Predicting the Risk of Hantavirus Infection in Beijing, People’s Republic of China

Wen-Yi Zhang, Li-Qun Fang, Jia-Fu Jiang, Feng-Ming Hui, Gregory E. Glass, Lei Yan, You-Fu Xu, Wen-Juan Zhao, Hong Yang, Wei Liu, and Wu-Chun Cao*

State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, People’s Republic of China; The Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Abstract. To understand the spatial distribution of hantavirus infections across landscapes and its influencing environment factors in Beijing, People’s Republic of China, rodents captured in this region were examined for hantavirus by reverse transcription–polymerase chain reaction. A total of 1,639 rodents were trapped at 86 randomly selected sites. The overall infection rate for hantavirus was 7.14% in the rodents. Multivariate logistic regression analysis indicated that the natural infection rates for hantavirus in rodents were significantly associated with rice agriculture (odds ratio [OR] = 2.88, 95% confidence interval [CI] = 1.27–43.55). A risk map was constructed on the basis of these significant factors to predict extension and transmission of infection with hantavirus in this region.

INTRODUCTION

Hemorrhagic fever with renal syndrome (HFRS) is a zoonosis caused by different species of hantavirus. This disease is highly endemic in the People’s Republic of China, which has 90% of the total reported cases of HFRS in the world.1 The causative agents of HFRS in China are predominately Hantaan virus (HTNV) and Seoul virus (SEOV), which result in case-fatality rates of approximately 10% and 1%, respectively.1 In 2003, Puumala-like hantavirus (PUUV) was detected in bank voles (Clethrionomys rufocanus) in northeastern China.2 Recently, a newly recognized hantavirus named Amur virus (AMRV) was found in Apodemus peninsulae.3,4 Although a vaccination program was initiated in the country 20 years ago and has effectively prevented transmission from animal hosts to humans, HFRS remains a significant public health problem with 20,000–50,000 human cases reported annually in mainland China.5 The incidence of HFRS greatly varies at provincial and county levels. Economic development, urbanization, human population fluctuation, and environmental and climate changes are thought to influence heterogeneity in the incidence and spatial distribution of the disease.6

In recent years, the incidence of HFRS has been increasing in some metropolitan and provincial capital cities.7 Beijing, the capital of China, had recently become a region in which hantavirus is endemic. The first reported HFRS case was imported to Beijing from another province in 1986.4 Thereafter, only few sporadic cases were reported. Because local transmission was established in 1996, HFRS incidence has increased annually and cases have been found in all the 18 districts of Beijing.9 In previous studies, hantavirus was detected in domestic rodents, among which substantial genetic heterogeneity of hantavirus was demonstrated by using molecular biologic approaches,8,9,10 which might be the biological aspect for the emergence of HFRS in this region. However, the exact geographic distribution of hantavirus infections in animal hosts and factors facilitating spread of the virus within a host population remain unclear. The objectives of this study were to understand the spatial distribution of hantavirus infection in rodent hosts, to identify landscape factors contributing to the presence of hantavirus in rodent populations, and to predict occurrence of HFRS for possible preemptive public health warnings.

MATERIALS AND METHODS

Data collection and management. One cloud-free LANDSAT-5 Thematic Mapper (TM) image on June 23, 2005 over the entire study area was acquired with a spatial resolution of 30 meters from a Chinese Remote Sensing Ground Station. The TM data for the study area were geometrically corrected by ground control points identified by global positioning system receivers (Garmin, Olathe, KS) by using environment for visualizing images image processing version 4.0 software (Research Systems Inc., Boulder, CO). In this study, all data were transformed to the UTM-WGS84 50N projection. Land use types were determined by maximum likelihood classification, which assumes that statistics for each classification in each band are normally distributed and calculates the probability of a given pixel belonging to a specific type. Each pixel was assigned to the class that had the highest probability. If the highest probability is smaller than a specified threshold, the pixel was defined as unclassified.11

A digital elevation model (DEM) was derived from shuttle radar topography mission data with a spatial resolution of 90 meters.12 According to the agriculture function region and population distribution in Beijing, elevation was divided into three levels. < 40 meters (average elevation), 40–80 meters (twice the average elevation), and > 80 meters.13 Normalized difference vegetation index (NDVI) is correlated with amount and productivity of vegetation, and is generated from a transformation of the near infrared (TM band 4) and red wavelengths (TM band 3) by using the equation:14

\[
\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}
\]

The NDVI values range from −1 to 1, with bare soil and impervious surface having values near 0, and high values indicating increasing green biomass and photosynthetic activity.15,16 The NDVI was used to estimate vegetation growth on the survey sites and was divided into four classes: < 0.1, 0.1–0.2, 0.2–0.3, and > 0.4.17

Sample collection. Rodents were captured at 86 sites in HFRS-epidemic seasons of 2005 and 2007 (Figure 1). The trapping sites were randomly selected with regard to land use.
and geographic locations to minimize spurious correlations related to possible seasonal effects. A standardized trapping protocol was used to sample the host rodents. A total of 100 traps with peanuts as baits were set for 3 consecutive nights as four parallel lines of 25 traps each. Traps were spaced at approximately 10-meter intervals. Longitude and latitude were recorded at the contralateral corners of the two outer trap lines by using a global positioning system receiver. After identification of species and sex, lung tissues were removed from the captured rodents, and stored in liquid nitrogen until tested. For unidentified species in the field, the crania were brought to the laboratory for further identification. Relative population density was measured as the number of animals captured divided by the number of trap nights at each site and used as an indicator of abundance.

Hantavirus detection. A reverse transcription–polymerase chain reaction (RT-PCR) was used to detect hantaviruses in rodents by using the methods previously described. Briefly, total RNA was extracted from lung tissues of rodents using an extraction kit (Invitrogen Life Technologies, Rockville, MD). Reverse transcription was conducted by using Avian myeloblastosis virus reverse transcriptase (Promega, Madison, WI) and random hexamers as specified by the manufacturer. Each sample was tested by RT-PCR for HTNV, SEOV, PUUV, and AMRV. The PCR products were separated by electrophoresis on a 1.2% agarose gel, stained with ethidium bromide, and visualized under ultraviolet light. To avoid contamination, DNA extraction, reagent setup, amplification, and agarose gel electrophoresis were performed in separate rooms. In parallel with each amplification run, a positive control and a negative control (distilled water) were included.

Data analyses. To test the geographic heterogeneity of rodent species composition across the land use types, the Shannon-Wiener diversity index was used to calculate species diversity for each land use type. The index is calculated as $H = \sum p_i \ln(p_i)$ ($i = 1, 2, 3 \ldots S$), where $S$ is species richness and $p_i$ is proportion of the total number of individuals belonging to species $i$. One-way analysis of variance was used to assess overall differences of species diversity among different land use types. The proportion of rodent species for each site was displayed on a map of Beijing with the administrative boundary using ArcGIS version 9.1 (Environmental Sciences Research Institute, Redlands, CA). Infection rate on each site was overlaid by land use map of Beijing by using ArcGIS version 9.1 software.

Univariate logistic analysis was conducted to determine the associations between the incidence of hantavirus infection and land use, NDVI, and elevation. A $P$ value < 0.05 was considered statistically significant. Odds ratios (ORs) were estimated by comparing infection rates between positive with negative sites. Multivariate logistic analysis was then performed by including variables with $P$ values < 0.10 from the univariate analyses as covariates. Condition index and variance decomposition proportions were used to test collinearity among the independent variables. The logistic model was evaluated with the Hosmer-Lemeshow goodness of fit statistics (C) for deciles of risk. To examine the accuracy of the model predictions, we evaluated the sensitivity and specificity by creating a receiver operator characteristic function. The function compares the true-positive rate (sensitivity) against the false-positive rate (1 - specificity) of a model by using various predicted values as thresholds of identifying positive and negative sites. All statistical analyses were performed using SPSS software (SPSS Inc., Chicago, IL).

**Risk mapping.** To develop a risk map for HFRS in Beijing, we used the land use classification and DEM layers to apply the logistic regression model across the coverage area using a raster calculator. The resulting layer depicted the probability of infection in rodents at a 90 × 90-meters resolution. Because coverage files to create the grid file had the same projection (UTM-WGS84 50N) and spatial resolution of 90 meters, the layers could be overlaid and data values corresponding to each layer were assigned to each cell. The final logistic function was then used to generate the probability of the presence of hantavirus infection within each cell by using the spatial analyst module in ArcGIS version 9.1 software. The risk map was divided into quartiles, which indicated different probability levels.

**RESULTS**

The land use map of Beijing produced by maximum likelihood classification (Figure 2) categorized the types of land use as forest, built-up land, farmland, rice agriculture, orchards, and water bodies (including rivers, reservoirs, and lakes). Classification accuracy was evaluated as 85.3% on the basis of 300 field observation points. Water bodies were excluded from the survey sites for rodents. Forests referred to areas with dense tree and scrub cover. Farmland was characterized either by the presence of agricultural crops or bare ground that had been allocated for planting crops. Orchards were areas producing fruits and raw materials for industry or for beverages. A total of 1,639 rodents, belonging to eight species from three families, were captured during the study at the 86 trapping sites (Figure 1). Of these, *Rattus norvegicus* (54.42%) and *Mus musculus* (19.34%) were the predominant species, and *A. agrarius* and *A. peninsulae* represented 8.85% and 9.88% of the captured rodents, respectively. Other rodent species included...
*Niviventer confucianus, Sciurotamias davidianus, Cricetus triton, and C. rufocanus* (Table 1).

Species diversity, as measured by Shannon-Wiener diversity index, were significant among different land use types ($F = 3.14, P < 0.05$). Several of the regularly captured species were restricted to a few land types. A higher proportion of *R. norvegicus* was found in the built-up land, most of which was located in downtown Beijing. However, *A. peninsulae* were more abundant in the northern area (Figure 3). There was also a significant difference in rodent densities among different land use types ($\chi^2 = 77.98$, degrees of freedom [df] $= 4$, $P < 0.001$). The highest rodent density was found in forest with a capture success of 10.74%. The lowest rodent density was found in orchards (5.30%).

One hundred seventeen rodents were positive for SEOV by RT-PCR, with an overall infection rate of 7.14% (95% confidence interval [CI] = 5.94–8.90%). The difference in infection rates among species was statistically significant ($\chi^2 = 53.61$, df $= 7$, $P < 0.001$) (Table 1). Only two dominant species, *R. norvegicus* and *M. musculus*, were infected by virus and had infection rates of 10.99% and 5.99%, respectively. Male rodents had a higher infection rate than the female rodents ($\chi^2 = 16.81$, df $= 1$, $P < 0.001$).

Univariate logistic regression analysis detected two variables, land use and elevation, significantly correlated with hantavirus infection in rodents (Table 2). Rice agriculture (OR = 3.27, 95% CI = 1.36–30.10) and orchards (OR = 1.36, 95% CI = 1.03–6.14) were favorable for hantavirus infection in rodent hosts than built-up land. Forest was protective against HFRS (OR = 0.18, 95% CI = 0.05–0.60). The medium level of elevation (40–80 meters) increased the chance hantavirus infection in rodents (OR = 9.00, 95% CI = 2.20–36.86). The NDVI was not associated with hantavirus infection in rodent hosts ($P = 0.11$ (Table 2).

Multivariate logistic regression analysis indicated that land use and elevation were significantly associated with the prevalence of hantavirus infection in rodents (Table 3). The final logistic regression function for predicting the risk areas was 

$$\text{Logit}(P) = 1.059 \times \text{rice agriculture} + 0.115 \times \text{orchards} + 2.285 \times \text{moderate elevation} - 1.909 \times \text{forest}.$$ 

Rice agriculture (adjusted OR = 2.88, 95% CI = 1.27–30.70), orchards (adjusted OR = 1.12, 95% CI = 1.02–5.96), and moderate elevation (adjusted OR = 9.83, 95% CI = 2.22–43.55) were risk factors for natural infection of hantavirus in rodent populations, and the forest (adjusted OR = 0.15, 95% CI = 0.03–0.72) was the protective factor against infection. The logistic regression analysis fit the observed data well (Hosmer Lemeshow C = 3.93; $P = 0.69$, df $= 6$), and none of the sites were obvious outliers. No significant collinearity was found among the independent variables. The receiver operator characteristic graph showed that

![Figure 2. Hantavirus prevalence in rodent population in different land use types, Beijing, People’s Republic of China.](image-url)
at least 80% of the positive sites were correctly identified until the proportion of negative sites correctly identified exceeded 71%. This threshold corresponded to a predicted value for hantavirus infection occurrence of 0.50. Thus, using a predicted log OR of 0.50 as a threshold for increased risk included 80% of the positive sites and excluded 71% of the negative sites.

The final logistic function was applied to each pixel in the study area to produce a predictive risk map (Figure 4).23 The higher probabilities indicated more chances of hantavirus infection in Beijing. The highest risk regions for hantaviruses in rodents mainly focused on downtown (Chongwen, Xuanwu, Xicheng, and Dongcheng districts) Beijing and several suburbs including most of Haidian and Fengtai, southeast of Changping, east of Shijingshan, west of Chaoyang, and the boundary of Shunyi, Miyun and Huairou districts. The boundary of Fangshan and Daxing districts also appeared to have a higher risk. The predictive risk map showed that low-risk regions (0.0–0.25) were located in the northwestern areas of Beijing (Figure 4).

**DISCUSSION**

As one of the recently developed spatial analysis techniques, spatial information technology has opened new avenues for ecoepidemiologic studies of infectious diseases by identifying and mapping the habitat of host species and their relationship to human settlements. Moreover, increasing availability of geographic information system–based environmental data can help to reveal associations between environmental factors and incidence, based on which potential risks for disease transmission can be predicted.23–26

This study showed the remarkable heterogeneity in the spatial distribution of rodent species in Beijing. Most *R. norvegicus* were distributed in the central region of Beijing, and *A. peninsulae* were largely distributed in the northern part of the region (Figure 3). Predominant land use types in the center and northern part of the city were built-up land and forests, respectively. Landscape features may have influenced rodent density

---

**Table 2**

Results of univariate logistic regression analysis for hantavirus infection in rodents from Beijing, People’s Republic of China*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Infection rate (%), (95% CI)</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Built-up land</td>
<td>8.26 (6.49–10.34)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>1.64 (0.53–3.78)</td>
<td>0.18 (0.05–0.60)</td>
<td>0.006</td>
</tr>
<tr>
<td>Farmland</td>
<td>7.94 (5.30–11.34)</td>
<td>0.35 (0.92–1.34)</td>
<td>0.125</td>
</tr>
<tr>
<td>Rice agriculture</td>
<td>10.20 (5.00–17.97)</td>
<td>3.27 (1.36–30.10)</td>
<td>0.029</td>
</tr>
<tr>
<td>Orchard land</td>
<td>9.84 (3.70–20.19)</td>
<td>1.36 (1.03–6.14)</td>
<td>0.007</td>
</tr>
<tr>
<td>Elevation, meters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 40</td>
<td>3.69 (1.60–7.13)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>40–80</td>
<td>9.43 (7.47–11.69)</td>
<td>9.00 (2.20–36.86)</td>
<td>0.002</td>
</tr>
<tr>
<td>&gt; 80</td>
<td>5.49 (3.86–7.56)</td>
<td>2.12 (0.54–8.26)</td>
<td>0.285</td>
</tr>
<tr>
<td>NDVI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 0.1</td>
<td>7.13 (5.40–9.21)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>0.1–0.2</td>
<td>10.03 (7.16–13.56)</td>
<td>1.89 (0.52–6.83)</td>
<td>0.335</td>
</tr>
<tr>
<td>0.2–0.3</td>
<td>6.06 (3.92–8.88)</td>
<td>0.41 (0.15–1.18)</td>
<td>0.100</td>
</tr>
<tr>
<td>&gt; 0.3</td>
<td>1.71 (0.21–6.04)</td>
<td>0.36 (0.05–2.41)</td>
<td>0.291</td>
</tr>
</tbody>
</table>

*OR = odds ratio; CI = confidence interval; NDVI = normalized difference vegetation index.

---

**Table 3**

Results of multivariate logistic regression analysis for hantavirus infection in rodents from Beijing, People’s Republic of China*

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Built-up land</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>0.15 (0.03–0.72)</td>
<td>0.018</td>
</tr>
<tr>
<td>Farmland</td>
<td>0.34 (0.08–1.50)</td>
<td>0.152</td>
</tr>
<tr>
<td>Rice agriculture</td>
<td>2.88 (1.27–30.70)</td>
<td>0.038</td>
</tr>
<tr>
<td>Orchards</td>
<td>1.12 (1.02–5.96)</td>
<td>0.008</td>
</tr>
<tr>
<td>Elevation, meters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 40</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>40–80</td>
<td>9.83 (2.22–43.55)</td>
<td>0.003</td>
</tr>
<tr>
<td>&gt; 80</td>
<td>5.09 (0.93–27.88)</td>
<td>0.061</td>
</tr>
</tbody>
</table>

*OR = odds ratio, CI = confidence interval.
and distribution. This finding was not surprising because one land
use type might be highly suitable for certain species of rodents. In
addition, some other undetected variables, each of which has its
contribution, might impact the heterogeneous distribution.\textsuperscript{27}

The current study found that only SEOV was found in rodent
populations in Beijing, which was consistent with results of previ-
ous studies.\textsuperscript{8,10} No HTNV was found in our study area, although
sufficient numbers of wild rodents such as \textit{A. agrarius} and
\textit{A. peninsulae}, which are the main reservoirs of HTNV in other
provinces of China, were detected for HTNV.\textsuperscript{28} The absence
of HTNV helps to explain the pattern of the risk of hantavi-
rus infection in the study area. Infection with SEOV in rodents
occurred mainly in the rural-urban fringe of Beijing (Figure 2),
where \textit{R. norvegicus} found suitable habitats. Maintenance of
virus in a region ultimately requires some connectivity between
populations, enabling reinfections to occur as virus is transmitted
between populations and infects previously unexposed persons
and populations.\textsuperscript{29} Thus, landscape features that influence popu-
lation connectivity can inevitably be related to viral dynamics.

Because \textit{R. norvegicus} had wide ecologic tolerances and
were often the most numerous rodent in many habitat types
in Beijing, their populations tended to be well connected, even
in heterogeneous landscapes. Land use and land cover could
influence the virus occurrence in hosts by controlling move-
ments of individuals and thus contact rate between infected
and susceptible rodents.\textsuperscript{27} Moreover, the average infection rate
(7.14\%) for hantavirus in rodents in Beijing was much higher
than that (2.75\%) in mainland China.\textsuperscript{30} It was most likely that
a higher SEOV prevalence among rodents contributed to the
rapid expansion and transmission of the epidemics in Beijing.

The significant environmental variables in the final logis-
tic regression model could offer a strong explanation for the
epidemiological features of HFRS.\textsuperscript{19} Rice agriculture and
orchards were appropriate habitats for rodent hosts. Farmers
and migrant workers had more chances to be exposed to con-
taminated urine and feces of infected rodents.\textsuperscript{19} As shown pre-
viously,\textsuperscript{23} NDVI was excluded from the multivariate logistic
regression model. And the failure of the association between
HFRS risk and vegetation growth, as measured by NDVI, may
indicate that the domestic rodent species, \textit{R. norvegicus} and
\textit{M. musculus}, were not, or rarely, influenced by environ-
ment determinants reflected by NDVI, because they rely a
lot on human food sources rather than merely on growing
vegetation.

By using a geographic information system technique, we gen-
erated a risk map (Figure 4) to predict hantavirus infections in
rodent hosts. The places with higher probability of hantavirus
infections in rodent hosts correspond to those with increased
incidence of HFRS cases and known HFRS-endemic areas in
Beijing.\textsuperscript{8} The extensive high-risk regions in Beijing (Figure 4)
could partially explain the rapid transmission and extension of
HFRS. In the risk map, the fringe of Shunyi, Miyun, Huairou
districts displayed higher risk (0.75–1.00). However, we did
not sample rodents in these areas in the present study. Further
studies are required to evaluate and refine the risk model
based on the data from these areas. When compared with pre-
vious studies,\textsuperscript{24,26} our risk map is based more on the distribu-
tion of the hantavirus infection rate than on the distribution of
vector/host species or cases, which ensured the higher credibil-
ity and validity for risk prediction.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Figure4.pdf}
\caption{Predicted risk map of hantavirus infection in Beijing, People’s Republic of China.}
\end{figure}
In the study, climatic factors were not considered in the logistic regression analyses. Some studies had shown that rainfall and temperature may play a significant role in transmission of HFRS. However, there are only two meteorologic stations in Beijing, and they are too sparse to obtain accurate interpolation in the wide study area. The economic conditions, population immunity, and immigration may also impact the transmission of the disease. Thus, these factors need to be considered in further research.

In conclusion, this model can be used to help determine the risk of acquiring HFRS by predicting the locations where hantavirus infection may occur among rodent populations in Beijing. This method can also be expanded and applied to other new epidemic foci in China. Further refinement based on findings from new areas could also improve application of the model. When facilitated with spatial information technologies such as geographic information systems and remote sensing, risk of HFRS transmission could be predicted. Informed HFRS control and prevention measures focused on high-risk areas identified in the current study could provide more effective interventions.

Received October 6, 2008. Accepted for publication January 6, 2009.

Financial support: This study was supported by grants from the National Science Fund for Distinguished Young Scientist (No. 30725032), the Natural Science Foundation of China (No. 30593074 and 30700682), and the Natural Science Foundation of Beijing (No. 7061005), and a special grant for the prevention and treatment of hemorrhagic fever with renal syndrome in China. Gregory E. Glass, The Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205.

Authors’ addresses: Wen-Yi Zhang, Li-Qun Fang, Jia-Fu Jiang, Peng-Ning Hui, Lei Yan, You-Fu Xu, Wen-Juan Zhao, Hong Yang, Wei Liu, and Wu-Chun Cao, Beijing Institute of Microbiology and Epidemiology, State Key Laboratory of Pathogen and Biosecurity, 20 Dong-Da Street, Fengtai District, Beijing 100071, People’s Republic of China. Gregory E. Glass, The Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205.

REFERENCES

1. Luo CW, Chen HX, 2003. Study on the factors influenced epidemio-


