Research Paper

Spatial Analysis of West Nile Virus: Rapid Risk Assessment of an Introduced Vector-Borne Zoonosis

JOHN S. BROWNSTEIN,¹ HILARY ROSEN,¹ DIANNE PURDY,¹ JAMES R. MILLER,²
MARIO MERLINO,² FARZAD MOSTASHARI,² and DURLAND FISH¹

ABSTRACT

The distribution of human risk for West Nile virus was determined by spatial analysis of the initial case distribution for the New York City area in 1999 using remote sensing and geographic information system technologies. Cluster analysis revealed the presence of a statistically significant grouping of cases, which also indicates the area of probable virus introduction. Within the cluster, habitat suitability for potentially infective adult mosquitoes was measured by the amount of vegetation cover using satellite imagery. Logistic regression analysis revealed satellite-derived vegetation abundance to be significantly and positively associated with the presence of human cases. The logistic model was used to estimate the spatial distribution of human risk for West Nile virus throughout New York City. Accuracy of the resulting risk map was cross-validated using virus-positive mosquito sample sites. These new epidemiological methods aid in rapid entry point identification and spatial prediction of human risk of infection for introduced vector-borne pathogens. **Key Words:** Cluster analysis—Geographic information system—Landscape epidemiology—Logistic model—Remote sensing—Risk map—West Nile virus. *Vector Borne Zoonotic Dis.* 2, 157–164.

INTRODUCTION

WEST NILE (WN) VIRUS is a vector-borne zoonotic disease that can cause acute febrile illness, encephalitis, and death in humans (Hayes 1989). The WN virus epidemic of August 1999 in the New York City (NYC) area represented the first documented introduction of this vector-borne pathogen in the Western Hemisphere (Lanciotti et al. 1999). Although most of the cases were found within a 2-mile radius of Queens County, NY (Nash et al. 2001), the spatial delineation of the case cluster and the full extent of human risk in the region were not immediately known. WN virus activity has since spread to 44 states and the District of Co-

lumbia, and a total of 3,389 human cases were reported in 2002 (Centers for Disease Control and Prevention 2002). A better understanding of the initial distribution of risk may have implications for emergency response strategies to future introductions of vector-borne pathogens in urban areas.

At least five mosquito species found in NYC have been shown to be vector-competent for WN virus (Turell et al. 2000), and virus has been detected in a total of nine species in New York State (Bernard et al. 2001, Nasci et al. 2001). The zoonotic reservoir hosts of the natural WN cycle are avian, and transmission to humans is accidental (Hulburt 1956, Tsai et al. 1998, Hubalek and Halouzka 1999, Savage et

¹Department of Epidemiology and Public Health, Yale School of Medicine, New Haven, Connecticut.

²New York City Department of Health and Mental Hygiene, New York, New York.

al. 1999). *Culex pipiens*, which feeds primarily on birds, has been shown to be an efficient zoonotic vector (Tempelis 1974). However, mosquito species that transmit zoonotic pathogens to humans must feed across zoological class lines in order to act as bridge vectors. Several *Culex*, *Aedes*, and *Ochlerotatus* species have this ability, making the identification of human vector species speculative (Turell et al. 2000).

Though the precise vectors for WN virus in the United States are imperfectly understood, most vectors of zoonoses are commonly tied to the landscape as environmental determinants control their distribution and abundance (Pavlovsky 1966, Fish 1996). Therefore, the dynamics of WN virus transmission can be understood as a function of a number of abiotic and biotic processes that control the survival and therefore the distribution of arthropod vectors (Hayes 1989, Kitron 1998). Elucidating the

relationship between landscape and vector facilitates the measurement of human risk in space. Subsequently, the landscape epidemiology of WN virus can be displayed in the form of a risk map for a specific region. Such maps are useful during outbreak situations, as they enable the focusing of surveillance and control efforts and the identification of the probable source or point of entry for exotic pathogens. Additionally, limited field resources can be vastly supplemented with satellite imagery in order to estimate risk at unsampled locations.

Remote sensing (RS) has been used to identify the landscape associations of several vector-borne pathogens by identifying suitable vector habitat (Pope et al. 1992, Wood et al. 1992, Beck et al. 1994, Dister et al. 1997). In the case of WN virus, the distribution of potentially infective mosquitoes can be considered to be dependent upon land use and landscape pat-

FIG. 1. Location of the WN virus outbreak cluster of 1999 in the NYC area. Geocoded WN cases ($n = 50$) are displayed. Cluster analysis was performed using the spatial scan statistic. Confounding due to age and the presence of mosquito habitat was accounted for by indirect standardization. One significant cluster of WN ($p < 0.001$) centered in Nassau County (40.820°N , 73.739°W) was found with a radius of 12.9 km and a relative risk of 4.2.

FIG. 2. Map of vegetation density in the study area. The NDVI from a Landsat TM image acquired July 31, 1999 with 30-m resolution was used to represent vegetation intensity.

terns that determine adult habitat. Vector mosquito habitat can be measured by the degree of vegetation cover, which provides both carbohydrate resources for flight energy and avian host blood resources for both reproduction and pathogen acquisition (Fish 1983). The simplest measure of vegetation density is the normalized difference vegetation index (NDVI), which has been previously shown to be highly predictive of the distribution of disease vectors, including tsetse flies (Robinson et al. 1997) and

ticks (Randolph 2000) as well as mosquitoes (Anyamba et al. 2001).

We have applied geographic information system (GIS) and RS technologies in combination with spatial statistics to generate a risk map of WN virus and aid in the identification of the entry point of WN virus into NYC. This analysis of the case distribution data demonstrates the application of these new technologies in facilitating a rapid response to vector-borne pathogen introductions.

MATERIALS AND METHODS

Data source

We created a WN virus risk map based upon geocoded case locations from the summer of 1999 provided by the NYC Department of Health. Six New York counties were included in the analysis: Bronx, Kings, Queens, New York, and portions of Nassau and Westchester. The 1990 census tracts of these counties provided by the U.S. Census Bureau were overlaid onto the image.

Water coverage removed portions of census tracts that are not land. A total of 50 cases found within 47 census tracts were reported in this region during the summer of 1999 (Fig. 1). We used the U.S. census tract as a base unit for mapping discrete geopolitical boundaries within the region. NDVI detected the density of vegetation for each 30- × 30-m pixel of a Landsat 7 Thematic Mapper (TM) image of New York acquired July 31, 1999 (Fig. 2). The mean NDVI for each census tract was calculated by taking the average pixel value within the tract.

Cluster analysis

A spatial scan statistic was used to test whether the WN virus cases were randomly distributed over space using SaTScan (Kulldorff et al. 1996, Kulldorff 1997). A Poisson-based model was chosen, where the number of cases in the area is Poisson-distributed under the null hypothesis (Hjalmars et al. 1996). The 1990 census data for each tract were used as the background population data. Because this is an analysis of point data, the geographic centroids for each of the tracts were calculated.

For each census tract centroid, the spatial scan statistic centers a circular window around it. This window varies continuously in size from 0 up to where it includes 50% of the total population at risk (Kulldorff 1997). The circles will therefore contain different sets of neighboring census tracts. This method tests the null hypothesis that disease risk inside the window is the same as outside for each of the circles at each of the centroids. The software calculates a likelihood ratio for each window with the assumption of a Poisson distribution to identify the most likely clusters (Kulldorff and Nargar-

walla 1995, Kulldorff et al. 1996). Thus, the presence of WN clusters can be assessed with the maximum likelihood ratio test statistic. The associated *p* value for the most likely cluster was calculated through a Monte Carlo simulation consisting of 999 random replications of the data set (Dwass 1957). The expected number of cases in a census tract were adjusted by indirect standardization according to potential categorical covariates, so that identified clusters exist beyond what results from the covariates alone. Given that most patients with WN virus are older adults (Centers for Disease Control and Prevention 2001), confounding due to age was removed by calculating incidence using the population of individuals ≥55 years for each census tract from the 1990 census. Confounding by mosquito habitat was accounted for by adjusting disease incidence according to the NDVI risk classification (Fig. 2).

Risk model

A single-factor analysis of variance (ANOVA) was performed to test whether census tracts with cases and without cases within the cluster result in different mean NDVI values. Logistic regression analysis was then used to explore the ability of mean NDVI to identify census tracts with cases within the cluster. A second-order interaction term was included in the model as goodness-of-fit testing revealed that the logit did not significantly differ from a quadratic relationship with mean NDVI. Logistic model success was judged by the receiver operating characteristic (ROC) plot. The ROC plot depicts the relationship between sensitivity and specificity across the range of possible cutoff probabilities. The area under the ROC curve (AUC) is a measure of the overall fit of the model.

Sensitivity analysis was used to identify high-risk census tracts by selecting the probability cutoff point that maximizes both sensitivity and specificity. The logistic model was then extrapolated to all census tracts of the study region to create a map of all high-risk census tracts for the entire NYC region. The threshold probability level was applied to all tracts, so that tracts above the cutoff were defined as high risk.

Model validation

The New York area risk map for WN virus, derived from human case data, was cross-validated using the locations of mosquito collections that were reported positive for WN virus (U.S. Department of Agriculture 1999). Mosquito light trap data from 75 locations during the summer of 1999 were used to validate the high-risk designated census tracts, by calculat-

ing the number of samples that intersect that high-risk region (Fig. 3).

RESULTS

Only one significant cluster of WN virus cases was found, with a relative risk adjusted for covariates of 4.2 ($p < 0.001$). The cluster was centered in western Nassau County (40.820°N,

FIG. 3. Risk map for the NYC area. Results were derived from logistic regression analysis that identifies the probability of a human case of WN virus within the previously defined cluster according to mean NDVI. Sensitivity analysis was used to select a probability cutoff ($p = 0.08$) point and define the high-risk census tracts. This relationship was extrapolated to all census tracts within the NYC area. Virus-positive mosquito locations used to cross-validate the model with an accuracy of 72% are also shown.

73.739°W) with a radius of 12.9 km, and included 43 of the 50 cases (Fig. 1).

Within the cluster, the mean NDVI for case census tracts was significantly greater than the mean value for control census tracts by ANOVA ($p < 0.05$). Logistic model fit was significant by the likelihood ratio test ($p < 0.05$) and yielded a risk probability of human cases within a census tract. The probability of a human case of WN virus increases up until a threshold NDVI of 0.22, and then decreases (Fig. 4). This decrease can be attributed to an inverse relationship between population density and proportion of vegetated area within the census tracts. Model accuracy evaluated by the AUC was 0.61 and differed significantly from chance ($p < 0.01$).

A cutoff probability for a human WN virus case of 0.08 was selected by sensitivity analysis. Using this threshold, the model identified 33 of the 40 case census tracts correctly, yielding a sensitivity of 83%. After applying the model to all census tracts of the study region, 33% (861 of 2,621) of the census tracts were classified as high risk (Fig. 3). This high-risk area covers 33% of the population, and 44% of the land surface area of the examined region.

Mapping WN virus-positive mosquito sites validated the risk model. Out of 75 virus-positive collection sites in the New York region, 54

locations (72%) were found within high-risk census tracts.

DISCUSSION

Our analysis provides insight into the possible location of WN virus introduction and its landscape association. A significant clustering of cases occurred beyond what would have been expected by chance, after accounting for the dependence of virus activity upon vegetated habitat and human population structure. Although traditional case reporting is used as a starting point for identifying the entry point of the virus, spatial statistics allows us to conclude with confidence that there was only a single source of introduction and identifies the original source location of WN virus in the NYC area (Fig. 1). The model highlights areas that are most likely to experience transmission of WN virus. A rapid understanding of where virus was introduced and why cases occurred in certain areas of New York City compared with others would aid greatly in the deployment of emergency surveillance and control efforts. Such a model can also be implemented in other areas of the country that experience WN virus introduction. Local public health agencies can employ this model to rapidly describe hu-

FIG. 4. Results of the logistic regression. The model describes the probability of a human case of WN virus (WNV) within the previously defined cluster according to mean NDVI ($p < 0.05$). The probability of a human case of WN increases with increasing NDVI up to a threshold NDVI of 0.22 and then decreases.

man risk when only sparse case reports are available.

The study also revealed NDVI to be an informative marker for WN virus vector mosquito habitat as it is significantly higher for census tracts with cases within the cluster. This emphasizes the importance of identifying and sampling in adult mosquito habitat as part of the vector surveillance effort. Until the specific vector species have been determined, NDVI can be used as a general predictor of vector mosquito distribution in urban areas. The analysis was performed within the cluster to control for the presence of virus, competent hosts, bridge vectors, and vulnerable humans. Human cases provided accurate data for the model because of the high sensitivity for detecting WN virus activity resulting from active case surveillance efforts throughout the cluster area. The model identified high-risk census tracts mainly in the less populated suburban regions of Queens, Bronx, Westchester, and Nassau Counties (Fig. 3). Virus-positive mosquito data cross-validated the model, which is based solely on human cases, by also correctly categorizing 72% of the positive mosquito collections (Fig. 3). Mosquito data are especially important for validating risk in areas that did not report human cases. The mosquito data are appropriate for validation because they involved an independent sampling of locations for virus activity.

This study provides an example of a multidisciplinary approach to rapidly identify the introduction site of an exotic infectious agent and to assess human risk in the surrounding community. The new technologies of RS and GIS, in conjunction with spatial statistics, offer effective tools for real-time surveillance and management of introduced vector-borne pathogens.

ACKNOWLEDGMENTS

This work was supported by NASA Headquarters under the Earth Science Fellowship Grant NGT5-01-0000-0205 (to J.S.B.), the National Science and Engineering Research Council of Canada (to J.S.B.), and USDA-ARS Cooperative Agreement 58-0790-2-072 (to D.F.).

ABBREVIATIONS

ANOVA, analysis of variance; AUC, area under the curve; GIS, geographic information system; NDVI, normalized difference vegetation index; NYC, New York City; ROC, receiver operating characteristic; RS, remote sensing; TM, Thematic Mapper; WN, West Nile.

REFERENCES

- Anyamba, A, Linthicum, KJ, Tucker, CJ. Climate-disease connections: Rift Valley fever in Kenya. *Cad Saude Publica* 2001; 17(Suppl):133-140.
- Beck, LR, Rodriguez, MH, Dister, SW, Rodriguez, AD, et al. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *Am J Trop Med Hyg* 1994; 51:271-280.
- Bernard, KA, Maffei, JG, Jones, SA, Kauffman, EB, et al. West Nile infection in birds and mosquitoes, New York State, 2000. *Emerg Infect Dis* 2001; 7:679-685.
- Centers for Disease Control and Prevention. Serosurveys for West Nile virus infection—New York and Connecticut Counties, 2000. *MMWR Morb Mortal Wkly Rep* 2001; 50:37-39.
- Centers for Disease Control and Prevention. Provisional surveillance summary of the West Nile virus epidemic—United States, January—November 2002. *MMWR Morb Mortal Wkly Rep* 2002; 51:1129-1133.
- Dister, SW, Fish, D, Bros, SM, Frank, DH, et al. Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. *Am J Trop Med Hyg* 1997; 57: 687-692.
- Dwass, M. Modified randomization tests for nonparametric hypotheses. *Ann Math Stat* 1957; 28:181-187.
- Fish, D. A systems approach to the control of Lacrosse virus. In: Calisher, CH, Thompson, WH, eds. *California Serogroup Viruses*. New York: Alan R. Liss; 1983:343-353.
- Fish, D. Remote sensing and landscape epidemiology. *Adv Astronaut Sci* 1996; 91:1057-1063.
- Hayes, C. West Nile fever. In: Monath T, ed. *Arboviruses: Epidemiology and Ecology*. Boca Raton, FL: CRC Press; 1989:59-88.
- Hjalmar, U, Kulldorff, M, Gustafsson, G, Nagarwalla, N. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Stat Med* 1996; 15:707-715.
- Hubalek, Z, Halouzka, J. West Nile fever—a reemerging mosquito-borne viral disease in Europe. *Emerg Infect Dis* 1999; 5:643-650.
- Hulburt, HS. West Nile virus infection in arthropods. *Am J Trop Med Hyg* 1956; 5:76-85.
- Kitron, U. Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *J Med Entomol* 1998; 35:435-445.

- Kulldorff, M. A spatial scan statistic. *Commun Stat-Theor M* 1997; 26:1481-1496.
- Kulldorff, M, Nargarwalla, N. Spatial disease clusters: detection and inference. *Stat Med* 1995; 14:799-810.
- Kulldorff, M, Rand, K, Williams, G. *SaTScan*, v1.0. Bethesda, MD: National Cancer Institute; 1996.
- Lanciotti, RS, Roehrig, JT, Deubel, V, Smith, J, et al. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* 1999; 286:2333-2337.
- Nasci, RS, White, DJ, Stirling, H, Oliver, J, et al. West Nile virus isolates from mosquitoes in New York and New Jersey, 1999. *Emerg Infect Dis* 2001; 7:626-630.
- Nash, D, Mostashari, F, Fine, A, Miller, J, et al. The outbreak of West Nile virus infection in the New York City area in 1999. *N Engl J Med* 2001; 344:1807-1814.
- Pavlovsky, EN. *Natural Nidality of Transmissible Diseases, with Special Reference to the Landscape Epidemiology of Zoonthropose*. Urbana, IL: University of Illinois Press; 1966.
- Pope, KO, Sheffner, EJ, Linthicum, KJ, Bailey, CL, et al. Identification of central Kenyan Rift Valley virus vector habitats with Landsat TM and evaluation of their flooding status with airborne imaging radar. *Remote Sens Environ* 1992; 40:185-196.
- Randolph, SE. Ticks and tick-borne disease systems in space and from space. *Adv Parasitol* 2000; 47:217-243.
- Robinson, T, Rogers, D, Williams, B. Univariate analysis of tsetse habitat in the common fly belt of southern Africa using climate and remotely sensed vegetation data. *Med Vet Entomol* 1997; 11:223-234.
- Savage, HM, Ceianu, C, Nicolescu, G, Karabatsos, N, et al. Entomologic and avian investigations of an epidemic of West Nile fever in Romania in 1996, with serologic and molecular characterization of a virus isolate from mosquitoes. *Am J Trop Med Hyg* 1999; 61:600-611.
- Tempelis, CH. Host-feeding patterns of mosquitoes, with a review of advances in analysis of blood meals by serology. *J Med Entomol* 1974; 11:635-653.
- Tsai, TF, Popovici, F, Cernescu, C, Campbell, GL, et al. West Nile encephalitis epidemic in southeastern Romania. *Lancet* 1998; 352:767-771.
- Turell, MJ, O'Guinn, M, Oliver, J. Potential for New York mosquitoes to transmit West Nile virus. *Am J Trop Med Hyg* 2000; 62:413-414.
- U.S. Department of Agriculture. Summary of West Nile Virus in the United States. 1999. Available at: <http://www.aphis.usda.gov/vs/ep/WNV/summary.html>. Accessed August 22, 2000.
- Wood, BL, Beck, LR, Washino, RK, Hibbard, KA, et al. Estimating high mosquito-producing rice fields using spectral and spatial data. *Int J Remote Sens* 1992; 13:2813-2826.

Address reprint requests to:

Dr. Durland Fish
 Department of Epidemiology and Public Health
 School of Medicine
 Yale University
 60 College Street
 P.O. Box 208034
 New Haven, CT 06520-8034

E-mail: durland.fish@yale.edu