Residence-Linked Human Plague in New Mexico:
A Habitat-Suitability Model


Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, P.O. Box 2087, Fort Collins, Colorado 80522; Zoonoses Program, New Mexico Department of Health, Santa Fe, New Mexico; Vector Control Program, New Mexico Environment Department, Santa Fe, New Mexico; Division of Epidemiology and Disease Prevention, Indian Health Services, 5390 Homestead Rd., NE, Albuquerque, New Mexico; City of Albuquerque Division of Environmental Health, Albuquerque, New Mexico

Abstract. Yersinia pestis, the causative agent of plague, has been detected in fleas and mammals throughout the western United States. This highly virulent infection is rare in humans, surveillance of the disease is expensive, and it often was assumed that risk of exposure to Y. pestis is high in most of the western United States. For these reasons, some local health departments in these plague-affected regions have hesitated to undertake surveillance and other prevention activities. To aid in targeting limited public health resources, we created a fine-resolution human plague risk map for New Mexico, the state reporting more than half the human cases in the United States. Our GIS-based model included three landscape features—a nonlinear relationship with elevation, distance to water, and distance to the ecotone between Rocky Mountain/Great Basin open and closed coniferous woodlands—and yielded an overall accuracy of \( \approx 80\% \). The model classified 17.25\% of the state as posing significant risk of exposure to humans on privately or tribally owned land, which suggests that resource requirements for regular surveillance and control of plague could be effectively focused on < 20\% of the state.

INTRODUCTION

Plague (Yersinia pestis infection) is a highly virulent disease of low frequency in humans. Humans most often become infected through the bites of infectious fleas, but infections also can be acquired through contact with infected animals or, much more rarely, through exposure to humans or cats with plague pneumonia and cough.\(^1\) Outcome of infection is improved by early diagnosis followed by appropriate antibiotic treatment.\(^2\) Raising awareness among health care providers, veterinary staff, and the public of areas posing high risk of exposure to Y. pestis could be useful in prevention and control of the disease. Within the United States, areas of risk are typically defined by identifying the numbers of human cases by year and by state and county of exposure.\(^3\)–\(^6\) It has been proposed that human cases are associated with epizootic activity, and, in the southwestern United States, epizootic activity is related to habitat type.\(^7\) However, within the western states where plague occurs, counties are typically large and ecologically highly diverse. If human plague risk is linked to environmental factors, then it is likely that risk of exposure to Y. pestis can range from minimal to high within a single county. Therefore, identifying high-risk areas at a fine geographical resolution would be useful for targeting limited public health resources toward those sites at greatest risk.

Although human cases have been identified from many western states, more than half of all U.S. cases were reported from New Mexico, and a high proportion were linked with residential exposure.\(^2\)–\(^5\)\(^,\)\(^6\)–\(^8\)\(^,\)\(^9\) Our study aimed to identify the landscape and ecological features associated with human risk of exposure to Y. pestis within residence-linked environments (e.g., within 2 km of a home site) and to create a fine-resolution risk map for the state of New Mexico.

MATERIALS AND METHODS

Construction of the model build set. Case points. From 1960 to 2003 in New Mexico, 224 human cases of plague were reported to the U.S. Centers for Disease Control and Prevention. Among these, 186 were determined to be exposed within 2 km of a home site and were considered “residence-linked.” Determination of residence-linked exposure was based on travel history, potential exposure (e.g., handling of dead animals), and on-site investigations aiming to identify epizootic activity near the home site. The 2-km exposure limit around case home sites (case patient’s or other) allows inclusion of exposure sites found on the actual case-patient’s property or within reasonable walking distance of the home. It excludes locations that were visited after transportation in vehicles, including those cases where exposure occurred during recreational activities (e.g., hunting) far from areas inhabited by humans. On the basis of case investigations in the field conducted at the time of illness, the latitude and longitude of probable sites of exposure for each patient were recorded, and these points were imported into a geographic information system (GIS). In our habitat-suitability analysis, case points represent geographic locations where epizootics are known to have occurred and are therefore considered high risk or highly suitable for epizootics.

Control points. We assume that, if a person in New Mexico was infected with Y. pestis, the etiological agent of plague, then the case was reported and included in the dataset; all other individuals in the population are assumed to be uninfected. Plague produces a severe clinical manifestation almost always requiring hospitalization. Mild or asymptomatic cases do not typically occur. Plague has been an internationally reportable disease for many years, and it is deemed highly unlikely that CDC’s dataset misses many cases.\(^8\) Because we are modeling human cases of plague in relation to environmental characteristics, the random control points represent a random sampling of the population rather than of the landscape. To generate this random sample of 186 points, we cre-
ated a raster layer based on 1990 Census block group data (people per square mile), where each 30-m cell within the New Mexico state boundary represents population density in 1990. Population density associated with the case points ranged from 0 to 1,904.5 people per square mile. We drew our random control sample from a similar population range (0.01 to 1,904.5 people per square mile), thus avoiding any bias that would result from sampling more densely populated urban areas. To avoid selecting control points from within the same grid cell, the minimum distance between control points was 1 km. It is unlikely that two or more of our 186 control points would co-occur in any one 30 m × 30 m cell with population density within the specified population density range because this represents a very large geographical area (≈ 303,250 km²). Nonetheless, we applied this restriction because, as is typical of habitat suitability modeling, the confidence in our ability to equate lack of human cases with unsuitability for epizootic activity was lower than for associating case locations with a high-risk classification; therefore, we avoided including any control location more than once in our regression analyses.

**Predictive landscape features.** Habitat type was determined based on classification of the New Mexico Gap analysis. Elevation was determined based on digital elevation models. Based on preliminary analysis, three habitat types were believed to be associated with human plague cases: Rocky Mountain lower montane conifer forests, Rocky Mountain/Great Basin closed conifer woodlands, and Rocky Mountain/Great Basin open conifer woodlands. We generated 30-m grid layers representing the minimum Euclidean distance to any of these habitats, or minimum Euclidean distance to the latter two habitats, or the minimum Euclidean distance to the ecotones representing the convergence of each of these habitats. Our ecotone layer was generated by creating 0.25-km buffers around Rocky Mountain/Great Basin closed conifer woodlands and Rocky Mountain/Great Basin open conifer woodlands and making a new layer that represents the convergence of these buffered areas. We also created a distance-to-water 30-m grid. All layers, including the case and control points, were projected to UTM Zone 13 North Clarke 1866.

**Construction and selection of landscape models.** Prior to constructing logistic regression models, we identified correlations between variables using Spearman correlations; correlated predictors (ρs > 0.55) were not included in the same model.

Logistic regression models were constructed to quantify the association between probability that an area is suitable with respect to plague habitat and landscape features. Seven candidate models were constructed using the build set (N = 372 points) (Table 1). The models are described by Eq 1:

\[
\text{Logit}(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k
\] (Eq 1)

where \(P\) is the probability of a cell being classified as suitable plague 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 Eq 1 using the following expression:

\[
P = \exp(\beta_0 + \beta_1 x_1 + \cdots + +\beta_k x_k)/[1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + +\beta_k x_k)]
\] (Eq 2)

To select the most parsimonious model with the best predictive power, Akaike’s information criterion (AIC) was used to rank each of the models. The model with the lowest AIC value was selected as the best. However, models within 2 AIC units of the minimum AIC (ΔAIC < 2) are considered competing with substantial support. To determine the amount of evidence in favor of a particular model, we calculated Akaikes weights (\(w_i\)) for each model \(i = 1, 2, \ldots, 7\). To determine if our models fit the data, we used a goodness-of-fit test. The goodness-of-fit analysis compares a pure-error negative log-likelihood with the fitted model log-likelihood. If the \(\chi^2\) test is not significant, then it supports the conclusion that sufficient data were included in the model.

Receiver operating characteristic (ROC) curves were used to assess the overall discrimination ability of each model, based on 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. The logit equation can be transformed into a prob-

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seven candidate models of human plague cases in New Mexico, based on 186 human cases and 186 random control points weighted by population density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model no.</th>
<th>(k^*)</th>
<th>Negative log-likelihood</th>
<th>AIC values†</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>Weight</th>
<th>Goodness of fit, (P) value</th>
<th>AUC‡</th>
<th>Independent model variables§</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>190.78</td>
<td>200.78</td>
<td>0</td>
<td>0.99</td>
<td>0.14</td>
<td>0.80</td>
<td>EL, EL², DistEco, DistWater</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>200.20</td>
<td>208.2</td>
<td>7.42</td>
<td>0.01</td>
<td>0.06</td>
<td>0.77</td>
<td>EL, EL², DistEco</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>206.77</td>
<td>212.77</td>
<td>11.99</td>
<td>0</td>
<td>0.01</td>
<td>0.74</td>
<td>EL, EL²</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>215.85</td>
<td>221.85</td>
<td>21.07</td>
<td>0</td>
<td>0.01</td>
<td>0.74</td>
<td>EL, EL², DistPJ</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>219.55</td>
<td>225.55</td>
<td>24.77</td>
<td>0</td>
<td>&lt; 0.01</td>
<td>0.74</td>
<td>DistKH</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>227.19</td>
<td>233.19</td>
<td>32.41</td>
<td>0</td>
<td>&lt; 0.01</td>
<td>0.73</td>
<td>EL, DistWater</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>237.50</td>
<td>241.50</td>
<td>40.72</td>
<td>0</td>
<td>&lt; 0.001</td>
<td>0.67</td>
<td>EL</td>
<td></td>
</tr>
</tbody>
</table>

* \(k\) = number of estimated parameters included in the model.
† AIC, Akaike information criterion value; ΔAIC = AIC of model – AIC of best model.
‡ AUC, area under ROC curve.
§ EL = elevation; EL² = elevation squared; DistEco, distance to the ecotone of Rocky Mountain/Great Basin closed conifer woodland and Rocky Mountain/Great Basin open conifer woodland; DistWater, distance to water; DistPJ, distance to Rocky Mountain/Great Basin closed conifer woodland or Rocky Mountain/Great Basin open conifer woodland; DistKH, distance to Rocky Mountain lower montane conifer forest, Rocky Mountain/Great Basin closed conifer woodland, or Rocky Mountain/Great Basin open conifer woodland.
ability following Eq 2. However, to evaluate the performance of the model using ROCs, this probability 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.\textsuperscript{11,15} All cells with a probability value at least equal to the optimal value were classified as high-risk plague habitats. All others were considered low risk in our evaluation matrix.

**RESULTS**

**Landscape features associated with residence-linked human plague cases in New Mexico.** Among the candidate logistic regression models, the best model (Table 1) for predicting areas where human plague cases were reported included elevation ($2.3 \times 10^{-2} \text{ m}^{-1} \pm 5.0 \times 10^{-3}$; $\beta \pm \text{SE}$), elevation\textsuperscript{2} ($-5.4 \times 10^{-6} \text{ m}^{-2} \pm 1.2 \times 10^{-6}$), distance to water ($-3.0 \times 10^{-4} \text{ m}^{-1} \pm 7.3 \times 10^{-5}$), and distance to the ecotone between Rocky Mountain/Great Basin open conifer woodlands and Rocky Mountain/Great Basin closed conifer woodlands ($-5.1 \times 10^{-5} \text{ m}^{-1} \pm 1.69 \times 10^{-5}$; intercept $-23.11 \pm 5.19$). The model indicated that suitability of habitat for plague cases increases up to an elevation of 2,129 m and declines as elevation increases thereafter.

Case points occurred closer to the Rocky Mountain/Great Basin open and closed conifer ecotone (mean 5,030 ± SD 6,817 m) than control points (19,075 ± 24,369 m). Similarly, cases were located closer to water (1,497 ± 1,694 m) than controls (2,562 ± 2,670 m). Accuracy of the best model, based on the area under the ROC curve, was 0.80. This value indicates that 80% of the time, randomly selected high- and low-risk pairs will be correctly ordered by their probability/habitat scores.\textsuperscript{11}

**Modeling suitable plague habitat within New Mexico.** Probability of suitability within the entire state of New Mexico was calculated based on the best model. Using the optimal cut-off probability value based on the ROC curve, 0.4616, the model predicted that 30.8% of the state is considered suitable plague risk habitat (Figure 1). However, most human plague cases occurred within 2 km of a home site. Therefore, we calculated the percentage of the state that was classified by the New Mexico Gap\textsuperscript{12} stewardship layer as private or tribal ownership and represented inhabited areas and had a model prediction probability of plague $\geq 0.4616$. Under these restrictions, 17.25% of the state is considered highly suitable habitat (Figure 2).

**Model evaluation.** A comparison of model predictions and actual point classifications (case or control) are presented in Table 2. Overall, user accuracies for correctly classifying highly suitable and less suitable plague habitats, based on a probability cut-off of $\geq 0.4616$, were 69.17% and 84.85%. This indicates that 69.17% of evaluated rasters classified by the model as highly suitable habitat contained human case points. In contrast, producer accuracies for correctly classifying highly suitable and less suitable habitats were 89.25% and 60.22%, respectively. In other words, 89.25% of actual case points were classified as highly suitable by the model, and 60.22% of control points were classified as less suitable. Errors of omission, where case points were misclassified as less suitable habitat, were rare (10.75%). In total, 39.78% of control points were classified as falling within highly suitable habitat. However, because plague is such a rare disease, it is not surprising to see such a high proportion of control points within highly suitable habitat.

**DISCUSSION**

Previously, evidence of *Y. pestis* has been reported from all but one (Hidalgo) of New Mexico’s 33 counties; human cases have been reported from 21 counties (CDC and New Mexico Department of Health, unpublished data). This broad geographic distribution has made it difficult to target control efforts. We created a GIS-based model of residence-linked human plague risk in New Mexico based on three landscape features: a nonlinear relationship with elevation, distance to water, and distance to the ecotone between Rocky Mountain/Great Basin open and closed coniferous woodlands. The model classified 30.8% of the state as posing a significant risk of exposure to humans. By adding land stewardship and extrapolating only within inhabited areas, we were able to further refine the area at risk to 17.25%. Within the counties where human cases have been identified, plague risk ranges from minimal to very high (Figures 1 and 2). Our model improves spatial accuracy in assessing human risk of exposure to the etiological agent of plague and can be useful for determining where limited public health resources should be targeted.

Prior to our study, it was proposed that human plague cases are closely associated with the distribution of piñon–juniper habitat, but quantitative data to support this assertion were not provided.\textsuperscript{2,7} Our model indicates that habitat suitability
for the occurrence of human plague increases where two types of piñon–juniper (Rocky Mountain/Great Basin open and closed conifer) converge. Human cases are typically associated with epizootic activity, and epizootics are more likely to occur when rodent and flea densities are high. Perhaps rodent or bridging vector densities and diversity are higher along these ecotones, thus increasing the likelihood of infecting multiple rodent and flea species once an infection becomes established in one habitat type or the other. Habitat suitability also increased with proximity to water. Water is scarce in the arid state of New Mexico; thus water sources could influence the distribution of key rodent species, as well as their fleas, and concentrate epidemiologically important hosts within the landscape, especially along riparian corridors. In addition, carnivores have been proposed as transport hosts for infected fleas and often move along riparian corridors, potentially increasing the likelihood of epizootics occurring near riparian areas and spreading along water courses. Determining why these landscape features emerged as risk factors is outside the scope of this study and will require field-based studies.

Among the 224 cases that were reported to the CDC from 1960 to 2003 in New Mexico, exposure sites were determined for 208. In total, 186 were exposed within 2 km of a home site. Therefore, our model focused exclusively on residence-linked risk. It is important to note that this model does not identify all areas in the state where Y. pestis is expected to occur. Epizootic or enzootic plague activity has been detected throughout the state, but the mere presence of Y. pestis in mammals and fleas does not necessarily indicate significant plague risks for humans. Our model yielded high (= 80%) overall accuracy in correctly classifying cases and controls. The exposure sites of = 89% of human cases occurred within areas predicted by our model as highly suitable or high risk. The overall accuracy was reduced because = 40% of controls were located within areas considered by the model to pose significant risk (Table 2). This result was expected because humans are incidental hosts of Y. pestis, resulting in very low incidence of human disease. Within suitable habitat, the likelihood of human exposure is defined by factors that cannot be modeled using a GIS, including behaviors related to pet care and handling, exposures to animals and insects, and rodent sanitation, the last of which influences the availability of harborage and food sources for plague-susceptible rodents around the home. These behavioral risk factors also could account for the 10.75% of cases that occurred within areas considered by our model to pose low risk.

Because we used privately or tribally owned land as a proxy for human-inhabited residential areas, and all of this land has not yet been developed for housing, the proportion of the state with actual high residence-linked risk is currently less than the 17.25% we estimated as having potential for significant risk. However, the privately or tribally owned areas not yet occupied by humans could present a significant risk if homes are built within these areas in the future or if nearby residents use these sites for recreational purposes, such as walking, or as a place for their pets to roam freely or on a leash. By narrowing our risk assessment to areas where people currently live or could reside in the future, we can target limited public health resources to the areas of highest risk.

Prevention measures should aim to educate health care providers, veterinary staff, public and environmental health workers, and the general public regarding exposure sites, manifestations, diagnosis of disease, and what steps can be taken to protect individuals and their families. Residents living within highly suitable plague habitat should be advised to eliminate harborage (e.g., piles of wood, brush, or debris) and food sources (e.g., pet food, garbage) for wild rodents and to eliminate fleas from pets. Because Y. pestis has been detected throughout the state, handling sick or dead animals ought to be avoided in all locations. In particular, hunters should be made aware of the risk of handling dead animals and be advised to use personal protection (e.g., latex gloves and eye protection) when handling animals. Public health officials must be aware that plague risk is not static. Within

**TABLE 2**

Evaluation matrix for the most parsimonious human plague risk model*

<table>
<thead>
<tr>
<th>Model classification†</th>
<th>Actual classification</th>
<th>% Correct‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human plague cases</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Highly suitable</td>
<td>166</td>
<td>74</td>
</tr>
<tr>
<td>Less suitable</td>
<td>20</td>
<td>112</td>
</tr>
<tr>
<td>‰</td>
<td>89.25</td>
<td>60.22</td>
</tr>
</tbody>
</table>

*The best predictive model was based on a nonlinear relationship with elevation, distance to Rocky Mountain/Great Basin closed conifer woodland, or Rocky Mountain/Great Basin open conifer woodland, and distance to water.
†Probability cut-off value used to classify a 30 m raster as suitable was based on the ROC optimal cut-off probability (P ≥ 0.4616).
‡User accuracy (commission error).
§Producer accuracy (omission error).
areas classified by our model as highly suitable, risk of human exposure varies temporally. Statewide, weather-related variables (e.g., winter precipitation and summer temperature) may be useful for predicting years with elevated epizootic activity.\textsuperscript{22,23} To identify epizootic plague activity at a fine geographic scale, animal-based surveillance focusing on areas classified by our model as high risk can be useful; when epizootics are detected, insecticides can be used to treat fleas within these limited geographic regions.\textsuperscript{7,16,21} However, plague is rare in humans, surveillance of the disease is expensive, and it often was assumed that risk of exposure to \textit{Y. pestis} is high in most of the western United States. For these reasons, some local health departments in these plague-affected regions have hesitated to undertake surveillance and other prevention activities. Our model, which identifies areas posing the highest risk, may help these programs to justify prevention activities and focus these in areas where they are most likely to lead to a reduction in human cases.

Received November 25, 2006. Accepted for publication March 12, 2007.

Acknowledgment: The authors thank L.G. Carter for technical support.

Authors’ addresses: Rebecca J. Eisen, Russell E. Enscore, Brad J. Biggerstaff, John A. Montenieri, and Kenneth L. Gage, Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, P.O. Box 2087, Fort Collins, CO 80522, Telephone: +1 (970) 266-3523, Fax: +1 (970) 225-4257, E-mail: dyn2@cdc.gov. Pamela J. Reynolds and Paul Biggerstaff, John A. Montenieri, and Kenneth L. Gage, Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases: A Modern Treatise of Infectious Processes, 4th ed. Philadelphia: J.B. Lippincott Company, 1050 1965–1976.

Reprint requests: Rebecca J. Eisen, Division of Vector-Borne Infectious Diseases, NCID/CDC, P.O. Box 2087, Fort Collins, CO 80522, Telephone: +1 (970) 266-3523, Fax: +1 (970) 225-4257, E-mail: dyn2@cdc.gov.

REFERENCES