Temporal Variations in the Distribution of West Nile Virus Within the United States; 2000–2008

Alexander Carnes · Yelena Ogneva-Himmelberger

Received: 13 June 2010 / Accepted: 17 May 2011 / Published online: 25 May 2011
© Springer Science+Business Media B.V. 2011

Abstract West Nile Virus (WNV) is a serious illness that has affected thousands of people in the United States. Over 1,000 disease related deaths have occurred since its introduction to American soil in 1999. Spatial statistics are used to analyze distributional trends of human WNV cases from 2000 to 2008 through four analyses: Weighted Mean Center, Standard Deviational Ellipses, Global Moran’s I, and Getis-Ord-Gi* statistic (hot spot analysis). We conclude that the directional trend in cases has been from East to West with the area affected increasing with time. Hot spot analysis reveals that recurring counties with a high number of human cases have been in the metro areas of large cities. However, normalized results indicate that the rate of humans showing symptoms of WNV is greatest in rural areas, particularly the Great Plains. These results provide a foundation for future research in analyzing the most persistent hot spots in more detail. Furthermore, these findings may aid decision makers in identifying areas to target for mitigation strategies such as spraying, larval control, and public awareness.

Keywords West Nile Virus · Spatial statistics · Hot spot analysis · United States · Health GIS

Problem Statement

West Nile Virus (WNV) is a vector borne disease that is contracted through the bite of a mosquito. The virus originated in Uganda in 1937 (Smithburn et al. 1940) and is
indigenous to Asia, Africa, Europe, and Australia (Campbell et al. 2002). The first human case of WNV in the United States occurred in 1999 in New York City, representing the first infection within the Western Hemisphere (Nash et al. 2001). The virus spread rapidly throughout the country in the following years. By 2002, WNV activity had been reported in all but four of the lower continental states (Mostashari et al. 2003). WNV is an unusual disease due to the fact that 80% of all people who contract the virus do not have symptoms (CDC 2009a). Of the 20% that do see symptoms, the effects are typically mild and include symptoms such as nausea, headache, fever, and vomiting. The symptoms typically last about one week. In contrast, about 1 in 150 people who contract the virus experience severe symptoms. These include encephalitis (acute inflammation of brain), meningitis (inflammation of membranes covering the brain and spinal cord), paralysis, high fever, vision loss, convulsions, and death. The death rate from the virus is estimated at around 4% with 1,131 deaths out of 28,691 total cases from 2000 to 2008 (CDC 2009a). Many studies including Campbell et al. (2002) and O’Leary (2004) have discovered that WNV incidence and death are strongly correlated with the elderly.

Given the severity of the acute symptoms associated with West Nile Virus, and the fact that there are millions of mosquitoes biting millions of Americans every summer, several questions can be asked about the distribution of human WNV cases. First, what has the spatial trend of WNV cases been over the last 10 years? Second, is the distribution of the cases clustered? And third, where are the statistically significant clusters (hot spots) of WNV activity located? WNV data has been already mapped by several authors and organizations at both local and national scales. Guptill et al. (2003) mapped cases of both avian and human WNV in the US. In Canada, the Integrated System for Public Health Monitoring of West Nile Virus (ISPHM-WNV) was developed in 2003 to provide real time information and mapping on WNV in the province of Quebec (Gosselin et al. 2005). Griffith (2005) mapped percentages of WNV cases resulting in death by state. Studies focusing on avian cases, however, are considered to be too general because sampling schemes vary by county. In addition, such schemes do not adequately describe the pattern of WNV because they merely signal the presence or absence of the disease, and not the severity (Cooke et al. 2006).

Mapping WNV is a good way to visualize patterns, but it does not provide any statistical backing to support a particular hypothesis. In many cases, the next step has then been to use spatial statistics (SS) to monitor clustering of West Nile Virus cases. Most SS studies have emphasized the disease in birds and focused on major cities as opposed to nationwide studies. New York City, the site of the first human WNV cases, has been the site of several studies including Brownstein et al. (2002) and Mostashari et al. (2003). Brownstein and colleagues performed cluster analysis of geocoded WNV cases within the 5 boroughs of New York City using Kulldorff’s scan test. The study then used remote sensing data from Landsat Thematic Mapper to produce maps of normalized difference vegetation index (NDVI) for the region. From these two input data sources, high risk census tracts were identified and mapped. Mostashari and colleagues, using a spatial scan statistic, examined dead birds in relation to human cases for the first two years after the initial outbreak. The goal of this study was to provide an early warning system for the year 2001 based on clustering of dead birds in 2000. In this case, the number of recently observed dead
birds (cases) inside and outside the potential cluster is compared with the expected number of cases inside of that circle. Using a Bernoulli-based likelihood derived from the observed vs. expected, the circle with the maximum likelihood is defined as the most unusual cluster, which is least likely to occur due to random chance (Mostashari et al. 2003). P values, indicating statistical significance, were derived and rank-ordered with respect to 999 randomized datasets. The results of the analysis found that Staten Island has been a persistent cluster, with 10 neurologic illnesses among humans (Kulasekera et al. 2001). This allowed for implementation of increased larval surveillance and abatement of standing water early in the 2001 season in the month of June (Mostashari et al. 2003). In addition, the New York City Department of Health (NYCDOH) developed a comprehensive program dedicated to the surveillance, control, and prevention of WNV through a citywide network of mosquito traps and dead bird sightings (Kulasekera et al. 2001). Geocoding of disease outbreaks cases coupled with spatial scan statistics have also been implemented by Wey et al. (2009) for WNV in South Dakota and Carrel et al. (2009) for cholera in Bangladesh. Similar to these previous works, the following study will incorporate the cluster identification methods and significance testing with the study area expanded to the lower 48 states. WNV clustering at the national scale has been analyzed by Sugumaran and colleagues (2009) by using Anselin’s Local Moran’s I and Kulldorff’s spatial scan statistic for rates of human WNV by county for 2002 to 2008. Our study follows a similar approach but uses the Getis-Ord Gi* statistic for both original count and normalized data. This statistic is potentially a more efficient method for cluster analysis because Local Moran’s I outputs do not specify whether a particular cluster consists of high or low values. The computation of hot spot persistence is touched upon briefly by Sugumaran but explained in more detail within this paper.

Alternatively, attempts to identify high risk areas and provide explanations for distributional patterns of WNV have been explored. Cooke et al. (2006) map areas of Mississippi that are environmentally suitable (and of high risk) for mosquito habitat using four variables: road density, stream density, slope, and NDVI. A similar study was performed for the Canadian province of British Columbia using basic geographic and temperature data (Tachiiri et al. 2006). Variables such as elevation, percentage of vegetation cover, land cover, temperature, housing density, and slope have been the most commonly used variables for such vulnerability analyses using regression techniques. A study on avian WNV distribution in Georgia using logistic regression reveals that the probability of a site to be WNV positive was highest in high density housing in urban/suburban areas (Gibbs et al. 2006). However, the reported values of the coefficients of determination were low (.239 in 2002, 0.028 in 2003, and 0.072 in 2004), showing a decrease in model performance over time. These variables explain less than 25% of the variance in the location of WNV cases. Landesman et al. (2007) notes that WNV incidence is correlated with precipitation from the previous year, and that outbreaks in the eastern half of the county are correlated with above average precipitation, while western outbreaks are associated with drought conditions (Landesman et al. 2007). However, this method was only applied for 2002–2004, and has not been verified with other years of data. Another study by DeGroote et al. (2008) found a relationship between WNV and below average precipitation in Iowa for certain years of their study, but found that in other
years, the relationship was not present. Ruiz and colleagues (2004) analyzed West Nile Virus in Chicago, IL. The Local Moran’s I was calculated for age adjusted rate of cases per 10,000 people in 2002 to identify clusters of WNV within Cook and DuPage Counties. A significance statistic was also calculated and confirmed that 5.6% (83 of the 1481) of the census tracts were hot spots at the 95% confidence level. This sort of significance testing is crucial for spatial statistics analysis. Lastly, two regression equations, using socioeconomic and physical environment variables, were used to predict the probability that a census tract would contain a cluster of WNV. Factors such as vegetation, age, income, race of the human population, distance to a WNV positive dead bird specimen, Mosquito Abatement Districts, and geology were including within the regression equation (Ruiz et al. 2004). Ruiz found that 53% of the existing clusters can be explained by these variables. While our method for investigating WNV cases is at a larger scale, the methods presented in the previous section should be considered for micro level studies in counties identified based on our nationwide hot spot results.

To reiterate, the purpose of this study is to examine the spatial and temporal distribution of West Nile Virus within the continental United States from the year 2000 to 2008 at the county level. The spatial characteristics of WNV are analyzed for both count and normalized (to 2000 population) data through the use of four spatial statistics: weighted mean center, standard deviational ellipse, global Moran’s I statistic and Getis-Ord Gi* statistic. The last statistic, Getis-Ord Gi*, is used to identify hot spots for each year of the analysis. Through an overlay analysis, the temporal aspect of WNV is determined based on the persistence of WNV in a particular county. Since county level analysis is not an appropriate scale for determining detailed areas for mitigation and allocation of funds, the results of this paper serve as a starting point for future in depth, local level research. The methods presented in the previous section, along with other explanatory WNV studies, are recommended when investigating finer scale WNV.

Data

WNV data for this study were found via five sources. First, data for West Nile Virus cases were found on the Centers for Disease Control and Prevention (CDC) and United States Geological Survey (USGS) websites. The CDC provided the counts of severe human WNV cases by state and year including the type of symptoms (fever, encephalitis, and meningitis) and number of deaths (CDC 2009b). Table 1 shows an aggregate of all states and all cases by year. The second source of WNV data was the USGS website, which provided the number of cases by year by county for all years from 2002 to 2008 (USGS 2009). Lastly, data for 2000 and 2001 were found on the West Nile Virus Surveillance page published by Nationalatlas.gov. Data for 1999 were excluded from the analysis because the disease was confined to New York City, while the data record for human WNV in 2009 was incomplete at the time of the study. A map of number of cases for all years in the study period is shown in Fig. 1. Likewise, the normalized rate of WNV infection per 10,000 people is shown in Fig. 2. Of particular note is the year 2003. This year represents the most widespread outbreak of the disease, with nearly 10,000 human cases. This coupled with the
outbreak of Severe Acute Respiratory Syndrome (SARS) marks an unprecedented time in the history of environmental epidemiology and public health (Allen and Wong 2006). While choropleth maps effectively show the spatial distribution of cases, they are not considered statistically significant ways to determine clusters because minor changes in the choice of class intervals can greatly change how a map is viewed (Unwin 1996).

Table 1  Human cases and death totals by year

| Year | Cases | Deaths |
|------|-------|--------|
| 1999 | 62    | 7      |
| 2000 | 21    | 2      |
| 2001 | 66    | 10     |
| 2002 | 4156  | 284    |
| 2003 | 9862  | 264    |
| 2004 | 2539  | 100    |
| 2005 | 3000  | 119    |
| 2006 | 4269  | 177    |
| 2007 | 3630  | 124    |
| 2008 | 1356  | 44     |
| Totals | 28961 | 1131 |

Fig. 1  The number of human WNV cases for 2000 to 2008
Methodology

The distribution of WNV cases was analyzed first with the original count data and then with normalized data. Data were normalized by 2000 census information to give a rate of WNV incidence per 10,000 people. This process removes the biases created by urban centers, which have higher counts solely as an artifact of higher population. Normalized data, however, are only used for the Getis-Ord Gi* statistic, which is explained later. All other analyses were computed with the original count data. First, to analyze the spread of WNV over time, the Weighted Mean Center and Standard Deviational ellipses were calculated for each year, with the number of occurrences of WNV in a given year used as the weight. The product of the weighted mean center analysis is a point representative of the center of nationwide WNV cases. Next, using the weighted mean centers as inputs, standard deviational ellipses were calculated by squaring the sums of the differences between the X and Y coordinates of each county’s centroid and the X and Y of the weighted mean center. The result is an ellipse that measures the degree to which features are concentrated or dispersed around the weighted mean (Mitchell 2005). If the data is normally distributed, one standard deviation ellipse will cover approximately 68% of the counties that had WNV in a given year.

Global Moran’s I and Getis-Ord Gi* statistics tests were performed next. The Global Moran’s I statistic determines if clustering exists in the dataset as a whole, while the Gi* identifies clusters of high and low values (Mitchell 2005).
demonstrate throughout this paper, this statistic has been proven to be able to reveal both the spatial structure of year to year disease outbreaks as well as identify the temporal stability of disease presence (Hinman et al. 2006). Prior to each test, the weight (or influence) of neighboring polygons (in this case, counties) on the result of the target polygon must be defined. These conceptualizations of spatial relationships can be broken into two major categories: contiguity based and distance based. Contiguity based conceptualizations focus on polygons that share a border with the target polygon. It is also possible to accommodate higher order contiguity (i.e. neighbors of neighbors). Polygons that do not share a border with the target polygon are not considered neighbors and are excluded from the analysis for that particular feature. In contrast, distance based applications use a threshold distance at which features stop influencing target feature. For this category of conceptualizations, however, several options exist for how neighboring features are weighted. In the case of fixed distance band, a user specified distance creates a moving window in which features within that window influence the result of the target feature. In contrast, inverse distance conceptualizations weight adjacent features more heavily, with a sharp decrease in influence with distance. In both cases, features beyond the threshold distance are excluded from the calculation. Alternatively, one can use a “zone of indifference” conceptualization, which combines the fixed distance and inverse distance methods. The main difference here is that the user specified threshold does not represent a cutoff, but rather signifies the distance at which the influence begins to drop off. Once the weights are calculated for each feature, it is important to row standardize them because each county has a different number of neighbors. For distance based neighborhoods, this is done by dividing the distance to each feature by the total distance to all neighbors for the row in the weight matrix. The effect of this is that distances are scaled to a range of 0 to 1 as opposed to absolute distances (Mitchell 2005).

The choice of conceptualization of spatial relationships proved to be a challenging part of the analysis. United States counties range from very small in the East to quite large in the West. A distance across a single county in the West may be large enough to cross five or six counties on the East Coast. For this reason, we considered distance based definition of neighborhood more appropriate than contiguity based. Based upon the literature we encountered, the fixed distance band was recommended because we had large variations in polygon size, and the fact that fixed distance is preferred for hot spot analysis (ESRI 2009). Upon using this method for our dataset, however, we discovered that the Gi* statistic found statistically significant hot spots in areas with very low numbers of WNV cases. In order to accommodate this problem, we found it intuitive to use the Inverse Distance conceptualization. The distance that produces the highest Z score for the global Moran’s I (representing the greatest spatial autocorrelation) is recommended to be used as the distance threshold for the hot spot tool (ESRI 2009). This was accomplished by calculating the global Moran’s I at different distances for each year, using Inverse Distance. The distance at which the global Moran’s I value peaked was used as the input threshold distance for the Gi* statistic.

Following this initial step, the statistics are run. For the Global Moran’s I tool, the outputs are a Moran’s I value and a Z score which tests for statistical significance. Moran’s I compares the differences from the mean for each pair of
features, producing a cross product. A high cross product indicates clusters of similar values. Next, the observed I is subtracted from the expected I and divided by square root of the variance in the entire dataset. The higher the Z value, the less likely the observed pattern is occurring due to random chance (Mitchell 2005). Clustering is more evident through the Z scores than the Moran’s I value itself. When a Z score is greater than 1.96, it indicates that there is less than a 5% chance that the clustering is due to chance. A score of 2.58 indicates that there is a 1% chance that the clustering is due to chance. However, unlike global measures such Moran’s I or Geary’s C, Getis–Ord’s Gi* is capable of revealing local clustering of specific phenomena (Getis and Ord 1992). In our case, Getis-Ord Gi* statistic was utilized to find hot spots (clustering of high counts and high rates of WNV) in the data for each year. The Gi* is calculated for each polygon by summing the values of the target-neighbor pairs (as defined in the previous paragraph) and dividing by the sum of all features’ values in the study area. The * indicates that the target feature is included in the analysis. Therefore, a high Gi* indicates a clustering of high values. A Z score output testing for statistical significance is again produced in the same manner as the Moran’s I test, where the difference between the expected Gi* and observed Gi* is divided by the square root of the variance for all features in the study area.

Results

Spatial Distribution of Cases

The weighted mean centers (Fig. 3) and standard deviational ellipses (Fig. 4) indicate that the trend in human WNV cases has moved from east to west. While it is not alarming that the disease spread across the country, the rate of westward expansion is noteworthy. In 2000, the center was located in the New York City area, followed by a southerly shift to North Carolina in 2001 as the virus spread throughout the East coast. The virus then spread rapidly westward during 2002 to 2004 from North Carolina to the Midwest, and ultimately the Four Corners region of Colorado, Utah, Arizona, and New Mexico. During this three year period, the X coordinate of the mean center shifted 1,597 miles westward (515 miles between 2001 and 2002, 620 miles 2002 to 2003, and 462 miles between 2003 and 2004), traversing about 3/5ths of the length of the country. Since 2005, the center has been situated in area of the continental divide. In 2004, this point was situated in northwestern New Mexico (in response to the first outbreak in California), but then moved to northeastern Colorado for three years between 2005 and 2007. The mean center for 2008 moved back into southwestern Colorado. Standard deviation ellipses have changed in size and transitioned from north-south elongation (2001) to west-east (2004-present). Table 2 shows the axes lengths, rotation, and area (square miles) of each ellipse. From this table one finds that the west-east axis is double to triple the length of the north-south axis by the end of the study period. The size of ellipses also increases over time from around 2 million square miles in 2001–2003 to between 4–5 million square miles between 2005 and 2008. Therefore, these ellipses confirm the suspected idea that the extent of WNV has increased over time and that the
Temporal Variations in the Distribution of West Nile Virus

Fig. 3 Weighted mean center of human WNV cases by year (with Trend Line)

Fig. 4 Standard deviational ellipses of WNV distribution (1 Std Dev) by year
distribution relative to the mean center is more focused on the east-west axis as opposed to the north-south axis.

Spatial Autocorrelation

The results of the analyses revealed some interesting results. First, the global Moran’s I statistic proved that human WNV cases in both datasets (number of cases and normalized) are clustered every year. All global Moran’s I Z scores are well above the 2.58 threshold (representing the 1% confidence interval), ranging from 14.01 in 2002 to 164.51 in 2003 for the number of cases and 44.08 in 2008 to 207.11 in 2003 for normalized results. This is true because in a given year, most counties had either zero or one case. Also, in places where there were higher numbers of human WNV cases, adjacent counties tended to have similar numbers of cases.

The hot spot analysis for the numbers of cases indicated that the disease is most prevalent in major cities (Fig. 5). There are two categories in this map—the category of greatest interest being that where clustering of high values is statistically significant, with a Z score greater than 2 (indicated in black). Hot spots start near New York City in 2000, and then spread throughout the Eastern U.S. in 2001, then target major cities in the Midwestern U.S. by 2002. The most affected counties during that year were Cook County (Chicago, IL), Cuyahoga County (Cleveland, OH), and Wayne County (Detroit, MI). Other cities with high numbers of cases are St. Louis, MO, Dallas, TX, and Houston, TX. From 2003 onward, the Great Plains (North Dakota southward to Texas) and eventually the Western U.S. become hot spots while the Eastern U.S. has very few hot spots. In 2003, four counties in Colorado (Larimer, Boulder, Weld, and Adams) had over 200 cases each. 2004 represents the first outbreak within the southwest. Maricopa County (Arizona) and Los Angeles County (California) have the highest numbers with 357 and 306 cases respectively during that year. 2005 represents the spread throughout California, particularly in the Central Valley (Sacramento, Stanislaus, Kern, and Fresno Counties). A very interesting hot spot appears in 2006 at the borders of Oregon,

Table 2 Characteristics of standard deviational ellipses by year

| Year | East-West Axis (Miles) | North-South Axis (Miles) | Direction (° from North) | Area (Sq. Miles) |
|------|------------------------|--------------------------|--------------------------|------------------|
| 2000 | 24                     | 41                       | 15                       | 3,122            |
| 2001 | 369                    | 1385                     | 15                       | 1,604,032        |
| 2002 | 679                    | 972                      | 40                       | 2,071,582        |
| 2003 | 1095                   | 717                      | 97                       | 2,462,703        |
| 2004 | 1789                   | 643                      | 90                       | 3,610,793        |
| 2005 | 2014                   | 836                      | 90                       | 5,287,922        |
| 2006 | 1707                   | 848                      | 107                      | 4,544,563        |
| 2007 | 1441                   | 915                      | 96                       | 4,142,055        |
| 2008 | 2133                   | 733                      | 85                       | 4,909,159        |
Idaho, and Nevada. Prior to and after 2006, this area is not a hot spot. The outbreak in 2007 is more widespread, occurring in scattered locations throughout Colorado, the Dakotas, California, and Arizona. Los Angeles and Maricopa Counties are again the top two counties for WNV in 2008.

Hot spot analysis for normalized data (WNV infection rate per 10,000 people) reveals that the highest rates of WNV are concentrated in rural areas, particularly the Great Plains (Fig. 6). Again, areas in black indicate hot spots and Z scores greater than 2. Areas such as the southwest (Los Angeles and Phoenix areas) and previously mentioned metropolitan areas do not exhibit significant clustering of high values. On a year to year basis, there is little variation between the 2000 and 2001 maps, due to the lack of WNV cases in most of the country during these two years. The number of hot spots increases dramatically in 2002 with the highest concentrations of WNV hot spots in Nebraska or South Dakota. Another noticeable hot spot occurs in Louisiana and Southern Mississippi. In 2003, one large hot spot dominates the states of Montana, North Dakota, South Dakota, Nebraska, and Colorado. The distribution of hot spots is more widespread in 2004, although this same general area retains hot spots in addition to areas of Texas and Western Colorado. Hot spots are most dominant in the Dakotas, Nebraska, and Colorado for 2005 and 2007. The large hot spot for the number of cases at the borders of Oregon, Idaho, and Nevada that is discussed earlier is again revealed with the normalized data. 2008 hot spots are again more scattered, although six of the top 10 rates of WNV during this year are found within counties in Kansas or Colorado. It can be seen through data normalization

Fig. 5 Hot spots of human WNV cases by year, 2000–2008
that the highest rates of human WNV are concentrated in rural areas, and not metropolitan areas.

Persistent Hot Spots

The final step of the analysis was to identify counties that were persistent hot spots for the entire study period. This was accomplished by establishing a Z score threshold for the Getis-Ord Gi* results. Counties that received a Z score above 2.0 were assigned a 1 to indicate that the county was a statistically significant hot spot for that year. This step was repeated for all years, with the results aggregated to reveal the number of years each county was a hot spot for the study period of 2000 to 2008. The results of this exercise are shown in Table 3 and Fig. 7. It can be seen through this figure that Southern California has been a hot spot (in terms of the number of cases) 4 years out of the 9 year study period. However, it is important to note that the first severe outbreak in California did not occur until 2004. Similarly, Harris County (Houston), TX was a hot spot 6 times out of 7 years (if we exclude 2000 and 2001 because WNV was found primarily in the Eastern U.S. during those years). Other metropolitan areas such as Denver, CO, Chicago, IL, Dallas, TX, and Jackson, MS have been hot spots for at least four years of the study period. A similar methodology was applied to the normalized data. Unlike the results for count data, the persistent normalized hot spots were located in rural areas, particularly in the Great Plains states of North Dakota, South Dakota, and Nebraska (Fig. 8). In fact, of

Fig. 6 Hot spots of normalized human WNV cases by year, 2000–2008
the 12 counties that were hot spots for 5 years out of the study period, 10 were located in these three states, with the exception of Phillips County (CO) and Goshen County (WY) (Table 4).

Table 3  Characteristics of the counties that are the most persistent hot spots based on absolute number of cases

| Name            | # Years hotspot | Pop 2000   | Total cases 2000–2008 |
|-----------------|-----------------|------------|-----------------------|
| Harris (TX)     | 6               | 3,302,269  | 301                   |
| Maricopa (AZ)   | 5               | 2,942,946  | 680                   |
| Weld (CO)       | 5               | 171,203    | 613                   |
| Larimer (CO)    | 5               | 242,402    | 725                   |
| Hinds (MS)      | 5               | 244,562    | 151                   |
| Pima (AZ)       | 4               | 816,221    | 100                   |
| Kern (CA)       | 4               | 649,311    | 317                   |
| Los Angeles (CA)| 4               | 9,424,833  | 553                   |
| Riverside (CA)  | 4               | 1,574,240  | 296                   |
| Boulder (CO)    | 4               | 278,798    | 622                   |
| Cook (IL)       | 4               | 5,192,969  | 940                   |
| Burleigh (ND)   | 4               | 67,930     | 240                   |
| Dallas (TX)     | 4               | 2,085,177  | 246                   |

Fig. 7  Number of years each county is a hot spot (Z score > 2) for count data
Discussion and Conclusion

Urban-rural differences were also discovered in Cooke’s study of West Nile Virus in Mississippi, where WNV counts were centered near the major urban centers of Jackson, Hattiesburg and Clarksdale. In contrast, normalized data revealed that the

Table 4 Characteristics of the counties that are the most persistent hot spots based on normalized data

| Name         | # Years hotspot | Pop 2000 | Total cases 2000–2008 |
|--------------|-----------------|----------|-----------------------|
| Phillips (CO)| 5               | 4,187    | 33                    |
| Morrill (NE)| 5               | 5,239    | 22                    |
| Boone (NE)  | 5               | 6,321    | 35                    |
| Phelps (NE) | 5               | 9,785    | 60                    |
| Furnas (NE) | 5               | 5,383    | 34                    |
| Mercer (ND) | 5               | 9,097    | 41                    |
| Morton (ND) | 5               | 24,607   | 98                    |
| Dickeys (ND)| 5               | 5,657    | 34                    |
| Brown (SD)  | 5               | 35,072   | 200                   |
| Kingsbury (SD)| 5             | 5,705    | 21                    |
| Tripp (SD)  | 5               | 6,539    | 26                    |
| Goshen (WY) | 5               | 12,577   | 113                   |
highest incidence rates (per 10,000 people) occurred in rural areas. Thus, it was speculated that a population bias existed within the data. However, a significance test examining the correlation between the number of human cases and human population was insignificant (Cooke et al. 2006). In a similar light, the hot spot results should not be interpreted as competing against one another, as they represent two distinctly different analyses: one to assess the absolute magnitude of the problem (with population bias) and another to assess the incidence rate (accounting for population). While no significance testing of the correlation between WNV counts and population was performed, it is clear that the virus is more prevalent in urban centers, solely because there are more people in these areas.

The cities and counties identified in this study, due to their persistent presence of the virus, should warrant more detailed studies (such as those described in the Problem Statement section) to identify high risk areas at a finer scale. Mitigation and funding will only come after further analyses that pinpoint specific areas (hot spots within the hot spot) of WNV activity in each county, along with an understanding of the factors that cause the hot spot to occur. Therefore, should further study be focused where the number of cases is highest? Or should the counties with the highest rate of human WNV be the main focus? Regardless, the media and local news affiliates have the power to enhance their public awareness efforts by explaining the dangers of mosquitoes and diseases including WNV. The allocation of funds and delineation of areas of feature study, on the other hand, is a more complicated issue. Undoubtedly, the most attention will be put on areas where there is a greater population at risk. In which case, urban areas will become the priority.

The South, which is very humid compared to the Midwest and the West, had very few cases and hot spots for the entire study period. This is particularly puzzling as mosquitoes prefer warm, moist environments. However, this may be reflective of the types of mosquitoes and avian species present in the South and their ability to transmit the disease, as vector proficiencies vary from region to region. In addition, the number of cases tends to be highest when a disease passes through an area for the first time. In later years, increased vigilance and control, along with resistance within the population will naturally diminish the severity of outbreaks over time. For example, in areas of Africa where the disease is prevalent, immunity has been reported in up to 90% of adults and 50% of children after an initial outbreak (Campbell et al. 2002). This would help to explain the lack of WNV instances throughout the eastern US at the end of the decade. As a result, existing WNV mitigation practices have resulted in a 5-fold decline in WNV neuroinvasive disease (from 1.02 to 0.23 cases per 100,000) from 2002 to 2008 (Hitt 2010), attributing declines to increasing immunity (in both people and animals) and the increased mosquito control (Doan 2006). Sardelis et al. (2001) notes in their study of vector competence for WNV, that there is considerable variation in vector competence between mosquito species. Two mosquito species, *Culex restuans* and *Culex salinarius*, which tested positive during the 2000 outbreak in New York City, were determined to be efficient laboratory vectors (between chickens and mosquitoes), with transmission rates between 34% and 55%. In contrast, *Culex quinquefasciatus* and *Culex nigripalpus*, found in Florida, were found to be competent but only fair to moderately efficient vectors, with transmission rates of 20% or lower (Sardelis et al. 2001). However,
the transmission to human beings remains unclear (Campbell et al. 2002). In addition, population density, host preference, feeding behavior, longevity, and seasonal activity of each mosquito species are factors that affect the rate of transmission (Sardelis et al. 2001). Therefore, a second area of future research is to determine the vector competence for mosquito species in the persistent hot spots.

Progress on the policy implementation front was demonstrated in 2007, when California governor Arnold Schwarzenegger allocated $10 million for WNV mitigation through mosquito abatement, surveillance, vegetation management, and expansion of public education (California Department of Public Health 2007). In addition, protocols for reporting, collecting, and testing of dead birds infected with the virus have been established in 48 states as well as the District of Columbia (Mostashari et al. 2003). In the future, Mostashari and colleagues note that identified counties with high rates of infection can be prioritized for spraying and larval control using EPA approved pesticides. Further mitigation efforts based on high risk areas include the state of Illinois, which supplied $200,000 of funding to WNV efforts in 2005 for the greater Chicago area (Illinois Department of Public Health 2005). The state of Pennsylvania will allocate $4.38 million to its West Nile program for fiscal year 2011 (Singleton 2010) to support spraying activities. However, this number is down from $5.1 million for fiscal year 2010 in response to the current recession. The state of Connecticut is another state experiencing funding cuts, as their funding was slashed by $100,000 in 2009 (Latina 2009). Despite these challenges, states are doing more with less. For example, the Pennsylvania DEP has done extensive work over the years prioritizing areas at a finer scale (as we recommend) in order to maximize resources (Singleton 2010). In addition to spraying and surveillance efforts, the media should broadcast public service announcements explaining the recommended precautions to take in order to minimize exposure to mosquito bites. These include wearing long sleeves and long pants with an effective mosquito repellent, avoiding post-dusk activity, and removal of stagnant water sources (such as bird baths and tires) (Illinois Department of Public Health 2005). A plethora of government websites have pages similar to those of Illinois explaining the risks and recommended precautions noted above.

As an ecological study, the preceding work has some limitations. While WNV reporting comes from reliable sources such as the CDC, it is important to remember that less than a quarter of individuals who obtain the virus show symptoms. Therefore, this study only presents the counties where symptomatic WNV has occurred. Furthermore, WNV data consistency may be an issue as reporting standards vary from state to state and county to county. In addition, the number of reported cases may reflect a reporting bias towards metropolitan areas. These areas have the most money, expertise, and potential to put surveillance programs in place. As a result, urban areas may be better suited to find instances of WNV as opposed to rural areas. Data normalization for all years was done with 2000 census data (as the census is done every ten years) and not with new population data for each year. This study does not take into account the vulnerability of certain bird species to WNV, or other demographic, social, and environmental factors that may contribute to a higher or lower prevalence of WNV in certain counties. County-level analysis is insensitive to small-area clustering (Mostashari et al. 2003) and thus overlooks clustering at finer scales. Variation in county size (counties in the east vs. a state such as Arizona)
effects the results of distance based spatial statistics as some counties have more neighbors than others.

WNV, despite the recent decline in the number of human cases, remains an annual threat to humans throughout the contiguous 48 states of America. This study has used spatial statistics to identify directional trends in the number of WNV cases for the years 2000 to 2008. In addition, hot spots of human WNV activity were identified for the same time period. To our knowledge, this paper is one of the few to analyze the temporal dynamic of WNV at the county level for the entire country. We have identified several counties which have been persistent hot spots throughout the entire study period. Clustering for the dataset as a whole was found to be significant; indicating that most counties have very low values (0–2 WNV cases) and that counties with higher occurrences tend to be found next to one another. Hot spot analysis of the numbers of WNV cases reveals that the Western U.S. (particularly Southern California and Arizona), Colorado, and specific metro areas such as Chicago, IL, Houston, TX, and Detroit, MI have been areas where the disease has been most infectious. However, normalized data reveal that the greatest concentrations of WNV rates have been in rural areas, particularly in the states of Montana, North Dakota, South Dakota, and Nebraska.

It is recommended that more detailed research be devoted to the identified counties. The use of methodologies (involving data such as geocoded cases, NDVI, socioeconomic variables, and terrain) such as those performed by Ruiz et al., Mostashari et al., Brownstein et al., Cooke et al., and Gibbs et al., should be considered in these persistent hot spots. Confirmation of high prevalence areas will require the detection of local hot spots within the identified counties. This will allow county officials to utilize their WNV funds more efficiently and concentrate mitigation efforts. Forecasting of future WNV outbreaks will undoubtedly continue to be a challenge over the coming years. Changes in vectors, avian hosts, human, and environmental factors make future prediction of WNV transmission levels difficult (Hitt 2010). It is important to recall that only 20% of the people who contract WNV show symptoms. Should this percentage increase, state and local officials would know where to prioritize their defenses against the disease in the event of an outbreak.

Acknowledgement We thank two anonymous reviewers for their detailed and insightful comments, which helped us to improve this paper.

References

Allen, T. R., & Wong, D. W. (2006). Exploring GIS, spatial statistics and remote sensing for risk assessment of vector-borne diseases: a West Nile virus example. *International Journal of Risk Assessment and Management, 6,* 253–275.

Brownstein, J. S., Rosen, H., Purdy, D., Miller, J. R., Merlino, M., Mostashari, F., et al. (2002). Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne Zoonotic Diseases, 2,* 157–164.

Campbell, G. L., Marfin, A. A., & Lanciotti, R. S. (2002). West Nile Virus. *Lancet Infectious Diseases, 2,* 519–529.

Carrel, M., Emch, M., Sneathfield, P. K., & Yunus, M. (2009). Spatio-temporal clustering of cholera: the impact of flood control in Matlab, Bangladesh, 1983–2003. *Health & Place, 15,* 771–782.
Smithburn, K. C., Hughes, T. P., & Burke, A. W. (1940). A neurotropic virus isolated from the blood of a native of Uganda. *American Journal of Tropical Medicine and Hygiene, 20*, 471–92.

Sugumaran, R., Larson, S. R., & Degroote, J. P. (2009). Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. *International Journal of Health Geographics, 8*(43).

Tachiiri, K., Klinkenberg, B., Mak, S., & Kazmi, J. (2006). Predicting outbreaks: a spatial risk assessment of West Nile virus in British Columbia. *International Journal of Health Geographics, 5*(21).

The California Department of Public Health. (2007) Governor Schwarzenegger Orders Enhancements to Fight West Nile Virus [http://www.westnile.ca.gov/news.php?id=86](http://www.westnile.ca.gov/news.php?id=86) (last accessed 21 February 2010).

Union of British Columbian Municipalities (2010). 2010 West Nile Virus funding now available. [http://ubcm.ca/EN/meta/news/wnv-funding.html](http://ubcm.ca/EN/meta/news/wnv-funding.html) (Last Accessed 13 June 2010).

United States Geological Survey. (2009). *West Nile Virus historical maps* [http://diseasemaps.usgs.gov/wnv_historical.html](http://diseasemaps.usgs.gov/wnv_historical.html) (last accessed 17 December 2009).

Unwin, D. J. (1996). GIS, spatial analysis and spatial statistics. *Progress in Human Geography, 20*, 540–551.

Wey, C. L., Griesse, J., Kightlinger, L., & Wimberly, M. C. (2009). Geographic variability in geocoding success for West Nile virus cases in South Dakota. *Health and Place, 15*(4), 1108–1114. doi:10.1016/j.healthplace.2009.06.001.