Predicting West Nile Virus in Chicago

DSI 5: Aakash Tandel, Chaim Gluck, Matt Terribile, Nellie Carr

August 11, 2017

Outline

- Scope of work
- Hypotheses of key contributors
- Models
- Cost-benefit analysis
- Conclusions

Scope of Work

Using data collected from 2007 through 2013 at 148 mosquito traps placed throughout the city of Chicago, we built a model to predict the locations at which traps were most likely to indicate occurrence of West Nile Virus.

Based on this analysis and further research, we developed recommendations and guidance addressing the most effective ways to structure future pesticide spraying efforts

Background research

Temporal and spatial patterns of precipitation and air temperature have a consistent and significant impact on the timing and location of increased mosquito infection

NIH Publication: Parasit Vectors. 2010 Mar 19;3(1):19. doi: 10.1186/1756-3305-3-19.

Effects of environmental factors on West Nile virus

Increased air temperature → Increased infection in Culex pipiens and Culex restuans mosquitoes

Cumulative high temperature → Higher mosquito infection and higher human illness rates

Drier conditions followed by wetter conditions → Increase in infection *in some years*

NIH Publication: Parasit Vectors. 2010 Mar 19;3(1):19. doi: 10.1186/1756-3305-3-19.

Vector Index

$$\text{Vector index} = \sum_{i=1}^{n} A_i P_i$$

- The Vector Index (VI) is an estimate of the abundance of infected mosquitoes in an area

- incorporates information describing the vector species that are present in the area, relative abundance of those species, and the WNV infection rate in each species into a single index (Gujaral et al. 2007, Bolling et al. 2009, Jones et al. 2011).

→ the sum across n mosquito species of the abundance, Ai (mosquitoes per trap-night), multiplied by the WNV infection prevalence, Pi, for that species, *i*, for that time period

Predicting Human West Nile Virus Infections With Mosquito Surveillance Data, *American Journal of Epidemiology*, Volume 178, Issue 5, 1 September 2013, Pages 29–835,https://doi.org/10.1093/aje/kwt046

Vector Index

- A vector index using a 2-week estimate of prevalence can be used to predict human infection up to 3 weeks later

- A “smoothed” or 3-week moving window of the vector index reduces the substantial week-to-week variability in mosquito abundance and prevalence and results in a more accurate risk index

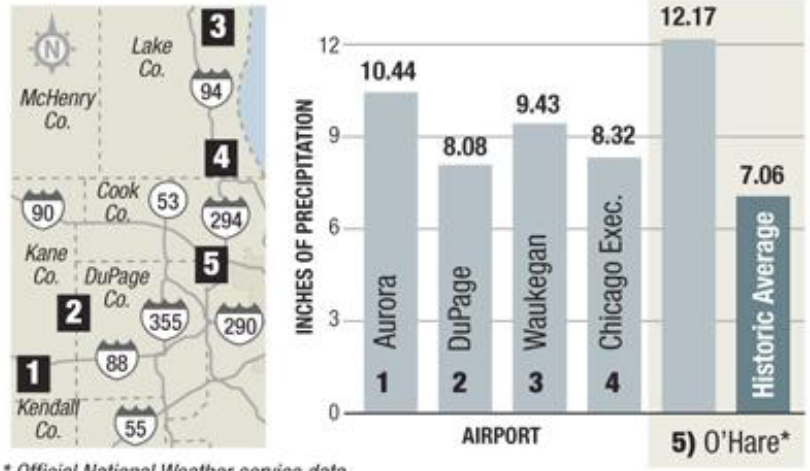
Predicting Human West Nile Virus Infections With Mosquito Surveillance Data, *American Journal of Epidemiology*, Volume 178, Issue 5, 1 September 2013, Pages 29–835,https://doi.org/10.1093/aje/kwt046

Our model included

Wet spring, buggy summer

With area rainfalls well above average, experts predict a bumper crop of mosquitoes in the next two weeks.

Total precipitation, April and May 2011

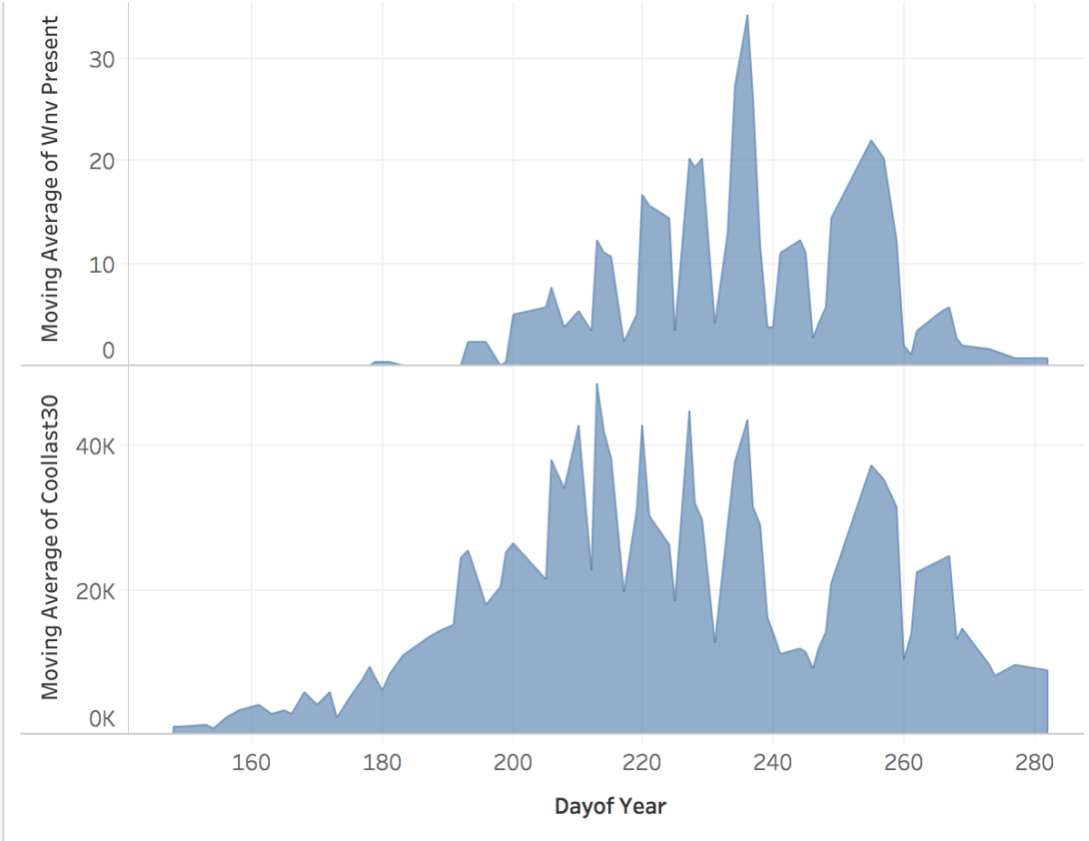


* Official National Weather service data

Source: Weather Underground - www.wunderground.com

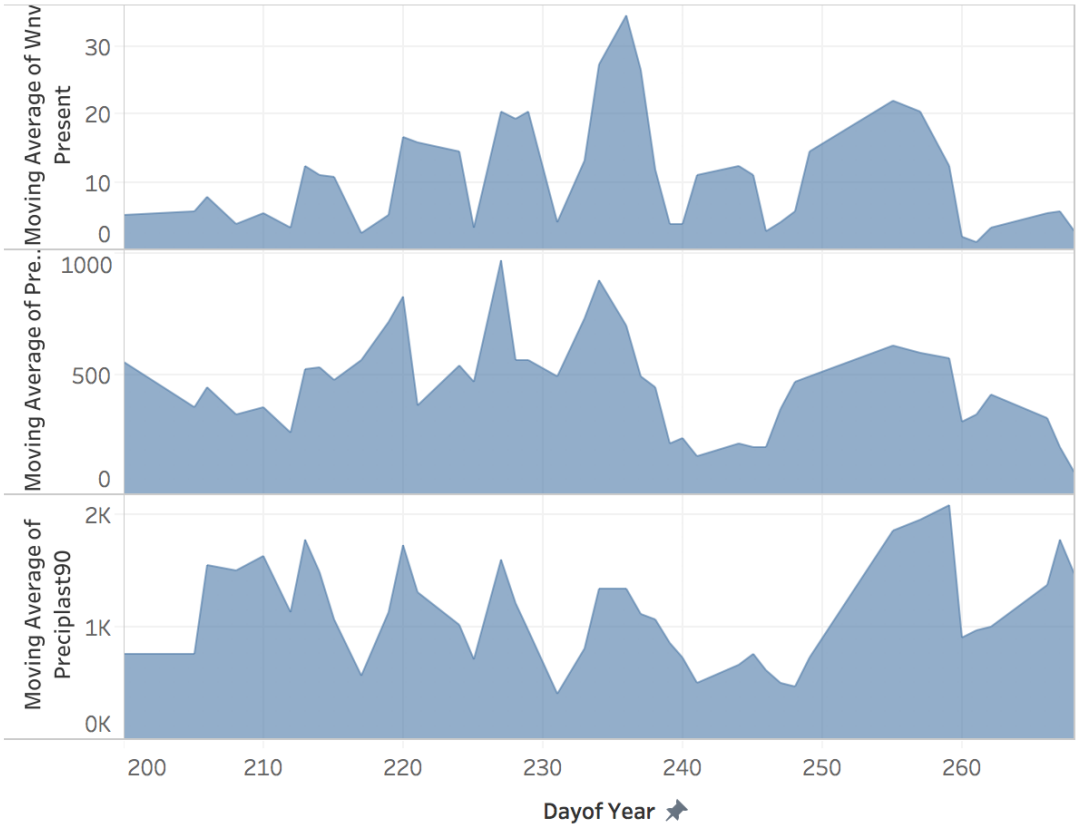
DAILY HERALD

WNV Presence vs. Cooling Degree Day History

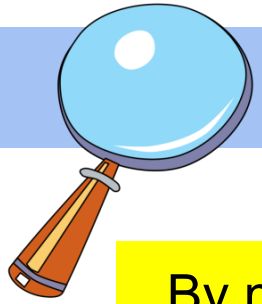


Data supports correlation between WNV spread and temperature in previous 30 days

WNV Presence vs. Precipitation History



Data supports correlation between WNV spread and precipitation in previous 30 and 90 days



Hypothesis

By modeling variations in temperature and moisture over time, we can better predict the volume of mosquitos and the likelihood of West Nile Virus

Key factors:

- Patterns of warm days → cooling degree days
- Patterns of increased moisture accumulating → precipitation

Modeling: Engineered features

Cooling Degree Days

Precipitation

Single day

7 days previously

8 to 30 days previously

31 to 90 days previously

Modeling: weather data

Data from 2 weather stations, covered warmer months

Weather data covered all dates in train and test data

LOCATION data:
- largely redundant (trap and lat/long are effectively 1:1)
- Categorical

COMBINED ROWS → one set of observations across both weather stations per day

Merged weather data into train and test data

MODELS

DECISION TREES

- Poor job of predicting

RANDOM FOREST

- Better than the single decision tree

MODELS

RANDOM FOREST & FEATURE IMPORTANCE

- Trap Level Data
- Species of Mosquito
- Day of Year
- Engineered Features - Cumulative Precipitation
- Engineered Features - Cumulative Cooling Days

MODELS

Boosting Methods

Utilizing Decision Stumps

AdaBoost & Gradient BOOST

- Performed worse than the random forest

XGBoost

- A Kaggle favorite

MODELS

XGBoost

- XGBClassifier vs. DMatrix
- Gridsearch and Default Parameters
  - Max Depth = 3
  - Learning Rate = 0.1
  - Number of Estimators = 100
  - Gradient Boosting Trees

MODEL: AUC / ROC

XGBoost with Random Forest Important Features

PUBLIC KAGGLE SCORE: 0.78550 (#379)

PRIVATE KAGGLE SCORE: 0.76924 (#360)

West Nile Virus in Cook County

YEAR	WNV Cook Cty	WNV Ilinois
2016	**90**	155
2015	**27**	77
2014	**N/A**	44
2013	**60**	117
2012	**174**	290
2011	**22**	34
2010	**30**	61
2009	**1**	5
2008	**9**	20
2007	**33**	101

Manifestation Rate of West Nile virus

On average, 2 in 10 people who are bitten by an infected mosquito will actually become ill.

Less than 1% of infected people develop a serious, sometimes fatal, neurologic illness.

In 2016, given that there were "***90 Lab Confirmed Cases of West Nile in Cook County***", we can conservatively assume:

- **90** became ill
- **450** were bitten
- **< 1** acquired serious/fatal illness

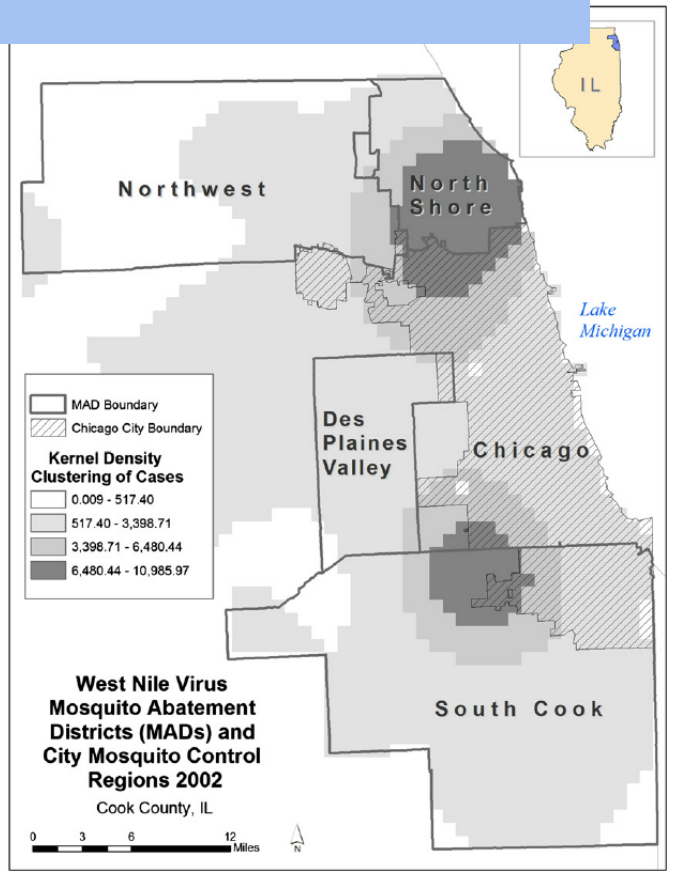
http://www.dph.illinois.gov/topics-services/diseases-and-conditions/west-nile-virus/surveillance/humanCases2016

RECOMMENDATIONS

Spend at least $1.5 mm on mosquito abatement.
Spraying likely to cost ~$2k per square mile.
City of Chicago is 234 square miles.
$500k to spray the entire city.

Comps:

- Northwestern Cook County at 242 square miles, spent > $2.3 million on mosquito abatement in 2015.
- South Cook County at 340 square miles spent ~$2.4 million on mosquito abatement.



The Cost of West Nile virus

“Consider limiting outdoor activity after dark (dusk to dawn), which is when Northern House Mosquito is most active.”

- City of Chicago guidance

Pupils in Arlington Heights School District 25 . . . 5,000 children were not allowed to go outside for recess or gym class Thursday after staff found the crow on the grounds of Dryden Elementary School

“West Nile virus scare keeps pupils indoors” September 2001, Chicago Tribune
http://articles.chicagotribune.com/2001-09-07/news/0109070047_1_west-nile-virus-crow-dusk-and-early-evening

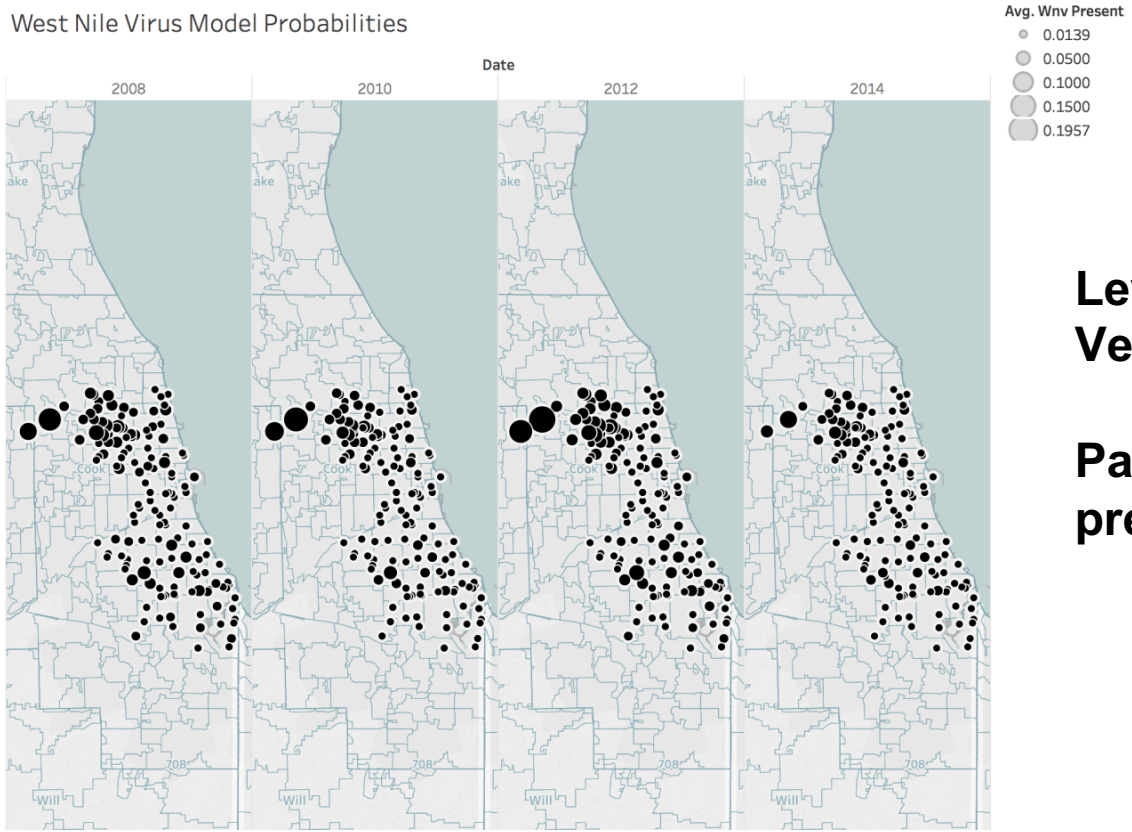
RECOMMENDATIONS

Disband the mosquito abatement districts ($260 / sq. mile v. $166k / sq. mile)

Emphasis on prevention and larvicide. While larviciding is a treatment that selectively kills mosquito larvae, adulticiding is a broad spectrum application that can kill beneficial insects as well as mosquitoes . Adulticiding is also much more expensive than larviciding and, to be effective, requires precise product and equipment selection and calibration, accurate, thorough application and timing, as well as favorable environmental conditions

Based on Vector Index research: conduct additional analysis addressing proximity to vegetation and bodies of water, to better predict locations of WNV.

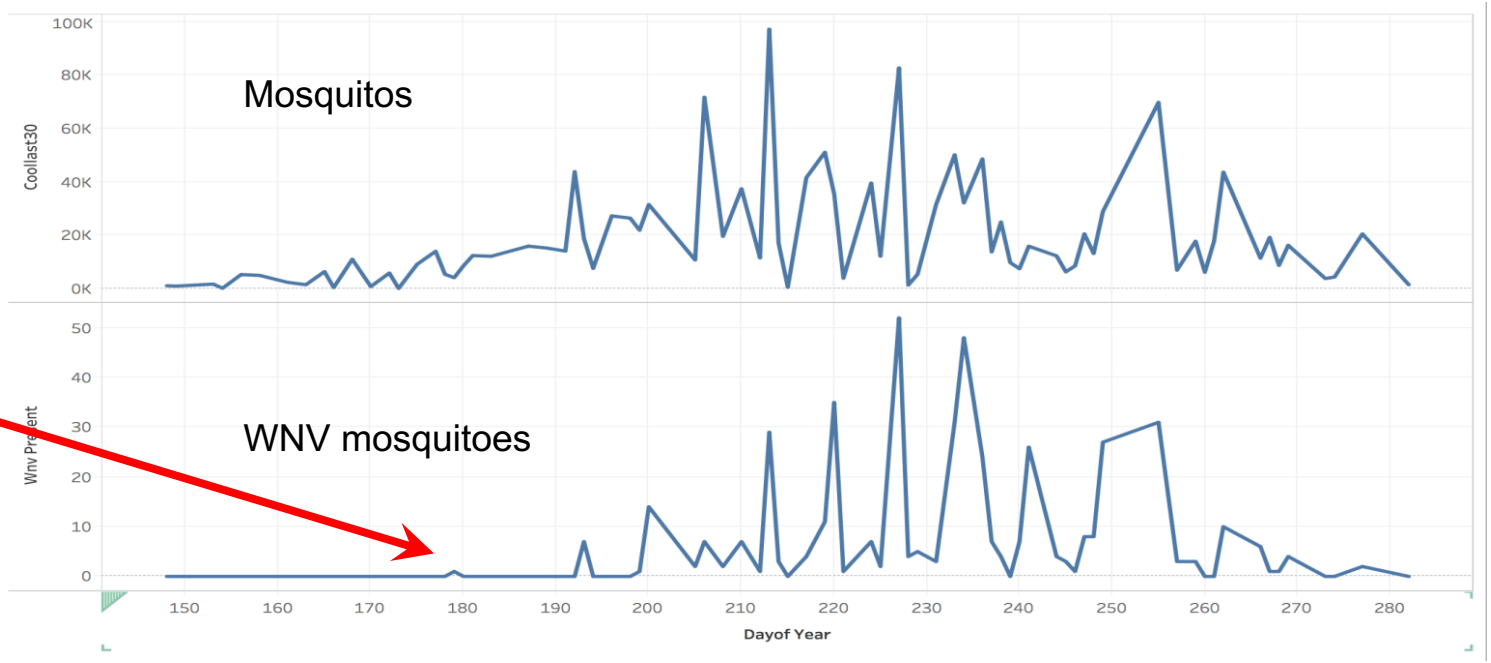
RECOMMENDATIONS: **WHERE** to spray



Leverage the utility of the Vector Index!

Past performance *does* predict future performance!

RECOMMENDATIONS: **WHEN** to spray



~ June 27

Spray in the evening when mosquitoes are active, temperature is between 60 F and 85 F, and there is little wind