A Spatial Model of Shared Risk for Plague and Hantavirus Pulmonary Syndrome in the Southwestern United States

Rebecca J. Eisen,* Gregory E. Glass, Lars Eisen, James Cheek, Russell E. Enscore, Paul Ettestad, and Kenneth L. Gage

Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne, Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado; The W. Harry Feinstone Department of Molecular Microbiology and Immunology, The Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado; Division of Epidemiology and Disease Prevention, Indian Health Services, Albuquerque, New Mexico; Zoonoses Program, New Mexico Department of Health, Santa Fe, New Mexico

Abstract. Plague and hantavirus pulmonary syndrome (HPS) are severe, often fatal diseases in humans that share a broad epidemiologic focus in the southwestern United States. Prevention of these diseases relies heavily on education and reducing rodent abundance in peridomestic environments. Resources for these activities are limited. Therefore, identifying areas sharing elevated risk for these two relatively rare but severe diseases could be useful for targeting limited public health resources. Using logistic regression and geographic information system–based modeling, we identified environmental predictors of elevated risk for plague (distance to piñon-juniper ecotones and amount of precipitation) and HPS (elevation and amount of precipitation) in northeastern Arizona and northwestern New Mexico. Our models accurately identified case locations as suitable (producer accuracies of 93% for plague and 96% for HPS) and indicated that approximately half of the coverage area was classified as suitable risk for either plague or HPS. The probability of a site being classified as suitable for plague was strongly correlated with its probability of being classified as suitable for HPS ($p_s = 0.88$). Increased risk for both diseases occurred for approximately 37% of the coverage area.

INTRODUCTION

Plague and hantavirus pulmonary syndrome (HPS) are severe and often fatal diseases in humans that share a broad epidemiologic focus in the southwestern United States.1–7 Plague, caused by the gram-negative bacterium Yersinia pestis, is transmitted to humans via infectious flea bites. Exposure can also occur by contact with infected animals or, much less commonly, by exposure to humans or cats with plague pneumonia and cough.2,8–11 In the southwestern United States, HPS is caused by Sin Nombre virus (SNV) (Bunyaviridae, Hantavirus).6,12 Deer mice (Peromyscus maniculatus) serve as the primary zoonotic reservoir of SNV.5,13 Viral infections in mice are chronic and persistent, and transmission may occur via direct contact (e.g., fighting).6,13–15 Transmission to humans is thought to most often occur through contact with infected mouse excreta and risk is positively associated with rodent density.5,17–19

Mortality rates for untreated plague range from 50% to 60% for the bubonic (flea-borne) form of the disease to nearly 100% for septicemic or pneumonic infection.4,19 For HPS, 35% of reported cases from 1993 through 2007 have been fatal (http://www.cdc.gov/ncidod/dises/hanta/hps/noframes/caseinfo.htm). When infections occur, early diagnosis followed by appropriate clinical treatment or supportive care improve outcomes for persons infected with either agent.1–5,20,22 The risk of both plague and HPS can be reduced, primarily through implementation of rodent sanitation measures, although flea control measures are also important for plague.1–5,9,15–21,23–26 Fine-scale identification of areas posing elevated risk of exposure to both disease agents followed by education campaigns aimed at raising awareness among health care providers and the public of these specific risk areas could be useful for prevention or improved disease management. Because the low frequency of occurrence of plague and HPS often translates to limited prevention and public health education funds being diverted to more common but less severe public health threats, improved recognition of localities at risk for these rare but severe rodent-associated diseases may help to target available funds to those populations most at risk. The objectives of this study were to 1) create fine-scale spatial risk models for plague and for HPS in an epidemiologic focus in northeastern Arizona and northwestern New Mexico and 2) identify areas that pose high risk of exposure to both pathogens.

MATERIALS AND METHODS

Study population and coverage area. The study area consisted of tribal or privately owned land in northeastern Arizona and northwestern New Mexico (Figure 1). Between November 1992 and November 1994, 27 confirmed cases of HPS occurred in this area. As described previously, sites of exposure were determined during case investigations, or retrospectively, and global positioning system coordinates for each exposure location were entered into a geographic information system (GIS).6 This data set represents exposure locations available from the initial HPS outbreak in the southwestern United States. Similarly, exposure sites identified during field investigations of 58 confirmed human plague cases occurring within the study area from 1965 through 2000 were incorporated into the GIS, as described elsewhere.27,28 When selecting control points, we aimed to include controls from within the same socioeconomic and geographic region as the plague and HPS cases.6 Because of the entitlement to free health care under the Indian Health Service system, virtually all individuals have a health record on file. The study area is divided into 10 service units to facilitate care delivery, with each service unit having an anchor hospital facility. A centralized

* Address correspondence to Rebecca J. Eisen, Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, 3150 Rampart Road, Fort Collins, CO 80522. E-mail: dyn2@cdc.gov

Copyright © 2007 by The American Society of Tropical Medicine and Hygiene


* Address correspondence to Rebecca J. Eisen, Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, 3150 Rampart Road, Fort Collins, CO 80522. E-mail: dyn2@cdc.gov
electronic data base is maintained for all beneficiaries. Case controls (n = 116 for plague and 54 for HPS) were randomly selected by service unit, with the number of controls chosen from each service unit on the basis of the service unit’s proportion of the total population rounded to the nearest whole number.

**Predictive landscape features.** Elevation was determined on the basis of a 1-km resolution digital elevation model (U.S. Geological Survey, Reston, VA). Thirty-year mean climate data (mean monthly and annual minimum, mean, and maximum temperature; mean monthly and annual cooling, heating degree days and growing degree days; mean monthly and annual precipitation) were derived from GIS-based data for 1961 through 1990 (2 × 2 km spatial resolution; Climate Source LLC, Corvallis, OR) using ArcGIS version 9.0 (Environmental Systems Research Institute, Redlands, CA). A 33-year average, rather than any single year, was used to reduce the likelihood of selecting a year with atypical weather patterns. On the basis of vegetation data from the southwest regional gap analysis,29 we created a grid layer (30 × 30 meter resolution) depicting the ecotone of intermountain basin juniper savannah and Colorado piñon juniper woodland. To accomplish this, we created 1-km buffers around each habitat type; the ecotone layer represents the intersection of these. We then generated 30 × 30 m grid layers representing the minimum Euclidean distance to any of these habitats, or minimum Euclidean distance to the ecotones. All layers, including the case and control points, were projected to North American Datum 1983 Albers.

**Construction and selection of landscape models.** Logistic regression models were constructed to quantify the association between the probability that an area is a suitable risk habitat for plague or HPS and landscape features. Candidate models were constructed separately for plague and HPS (Tables 1 and 2). The models are described by the equation

\[
\text{Logit}(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k
\]

where \(P\) is the probability of a cell being classified as suitable plague or HPS habitat, and \(\beta_0\) is the intercept. The values \(\beta_1, \ldots, \beta_k\) represent the coefficients assigned to each independent variable included in the regression and \(x_1, \ldots, x_k\) symbolize the independent variables. The probability that a particular cell in the GIS is classified as a suitable habitat can be derived from eqn 1 using the expression

\[
P = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}
\]

To select the most parsimonious model with the best predictive power, Akaike’s Information Criterion (AIC)30 was used to rank each of the models. The one with the lowest AIC value was selected as the best. However, models within two AIC units of the minimum AIC (\(\Delta\text{AIC} < 2\)) are considered competing with substantial support.31 To determine if our models fit the data, we used a goodness of fit test. This analysis compares a pure-error negative log-likelihood with the fitted model log-likelihood. If the chi-square test result is not significant, it supports the conclusion that sufficient data were included in the model. The spatial relationships of the residuals from the two models were examined for autocorrelation. Under the assumption of an appropriate model specification, we expected the residuals to be spatially independent. We used geoR (The R-Development Core Team, http://www.r-project.org) to estimate the semivariogram of the residuals.

Receiver operating characteristic (ROC) curves were used to assess the overall discrimination ability of each model on the basis of the area under the ROC curve (AUC), and to determine the optimal probability cut-off for characterization of habitat suitability. An ROC curve plots all true positive fractions (sensitivity values) obtained from the model build set on the vertical axis against their corresponding equivalent false-positive fraction values (1 — specificity) for all available thresholds on the horizontal axis. The AUC provides a threshold-independent measure of the overall accuracy of the model. This value ranges from 0.5 to 1, where a value of 1 indicates that all points in the build set were correctly classified by the model.32 The logit equation can be transformed into a probability following equation 2. However, to evaluate the performance of the model using ROC, this probability

**Table 1**

<table>
<thead>
<tr>
<th>Model no.</th>
<th>(k)*</th>
<th>Negative log-likelihood</th>
<th>AIC</th>
<th>AIC value</th>
<th>(\Delta\text{AIC})</th>
<th>Goodness of fit (P) value</th>
<th>AUC§</th>
<th>Independent model variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>87.34</td>
<td>93.34</td>
<td>0</td>
<td>0.28</td>
<td>0.80</td>
<td>Dist-PJEco, Ppt_ann</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>84.48</td>
<td>94.48</td>
<td>1.14</td>
<td>0.42</td>
<td>0.81</td>
<td>EL, EL², Dist-RMPP, Ppt-Feb</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>87.02</td>
<td>97.02</td>
<td>3.68</td>
<td>0.33</td>
<td>0.80</td>
<td>Dist2PJeco, Ppt_ann, EL, EL²</td>
<td></td>
</tr>
</tbody>
</table>

* EL = elevation; EL² = elevation squared; Dist-RMPP = distance to Rocky Mountain Ponderosa Pine; Ppt-Feb = average precipitation in February 1961–1990; Dist-PJEco = distance to the ecotone of intermountain basin juniper savannah and Colorado pinyon juniper woodland; Ppt_ann = average annual precipitation 1961–1990.

* \(k\) = number of estimated parameters included in the model.

* AIC = Akaike information criterion value; \(\Delta\text{AIC}\) = AIC model – AIC of best model.

* AUC = area under receiver operating characteristic curve.
must be converted to a binary value (e.g., a grid cell is considered suitable or it is not). The optimal probability cut-off value was chosen by maximizing sensitivity and specificity simultaneously. All cells with a probability value at least equal to the optimal value were classified as high risk for plague or HPS. All others were considered low risk in our evaluation matrix.

To test the correlation between areas considered suitable for plague and HPS, we generated a random sample of 500 points contained within the coverage area. For each point, the probability of suitability for plague and HPS based on equation 2 were extracted. The correlation between these probabilities was tested using Spearman correlations.

RESULTS

Environmental predictors of plague and HPS risk. Among the candidate models (Table 1), suitable plague habitat was best predicted by distance to the ecotone of intermountain basin juniper savannah and Colorado piñon juniper woodland ($-5.8 \times 10^{-5} \text{ m}^{-1} \pm 3.39 \times 10^{-5}, \beta \pm \text{SE}$), and 30-year mean annual precipitation ($0.0239 \text{ mm}^{-1} + 5.15 \times 10^{-5}, \text{ intercept} -7.08 \pm 1.511$). Using the optimal cut-off value obtained from the ROC curve, 0.2579, and reclassifying the risk model, we determined that approximately 46% of the coverage area represents locations suitable for plague risk (Figure 2). Although model 2 was considered competing, producer accuracy was lower for this model than for the selected model. The most parsimonious HPS model was based on a non-linear relationship with elevation (Table 2). However, two models were considered competing with $\Delta \text{AIC}$ values < 2. Because user and producer accuracies were better and the predictive variables were consistent with previously published studies, model 3 (Table 2) was selected as the best. This model included elevation ($0.069 \text{ m}^{-1} \pm 0.055$), elevation squared ($-1.56 \times 10^{-5} \text{ m}^{-2} \pm 1.35 \times 10^{-5}$), and 30-year mean precipitation in August ($0.011 \pm 0.035$, intercept $-76.30 \pm 56.30$). The model indicated that suitability for HPS increased up to an elevation of 2,205 meters and decreased as elevation increased thereafter. On the basis of reclassification of the model using a probability of 0.2586, approximately 43% of the area was classified as suitable HPS risk (Figure 2). Risk for plague and HPS were positively associated (Spearman’s correlation $r_s = 0.88, P < 0.0001$). Approximately 37% of the coverage area had elevated risk for both HPS and plague (Figure 2). Risk for plague and HPS were positively associated (Spearman’s correlation $r_s = 0.88, P < 0.0001$). Approximately 37% of the coverage area had elevated risk for both HPS and plague (Figure 2). Risk for plague and HPS were positively associated (Spearman’s correlation $r_s = 0.88, P < 0.0001$). Approximately 37% of the coverage area had elevated risk for both HPS and plague (Figure 2). Risk for plague and HPS were positively associated (Spearman’s correlation $r_s = 0.88, P < 0.0001$).

Evaluation of plague and HPS models. Comparisons of model predictions and actual cases and controls are shown in

![Figure 2](https://www.ajtmh.org).
Tables 3 and 4. For plague, producer accuracies for categorizing case locations as suitable or control locations as less suitable were 93.10% and 56.89%, respectively. In contrast, user accuracy for observing a plague case in locations considered by the model as suitable was 51.92%. Locations classified as unsuitable contained controls, rather than cases, 94.29% of the time.

For HPS, producer accuracies for cases and controls were 96.30% and 42.59%, respectively. This indicates that 96.30% of HPS case locations were classified as highly suitable whereas 42.59% of controls were classified as less suitable. For locations classified by the model as highly suitable, 45.60% contained HPS cases and for locations classified as less suitable, 95.83% contained controls. Because HPS and plague are uncommon diseases, it is not surprising to see such a high proportion of controls within areas considered highly suitable.27,28,32

DISCUSSION

Occurrences of HPS and plague in humans are reported to Centers for Disease Control and Prevention by state health agencies. Such reporting has shown that the major epidemiologic foci for both diseases is in the southwestern United States,1,2,4,6 but it was not clear to what extent fine-scale spatial patterns of high risk for plague and HPS overlapped. Our models indicated that within the regional focus, approximately half the coverage area posed an elevated risk for either one of the diseases and one-third of the coverage area was considered a suitable risk habitat for both plague and HPS. Furthermore, there was a strong positive correlation between areas with elevated risk of exposure to SNV and those at risk for Y. pestis in northeastern Arizona and northwestern New Mexico. This is significant because it suggests that limited public health resources can be shared to implement risk reduction measures that can be effective for both diseases.

Factors predictive of elevated risk for plague (e.g., distance to piñon-juniper ecotones and amount of precipitation) and HPS (e.g., elevation and amount of precipitation) within our study area are consistent with previously published models.6,27,28,34,35 It was proposed previously that human plague cases are closely associated with piñon-juniper habitat1,3,27,28 and that above average precipitation during the winter and

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Evaluation matrix for the most parsimonious model for human plague cases*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model classification</td>
<td>Actual classification</td>
</tr>
<tr>
<td>Highly suitable</td>
<td>Human plague case</td>
</tr>
<tr>
<td>54</td>
<td>50</td>
</tr>
<tr>
<td>Less suitable</td>
<td>4</td>
</tr>
<tr>
<td>% correct</td>
<td>93.10</td>
</tr>
</tbody>
</table>

* Logit(P) = (0.24 × Ppt_ann) - (5.8 × 10^-1 × Dis × PJEco) - 7.08. For definitions of abbreviations, see Table 1.
† Probability cut-off value used to classify a 30-meter raster as suitable was based on the receiver operating characteristic optimal cut-off probability (P = 0.2579).
‡ User accuracy (commission error).
§ Producer accuracy (omission error).

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Evaluation matrix for the best competing model for hantavirus pulmonary syndrome cases*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model classification</td>
<td>Actual classification</td>
</tr>
<tr>
<td>Human plague case</td>
<td>26</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
</tr>
<tr>
<td>% correct</td>
<td>96.30</td>
</tr>
</tbody>
</table>

* Logit(P) = (0.0688 × EL) - (1.56 × 10^-7 × EL^2) + (0.011 × Ppt_Aug) - 76.295. For definitions of abbreviations, see Table 1.
† Probability cut-off value used to classify a 30-meter raster as suitable was based on the receiver operating characteristic optimal cut-off probability (P = 0.2052).
‡ User accuracy (commission error).
§ Producer accuracy (omission error).
spring cases.

36,37 Our model indicated that the probability of an area being classified as a suitable plague risk was higher in close proximity to the ecotone of intermountain basin juniper savannah and Colorado piñon juniper woodland and areas with higher annual precipitation were more likely to be classified as suitable than relatively drier areas. Likewise for HPS, previous models have shown that elevation is a good predictor of risk. 6,34 Most likely because human risk of exposure to SNV is related to deer mouse abundance, 17 and abundance of infected deer mice has been correlated with elevation. 15 Mills and others noted that sites they sampled at the extremes of their elevation gradient often lacked P. maniculatus infected with SNV. 15 Our model predicted a non-linearity in the relationship between HPS risk and elevation. Risk increased up to 2,205 meters and decreased as elevation increased thereafter. This differs from a previous analysis, 6 which indicated risk at higher elevations, but most likely reflects differences in methodologic approaches. The previous analyses simply dichotomized elevation at the mid-point of the elevation ranges of the data sample rather than testing for potential nonlinearities in the relationships. More recent work in this geographic region concluded that refugia sites (i.e., sites persistently yielding SNV-infected deer mice) may exist above 2,130 meters but below 3,200 meters. 34,38

Clearly our study did not address why these factors emerged as predictors. Others have suggested that following the trophic cascade hypothesis, increased precipitation (a predictor in both models) leads to increased soil moisture, which triggers increased ecosystem productivity (vegetative growth). The increased vegetative growth provides food resources for rodents, which results in a population increase and greater transmission potential for Y. pestis and SNV. 36–40 Alternatively, the association between plague risk and precipitation could be related to moisture requirements for vector flea survival, especially in off-host environments such as rodent burrows. 36,37,41

For both models, producer accuracy for classifying locations where HPS or plague cases occurred as suitable risk was high (96.3% and 93.1%, respectively). Rarely were HPS (3.7%) or plague (6.9%) cases exposed in locations classified as less suitable. When an area was considered by the models to be less suitable, approximately 95% of the time controls, rather than cases, were present. Plague and HPS are diseases of low frequency in humans and exposure is at least partly dependent on behavioral factors (e.g., actions taken to reduce rodent abundance around homes, pet care and handling, exposure to sick or dead animals 5,17,21,42,43) that cannot be identified by a GIS. If proper prevention guidelines are implemented, we would expect to see a high proportion of uninfected controls living within areas considered by the model to represent suitable risk. Indeed, 48% of controls were located in areas considered suitable risk for plague and 55% within HPS risk areas. Similarly, high risk behavior (e.g., handling sick or dead animals, entering closed areas that are heavily infested with rodents without personal protective equipment) within sites where risk is considered low, could result in exposure to SNV or Y. pestis.

For plague and HPS, reduction of rodent abundance in peridomestic environments and avoidance of handling sick or dead animals are important prevention strategies. To reduce rodent abundance, harborage (e.g., piles of wood, brush, or debris), and food sources (e.g., pet food, garbage) for wild rodents should be removed. 5,19,21 In situations where rodent infestations are severe, rodents should be removed from homes, decontamination protocols should be followed, and these structures should be rodent-proofed following previously published guidelines. 5 Removal of rodents in the absence of rodent-proofing appears to be ineffective. To reduce the risk of human exposure to fleas potentially infected with the plague bacterium, use of insect repellent containing N,N-diethyl-m-toluamide (DEET) is recommended when handling dead rodents and rodent traps. 9 If rodenticides are used, fleas should be removed using insecticides prior to rodent eradication. Furthermore, to reduce plague risk, pets should be kept free of fleas. 19,21

Received June 20, 2007. Accepted for publication July 31, 2007.

Acknowledgments: We thank James N. Mills for helpful comments on the manuscript, the Diagnostics and Reference Laboratory (Bacterial Diseases Branch, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, Fort Collins, CO) for technical support, and the Special Pathogens Branch (Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA) for hantavirus serologic testing.

Authors’ addresses: Rebecca J. Eisen, Russell E. Enscore, and Kenneth L. Gage, Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne, Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, 3150 Rampart Road, Fort Collins, CO 80522. James Cheek, Division of Epidemiology and Disease Prevention, Indian Health Services, 5300 Homestead Rd., NE, Albuquerque, NM 87110. Paul Ettestad, Zoornoses Program, New Mexico Department of Health, Santa Fe, NM 87505.

REFERENCES


