

# Towards an Early Warning System for Forecasting Human West Nile Virus Incidence

March 6, 2014 · Research Article

## Citation

Manore CA, Davis JK, Christofferson RC, Wesson DM, Hyman JM, Mores CN. Towards an Early Warning System for Forecasting Human West Nile Virus Incidence. PLOS Currents Outbreaks. 2014 Mar 6 . Edition 1. doi: 10.1371/currents.outbreaks.ed6f0f8a61d20ae5f32aaa5c2b8d3c23.

## Revisions

This article is either a revised version or has previous revisions

Edition 1 - March 6, 2014 ▼

## Authors

[Carrie A. Manore](#)

Center for Computational Science, Tulane University, New Orleans, Louisiana, USA.

[Justin K. Davis](#)

School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, USA.

[Rebecca C. Christofferson](#)

Department of Pathobiological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA.

[Dawn M. Wesson](#)

School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, USA.

[James M. Hyman](#)

Department of Mathematics, Tulane University, New Orleans, Louisiana, USA.

[Christopher N. Mores](#)

Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, Louisiana, USA.

## Abstract

We have identified environmental and demographic variables, available in January, that predict the relative magnitude and spatial distribution of West Nile virus (WNV) for the following summer. The yearly magnitude and spatial distribution for WNV incidence in humans in the United States (US) have varied wildly in the past decade. Mosquito control measures are expensive and having better estimates of the expected relative size of a future WNV outbreak can help in planning for the mitigation efforts and costs. West Nile virus is spread primarily between mosquitoes and birds; humans are an incidental host. Previous efforts have demonstrated a strong

correlation between environmental factors and the incidence of WNV. A predictive model for human cases must include both the environmental factors for the mosquito-bird epidemic and an anthropological model for the risk of humans being bitten by a mosquito. Using weather data and demographic data available in January for every county in the US, we use logistic regression analysis to predict the probability that the county will have at least one WNV case the following summer. We validate our approach and the spatial and temporal WNV incidence in the US from 2005 to 2013. The methodology was applied to forecast the 2014 WNV incidence in late January 2014. We find the most significant predictors for a county to have a case of WNV to be the mean minimum temperature in January, the deviation of this minimum temperature from the expected minimum temperature, the total population of the county, publicly available samples of local bird populations, and if the county had a case of WNV the previous year.

## ***Funding Statement***

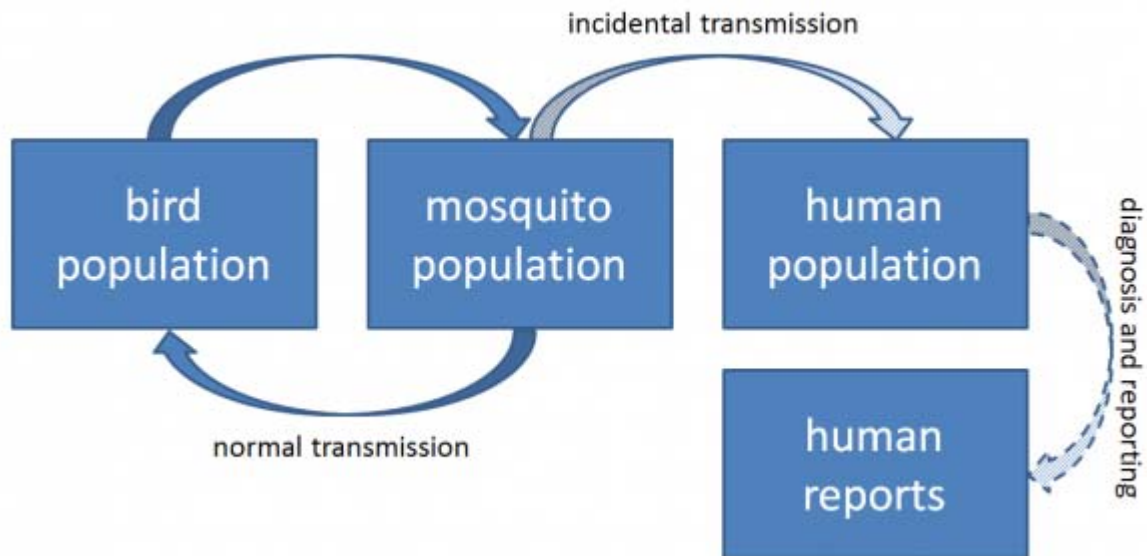
This work was funded through NIH/NIGMS U01GM097661 and NSF/CHE-1314029. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors have declared that no competing interests exist.

## ***Introduction***

West Nile virus (WNV) is a mosquito-borne flavivirus first identified in Uganda in 1937. WNV continues to be a public health hazard in the continental United States 13 years after its introduction into New York City in 1999 where it initially caused significant bird mortality and neuroinvasive disease (NID) in humans. In 2012, the USA experienced one of the worst years on record for human WNV <sup>1</sup>, with a higher than expected number of 5,674 cases reported to the CDC <sup>2</sup>; this is in contrast to the three-year period from 2009 to 2011, during which a total of 3,809 cases were reported. In 2012, 2,873 NID cases were reported from 976 counties across the 48 contiguous states, the District of Colombia and Puerto Rico. There were 270 fatalities among the NID cases <sup>3</sup>. Only about 20-30% of the people infected with WNV develop symptoms and less than 1% of people infected develop NID. The states with the highest incidence of WNV in 2012 were North and South Dakota, Louisiana, Texas, and Mississippi, indicating a large geographic range of intense transmission that year <sup>3</sup>.

Recent studies indicate that the genetic makeup of the virus isolated from the 2012 outbreak in Dallas is not significantly different from previously circulating isolates, and that all are of the WN02 genotype <sup>4</sup>. This suggests that a change in viral efficiency or virulence is not a major contributing factor of the 2012 outbreak. The increased incidence could be caused by extrinsic factors, such as extreme environmental conditions and the potential for these conditions to drive changes in the interaction or coincidence of bird, mosquito, and human populations. Some studies have correlated drought conditions to higher concentrations of mosquitoes and early migration of birds <sup>5</sup>. Migratory birds have been implicated as important in the spread of WNV virus <sup>6</sup>, while other studies have shown the highest seroprevalence in resident bird species <sup>7,8</sup>.

---



**Fig. 1: WNV Transmission Cycle**

Simplified model of West Nile virus transmission cycle. Human WNV activity is tangential to the enzootic, avian-centric transmission cycle. Reported cases of human WNV represent only a portion of the total transmission in the human population.

---

The WNV cycle is characterized in Figure 1. Birds and mosquitoes amplify and pass on the virus regularly. Humans are only inoculated by mosquitoes and do not, under normal circumstances, reach high enough viremia levels to pass virus back into the system <sup>9</sup>. Except from the perspective of human health, an examination of infection in the enzootic cycle—birds and associated mosquitoes—is likely more informative than any measure of human cases <sup>10</sup>. This is especially true as mosquitoes can be collected and tested in large batches. Indeed, as WNV spread across the country after introduction in 1999, mosquito abatement and control districts responded by partnering with state and district laboratories to test *Culex spp.* mosquitoes for the presence of the virus. Wild birds are less easily sampled in large numbers for WNV infection and thus mosquito infections and human WNV disease cases are often the only evidence of local WNV transmission.

Despite the availability of these samples, increases in surveillance efforts, and the renewed interest in WNV transmission, reliable predictive methods for human disease outbreaks remain elusive, owing to the milieu of interacting factors affecting overall transmission of WNV. We note specifically the review and novel results of a study in which incidence rates per county in years 2003 to 2008 were characterized by rule-based predictive models conditioned on meteorological and GIS data <sup>11</sup>. The study replicated results from locally conditioned studies and concluded that remotely sensed environmental variables could, with caveats, be used to predict WNV on a national scale.

The study noted a cluster of errors in the Northern Great Plains (NGP) region, which grew in magnitude when a locally conditioned model was attempted <sup>11</sup>. We suggest that a study of incidence without reference to measures of healthcare or similar intangibles will be susceptible to systematic biases (e.g. differences in testing and reporting practices) and that this was the case in that study. We note for example that the highest model errors corresponded to the areas of highest unemployment, which is likely a barrier to diagnosis.

The effect of these factors and their interaction on transmission—especially on tangential transmission into human populations—remains poorly understood. We create a predictive framework based on the statistical correlations among:

- the structure of the enzootic host community
- anthropological factors that
  - define differential contact rates of infectious mosquitoes within and tangential to the enzootic transmission cycle
  - influence the probability of diagnosing and reporting human cases
- environmental and weather conditions.

#### *Structure of enzootic host community*

The enzootic cycle of WNV is maintained by avian species and associated mosquitoes, especially *Culex* spp. Transmission characteristics of the avian hosts include host competence variability (relative infectiousness of bird species), susceptibility profiles of the avian communities, and their relative coincidence with human populations <sup>12,13</sup>. For example, small songbirds were found to be highly competent for WNV, indicating their role as likely amplification hosts <sup>14</sup>. Variability of viremia and competence among three bird species was observed in Louisiana, with sparrows having a higher competence for WNV than either cardinals or mockingbirds <sup>15</sup>. Also, WNV transmission may be different for migratory and residential bird populations. In India, wild resident water birds were infected with WNV more frequently than wild migratory birds <sup>16</sup>. In Virginia, the peak in human incidence occurred after the peak in the enzootic cycle in successive years, indicating the potential for a pattern between the enzootic cycle and tangential transmission into the human population <sup>17</sup>. In Los Angeles, a loss of resident avian immunity was associated with human outbreaks <sup>7</sup>. We include a broad diversity of bird species in modeling efforts.

#### *Anthropological factors*

Contact between infectious mosquitoes and susceptible humans is a function of many factors, several anthropological in nature. Human population density relative to mosquito population size (mosquito density) is included in the measure of arbovirus transmission, vectorial capacity<sup>18</sup>. In addition, one of the most mathematically powerful parameters of transmission is the biting rate, a direct estimate of mosquito-to-host contact<sup>19</sup>. In the context of human population risk, the likelihood of contact between humans and mosquitoes can be affected by many things, including demographic and socioeconomic factors, as well as the proximity of the enzootic cycle<sup>17</sup>. For example, researchers have associated hot spots of WNV transmission with—in addition to ecological parameters—demographic variables, specifically human population characteristics and per capita household income<sup>20,21,22</sup>. The age of housing structures and race were significant factors associated with WNV disease in urban Chicago and Detroit<sup>23</sup>. In terms of disease reporting as a covariate, a recent study demonstrated that reliance on NID reporting alone did not accurately predict the magnitude of the 2012 outbreak<sup>2</sup>. Thus we used all reported cases, including both NID and non-NID.

On a smaller ecological scale, land cover is associated with differences in presumed transmission. In Louisiana, the probability of a mosquito pool being positive is greater in wetland or forested areas than in developed areas<sup>24</sup>. In South Dakota, landscape-scale environmental factors are associated differentially with degree of urbanization. In less developed Aberdeen, WNV activity is positively correlated with grass/hay land cover and emergent wetlands while in Sioux Falls, WNV risk is higher in surrounding, relatively rural communities rather than urbanized cityscape<sup>25</sup>. Further, agricultural areas are at higher risk in Iowa for WNV than are urbanized areas<sup>26</sup>. However, there is no single modeling effort that accounts for the enzootic cycle, its interaction with human population and abiotic factors in order to make national-level predictions about WNV transmission, though subsets of these factors have led to predictive national models. For example, a national study found that the relationship between land use, demography, and WNV incidence vary for different regions of the U.S. as defined by mosquito species distributions<sup>27</sup>. Indeed, researchers have asserted that the association of land use and WNV incidence is regionally specific<sup>28</sup>, though this model did not make use of weather data as we propose herein. Thus we considered a number of demographic and land use variables in our national model.

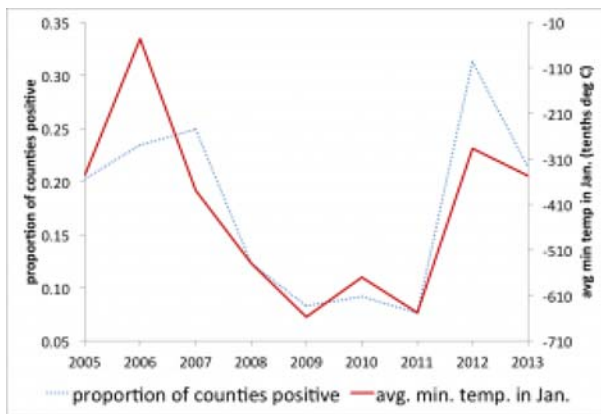
#### *Environmental and weather conditions*

Arbovirus transmission is altered by several environmental and weather conditions. For example, temperature is a known modifier of WNV efficiency, directly reducing the time necessary for the mosquito to become infectious after exposure or increasing the number of mosquitoes ultimately infectious<sup>29,30,31</sup>. Further, WNV incidence in Georgia was associated with increases in minimum temperature early in the calendar year<sup>32</sup>. Recently, drought conditions (increased temperature, decreased precipitation) were positively associated with increases in WNV transmission<sup>5,33</sup>. Other studies have shown the positive correlation between WNV transmission and temperature, as well as an inverse or no relation between WNV and precipitation levels<sup>34,35,36,37,38,39</sup>. While other environmental conditions (or different measures thereof<sup>36</sup>) may contribute to changes in transmission, temperature and precipitation are by far the most commonly used abiotic metrics.

#### *Model Rationale*

Anecdotal evidence from mosquito abatement experts suggests that drought conditions may have contributed to the increase in human WNV activity in 2012, with published research supporting this hypothesis<sup>5</sup>. Our goal was to determine if there are predictive variables for the risk for increased human WNV infections early enough to support potential large scale (regional or national) preparedness. As mentioned above, a study in Georgia associated WNV transmission with minimum temperature in January for 2002-2004<sup>32</sup> and high temperatures in winter were associated with the 2012 Dallas outbreak<sup>40</sup>. Preliminary investigation of this relationship on the national scale and over several years (2005-2013) showed a continued strong correlation between the deviation from usual minimum temperature in January and reports of WNV in that year (Figure 2).

---



**Fig. 2: Minimum temperature predicts reporting**

Relationship between the mean minimum temperature in January and the proportion of counties reporting West Nile Virus later that same year.

---

Thus, we continued this line of investigation and included several variables related to the enzootic cycle, human demographic data, and other environmental variables. The study relies on historical and publicly available data sets that are available early in the year in question (including the previous year's WNV report, remote sensing data, environmental parameters, etc.), so that the model may be employed as an early warning system and results used to support policy decisions and resource allocation.

## ***Methods***

We developed a statistical model to predict yearly national human WNV risk with data collected early in the year and available for all counties in the contiguous United States. We hypothesized that although risk factors vary by region, deviation from minimum temperature in the winter is associated with WNV risk across regions. Based on previous studies and the ecology of WNV, we included multiple demographic, ecological, and climatological variables to build our statistical model with the goal of finding a simple cohort of variables that accurately predict increases in human WNV risk.

### *National geography and WNV reporting*

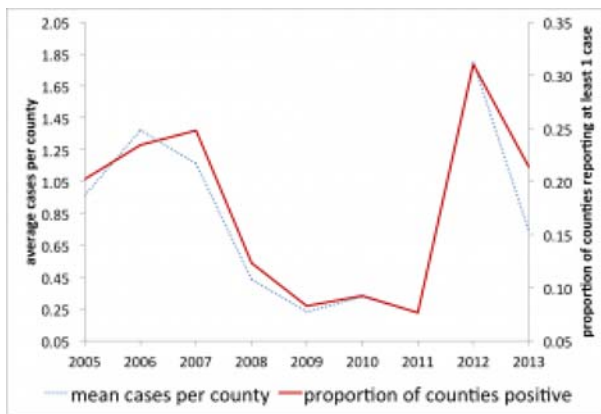
We obtained county boundaries for the 48 contiguous US states from the SAS base maps file (SAS 9.3, Cary, NC). The coordinates of the county centroids were approximated by the mean of points on the boundary of each county. Cumulative yearly counts of all reported WNV cases in humans reported to the CDC by county were obtained from the USGS summary for 3,111 counties from 2005 to 2013 <sup>41</sup>.

We believe that attempts to characterize incidence within a county expose the model to the problem of small numbers. In this case, the number of samples (the population) within a county is usually large and the number of positives is relatively small. Young et al. discusses three methods to address the problem: spatial smoothing, aggregation over geographic areas, or aggregation over time <sup>11</sup>. As in that study, we take the county and year as our basic units.

The Young et al. study also concluded that their attempt had not mitigated the small numbers problem. While we suspect, as the authors did, that confounding factors were at least partially responsible, we also suspect that characterization of incidence by incidence rate and multivariate linear models is likely inappropriate <sup>11</sup>. We follow the arguments of O'Hara and Kotze (2010) in which it is argued that count data are best modeled by Poisson or negative binomial distributions, rather than by transformation and approximation by a normal random variable <sup>42</sup>. This is especially true in the current data, with the majority of the reported rates being zero.

Here, if a county had reported any human WNV cases, it was coded as 1 and 0 otherwise. We argue that variations in this (0/1) metric capture the majority of variation in the total number of WNV cases across space and time and a more detailed model is not currently indicated (e.g. see Figure 3). This dichotomous method has been successfully employed at a regional level <sup>43</sup>.

---



**Fig. 3: WNV human cases and reporting per county**

Yearly comparison of the average cases of WNV per county (total WNV/counties reporting) to the proportion of counties reporting at least one case of WNV that year, 2005-2013.

---



## *Covariates*

Bird community. We used the Breeding Bird Survey (BBS) to approximate numbers of birds, summarized by order, in an area. Surveys are comprised of counts made along distinct 24.5-mile routes in .5-mile intervals. Each survey was summarized by year conducted, route sampled, and order of bird species identified. Twenty-seven orders were represented and zeros were imputed when no examples of an order were reported along a particular survey route. Two attempts were made to reduce the dimensionality of the set. First, the total number of birds identified in a survey was log-transformed; this correlated poorly with WNV reports and was not further investigated. Next, log-transformed counts by order were subjected to a principal component analysis; the principal components led to a relatively compact description of the BBS. The principal components were computed for each year per route and represent a linear combination of the log (order count) for orders 1-27.

It is important to note that the BBS data set was included here not so that individual species or orders could be associated with WNV, but as a means to account for unique regional ecologies and bird community biodiversity. Thus, the parameter estimates for principal components do not correspond to any specific characteristic of the bird community and are meant as markers for underlying environmental intangibles.

We extrapolated the set of principal components to county centroids with a power semivariogram model (proc variogram and proc krige2d in SAS 9.3) (SAS®, Cary, NC). Estimates were conditioned on the nearest 50 points to avoid prohibitive computational complexity. Similar semivariogram models were considered and most yielded qualitatively similar results; the power model was chosen for its smaller number of parameters.

Anthropological data. We used the 2010 census demographic data on a per-county basis. The data included estimates of the total population, median total income, proportion of the population in the labor force, and proportion of the population by race. The first two variables were log transformed and log-population was included as a second-degree polynomial (the quadratic form has been used successfully in other local studies <sup>32</sup>). A nonlinear relationship was observed in the marginal distribution, with low and high population counties more frequently reporting WNV than mid-size counties. Measures of density, e.g. persons per square mile, did not improve on the models and were excluded from consideration.

The National Land Cover Database 2006 (NLCD2006) from the Multi-Resolution Land Characteristics Consortium (MRLC) classifies land areas into one of 16 categories. We summarized each county by the proportion of pixels within its boundary classified as developed (types 21-24) in Quantum GIS (QGIS) <sup>32</sup>.

Finally, we investigated the potential for county-level reporting biases by including WNV reporting in the previous year and, in separate models, the previous two years. We reasoned that biases not otherwise addressed might be corrected by continued deviation from expectation. Both year and previous two years were significant predictors of WNV status ( $p < .05$ ) but were also highly correlated. In the final model, we determined that the previous year's report was sufficient.

Environment and weather. Our models initially included maximum temperature, minimum temperature, precipitation, and deviation from the average for each of these variables. While the WNV cycle depends on covariates throughout the year, mean minimum temperatures in January alone were retained from the Global Historical Climatology Network monthly summary (GHCN-M) for all available stations; stations with any missing values in January 2005 to 2011 were excluded. Expected mean minimum temperatures for a station for January were defined as the average of that month's data from 2005 to 2011. Various periods were considered (e.g. 1970 to present) to define the expected temperatures but yielded no qualitative differences, indicating that 2005 to 2011 was in some way representative. These indices were extrapolated to county centroids by weighing the nearest 50 stations by distance from the centroid.

Maximum and average temperatures were individually significant predictors, but were not as powerful as minimum temperature and deviation from minimum temperature and did not improve on models already

containing minimum temperature. We made an attempt to include precipitation, but the semivariogram fits were poor and extrapolation to county centroids was questionable. This seemed to indicate that precipitation was a local phenomenon with scale smaller than we were able to observe. Thus maximum temperature, average temperature, and precipitation are excluded from the final model.

### *Statistical analysis*

We carried out logistic regression analyses in SAS 9.3 using the dichotomous (0/1) variable indicating WNV activity for each county-year. A stepwise procedure with entry/exit  $p < 0.10$  (proc logistic) was used to determine which variables best fit the data. We sought influential observations in the selected model by DFBETA; exclusion of the most influential observations produced no qualitative changes and no observations were deemed outliers.

Because we repeatedly sampled the same counties over time, we do not model all samples as independent. Two samples from the same county were assumed to correlate even after the influence of all covariates had been removed. In the final model, we employed a generalized estimating equation (GEE) regression with an exchangeable working correlation structure (proc genmod in SAS 9.3) and estimated this correlation coefficient as  $\rho = .1041$  for 2005-2011 data. This indicates that when modeled effects are considered, the remaining, unexplained correlation among WNV reports within a county is not prohibitively large. Regular and standardized regression coefficients and chi-square statistics were calculated for the resulting model.

## **Results**

We recall that the model is conditioned on WNV human reported incidence data from 2005-2011 and we generated predictions for 2012-2013 without training on those years. The area under the receiver operating characteristic curve (AUC ROC or c) and the Hosmer-Lemeshow test statistic were calculated to measure the fit of the model for the 2012 data <sup>44</sup>.

The predictors included in the final model for 2005-2011 (Table 1) are ranked by the magnitude of the standardized parameter estimate, in which larger values indicate more sensitivity of WNV status to changes in the covariate. The five most powerful predictors of WNV reporting in a county in order are

1. the expected mean minimum temperature in January
2. the actual mean minimum temperatures in January
3. the total population of the county per the 2010 census
4. the first principal component of bird data
5. the WNV status of the county in the previous year

In 5,000 simulated case-control studies, in which 25% of all positive and 25% of negative counties from years 2005-2012 were randomly chosen for training, the order of these components was unchanged in only 21% of trials. When population is omitted, this rose to 98% and none of the coefficient estimates were ever observed to change sign. This indicates that population is primarily important in differentiating relatively large, rare counties, which might not be present in random samples, from smaller, more numerous counties. This also indicates that the primary hypothesis, dependence on mean and actual temperature in January, is robust against missing data and selection of different training sets.

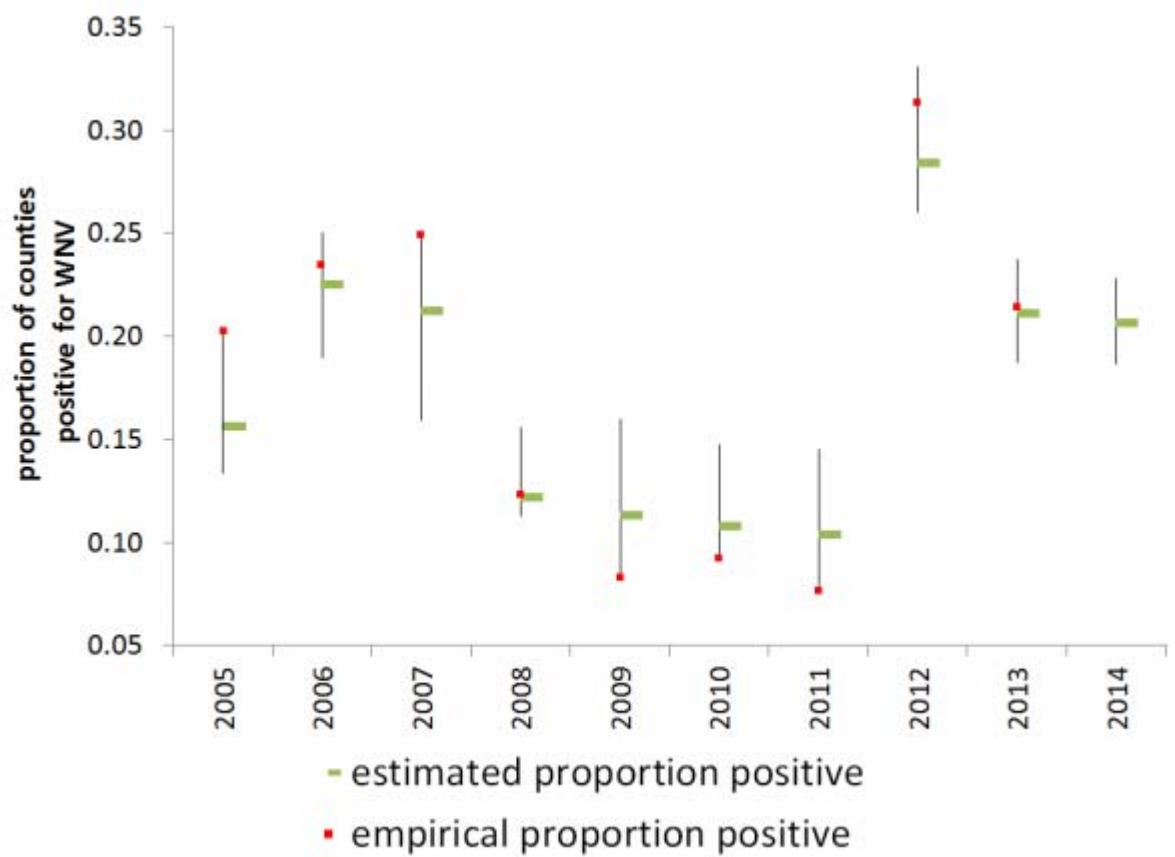
## Model for predicting human WNV activity

Model variables and statistical summary for prediction of human WNV activity, 2005-2011. "BP" = Bird Principal component for a particular year.

parameter	estimate (log odds)	standardized estimate (c/ $\sigma$ )	std. error	p>X <sup>2</sup>
intercept	-0.547		1.4253	0.7012
expected mean minimum temperature in January	-0.00224	-0.6185	0.000107	<0.0001
mean minimum temperature in January	0.00185	0.5157	0.00008	<0.0001
(log total population) <sup>2</sup>	0.0284	0.4673	0.00113	<0.0001
first principal component of bird data (BP1)	0.2714	0.3615	0.0164	<0.0001
WNV present in previous year? (Yes=1)	1.0758	0.2606	0.0503	<0.0001
BP3	-0.244	-0.2365	0.0166	<0.0001
BP2	0.1981	0.1810	0.0208	<0.0001
% Black or African American	0.0144	0.1372	0.0021	<0.0001
BP20	0.4544	0.1130	0.0512	<0.0001
BP12	0.0751	0.1043	0.0379	0.0475
BP18	-0.3397	-0.0886	0.0518	<0.0001
BP5	-0.2767	-0.0807	0.0242	<0.0001
BP7	-0.1660	-0.0745	0.0292	<0.0001
% employed (Census 2010)	2.1885	0.0686	0.4636	<0.0001
(proportion developed) <sup>2</sup>	1.278	0.0667	0.5282	0.0155
log (total income and benefits)	-0.6033	-0.0623	0.1488	<0.0001
BP6	0.1508	0.0556	0.0208	<0.0001
BP16	-0.1021	-0.0500	0.0453	0.0240
proportion of county land developed	-0.7137	-0.0454	0.2931	0.0149
BP13	-0.2086	-0.0366	0.0360	<0.0001
BP10	-0.1220	-0.0342	0.0314	0.0001
BP15	-0.0794	-0.00937	0.0430	0.0649
BP19	-0.1294	-0.00609	0.0525	0.0137
BP21	0.2658	0.0045	0.0603	<0.0001
BP9	-0.1046	-0.0042	0.0348	0.0027

Additionally, WNV is positively associated with employment levels in a county while actual income levels are negatively associated with WNV reported human cases. The percent of African Americans in a county is also a positive predictor for WNV reporting.

Observed and estimated proportion of counties reporting positive per year are shown in Figure 4. When conditioned on 2005-2011 reported human WNV cases, the model prediction for 2012 had c-value of .849, which Hosmer and Lemeshow consider to show excellent but not outstanding discrimination <sup>44</sup>. In the Hosmer and Lemeshow goodness-of-fit test, the model was well-calibrated at the 10% (but not 5%) level of significance (p = .0780).



**Fig. 4: Estimated rate of WNV human activity**

The empirical (red square) and estimated (green bar) rate of presence of WNV activity with the 95% confidence intervals of the statistical prediction.

We present the empirical and estimated rates of positivity for 2013, but note that the empirical proportion is likely an underestimate, as some states are at the time of writing only reporting on a state-wide level and not all reporting has been completed. Data in 2013 were used neither to condition nor to validate the model. So far, 2,374 human cases were reported in 2013 with 1,205 NID cases reported in the United States. The Hosmer and Lemeshow test for 2013 is well calibrated at the 10% level of significance ( $p=0.0818$ ). We also present predictions for 2014 based on the available preliminary 2013 WNV human data (missing 2013 county-level data was replaced with 2012 data) and temperature data for January 2014. Since winter weather in 2013-2014 was anomalous for many regions of the United States, the 2014 year should present a particularly interesting test for the model.

## ***Discussion and Conclusion***

Several modeling, laboratory and field studies have shown that WNV transmission is associated with environmental factors such as temperature, precipitation, drought, and land use, as well as biological factors such as bird community structure and anthropological variables, including urbanization and human density <sup>8,45</sup>. However, most of the studies focused exclusively on particular states, counties, or regions based on the data available to the researchers. There is a need for a concise, relatively simple model that can predict early in the year when risk of human WNV may be elevated in order to inform public health decisions, resource allocation, and public education.

By focusing on the development of an early warning system from a national perspective, we directly speak to this need for better predictive tools to guide decisions regarding control and mitigation. Our model shows that in the contiguous United States, bird presence estimated from the previous year, deviation from minimum temperature in January of the year in question, human density estimates from census data, and history of WNV reporting are sufficient for prediction of risk of WNV human activity later in the year in question.

In fact, conditioned on the 2005-2011 data, our model predicted the large upswing in human WNV in 2012, requiring only data from the previous year and temperature data from January 2012. To the extent that the model holds, none of the years in question are anomalous (Figure 4).

Specifically, we found that the most significant predictors of WNV were the expected and actual minimum temperatures in January of the year in question. These do not seem to admit interpretation as biases and likely have straightforward biological interpretations. For example, unusual, positive deviations in temperature early in the year may increase a mosquito's probability of surviving the remainder of the winter, may shorten the extrinsic incubation period, or encourage earlier and higher rates of mosquito breeding, and have been associated with heightened bird activity such as greater numbers of resident birds nesting and migration of non-resident birds <sup>31,38,40</sup>. Deviation from minimum temperature accounts for relative variations in climate and for local adaptations of mosquitoes and birds to their environment, which actual minimum temperature alone cannot capture.

Higher county-level income measures are negatively associated with reported WNV human incidence. This is supported by several studies, which have shown that higher housing density (as seen in lower socio-economic statuses) and increased urbanization are positively associated with WNV in humans <sup>46,47</sup>. Further, the proportion of African Americans in a population was a positive indicator of human WNV activity. The significance and interpretation of this factor is likely regionally specific and requires a more anthropological and socio-economically targeted study to determine context.

The dual importance of human density and general bird presence is intuitive, given that the spillover into humans requires proximity to the otherwise bird-driven WNV transmission cycle. Again, specific orders of birds or specific species composition could be important on a local or regional level. In this model, however, bird

orders serve more as proxies for the entirety of the underlying system and do not admit direct interpretation of the specific makeup of avian communities.

The idea of birds as proxies is supported by studies that have shown different bird species to have higher seroprevalence than others (e.g. northern cardinal in Georgia <sup>48</sup>, robins <sup>49</sup>, house finch and house sparrow in California <sup>7</sup>). Likely, regional iterations of this model will show more consistency with regard to specifics about the bird community, as this is ecologically driven, although the results of Young et al. call into question this intuition <sup>11</sup>.

This model can be used as an additional tool for public health institutions in the United States, as a national indication of the potential for human WNV activity increases, and eventually at the regional or local level, to inform surveillance, mitigation, and resource allocation efforts. For example, once on high alert, counties or regions can consider surveillance data, specific climate data, mosquito ecology for the region, land use, bird ecology, etc. and develop models such as ours for their particular regions, allowing them to continue to monitor risk and implement mitigation in a timely manner. In future work, this model will be applied regionally to show how particular regions can use additional weather data (precipitation, soil moisture levels, etc.), land use data, bird data and mosquito surveillance to determine further risk that year.

The model uses data from the first month of the year in an effort to provide some confidence in its performance going forward and to provide earliest possible warning of heightened WNV risk. Since this model is based in reported human cases, its predictive capability is limited to just that. Thus, it should not be used to make any assumptions about the total volume of virus in the enzootic cycle, or the true transmission intensity, as most human cases are asymptomatic. This model is proposed on a national scale, and the intention of its use is to raise awareness of a potential heightened WNV season in time for any resulting reactions to have transmission mitigation efficacy. We envision that this model will be the first step in a series of spatially tiered warning systems that will increase in regional specificity and timing.

Since the primary vectors of WNV are regionally dependent and given the spotty nature of consistent county-level mosquito control and surveillance data, mosquitoes were not explicitly considered in this model. Additionally, it is difficult to quantify the impact of mitigation strategies such as spraying on human incidence over multiple years without bird and mosquito surveillance data. This speaks to the lack of consistent data of both bird and mosquito populations in many regions, which would improve the predictive capabilities of any model, the regional scale at which a model could be formulated, and our understanding of WNV dynamics. On county and regional levels, models with access to detailed bird and mosquito surveillance data have had success in explaining WNV incidence in the enzootic and tangential cycles based on ecological and environmental factors <sup>5,7,22,31,33,37,40,45,50,51,52</sup>. Delving into the specifics of regional climates, and into the social and land planning factors that increase risk will also be important for future efforts. Bird diversity can also be an important factor in national WNV spread and human risk <sup>53</sup>. Additional data sources (seroprevalence from bird populations, heterogeneity of infectiousness among bird populations, the sero-conversion rate of sentinel chickens, and mosquito populations) are needed to truly understand the transmission cycle of WNV. This understanding would lead to better predictions of transmission intensity and thus inform mitigation strategies and ultimately reduce human cases. However, in the absence of these explicit models, our model is a tool for early warning and policy decision support.

## ***Acknowledgements***

We would like to thank the National Oceanic and Atmospheric Administration (NOAA), especially Ms. Wendy Marie Thomas, for her assistance in procuring and navigating the climatological and weather data for this study.

## ***References***



1. Beasley DW, Barrett AD, Tesh RB (2013) Resurgence of West Nile neurologic disease in the United States in 2012: what happened? What needs to be done? *Antiviral Res* 99: 1-5.
2. Lindsey NP, Staples JE, Delorey MJ, Fischer M (2014) Lack of evidence of increased west nile virus disease severity in the United States in 2012. *Am J Trop Med Hyg* 90: 163-168.
3. Centers for Disease C, Prevention (2013) West Nile virus and other arboviral diseases--United States, 2012. *MMWR Morb Mortal Wkly Rep* 62: 513-517.
4. Duggal NK, D'Anton M, Xiang J, Seiferth R, Day J, et al. (2013) Sequence analyses of 2012 West Nile virus isolates from Texas fail to associate viral genetic factors with outbreak magnitude. *Am J Trop Med Hyg* 89: 205-210.
5. Johnson BJ, Sukhdeo MV (2013) Drought-induced amplification of local and regional West Nile virus infection rates in New Jersey. *J Med Entomol* 50: 195-204.
6. Dusek RJ, McLean RG, Kramer LD, Ubico SR, Dupuis AP, 2nd, et al. (2009) Prevalence of West Nile virus in migratory birds during spring and fall migration. *Am J Trop Med Hyg* 81: 1151-1158.
7. Kwan JL, Kluh S, Reisen WK (2012) Antecedent avian immunity limits tangential transmission of West Nile virus to humans. *PLoS One* 7: e34127.
8. Vuong HB, Caccamise DF, Remmenga M, Creamer R (2012) Ecological associations of West Nile virus and avian hosts in an arid environmnet. In: Paul E, editor. *Emerging avian disease: Studies in Avian Biology*. Berkely, Ca: University of California Press. pp. 3-22.
9. Petersen LR, Brault AC, Nasci RS (2013) West Nile virus: review of the literature. *JAMA* 310: 308-315.
10. Levine RS, Mead DG, Kitron UD (2013) Limited Spillover to Humans from West Nile Virus Viremic Birds in Atlanta, Georgia. *Vector Borne Zoonotic Dis*.
11. Young SG, Tuliis JA, Cothren J (2013) A remote sensing and GIS-assisted landscape epidemiology approach to West Nile virus. *Applied Geogrphay* 45: 241-249.
12. Komar N, Langevin S, Hinten S, Nemeth N, Edwards E, et al. (2003) Experimental infection of North American birds with the New York 1999 strain of West Nile virus. *Emerg Infect Dis* 9: 311-322.
13. McKenzie VJ, Goulet NE (2010) Bird community composition linked to human West Nile virus cases along the Colorado front range. *Ecohealth* 7: 439-447.
14. Kilpatrick AM, Peters RJ, Dupuis AP, 2nd, Jones MJ, Marra PP, et al. (2013) Predicted and observed mortality from vector-borne disease in small songbirds. *Biol Conserv* 165: 79-85.
15. Komar N, Panella NA, Langevin SA, Brault AC, Amador M, et al. (2005) Avian hosts for West Nile virus in St. Tammany Parish, Louisiana, 2002. *Am J Trop Med Hyg* 73: 1031-1037.
16. Mishra N, Kalaiyarasu S, Nagarajan S, Rao MV, George A, et al. (2012) Serological evidence of West Nile virus infection in wild migratory and resident water birds in Eastern and Northern India. *Comp Immunol Microbiol Infect Dis* 35: 591-598.
17. Liu H, Weng Q, Gaines D (2011) Geographic incidence of human West Nile virus in northern Virginia, USA, in relation to incidence in birds and variations in urban environment. *Sci Total Environ* 409: 4235-4241.
18. Christofferson RC, Mores CN (2011) Estimating the magnitude and direction of altered arbovirus transmission due to viral phenotype. *PLoS One* 6: e16298.

19. Chitnis N, Hyman JM, Cushing JM (2008) Determining Important Parameters in the Spread of Malaria Through the Sensitivity Analysis of a Mathematical Model. *Bulletin of Mathematical Biology* 70: 1272-1296.
20. Harrigan RJ, Thomassen HA, Buermann W, Cummings RF, Kahn ME, et al. (2010) Economic conditions predict prevalence of West Nile virus. *PLoS One* 5: e15437.
21. Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, et al. (2002) Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne Zoonotic Dis* 2: 157-164.
22. Liu A, Lee V, Galusha D, Slade MD, Diuk-Wasser M, et al. (2009) Risk factors for human infection with West Nile Virus in Connecticut: a multi-year analysis. *Int J Health Geogr* 8: 67.
23. Ruiz MO, Walker ED, Foster ES, Haramis LD, Kitron UD (2007) Association of West Nile virus illness and urban landscapes in Chicago and Detroit. *Int J Health Geogr* 6: 10.
24. Christofferson RC, Roy AF, Mores CN (2010) Factors associated with mosquito pool positivity and the characterization of the West Nile viruses found within Louisiana during 2007. *Virol J* 7: 139.
25. Chuang TW, Hockett CW, Kightlinger L, Wimberly MC (2012) Landscape-level spatial patterns of West Nile virus risk in the northern Great Plains. *Am J Trop Med Hyg* 86: 724-731.
26. DeGroot JP, Sugumaran R, Brend SM, Tucker BJ, Bartholomay LC (2008) Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA. *Int J Health Geogr* 7: 19.
27. DeGroot JP, Sugumaran R (2012) National and regional associations between human West Nile virus incidence and demographic, landscape, and land use conditions in the coterminous United States. *Vector Borne Zoonotic Dis* 12: 657-665.
28. Bowden SE, Magori K, Drake JM (2011) Regional differences in the association between land cover and West Nile virus disease incidence in humans in the United States. *Am J Trop Med Hyg* 84: 234-238.
29. Bowden SE, Magori K, Drake JM (2011) Regional differences in the association between land cover and West Nile virus disease incidence in humans in the United States. *Am J Trop Med Hyg* 84: 234-238.
30. Reisen WK, Fang Y, Martinez VM (2006) Effects of temperature on the transmission of west nile virus by *Culex tarsalis* (Diptera: Culicidae). *J Med Entomol* 43: 309-317.
31. Richards SL, Mores CN, Lord CC, Tabachnick WJ (2007) Impact of extrinsic incubation temperature and virus exposure on vector competence of *Culex pipiens quinquefasciatus* Say (Diptera: Culicidae) for West Nile virus. *Vector Borne Zoonotic Dis* 7: 629-636.
32. Hartley DM, Barker CM, Le Menach A, Niu T, Gaff HD, et al. (2012) Effects of temperature on emergence and seasonality of West Nile virus in California. *Am J Trop Med Hyg* 86: 884-894.
33. Gibbs SE, Wimberly MC, Madden M, Masour J, Yabsley MJ, et al. (2006) Factors affecting the geographic distribution of West Nile virus in Georgia, USA: 2002-2004. *Vector Borne Zoonotic Dis* 6: 73-82.
34. Deichmeister JM, Telang A (2011) Abundance of West Nile virus mosquito vectors in relation to climate and landscape variables. *J Vector Ecol* 36: 75-85.
35. Walsh MG (2012) The role of hydrogeography and climate in the landscape epidemiology of West Nile virus in New York State from 2000 to 2010. *PLoS One* 7: e30620.
36. Chen CC, Epp T, Jenkins E, Waldner C, Curry PS, et al. (2012) Predicting weekly variation of *Culex tarsalis* (Diptera: Culicidae) West Nile virus infection in a newly endemic region, the Canadian prairies. *J Med Entomol* 49: 1144-1153.



37. Crowder DW, Dykstra EA, Brauner JM, Duffy A, Reed C, et al. (2013) West nile virus prevalence across landscapes is mediated by local effects of agriculture on vector and host communities. PLoS One 8: e55006.
38. Paz S, Malkinson D, Green MS, Tsioni G, Papa A, et al. (2013) Permissive summer temperatures of the 2010 European West Nile fever upsurge. PLoS One 8: e56398.
39. Wimberly MC, Hildreth MB, Boyte SP, Lindquist E, Kightlinger L (2008) Ecological niche of the 2003 west nile virus epidemic in the northern great plains of the United States. PLoS One 3: e3744.
40. Chung WM, Buseman CM, Joyner SN, Hughes SM, Fomby TB, et al. (2013) The 2012 West Nile encephalitis epidemic in Dallas, Texas. JAMA 310: 297-307.
41. Survey USG (2013) West Nile Virus Human. Disease Maps: USGS.
42. O'Hara RB, Kotze DJ (2010) Do not log-transform count data. Methods in Ecology and Evolution 1: 118-122.
43. Ruiz MO, Tedesco C, McTighe TJ, Austin C, Kitron U (2004) Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int J Health Geogr 3: 8.
44. Hosmer DW, Lemeshow S (2000) Applied logistic regression. New York: Wiley. xii, 375 p. p.
45. Swaddle JP, Calos SE (2008) Increased avian diversity is associated with lower incidence of human West Nile infection: observation of the dilution effect. PLoS One 3: e2488.
46. Brown HE, Childs JE, Diuk-Wasser MA, Fish D (2008) Ecological factors associated with West Nile virus transmission, northeastern United States. Emerg Infect Dis 14: 1539-1545.
47. LaDeau SL, Calder CA, Doran PJ, Marra PP (2011) West Nile virus impacts in American crow populations are associated with human land use and climate. Ecological Research 26: 909-916.
48. Bradley CA, Gibbs SEJ, Altizer S (2008) URBAN LAND USE PREDICTS WEST NILE VIRUS EXPOSURE IN SONGBIRDS. Ecological Applications 18: 1083-1092.
49. Beveroth TA, Ward MP, Lampman RL, Ringia AM, Novak RJ (2006) Changes in seroprevalence of West Nile virus across Illinois in free-ranging birds from 2001 through 2004. Am J Trop Med Hyg 74: 174-179.
50. Kilpatrick AM, Pape WJ (2013) Predicting human west nile virus infections with mosquito surveillance data. Am J Epidemiol 178: 829-835.
51. Chuang TW, Wimberly MC (2012) Remote sensing of climatic anomalies and West Nile virus incidence in the northern Great Plains of the United States. PLoS One 7: e46882.
52. Johnson BJ, Munafo K, Shappell L, Tsipoura N, Robson M, et al. (2012) The roles of mosquito and bird communities on the prevalence of West Nile virus in urban wetland and residential habitats. Urban Ecosystems 15: 513-531.
53. Allan BF, Langerhans RB, Ryberg WA, Landesman WJ, Griffin NW, Katz RS, Oberle BJ, Schutzenhofer MR, Smyth KN, de St Maurice A, Clark L, Crooks KR, Hernandez DE, McLean RG, Ostfeld RS, Chase JM. Ecological correlates of risk and incidence of West Nile virus in the United States. Oecologia. 2009 Jan;158(4):699-708. PubMed PMID:18941794.