forest cover, and (d) white-tailed deer density.

deciduous forest, swamp forest and/or landscapes historically dominated by prairie grasslands converted into agricultural land in the other Missouri ecoregions), a relatively long frost-free period (140–230 and 165–235 days, respectively, potentially increasing over-winter survival for *A. americanum*) and occurring on a boundary between a humid continental climate and a humid subtropical climate [24]. Recent studies concerning *A. americanum* and other tick species have shown support for biogeographical-scale determinants of tick-borne disease risk and incidence [35]. A statewide survey in Tennessee determined that *E. chaffeensis*-infected *A. americanum* were only present in two of the seven ecoregions [36], while a field survey in southern Missouri found the presence and abundance of *A. americanum* to be positively associated with forest cover and increasing relative humidity during June [17]. A constellation of biogeographical factors may ultimately determine disease risk at these large spatial scales, but empirical support has been demonstrated for factors such as elevation, vapour pressure deficit, and forest area [11].

Results from the multivariate analysis predicting membership of a LISA cluster revealed that lower population density, higher deer density, and greater proportions of vacant houses and deciduous forest cover were associated with elevated incidence. These factors taken together may be indicative of forces that drive higher incidence in rural portions of the state, where a convergence of demographic and environmental characteristics across different spatial scales may be occurring. The overlap between these factors suggests that, within biogeographical regions that support deciduous forest growth and maintenance of large deer populations, the risk of disease transmission may be higher where population density is low and the proportion of vacant houses is high. However, we found further regional differences in many of the factors significantly associated with LISA clusters (Table 3), suggesting that the strength of these relationships vary with changes in biological and socioeconomic conditions. Identifying these clusters may provide public health officials with a new set of metrics to implement proactive, early detection and control initiatives.

A novel finding of this analysis was that an increased proportion of vacant housing units per zip code was consistently associated with elevated incidence of ehrlichiosis. This factor may serve as a proxy for anthropogenic influences contributing to focal points for high transmission risk in the

peridomestic environment. Previous studies on the socioeconomic risk factors associated with risk for other vector-borne diseases have suggested that poor maintenance associated with these housing characteristics may result in higher levels of established vegetation cover, resulting in micro-level vector propagation or host aggregation [37]. Therefore, vacant housing units may represent a new measure for identifying regions where socioeconomic conditions result in high-risk land-use practices. This finding may be used to direct future entomological sampling efforts and serve as a beacon for targeted, active surveillance efforts in areas that may now be considered focal points for high vector–human contact.

A negative correlation between tick-borne disease incidence and human population density has been observed previously for ehrlichiosis [38]. However, this association is often attributed to environmental factors such as increased host–vector habitat or density. Here, we control for the contributions of several social-ecological factors and still find a widespread negative association with human population density. This may be indicative of behavioural factors that associate with low population density, such as an increase in behaviours that contribute to risk of exposure. To wit, a survey of human behaviours spanning an urban–rural human land-use gradient in Missouri found respondents from rural areas significantly less likely to engage in tick-borne disease prevention behaviours [15]. Our results may therefore be of use to public health agencies tasked with promulgating disease prevention messages in areas we now show to be epidemiologically important.

Using global and locally varying regression methods, we provide a framework for identifying factors that contribute to elevated risk for disease in susceptible human populations spanning numerous biological and socioeconomic gradients. Multiple factors proved to be important in determining human incidence of ehrlichiosis in Missouri, and for several of these factors management would fall outside the traditional institutional boundaries of public health agencies (e.g. density of white-tailed deer). Therefore, our results may be of interest to multiple stakeholders with a vested interest in enhanced spatial understanding of the transmission dynamics of ehrlichiosis in this region. Based on these results, we propose an integrated approach to reducing ehrlichiosis emergence in Missouri by refining wildlife management programmes, improving vector and medical surveillance



efforts, and targeting health education campaigns to high-risk areas based upon the combination of factors that associate with high human incidence. Our findings may also be used to increase physician awareness in specific areas, which may improve rates of true positive diagnoses, ultimately resulting in more precise surveillance estimates and timelier administration of antibiotic therapies to those most at risk of serious disease complications (e.g. the elderly or immunocompromised) [13, 14].

It is important to note potential sources of bias in the data used for these analyses. For example, incidence data are subject to surveillance bias due to underreporting or misdiagnoses of cases, particularly for newly emerging diseases. A prospective study in rural southeastern Missouri suggested that cases may be significantly underreported in this region [39]. Further, geographical estimates of incidence are based on the residence of each individual, rather than knowledge of the location where disease transmission occurred. However, studies have suggested that transmission risk for tick-borne diseases may be highest in the peridomestic environment [40]. Finally, estimates of spatial variation in the abundance of *A. americanum* may have improved our model estimates of key environmental factors that drive ehrlichiosis incidence. However, as no statewide survey data currently exist, and tick abundance has been shown to correlate with broad-scale biogeographical factors in numerous recent studies [17, 18], we instead utilized available environmental data in place of estimates of tick abundance.

Our findings illustrate the value of synthesizing social-ecological factors to understand spatial variability in the incidence of a vector-borne, zoonotic disease. The manifestation of this disease in human populations is a product of both ecological factors that drive pathogen amplification within the enzootic transmission cycle, and the socioeconomic and demographic conditions that influence which human populations are at greatest risk of exposure. These results provide an increased knowledge base of the spatial distribution of ehrlichiosis in Missouri, and represent several opportunities for greatly improved surveillance and targeted health interventions.

## SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0950268813002951>.

## ACKNOWLEDGEMENTS

We thank Karen Yates (MDHSS) and the MDC for their assistance with access to and interpretation of the databases used for this research. Funding was provided by Environmental Protection Agency grant no. 834495 and by the North Central Integrated Pest Management Center of the United States Department of Agriculture.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. Childs JE, Paddock CD. The ascendancy of *Amblyomma americanum* as a vector of pathogens affecting humans in the United States. *Annual Review of Entomology* 2003; **48**: 307–337.
2. Centers for Disease Control and Prevention. Summary of notifiable diseases—United States, 2010. *Morbidity and Mortality Weekly Report* 2012; **59**: 1–111.
3. Dahlgren FS, et al. Increasing incidence of Ehrlichia chaffeensis and Anaplasma phagocytophilum in the United States, 2000–2007. *American Journal of Tropical Medicine and Hygiene* 2011; **85**: 124–131.
4. Paddock CD, Yabsley MJ. Ecological havoc, the rise of white-tailed deer, and the emergence of *Amblyomma americanum*-associated zoonoses in the United States. *Current Topics in Microbiology and Immunology* 2007; **315**: 289–324.
5. Killilea ME, et al. Spatial dynamics of lyme disease: a review. *Ecohealth* 2008; **5**: 167–195.
6. Wimberly MC, et al. Enhanced spatial models for predicting the geographic distributions of tick-borne pathogens. *International Journal of Health Geographics*. Published online: 15 April 2008. doi: 10.1186/1476-072X-7-15.
7. Mount GA, et al. New version of LSTSIM for computer simulation of *Amblyomma americanum* (Acari: Ixodidae) population dynamics. *Journal of Medical Entomology* 1993; **30**: 843–857.
8. Paddock CD, Childs JE. Ehrlichia chaffeensis: a prototypical emerging pathogen. *Clinical Microbiology Reviews* 2003; **16**: 37–64.
9. Allan BF, et al. Blood meal analysis to identify reservoir hosts for *Amblyomma americanum* ticks. *Emerging Infectious Diseases* 2010; **16**: 433–440.
10. Allan BF, et al. Invasive honeysuckle eradication reduces tick-borne disease risk by altering host dynamics. *Proceedings of the National Academy of Sciences USA* 2010; **107**: 18 523–18 527.
11. Diuk-Wasser MA, et al. Human risk of infection with *Borrelia burgdorferi*, the Lyme disease agent, in eastern United States. *American Journal of Tropical Medicine and Hygiene* 2012; **86**: 320–327.



12. **Standaert SM, et al.** Ehrlichiosis in a golf-oriented retirement community. *New England Journal of Medicine* 1995; **333**: 420–425.
13. **Dumler JS, et al.** Ehrlichioses in humans: epidemiology, clinical presentation, diagnosis, and treatment. *Clinical Infectious Diseases* 2007; **45** (Suppl. 1): S45–S51.
14. **Demma LJ, et al.** Epidemiology of human ehrlichiosis and anaplasmosis in the United States, 2001–2002. *American Journal of Tropical Medicine and Hygiene* 2005; **73**: 400–4099.
15. **Bayles BR, Evans G, Allan BF.** Knowledge and prevention of tick-borne diseases vary across an urban-to-rural human land-use gradient. *Ticks and Tick-borne Diseases* 2013; **4**: 352–358.
16. **Ostfeld RS, Glass GE, Keesing F.** Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology and Evolution* 2005; **20**: 328–336.
17. **Brown HE, et al.** An acarologic survey and *Amblyomma americanum* distribution map with implications for tularemia risk in Missouri. *American Journal of Tropical Medicine and Hygiene* 2011; **84**: 411–419.
18. **Yabsley MJ, et al.** Spatial analysis of the distribution of *Ehrlichia chaffeensis*, causative agent of human monocytotropic ehrlichiosis, across a multi-state region. *American Journal of Tropical Medicine and Hygiene* 2005; **72**: 840–850.
19. **McQuiston JH, et al.** The human ehrlichioses in the United States. *Emerging Infectious Diseases* 1999; **5**: 635–642.
20. **Centers for Disease Control and Prevention.** Changes in the case definition for human ehrlichiosis, and addition of a new ehrlichiosis category as a condition placed under surveillance according to the National Public Health Surveillance System (<http://www.cste.org/ps/2000/2000-id-03.htm>). Accessed 7 August 2012.
21. **Council of State and Territorial Epidemiologists.** Revision of the national surveillance case definition for ehrlichiosis (ehrlichiosis/anaplasmosis). (<http://www.cste.org/ps/2007ps/2007psfinal/id/-7-id-03.pdf>). Accessed 7 August 2012.
22. **US Census Bureau.** American Fact Finder. 2000 (<http://factfinder.census.gov/>) Accessed 3 July 2012.
23. **Homer CG, et al.** Development of a 2001 National Landcover Database for the United States. *Photogrammetric Engineering & Remote Sensing* 2007; **73**: 829–840.
24. **Ricketts TH, et al.** *Terrestrial Ecoregions of North America. A Conservation Assessment.* USA: Island Press, 1999.
25. **Rushton G, Lolonis P.** Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine* 1996; **15**: 717–726.
26. **Moran PA.** Notes on continuous stochastic phenomena. *Biometrika* 1950; **37**: 17–23.
27. **Anselin L.** Local indicators of spatial association-LISA. *Geographical Analysis* 1995; **27**: 93–116.
28. **Anselin L, Syabri I, Kho Y.** GeoDa: an introduction to spatial data analysis. *Geographical Analysis* 2006; **38**: 5–22.
29. **Legendre P.** Spatial autocorrelation: Trouble or new paradigm? *Ecology* 1993; **74**: 1659–1673.
30. **Brunsdon C, Fotheringham AS, Charlton ME.** Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis* 1996; **28**: 281–298.
31. **Glattig G, et al.** Choosing the optimal fit function: comparison of the Akaike information criterion and the F-test. *Medical Physics* 2007; **34**: 4285–4292.
32. **Fotheringham AS, Brunsdon C, Charlton M.** *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* England: Wiley and Sons, 2002.
33. **Fotheringham AS, Charlton ME, Brunsdon C.** Spatial variations in school performance: a local analysis using geographically weighted regression. *Geographical and Environmental Modelling* 2001; **5**: 43–66.
34. **Killilea ME, et al.** Spatial dynamics of Lyme disease: a review. *EcoHealth* 2008; **5**: 167–195.
35. **Cohen SB, et al.** Host associations of *Dermacentor*, *Amblyomma*, and *Ixodes* (Acari: Ixodidae) ticks in Tennessee. *Journal of Medical Entomology* 2010; **47**: 415–420.
36. **Cohen SB, et al.** Prevalence of *Ehrlichia chaffeensis* and *Ehrlichia ewingii* in ticks from Tennessee. *Vector-Borne and Zoonotic Diseases* 2010; **10**: 435–440.
37. **Reisen WK, et al.** Delinquent mortgages, neglected swimming pools, and West Nile virus, California. *Emerging Infectious Diseases* 2008; **14**: 1747–1749.
38. **Cromley EK, et al.** Residential setting as a risk factor for Lyme disease in a hyperendemic region. *American Journal of Epidemiology* 1998; **147**: 472–477.
39. **Olano JP, et al.** Human monocytotropic ehrlichiosis, Missouri. *Emerging Infectious Diseases* 2003; **9**: 1579–1586.
40. **Connally NP, et al.** Peridomestic Lyme disease prevention: results of a population-based case-control study. *American Journal of Preventive Medicine* 2009; **37**: 201–206.