**Title:** West Nile Virus in Colorado: Who, What, When, Where, Why, and How

**Abstract**

West Nile Virus found its way to North America in 1999, starting with the diagnosis of two cases of encephalitis in the Queens borough of New York City. WNV had found its way to Colorado by 2002. This paper will address the who, what, when, where, why, and how of WNV in the counties of Colorado.

* Who: The mosquito, primarily the *Culex* species
* What: West Nile Virus spread in Colorado
* When: Time frame of study is 2005 to 2021
* Where: The 64 counties of Colorado
* Why and How: Will be addressed in this paper

It is shown, from historical data along with qualitative, quantitative, and geospatial methods that the primary variables behind West Nile Virus cases by county in Colorado are urban/rural classification, water area in county (in square miles), and if it is an El Niño year or not. Other variables are discussed and their merit in a model presented. Mapping tools are used to illustrate the presence of West Nile Virus as well as its spread, over time, through the counties in Colorado.

**Introduction**

West Nile Virus (WNV), a mosquito transmitted virus, was first identified in 1937 in the West Nile province of Uganda (Kramer, Li, & Shi, 2007). WNV emerged into North America, and in particular, New York in 1999 and has since spread throughout the USA, Canada, and Mexico (Petersen, 2019; Kramer, Li, & Shi, 2007). Colorado recorded its first case of WNV in 2002 (CDPHE, 2022) and saw a large number of cases in 2003 due, it is believed, to the initial large-scale testing for WNV, and the lack of immunity in the population (Marzec, 2022).

WNV can infect birds, which also serve as a reservoir and diffuser for the virus, horses, and humans, with the latter two being the most negatively affected by the virus (Peterson, 2019; Kramer, Li, & Shi, 2007). WNV is spread, primarily, by the enzootic cycle between birds and mosquitoes. There are over three hundred species of birds that have been identified as viable WNV hosts and generally the mosquito genus *Culex* is the transmission vector to humans (Ciota, 2017). Most humans show no symptoms upon being infected with WNV (80%), while approximately 20% exhibit flu-like symptoms, and less than 1% become seriously ill with arboviral encephalitis (Ciota, 2017; Wildlife Futures Team, 2021). In Colorado, there have been over 5800 cases of WNV identified in humans since WNV arrived to the state in 2002 (CDPHE, 2022). Extrapolating, this implies that there could have been over 23,000 undetected cases with up to 290 cases of encephalitis, making this a public health issue involving hospital stays, transplantation medicine, and transfusion concerns (Petersen, 2019; Ronca, Ruff, & Murray, 2021). While there are vaccines against WNV for horses, first approved in 2005, there are no human vaccines that have gone beyond Phase I/II trials, indicating a need for other mitigation strategies for the virus and its vectors (Ronca, Ruff, & Murray, 2021).

Prevention of WNV through target mitigation strategies can assist the state of Colorado in preserving the health of the population, as well as avoidance of medical costs and lost productivity costs which can run into the tens of millions of dollars (Ronca, Ruff, & Murray, 2021).

Observations and studies from Chicago to Texas have suggested that “enhancement of surveillance and vector control in limited geographic areas could produce and outsized impact on WNV incidence nationwide” (Petersen, 2019, p. 1457); while Hadfield et al. (2019) state that a coordinated effort between state health departments could be used as vector control by “strategically focusing resources at a precise time and location to limit potential outbreaks” (p.10).

Potential variables affecting the transmission of WNV have been discussed in the literature and include:

* temperature (Ciota, 2017; Hadfield, et al., 2019; Paz, 2015; Garcia-Carrasco, et al. 2021)
* precipitation (Ciota, 2017; Hadfield, et al., 2019; Paz, 2015)
* humidity (Paz, 2015)
* urban/rural (Ciota, 2017)
* land use (Hadfield, et al., 2019; Paz, 2015; Garcia-Carrasco, et al. 2021)
* presence of water (Garcia-Carrasco, et al. 2021)
* elevation (Hadfield, et al., 2019; Garcia-Carrasco, et al. 2021)
* seasonality (Paz, 2015)
* wind (Paz, 2015)
* mosquito abundance (Ronca, et al., 2021)
* bird populations (Ronca, et al., 2021)

Other variables added to this model for their perceived usefulness for multiple types of analysis include:

* year
* county name
* county population
* number of WNV cases per county
* El Niño, La Niña years
* land area of county
* severity of WNV in county (primarily as a filter for geospatial analysis)

Variables which could be found are illustrated in Table 1, along with references or links to where to find the original data source.

Table 1

Variables for the Colorado WNV Models

|  |  |
| --- | --- |
| **Variable** | **URL** |
| Calendar year | <https://data.colorado.gov/Demographics/Total-Population-by-County-by-Year/9dd2-kw29> |
| County name in Colorado |
| Population (2020 Census) of county |
| Number of WNV cases in county for given year | <https://cdphe.colorado.gov/animal-related-diseases/west-nile-virus/west-nile-virus-data> |
| Relative measure of WNV cases against county population |
| Rainfall in county | <https://www.ncdc.noaa.gov/> |
| Urban: 1 if population of county is greater than 150,000; 0 otherwise | <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html> |
| Minimum county elevation, ft | <https://en.wikipedia.org/wiki/List_of_counties_in_Colorado> |
| Maximum county elevation, ft |
| Difference in elevation, Maximum - Minimum, ft |
| Water area of county, square miles | <https://en.wikipedia.org/wiki/List_of_counties_in_Colorado#County_data> |
| Percent coverage of county by water |
| Number of people in county per square mile | derived from population and area variables |
| 1 if El Nino year, 0 otherwise | <https://ggweather.com/enso/oni.htm> |
| 1 if La Nina year, 0 otherwise |
| Area of county, square miles | <https://en.wikipedia.org/wiki/List_of_counties_in_Colorado#County_data> |
| 1 if mean elevation of county is over 7000 ft, 0 otherwise | <http://www.cohp.org/records/mean_elevation/mean_elevations.html> |
| Average temperature per county for the four-month period (July – Oct.), Fahrenheit | <https://www.ncdc.noaa.gov/cag/county/time-series> |
| Elevation at the geographical center of the county, ft | <https://www.usgs.gov/products/data/all-data> |
| Rainfall average per county for the four-month period (July – Oct.), inches | <https://www.ncdc.noaa.gov/cag/county/time-series> |
| Qualitative rating of WNV historical case count per county (High, Medium, Low) | derived from WNV case data from 2005 to 2021 |

Some variables from the original list were difficult or otherwise unobtainable for this study. These include humidity at a county level, mosquito, bird, and wind data at a county level. Other variables that proved difficult included the number of irrigated acres per county (only found 2007 and 2012 data sets) and a decent quality elevation variable.

**Problem Statement**

This research seeks to determine what variables influence WNV presence in the counties of Colorado. By applying techniques such as:

* descriptive statistics
* correlation analysis
* hypothesis testing
* multiple linear regression analysis
* logistic regression analysis
* cluster analysis
* geospatial analysis
* graph analysis

salient variables will be selected for modeling purposes and mapping techniques will be used as an illustrative guide for deploying resources at a county level. This analysis will aid the decision maker (hospital, EMS, medical community, county/state level medical personnel) in when and where to apply mitigation measures for WNV, when and where to deploy limited personnel across the state of Colorado, where to locate sentinel devices for WNV, and who and when to notify of WNV activity.

**Methodology and Results**

The first finding is that elevation played no significant part in the analysis, even though six different elevation variables (maximum, minimum, difference, qualitative over 7000 ft., elevation at center point of county, and high qualitative (maximum over 9000 ft.)) were tried.

Performing both multiple linear regression (y1 = number of WNV cases in county, y2 = number per 100,000 people in county) and logistic regression ( = presence of WNV in county) with the data set excluding elevation, the best models (using backward elimination method) and their R2 values are given in Table 2.

Table 2

Best models for WNV

|  |  |
| --- | --- |
| **Model** | **R2** |
|  | 25.7% |
|  | 11.2% |
| , where | **\*Pseudo R2** |

**\***  see the Real Statistics help page at: <https://www.real-statistics.com/logistic-regression/significance-testing-logistic-regression-model/> (Zaiontz, 2022)

Evaluating the explanatory power of the three models in Table 2 leads to removal of y2 and the construction of Diagram 1, a Venn diagram of the statistically significant variables from y1 and . Diagram 1 shows the core variables of the model and the peripheral variables to be used for the rest of the models and analysis.

Diagram 1

Significant variables for WNV models

variables from multiple linear regression, y1 variables from logistic regression,



population density

AvgTempJulyOct

AvgRainfallJulyOct

urban/rural

water area mi2

El Niño

The variables identified in Diagram 1 can be further defined here:

**Core Variables**

* urban/rural – a qualitative variable, 1 if the population of the county is greater than 150,000 people. This variable will be used to compare the two data sets for differences (urban counties versus rural counties) and as a filter for visualizations.
* El Niño – a qualitative variable, 1 if it is an El Niño year (which occurred in 2005, 2007, 2010, 2015, 2016, and 2019 for this study). This variable will be used to compare the two data sets for differences (El Niño year versus not an El Niño year) and as a filter for visualizations.
* water area – a quantitative variable for the percent of each county that is covered by water. An important variable, in that mosquitoes (the WNV vector from birds -> humans) need water to deposit their eggs.

**Peripheral Variables**

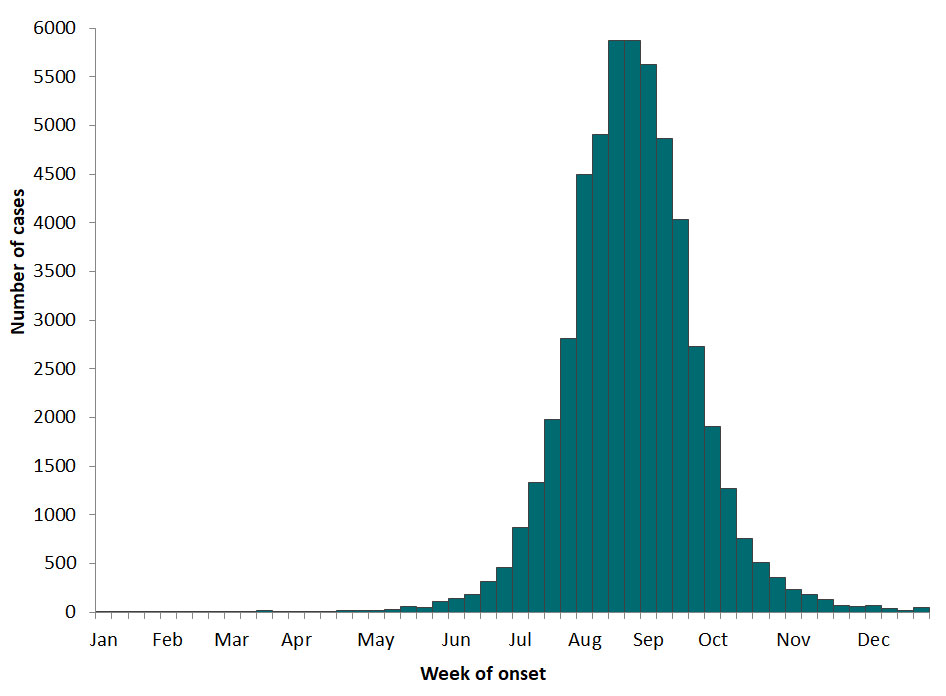
* population density – number of people per square mile. With any disease of the human, population density can play a factor
* AvgTempJulyOct and AvgRainfallJulyOct – as named, these variables measure the average temperature in Fahrenheit and the average rainfall in inches over the four-month period of July to October. This is the period in which 80-90% of the WNV cases occur as can be seen in Diagram 2 from the CDC. Due to the short life cycle of the mosquito (1~2 weeks) looking at the environmental variables during this period would make sense.

**Time Period**

* WNV arrived to New York State in 1999, then spread throughout the country. WNV made it to Colorado in 2002, and being a new virus there was a lot of testing and no natural immunity (Marzec, 2022), causing a large spike in WNV numbers for 2003 and 2004. In 2005, the pattern or WNV settled down, exhibiting peaks and valleys, but not as severe as in 2003 - 2004. This model uses the WNV case count, by Colorado county, for the time frame 2005 – 2021 assuming these years to be more representative of the WNV numbers in the state moving forward.

Diagram 2

WNV during the year

(CDC, 2021)

Analyzing the counties longitudinally for the percentage of years the county has had WNV cases present over the study period, Graph 1 is obtained.

Graph 1

Percent of years county has WNV, 2005-2021

Graph 1 points at a couple of items, three counties have WNV every year (Denver, Boulder, Larimer) and three counties have a lot of WNV (Boulder, Larimar, Weld). Graph 1 also facilitates the development of a qualitative variable for a county’s WNV status.

CountyLongitudinalWNVStatus =

It can be calculated that the HighRiskCounty group, when summed, represents 76.3% of the cases, leaving only 23.7% of the cases in the medium and low risk categories, roughly aligning with the Pareto Principle. This variable will be used in the visualizations as well as for separating the data set into multiple pieces for some descriptive analysis and interpretive guidance on differences between the risk groups. This analysis is presented in Table 3.

Table 3

Statistically significant differences when dividing the data set along qualitative lines (high, medium, low and El Niño dimensions)



Overall, Table 3 (Excel, 2022) tells us that there will be more WNV cases in El Niño years when there is less rainfall, warmer temperatures, higher populated, and lower elevation counties with more surface water. The orange-colored cells indicate a non-significant result with a possible implication and that the data shows more WNV in years with less rainfall overall, an unexpected result, as mosquitos need water to lay their eggs in.

Cluster analysis (k = 3 clusters) also indicated that counties with high WNV cases clustered around:

* counties with more water area
* counties that have denser population
* counties with a higher AvgTempJulyOct
* counties lower in elevation.

This helps to identify salient variables and well-designed filters with which to build visualizations to illustrate these primarily analytical results.

**Visualizations**

Visualization 1 is a Tableau dashboard tracking WNV cases against four of the independent variables of the model.

Visualization 1

Tracking WNV based on salient variables

Chart, scatter chart

Description automatically generated

Denver

Visualization 1 gives us some information not only about the current status of the WNV model, but information on future directions as well. The Colorado WNV “Goldilocks Zone” for mosquitoes has been addressed in news articles (Bailey, 2022) where it is stated “Spring rain, summer drought, and heat created ideal conditions for mosquitoes to spread the West Nile Virus through Colorado last year” (p. 1). Visualization 1 yields information on the Goldilocks zone for “High” WNV counties (this is the filter applied in Visualization 1) as:

* Temperature: 57 – 67 degrees Fahrenheit
* AvgRainfallJulyOct 5.5 – 7 inches

which, according to the article by Bailey might also need a spring rain variable and a drought variable.

Visualization 1 also indicates that Denver County (formally the City and County of Denver) might be an outlier and could be removed for future analysis. It would also be prudent to address the other city/county designated area: City and County of Broomfield for potential removal as well. These large urban areas could also be importing the virus from camping or other “out of area” activities.

Visualization 1 also indicates that the number of WNV cases increase as the area of surface water increases in the county. This indicates that minimization of water area (particularly in urban counties) could lead to a lessening of the WNV impact in urban areas. This would also align with the West’s increasing water conservation issues – reduction of water features in golf courses, urban parks, and other (arguably) unnecessary water features in the urban landscape.

While Visualization 1 focused on “High” WNV counties (counties that have had WNV present for more that 80% of the years in this study), Visualization 2 explores, geospatially, the impact of urban (greater than 150,000 population) counties in Colorado. It should be noted that over the time span of this study, two counties, Mesa, and Pueblo, changed from rural to urban (Mesa eclipsed 150,000 in 2015 and Pueblo in 2006), noting the dynamic nature of studies of this type.

Visualization 2

Geospatial analysis of the WNV load in Colorado

Graphical user interface, chart, treemap chart

Description automatically generated

Visualization 2 indicates that urban counties follow the I-25 and I-70 corridors (except in the mountains), and that WNV also follows the urban corridors particularly in the counties that have more surface water area. Additionally, the population density choropleth verifies that Denver County could be a candidate for removal from this study. These maps could be utilized to effectively direct mosquito and WNV mitigation efforts.

Interactive visualizations can be found in the .zip folder submitted for the GoCode Colorado Competition.

**Findings**

WNV follows the urban corridor throughout Colorado, occurring more in urban counties with large surface areas of water. WNV mosquitoes enjoy certain temperature ranges and precipitation conditions. WNV occurs during the time period July – October and occurs more in an El Niño year. Counties can be classified as high, medium, or low depending on their WNV load, directing mitigation efforts to areas most in need. It appears that any major WNV events would be located in one of Boulder, Larimar, or Weld counties.

WNV is in Colorado and will be with us for the foreseeable future. Major findings include the positive correlation of WNV with temperature, the negative correlation with precipitation, and the positive correlation with El Niño years. Current climate change predictions coupled with the results of this study indicate that WNV cases in select counties in Colorado can be expected to increase, indicating a need for mitigation strategies and WNV preparedness of medical personnel.

**Business Benefit**

Hospitals, EMS, state, county, and local officials (health and water agencies) can all benefit from this analysis.

Medical personnel can benefit by being aware of high-risk counties for WNV and being alert to the symptoms. State and county officials can direct limited resources to those counties most in need of mosquito mitigation strategies, surveillance activities, and water conservation efforts (in particular, surface water features). This can protect the citizens of Colorado and protect businesses from lost employee productivity.

**Summary and Future Direction**

Addressing WNV in Colorado, its causes and primary locations can protect the citizens and the labor force of Colorado. Mitigation strategies should be directed to identified high risk counties of the state. Early warning for medical personnel can be based on weather patterns observed and/or other indicator variables.

Future research would include variables not available or not gathered in the current study, including:

* spring rainfall data
* a drought variable
* WNV patient latitude and longitude (or address or zip code) (for drill down)
* average humidity per county
* average elevation per county
* mosquito data (location, population density)
* bird data (location, population density, migration patterns)
* garden zone (planting guidelines, might be based on elevation)
* agricultural variables (water usage)
* monthly data rather than annual data (for drill down)

Analysis needed to enhance the model would include addition of identified variables, drill down analysis, and interaction effects (in particular between temperature, humidity, and elevation – if available – in pursuit of the elusive “mosquito line” – a hypothesized but never derived “contour” line (iso-line) that would divide the state into two zones – mosquito and mosquito-free). A variable map dividing the variables into categories (environmental, physical, human-caused, filters) which could be used to gage the impact of each of the categories.

**References**

Bailey, M. (2022). Climate Change May Push the US Toward The ‘Goldilocks Zone’ For West Nile Virus. Retrieved from: <https://techilive.in/climate-change-may-push-the-us-toward-the-goldilocks-zone-for-west-nile-virus/>

Centers for Disease Control (CDC). (2021). West Nile virus disease cases reported to CDC by week of illness onset, 1999-2020, graphical result. Retrieved from: <https://www.cdc.gov/westnile/statsmaps/cummapsdata.html>

Ciota, A. (2017). West Nile virus and its vectors. Current Opinion in Insect Science. 22, p. 28-36.

Colorado Department of Public Health and Environment (CDPHE). (2022). West Nile Virus Data. Retrieved from: <https://cdphe.colorado.gov/animal-related-diseases/west-nile-virus/west-nile-virus-datahttps://cdphe.colorado.gov/animal-related-diseases/west-nile-virus/west-nile-virus-data>

Excel. (2022). Software package. Microsoft Corporation.

Garcia-Carrasco, J., Muñoz, A., Olivero, J., Segura, M., and Real, R. (2021). Predicting the spatio-temporal spread of West Nile Virus in Europe. PLoS Neglected Tropical Disease, 15(1).

Hadfield, J., Brito, A., Swetnam, D., Vogels, C., Tokarz, R., Andersen, K., Smith, R., Bedford, T., and Grubaugh, N. (2019). Twenty years of West Nile virus spread and evolution in the Americas visualized by Nextstrain. PLOS Pathogens. 15(10). p. 1-18.

Kramer, L., Li, J., and Shi, P. (2007). West Nile Virus. The Lancet Neurology. 6(2), p. 171-181.

Marzec, N. (2022). Personal email communication. MPHZoonotic Disease Unit Manager

Communicable Disease Branch, CDPHE.

Petersen, L. (2019). Epidemiology of West Nile Virus in the United States: Implications for Arbovirology and Public Health. Journal of Medical Entomology. 56(6). p. 1456-1462.

Ronca, S., Ruff, J. and Murray, K. (2021). A 20-year historical review of West Nile virus since its initial emergence in North America: Has West Nile virus become a neglected tropical disease? PLoS Neglected Tropical Diseases. 15(5). p. 1-14.

Tableau. (2022). Software package. Retrieved from: <https://www.tableau.com>

Wildlife Futures Team. (2021). West Nile Virus. Retrieved from: <https://www.vet.upenn.edu/research/centers-laboratories/research-initiatives/wildlife-futures-program/resources/fact-sheets/fact-sheet-detail/west-nile-virus#:~:text=West%20Nile%20virus%20infects%20over,bite%20of%20an%20infected%20mosquito>

Zaiontz, C. (2022). Real Statistics software add-in for Excel. Available from: <https://www.real-statistics.com/>