Identifying Areas at Greatest Risk for Recent Zika Virus Importation — New York City, 2016

Sharon K. Greene, PhD, MPH, Sungwoo Lim, DrPH, and Annie D. Fine, MD
New York City Department of Health and Mental Hygiene

Motivation

- The New York City (NYC) population includes many travelers and recent immigrants and is at risk for travel-associated communicable diseases, including Zika virus (ZIKV) disease
- Infected persons who acquire ZIKV while traveling and are viremic in NYC might be bitten by *Aedes albopictus* vectors, introducing an unknown but presumably low risk of local transmission
- Throughout the 2016 mosquito season, we sought to identify areas with persons with ZIKV viremia — locations where mosquitoes could become infected with ZIKV — to target public education and trapping and controlling *Aedes* spp. mosquitoes

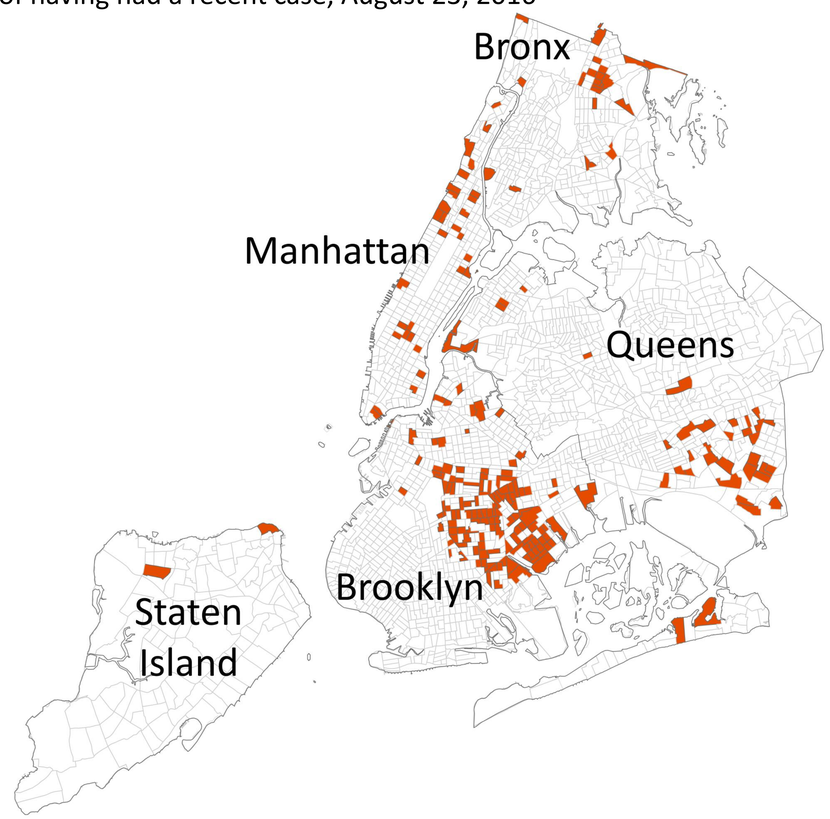
Table 1. Considerations in analysis design

Challenge	Analytic Solution
Reported ZIKV cases incomplete and spatially non-representative, because of asymptomatic infections and under-testing	Applied regression modeling to available data to estimate areas at risk
Define at-risk areas at high geographic resolution to inform programmatic activities	Unit of analysis: census tract (N=2,123 with >25 residents)
Distribution of population at-risk (travelers to ZIKV-affected countries/territories) unknown	Used area-based sociodemographic characteristics potentially associated with ZIKV importation and testing
Areas at-risk change over time: changing incidence in ZIKV-affected countries/territories affects communities concentrated in different areas of NYC	Updated regression model weekly, using latest available data on ZIKV testing and cases
Barriers to seeking care and receiving ZIKV testing in some areas	Fit model restricting to census tracts with any recent testing (i.e., excluded census tracts where having zero observed cases attributable to no testing); for all census tracts, outputted model-predicted individual and cross-validated probabilities of any recent cases
Assessing model performance, i.e., discriminatory ability to predict census tracts with recent cases	Used receiving operator characteristic (ROC) contrast test to compare area under the ROC curve (AUC) for fitted model with cross-validation vs. intercept-only model

Table 2. Variables used weekly in logistic regression model

Data Source	Variable	Role in Model
Communicable disease surveillance database, NYC Department of Health and Mental Hygiene	Any ZIKV cases in prior month (Y/N)	Outcome variable
	Any ZIKV tests in prior month (Y/N)	Restrict to "yes" for model fitting
	ZIKV cases >1 month ago	Independent variables
	Dengue and chikungunya cases since 2013 with travel to areas with local ZIKV transmission	
American Community Survey, 2010–2014	% living below federal poverty level	
	Number with ancestry in areas with local ZIKV transmission (quartile)	
	Number born in areas with local ZIKV transmission (quartile)	
Census 2010	% Hispanic ethnicity	
	% women of childbearing age	
	Total population size (quartile)	

Figure. Census tracts in highest decile of model-predicted probability of having had a recent case, August 23, 2016



Results

- Model adequately discriminated between census tracts with and without recent cases for 8 of 14 weekly analyses, June–September 2016
- No independent variables were consistently associated with presence of recent cases
- Illustrative results, August 23, 2016:
  - Of 752 census tracts with any recent testing, 105 (14%) had any observed recent ZIKV cases
  - Variables independently associated with census tract-level presence of recent ZIKV cases:
    - % women of childbearing age ($P = 0.0007$)
    - No. with ancestry in areas with ZIKV transmission ($P = 0.03$)
    - % Hispanic ethnicity ($P = 0.07$)
  - AUC for fitted model with cross-validation: 0.66
    - ROC contrast test with intercept-only model (AUC=0.50): $P < 0.0001$
  - In a defined Brooklyn neighborhood of 33 census tracts, 3 had recent observed cases, but 23 were in the highest decile of modeled risk, suggesting expanding public health activities in this area might be warranted
  - Areas at highest risk varied over time; e.g., model-predicted probabilities for June 29 vs. August 23: Spearman correlation coefficient: 0.59, $P < 0.0001$

Conclusions

- We used observable characteristics of areas with recent, known travel-associated ZIKV cases to identify similar areas with no observed cases that might also be at-risk in any given week
- Model performance would likely improve if census tract-level data on persons arriving from ZIKV-affected countries were readily available to health departments
- These methods could be applied to any disease where cases are incompletely ascertained and knowledge of how cases are geographically distributed at any moment is important for targeting public health activities

Questions? Suggestions?
Interested in the SAS code?
sgreene4@health.nyc.gov