PREDICTION ON WEST NILE VIRUS IN CHICAGO

DSI-15 PROJECT 4 GROUP 2

Chua Hung Boon

Dominic Teo

Ray Chia

GENERAL INFORMATION ON WEST NILE VIRUS

- 80% of people are asymptomatic
- Older people and those with weakened immune systems are most at risk.
- Vaccines are not yet available for people

BACKGROUND INFORMATION

With climate change happening in the world, there is an increase of spread of mosquitoes thriving under warmer temperatures

Faced with an epidemic of West Nile Virus here in the Chicago

Department of Public Health has set up a surveillance and control system

Disease And Treatment Agency, division of Societal Cures In Epidemiology and New Creative Engineering (DATA-SCIENCE).

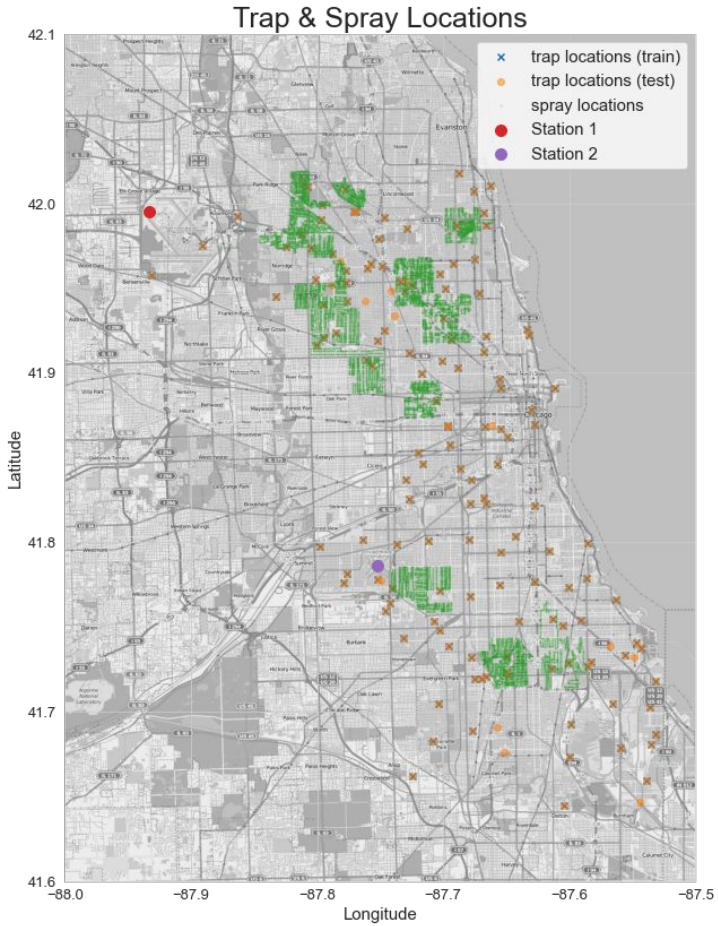
PROBLEM STATEMENT

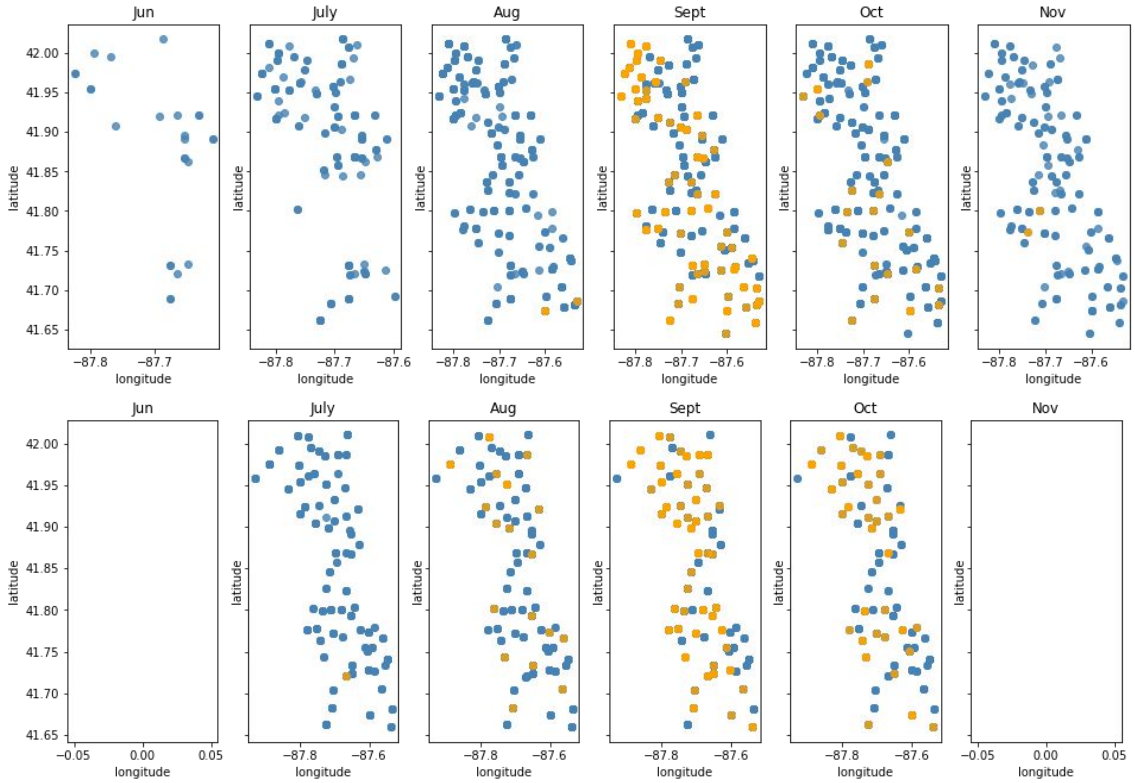
- Predicting whether or not West Nile Virus is present, for a given time, location, and species.
- Cost projections for spraying pesticide VS effects and benefits
- Receiver Operating Characteristic (ROC) Area Under Curve (AUC) score > 0.7
- ↑ Recall score - real wnv cases wrongly classified as non case
- Accurately predict outbreaks and help the City of Chicago to efficiently allocate resources towards preventing transmission of WNV

Model Selection

Data Collection	Data Cleaning EDA	Data Preprocessing	Modelling	Production Model Evaluation
mosquito and WNV occurrences in Chicago between 2007 to 2014 [Chicago Department of Public Health]	Remove duplicates	Combine WNV occurrences data against local weather data.	Setup training and validation splits dataset	Determine Production model and parameters
Chicago Weather Data between 2007 to 2014	Identify key information columns	Feature Engineering for modelling	Explore **Logistic Regression, Random Forest and Extremely Randomized Trees Models**	Review Production Model metrics (Accuracy, Specificity, Precision, Sensitivity, ROC-AUC)
GIS data for their spray efforts in 2011 and 2013			Construct & Run Gridsearch to optimize hyperparameters	Review Feature Importance

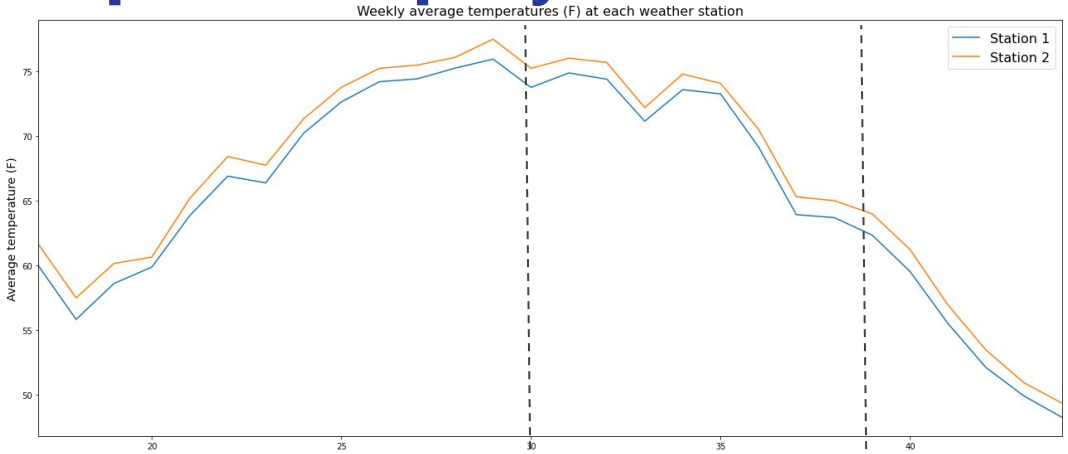
Traps and Spray Activities locations

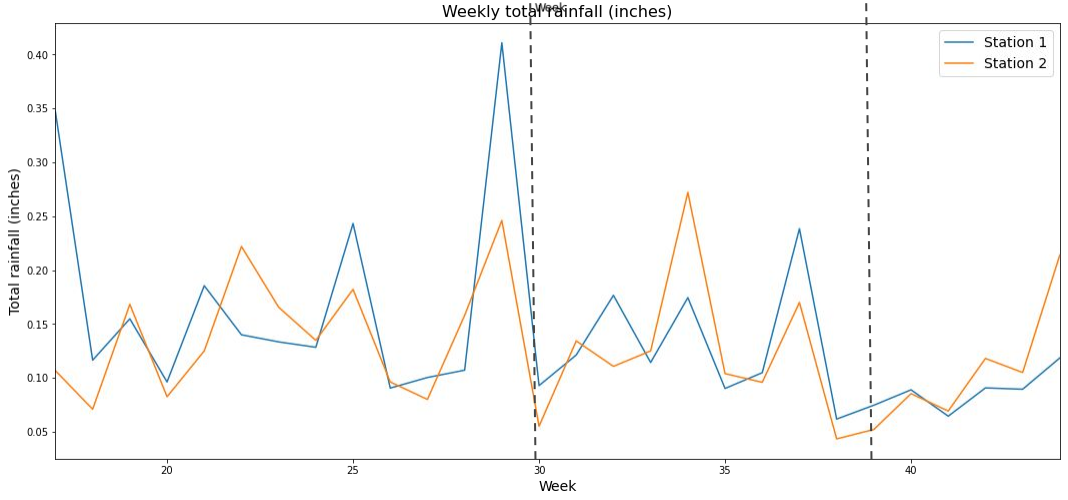




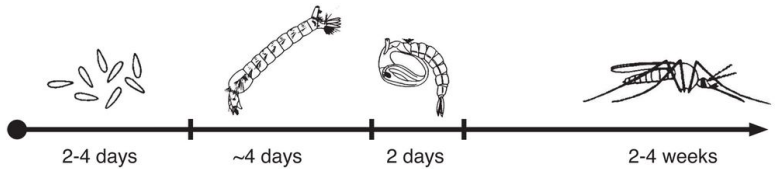
- Trap locations found with WNV borne mosquitos
- Between 2007 to 2013, months of Aug, Sep Oct most occurrences

Traps and Spray Activities locations

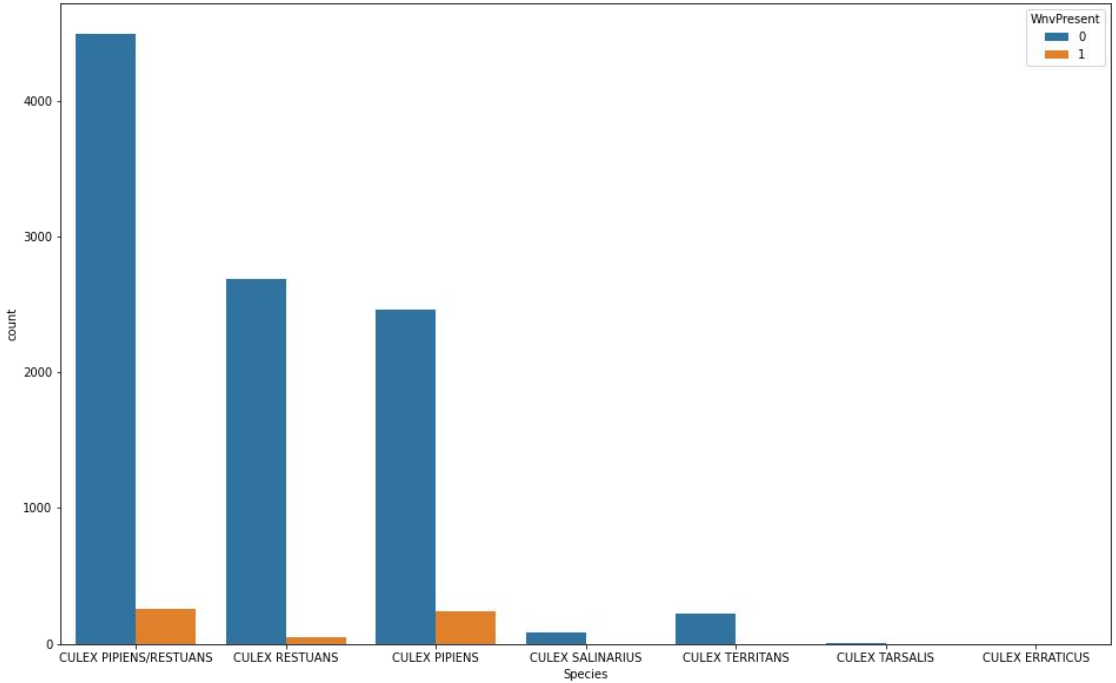




- Between 2007 to 2013, months of Aug, Sep Oct most occurrences
- From eggs to adult mosquito takes about 8 to 12 days
- Adult mosquito may survive up till to 14-28 days
- High rainfall and high temperature maybe the prelude to high breeding

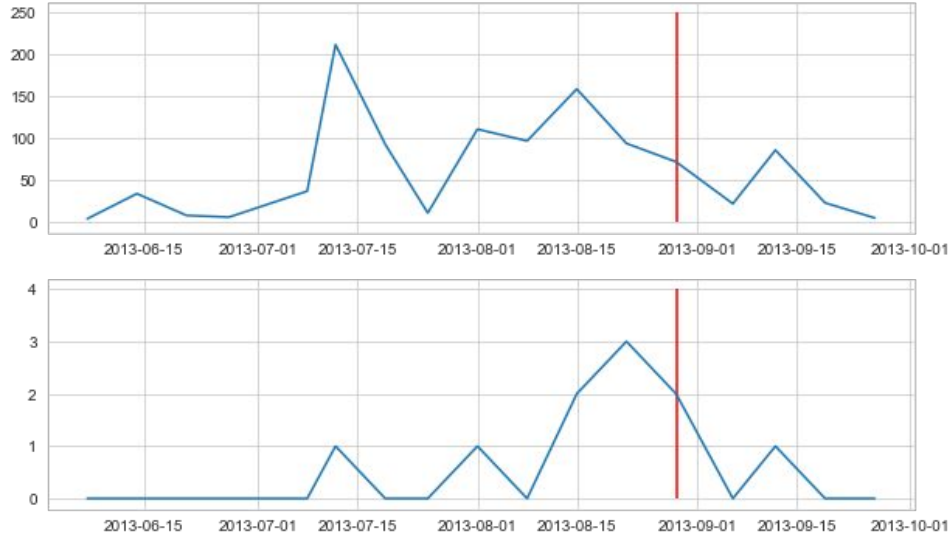


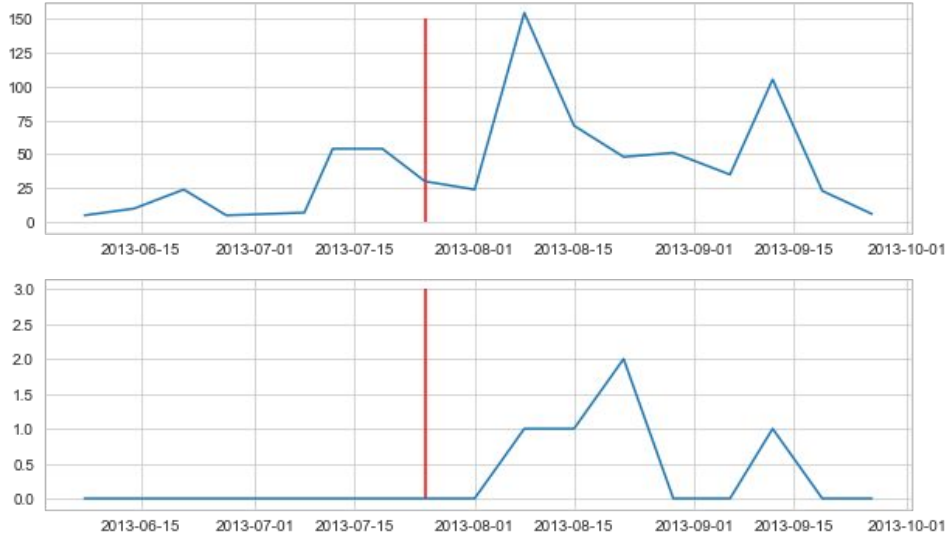
WNV borne Species



- Culex Pipens and Culex Restuans are the WNV borne species

Spraying Efficacy





- Inconsistent observations
- Spraying efforts not always able to control mosquito counts and WNV occurrences

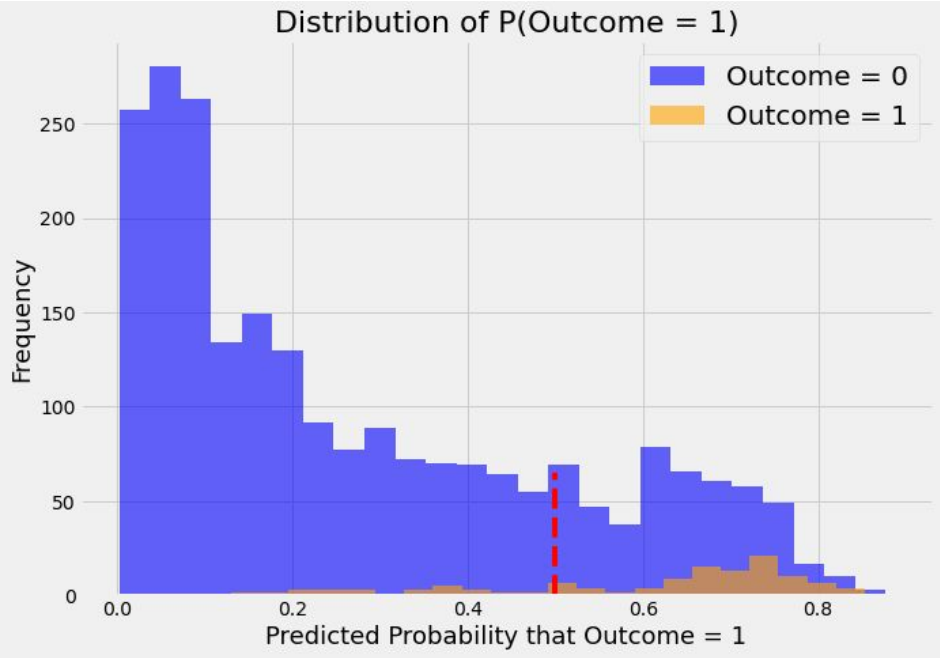
Model Selection

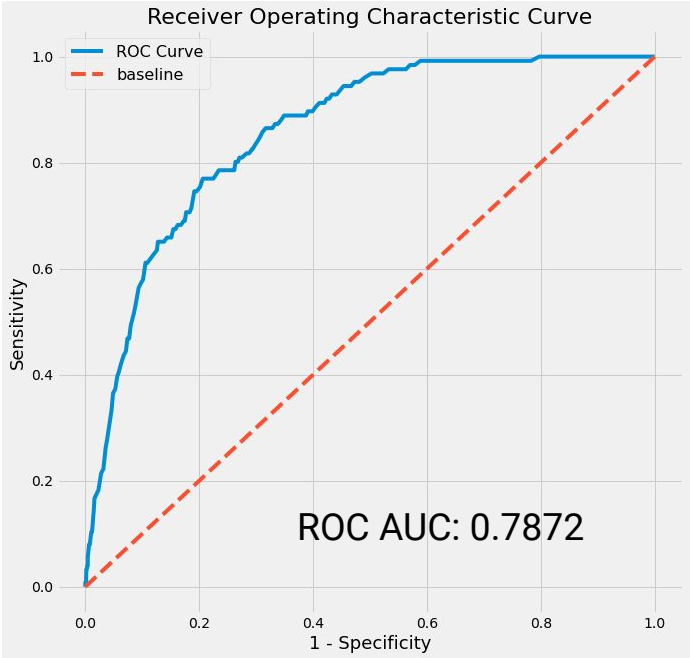
Scores/Model	Logistic Regression	Random Forest	Extreme Randomized Trees
CV Best Score	0.7673	0.836	0.8294
CV Best Score std	0.0186	0.0238	0.0244
Train Set ROC-AUC	0.8361	0.9183	0.9445
Validation Set ROC-AUC	0.7802	0.8637	0.8571
Validation Set Recall	0.7381	0.746	0.7698
Validation Set Precision	0.1144	0.1744	0.1675
Validation Set Accuracy	0.6894	0.8032	0.7892

- Objective is to FIRST create a model that can generalize well followed by best ROC-AUC and Recall scores
- Extreme Randomized Trees and Random Forest models have higher spread among it's cross validation tests scores.
- Logistic Regression has lowest spread among it's cross validation test scores
- Extreme Randomized Trees has highest difference between Train and Validation set scores
- Logistic Regression and Random Forest have similar Train-Validation test scores difference
- Although not Logistic Regression do not have the best scores, it is more likely to generalize unseen data

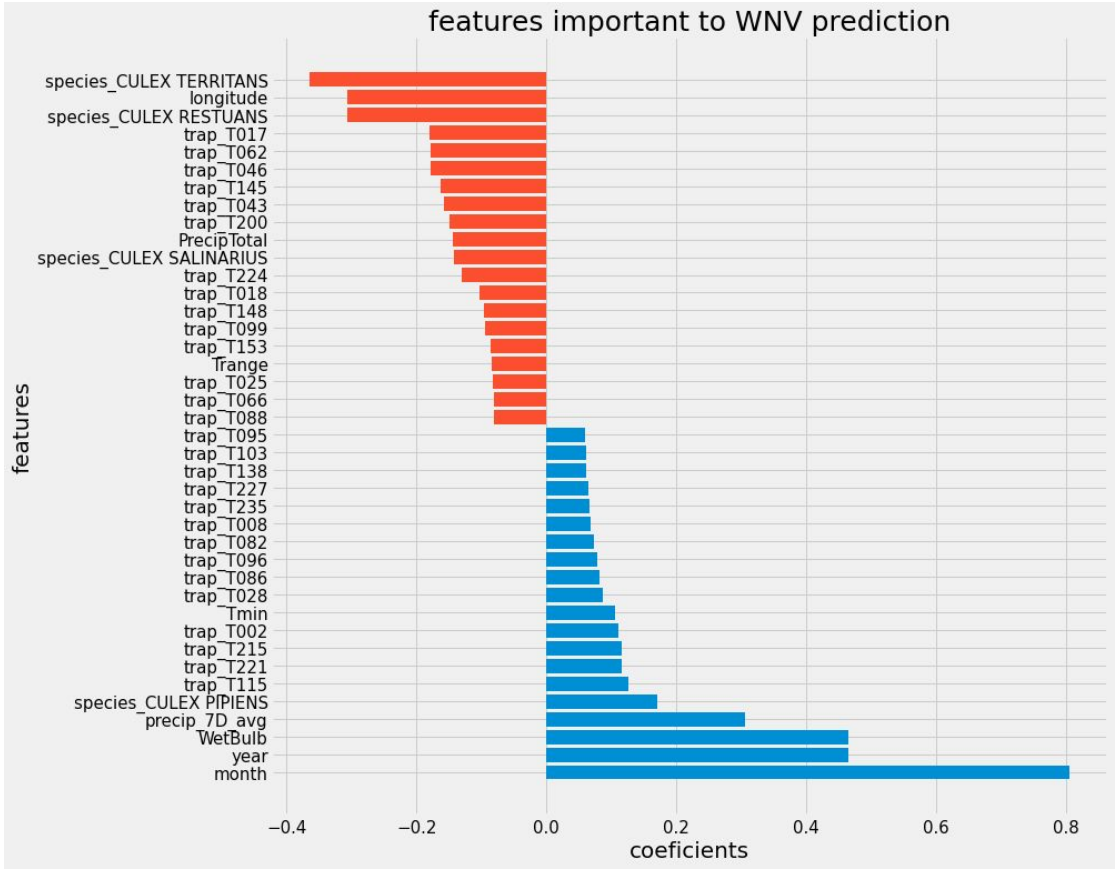
Production Model

	precision	recall	f1-score	support
WNV=0	0.98	0.72	0.83	2298
WNV=1	0.13	0.75	0.22	126
accuracy			0.72	2424
macro avg	0.55	0.73	0.52	2424
weighted avg	0.94	0.72	0.80	2424





Feature Importance



WNV occurrences is highly related to

- month of the year. Likely due to higher temperature, rainfall between Aug and Oct
- Wetbulb, precip 7D ave: Humidity related
- Mosquito Species present
- trap ids, probably location related (e.g. locations with many stagnant water traps)

Cost-Benefit Analysis

Symptoms

20–30% of cases develop influenza-like illness, defined as West Nile fever (WNF),

<1% of infected individuals, mainly elderly and immunocompromised patients, develop neuroinvasive disease (WNND)

Taken from: MDPI--Human West Nile Virus Lineage 2 Infection: Epidemiological, Clinical, and Virological Findings

Cost of WNV case study: Sacremento county, California, 2005

	symptom/diagnosis	count	count_distribution(%)	productivity cost	medical cost	cost_per_patient
0	Asymptomatic	772	82.6	0	0	0.0
1	West Nile Fever	117	12.5	101505	19539	1035.0
2	West Nile Neuroinvasive Disease	46	4.9	484800	2140409	57070.0

Calculated using figures from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3322011/

Aerial spray response to the WNV epidemic cost a total of **$701,790**

Economic breakeven If **>=15 WNND** patients Are prevented

Benefit of spraying

"Carney et al. documented 18 (13 WNF, 5 WNND) total WNV disease cases outside the spray area after the Sacramento County emergency spray and ***no cases*** *within the spray area..."*

What about Chicago?

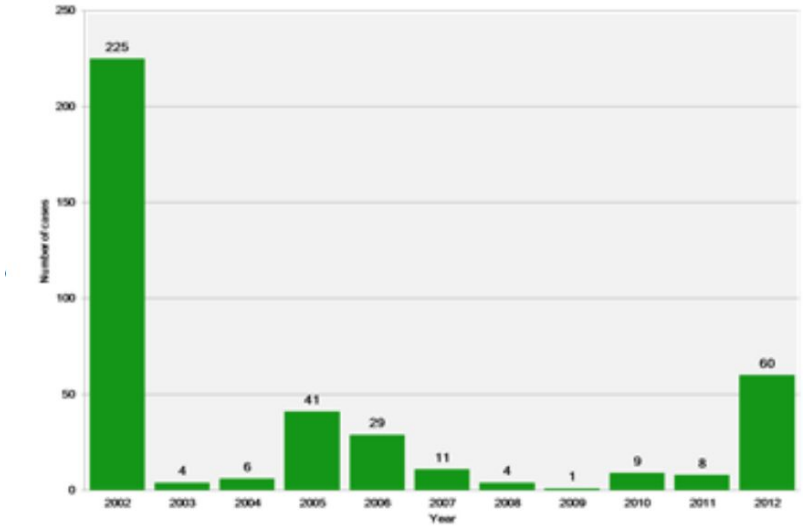
477km2 (spray area) → $700,000 worth of vector control (Sacramento county)

1km2 → ~$1467

39km2 → ~$57000 (cost of one WNND patient)

What about Chicago?

Figure 1. Number of reported confirmed and probable cases of West Nile virus among Chicago residents by year, 2002-2012.



Source: https://www.chicago.gov/content/dam/city/depts/cdph/statistics_and_reports/CDInfo_2013_JULY_WNV.pdf

Recommendations

Preventive: Larvicide treatments (model prediction)
Reactive: Adulticide treatments (ie. Sacramento)

"[Larviciding is] the most cost-effective way of attacking the mosquito population."

-Dennis Wallette,
director of the Tangipahoa Mosquito Abatement District in Hammond

Thank you
Any questions?