Detecting Emerging COVID-19 Community Outbreaks

CSTE Spatial Analysis Workgroup

July 22, 2020

Presenters: Sharon K. Greene, Martin Kulldorff

Objectives

- General: to prioritize case investigations and contact tracing efforts, to allocate testing and community engagement resources, and to protect the elderly and other vulnerable populations

- Specific: To determine where SARS-CoV-2 percent positivity is increasing faster than jurisdiction-wide trend, or decreasing slower than jurisdiction-wide trend

- Why SaTScan?
  - Scan statistics can detect increased disease activity anywhere at any time, without a priori specification of temporal period or geographic location or size

Do we want to detect increases in case counts or in percent testing positive?

- Why might COVID-19 case counts increase?
  - Ongoing disease transmission requiring public health investigation
  - Proportionally more patients with SARS-CoV-2 are tested
- Why might the percent positive of SARS-CoV-2 tests increase?
  - Ongoing disease transmission requiring public health investigation
  - Proportionally fewer patients without SARS-CoV-2 are tested
- True disease transmission most apparent when both case counts and percent positive tests increase

EMERGING INFECTIOUS DISEASES®

EID Journal > Volume 22 > Number 10—October 2016 > Main Article

Volume 22, Number 10—October 2016

Dispatch

Daily Reportable Disease Spatiotemporal Cluster Detection, New York City, New York, USA, 2014–2015

Sharon K. Greene, Eric R. Peterson, Deborah Kapell, Annie D. Fine, and Martin Kulldorff

On This Page

Research

PUBLIC HEALTH *Reports*

A Large Community Outbreak of Legionnaires' Disease Associated With a Cooling Tower in New York City, 2015

Public Health Reports
2017, Vol. 132(2) 241-250

DOI: 10.1177/0033354916689620
journals.sagepub.com/home/phr
SAGE

Morbidity and Mortality Weekly Report (*MMWR*)

CDC

Salmonellosis Outbreak Detected by Automated Spatiotemporal Analysis — New York City, May–June 2019

Weekly / July 3, 2020 / 69(26);815–819

Julia Latash, MPH[1]*; Sharon K. Greene, PhD[1]*; Faina Stavinsky, MS[2]; Sandy Li[3]; Jennifer A. McConnell, MS[3]; John Novak, PhD[3]; Teresa Rozza[3]; Jing Wu, PhD[3]; Enoma Omoregie, PhD[3]; Lan Li, MPH[1]; Eric R. Peterson, MPH[1]; Bruce Gutelius, MD[1]; Vasudha Reddy, MPH[1] (View author

Practice Full Report

SDC

Detecting Drop-offs in Electronic Laboratory Reporting for Communicable Diseases in New York City

Sharon K. Greene, PhD, MPH; Erin M. Andrews, DrPH; Pamela Evans Lloyd, MPA; Jennifer Baumgartner, MSPH; Eric R. Peterson, MPH

Why different from existing aberration detection processes?

- Different goals:
  - Detect new outbreak in context of minimal or stable disease incidence

  vs.
  - Monitor ongoing outbreak

- Case count-only analyses are limited by temporal and spatial shifts in SARS-CoV-2 testing availability and uptake

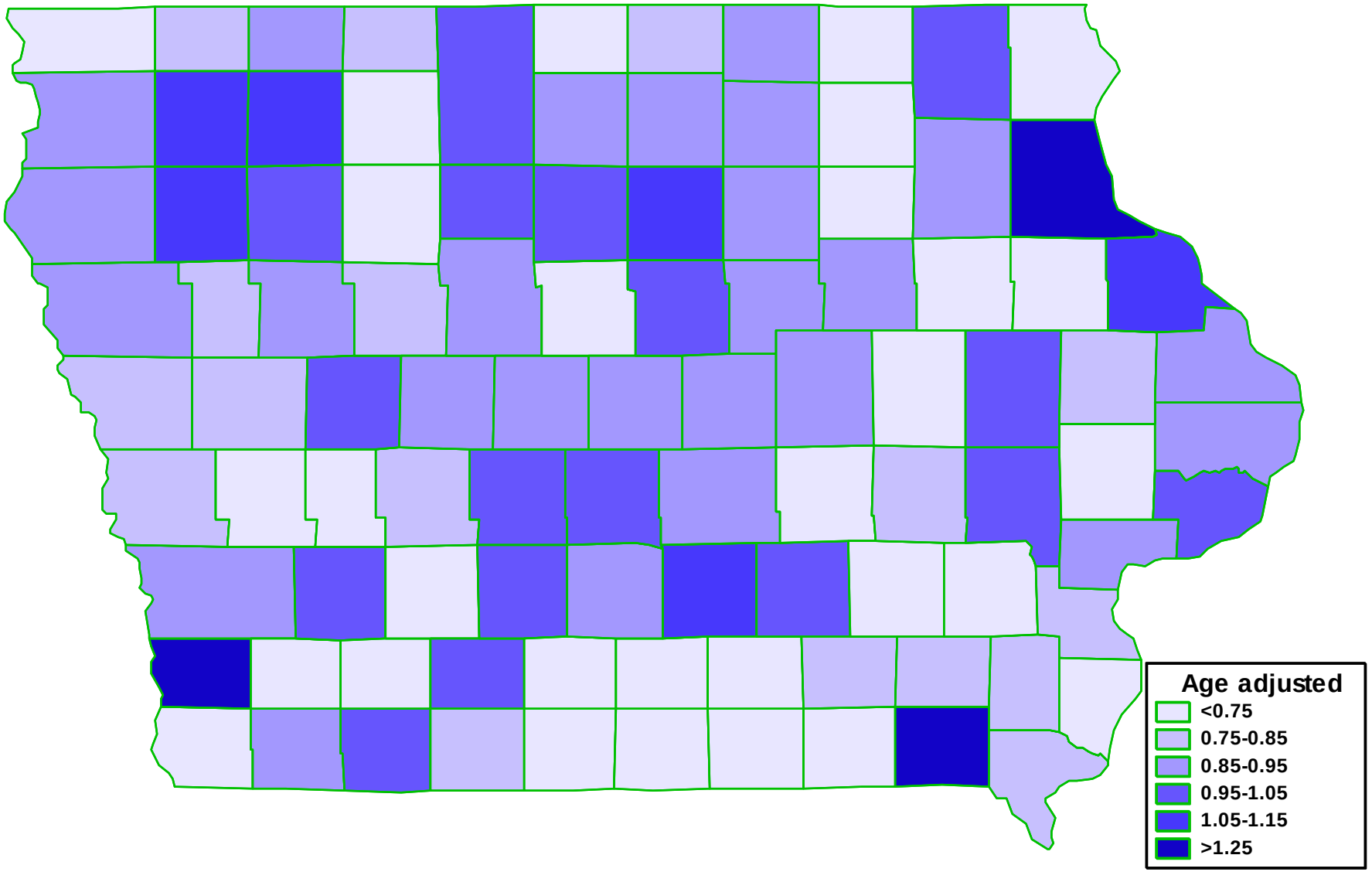
Strategy

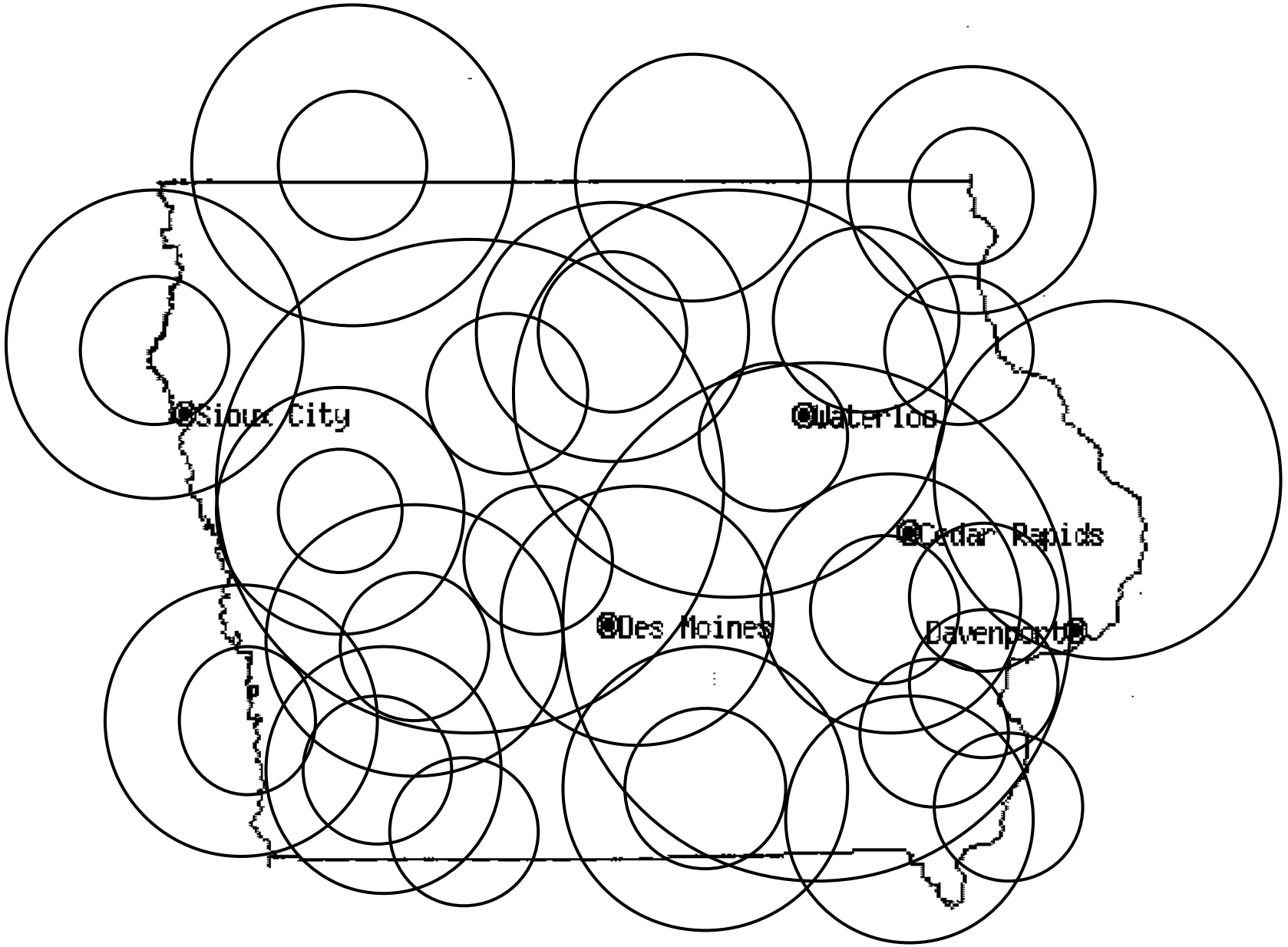
- Automate daily analyses of geocoded lab data to identify community clusters of increased test positivity
- Prospective discrete Poisson-based space-time scan statistic
  - “Population” = unique persons receiving PCR testing
- Adjust for purely spatial variation nonparametrically using spatial stratified randomization
- Adjust for purely temporal trends using log-linear trend adjustment
- Details on medRxiv

Statistical Analysis

- Spatial Scan Statistic
- Space-Time Scan Statistic
- Different Probability Models
- Adjusting for Covariates, Geographical Variation and Temporal Trends
- Free SaTScan software

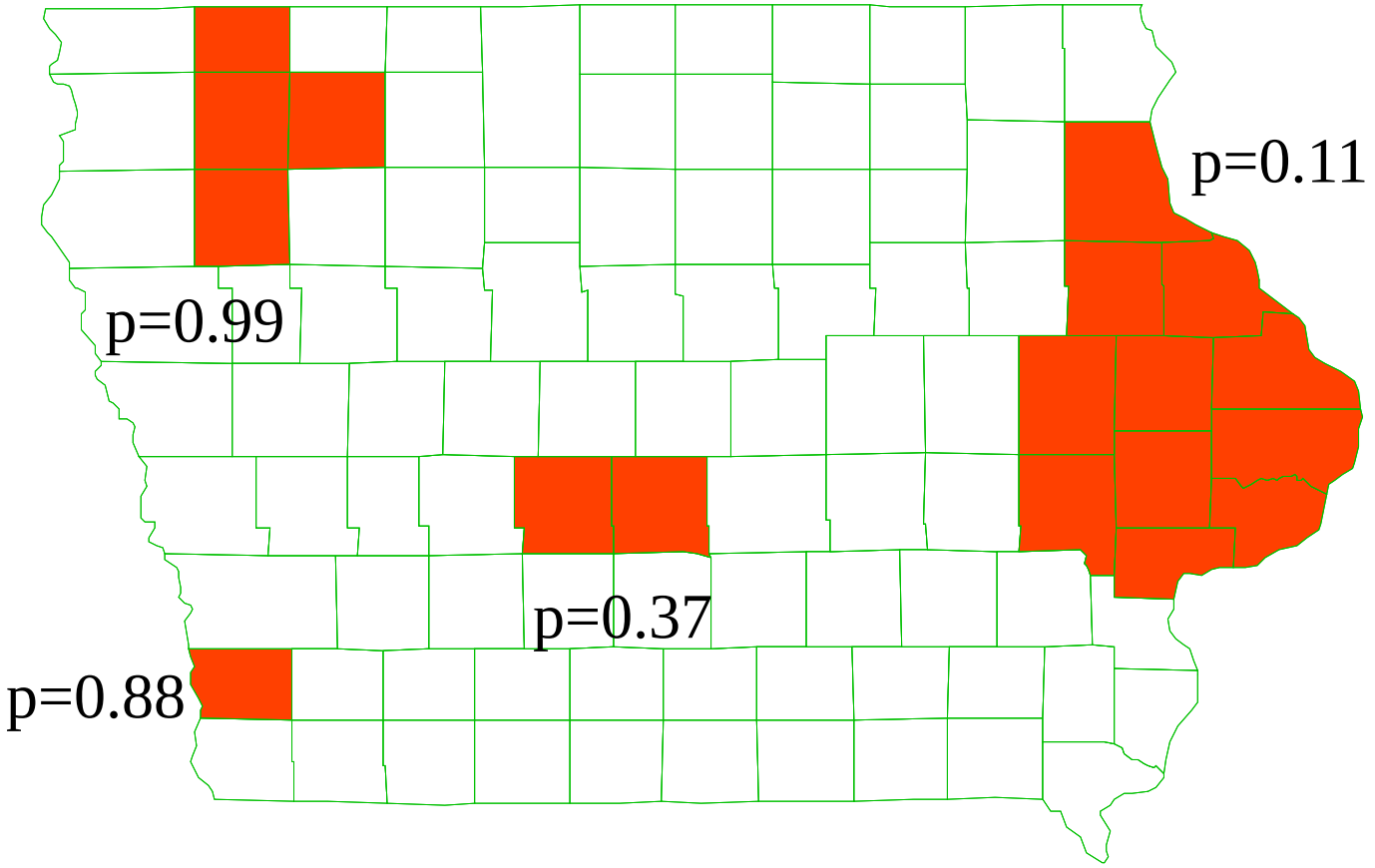
Iowa Breast Cancer Incidence
Age-Adjusted Relative Risks

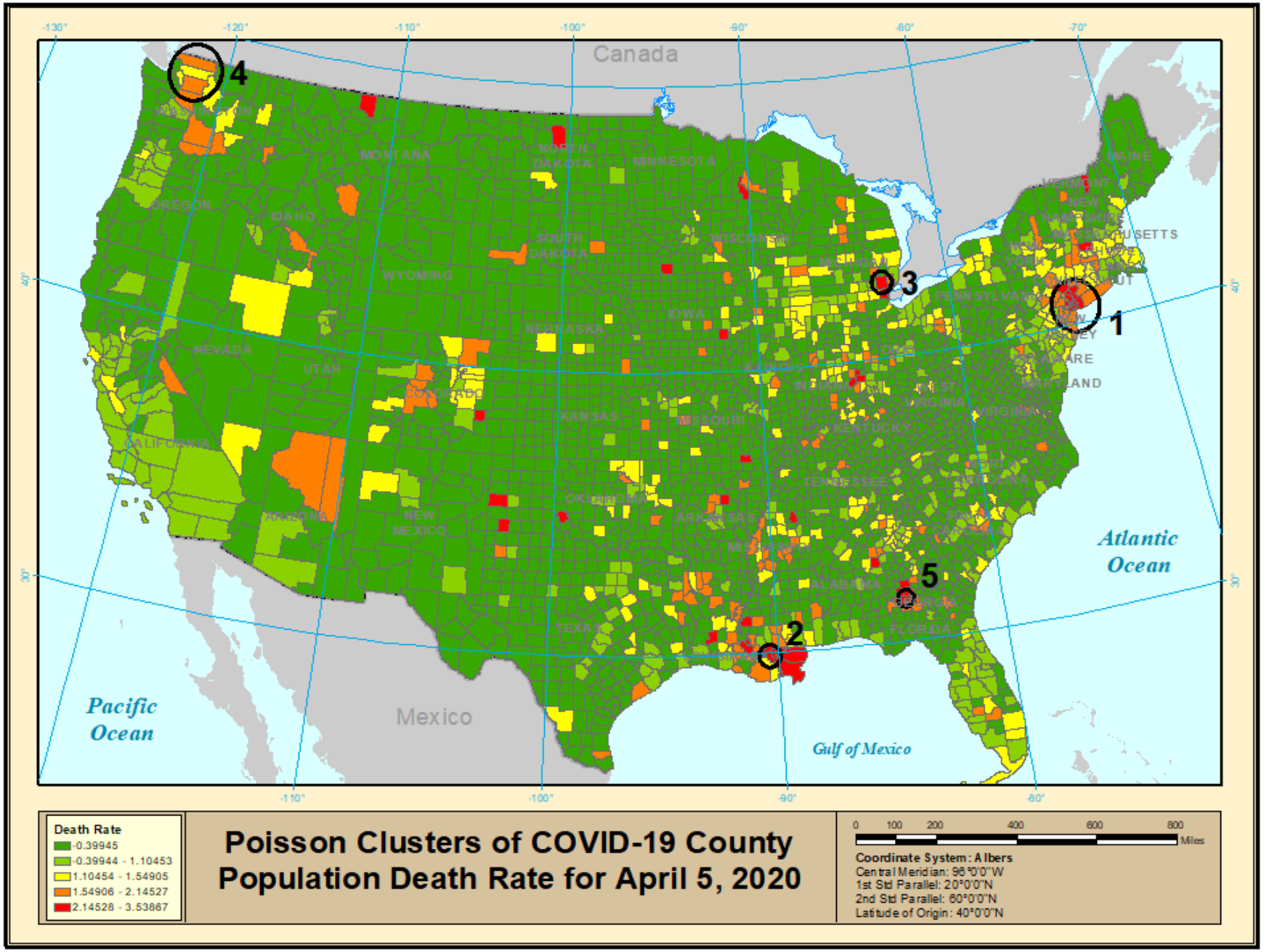




Spatial Scan Statistic: A small sample of the circles used

Four Most Likely Clusters





Raid Amin et al, Univ West Florida

Space-Time Scan Statistic

- Use a cylindrical window, with the circular base representing space and the height representing time.
- Retrospective surveillance: One analysis on a static data set, to detect historical and/or current clusters.
- Prospective surveillance: Daily or weekly analyses as new data accrues, to quickly detect emerging clusters. Only consider cylinders that reach the present time.

For each cylinder:

– Obtain actual and expected number of cases inside and outside the cylinder.

– Calculate likelihood function.

Compare Cylinders:

– Pick cylinder with highest likelihood function as Most Likely Cluster.

Inference:

– Generate random replicas of the data set under the null-hypothesis of no clusters. May e.g. use 9,999 random replicas.

– Compare most likely clusters in real and random data sets (Likelihood ratio test). If R=rank of the real data set, then p=R/(1+9999)

Space-Time Scan Statistic: Properties

- Simultaneously tests for outbreaks of any size at any location, by using a cylindrical windows with variable radius and height.
- Accounts for multiple testing.
- Aggregated or non-aggregated data, such as counties, zip-codes, census tracts, hospitals or residences.

Probability Models

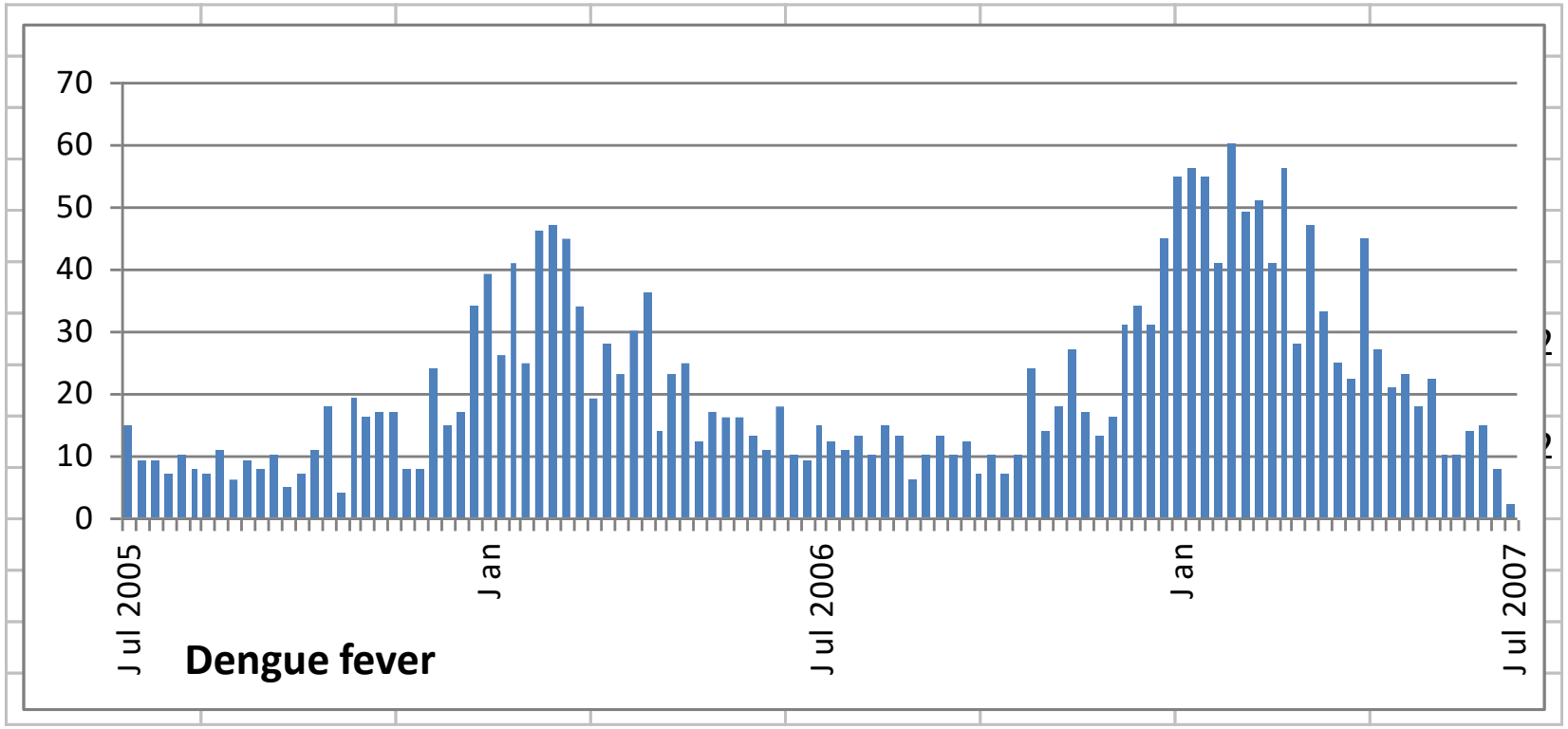
- **Bernoulli Model** (binomial): For 0/1 count data, such as tested positive vs negative.
- **Poisson Model**: For count data with expected counts under the null, such as incidence and age-adjusted population based expected counts.
- **Space-Time Permutation Model:** For count data without a denominator, automatically adjusting for purely spatial variation and purely temporal trends
- ***Normal Model:*** *For continuous data, such as birth weights*
- ***Exponential Model****: For time-to-event data, such as cancer survival*
- ***Ordinal Model:*** *For ordered categorical data with 3+ categories, such as cancer stage*
- ***Multinomial Model:*** *For un-ordered categorical data with 3+ categories, such as salmonella serotypes*

Adjusting for Covariates

- Done in different ways for different probability models
- May or may not be appropriate or important
- When looking for unknown risk factors, adjust for age, smoking, etc, to find clusters not explained by those variables
- When allocating public health resources, do not adjust for age, as interest is in finding areas with more cases
- When implementing public health education, do not adjust for e.g. smoking, in order to find areas with a high-risk population

Adjusting for Purely Temporal Trends

- Adjust if there is no interest in detecting region-wide temporal changes
- Non-parametric adjustments
- Log-linear adjustments, for a specific percent increase or decrease



- Log-linear adjustments may be calculated from the data, or provided by user from external information

Adjusting for Purely Geographical Variation

- To adjust for background population, when census data is lacking
- To adjust for inherent differences in data collection, such as geographical variation in screening or testing
- When there is only an interest in emerging problem, rather than long-term geographical differences

- Adjustment is normally done non-parametrically
- Also possible to adjust for known geographical differences in the relative risks.

Adjusting for both Geography and Time

- Possible to adjust the Poisson model for geography non-parametrically and time with a log-linear trend.
- Adjusting for both space and time non-parametrically, requires use of the space-time permutation model
- Missing data can be adjusted for by setting RR=0 for the relevant locations and time periods. Should never be used when the true counts are zero. Only when the count is unknown.

SaTScan Software

- SaTScan is a software that analyzes spatial, temporal and space-time data using the spatial, temporal, or space-time scan statistics.
- Free
- Download: www.satscan.org
- Thousands of users, including many at CDC and at least 44 of the 50 state health departments
- Funded by NCI, NICHD, NIGMS, CDC, Sloan Foundation
- Upcoming COVID-19 improvements funded by Alfred P. Sloan Foundation, CDC Foundation, ELC CARES, and Open Society Foundations
  - With support of Fund for Public Health in New York City

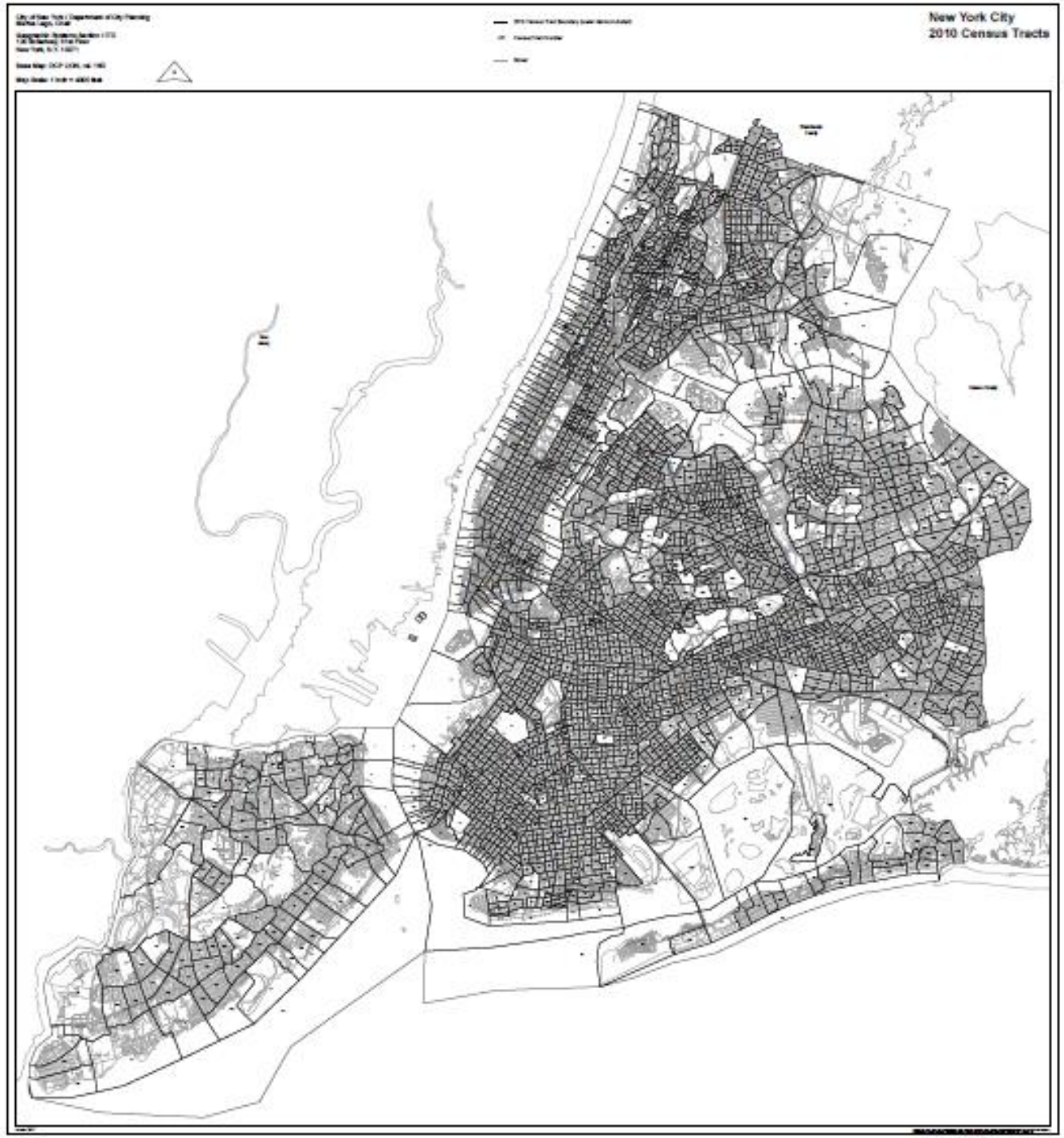
Input files

- Numerator: Confirmed COVID-19 cases
- Denominator: Unique persons in electronic laboratory reports with PCR test for SARS-CoV-2
- Study period:
  - To prioritize case investigations daily: 21 days
  - To allocate testing and community engagement resources: since June 1, or shorter if discontinuity in citywide percent positivity trend
- Temporal element: specimen collection date
- Lag for data accrual: 3 days

Input files

- Exclusions, to focus on recent community transmission:
  - Residents of long-term care facilities, correctional facilities, facilities housing people with developmental disabilities, homeless shelters
  - Home address matches hospital or laboratory
  - Diagnosed in 14 days prior to a more recent case residing in same building
  - Illness onset (where available from patient interview) prior to first date of study period

- Spatial resolution: Census tract 2010 (N=2165)
  - Automated geocoding process


New York City
2010 Census Tracts

Analyses

- Detect disease clusters in space–time cylinders centered on every census tract centroid
- Circular base: represents space
  - Maximum geographic cluster size: 50% of persons tested
- Height: represents time
  - To prioritize case investigations daily:
    - Min–max temporal window: 3–7 days
    - Study period 21 days
  - To allocate testing and community engagement resources:
    - Min–max temporal window: 7–21 days
    - Study period: since June 1, or shorter if discontinuity in citywide percent positivity trend
- Minimum number of cases: 5

Adjust for purely spatial variation

- To detect areas with relative increases from baseline, even if still lower than the citywide average
- Will not detect areas with consistently high percent positivity
- Spatial adjustment: nonparametric, with spatial stratified randomization

Adjust for purely temporal trends

- If test % positivity is decreasing in NYC overall, detect areas where the decrease is less than citywide average

- If test % positivity is increasing in NYC overall, detect areas where the increase is more than citywide average

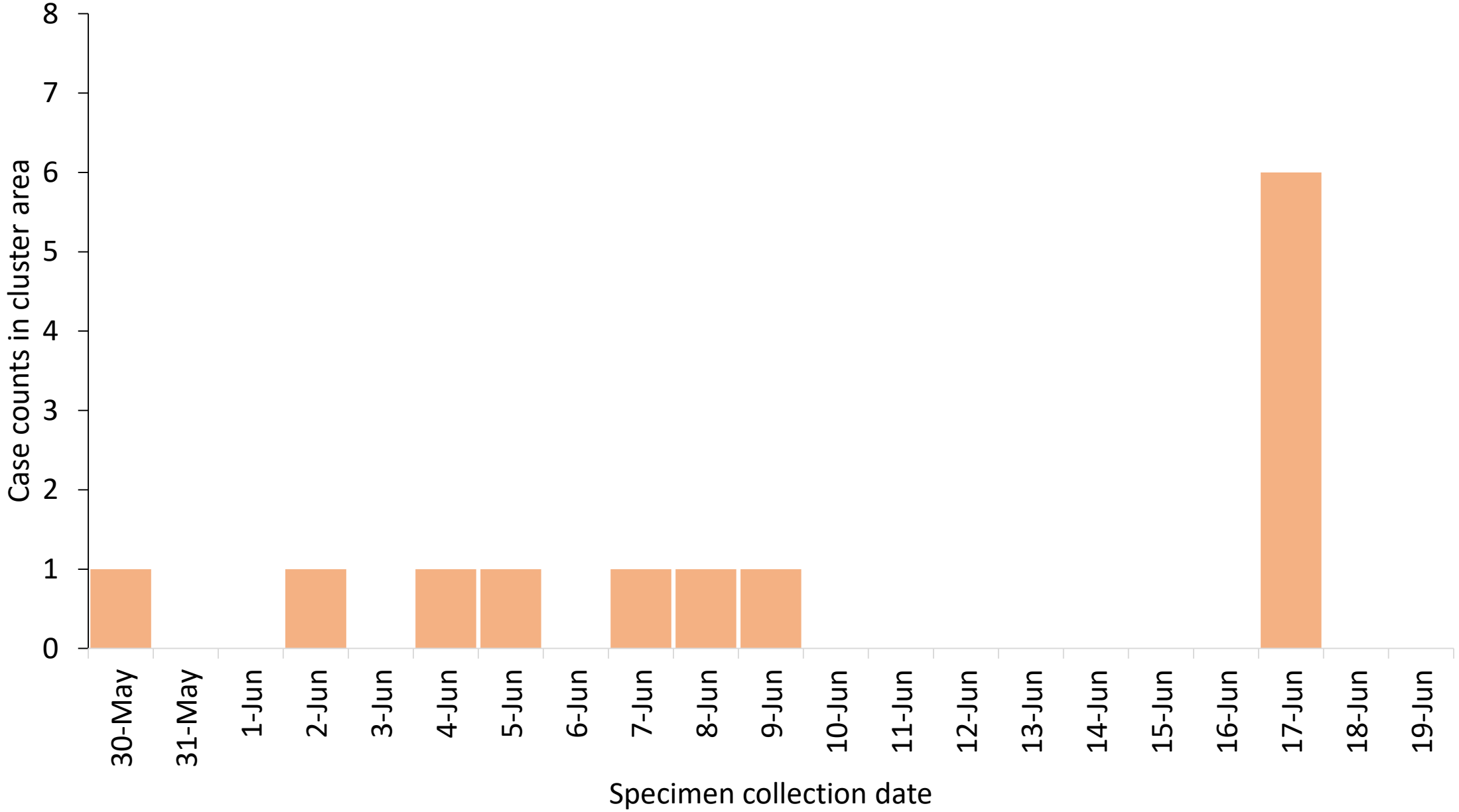
- Temporal adjustment: log-linear with automatically* calculated trend
  - *Specify input files in years, not days
  - SaTScan v9.6 estimates temporal trends at annual time scale
  - Pending SaTScan enhancement to accommodate rapidly changing trend

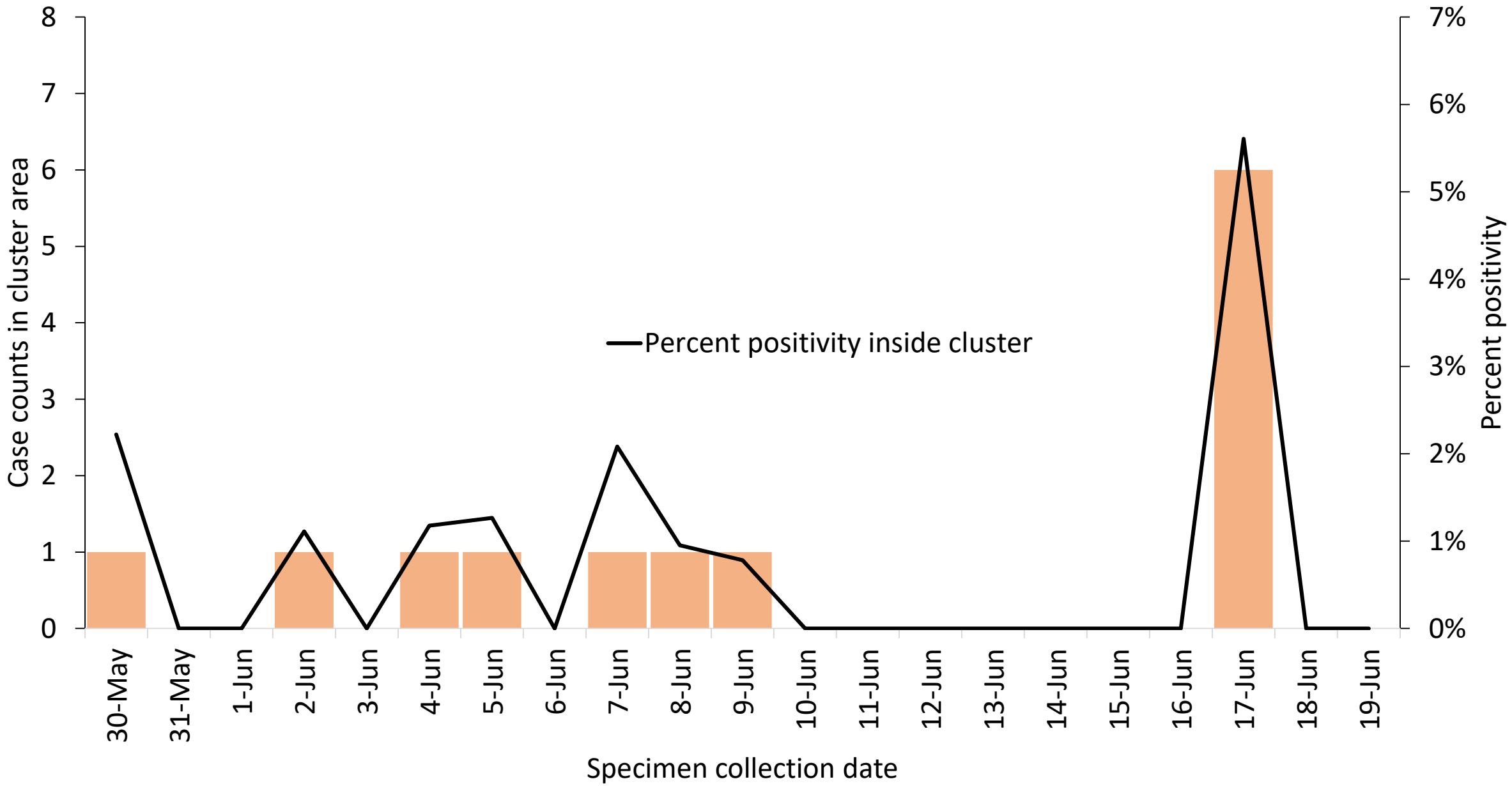
Percent positivity spatiotemporal cluster detected June 22, 2020

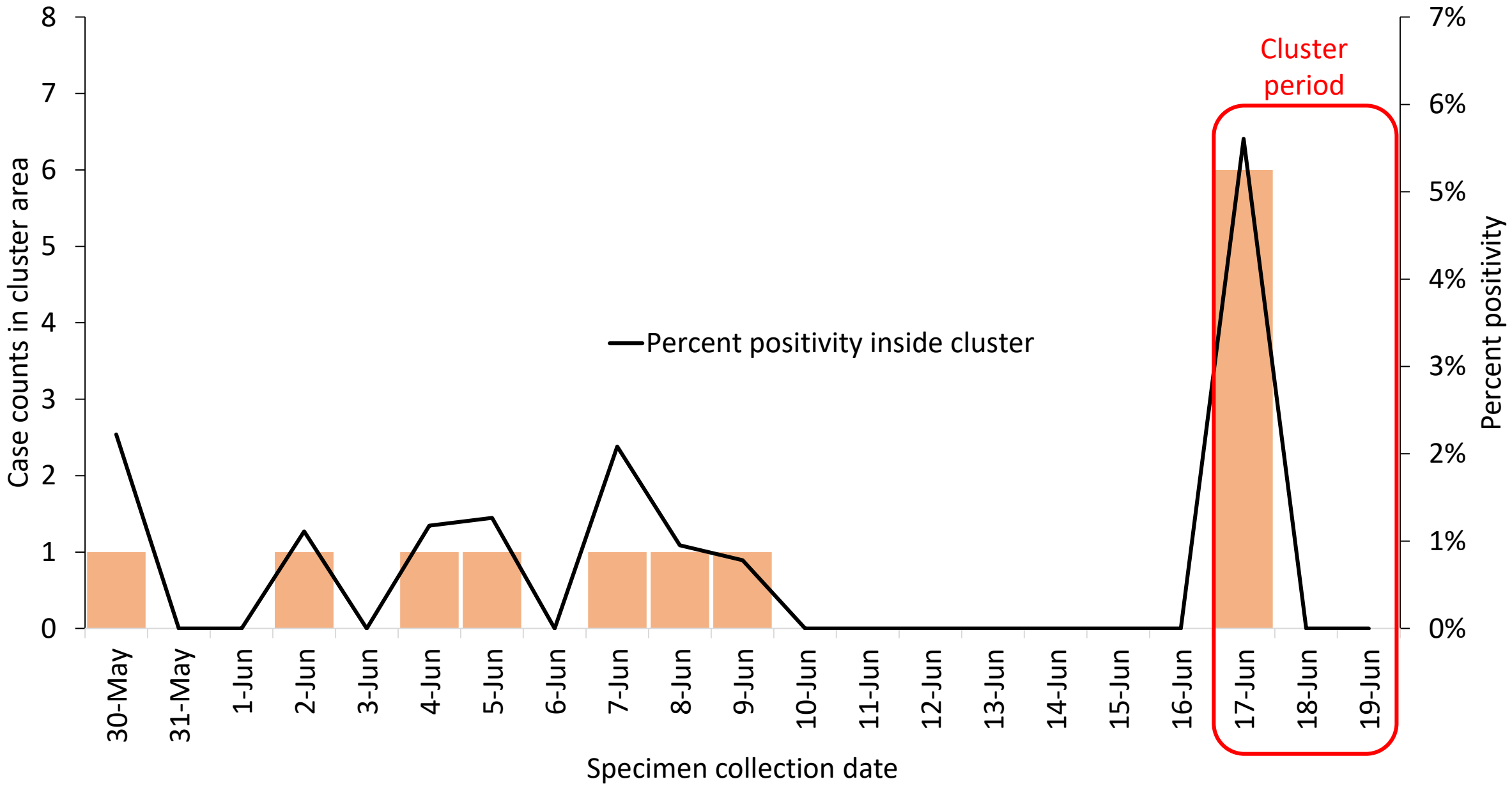
- During June 17–19:

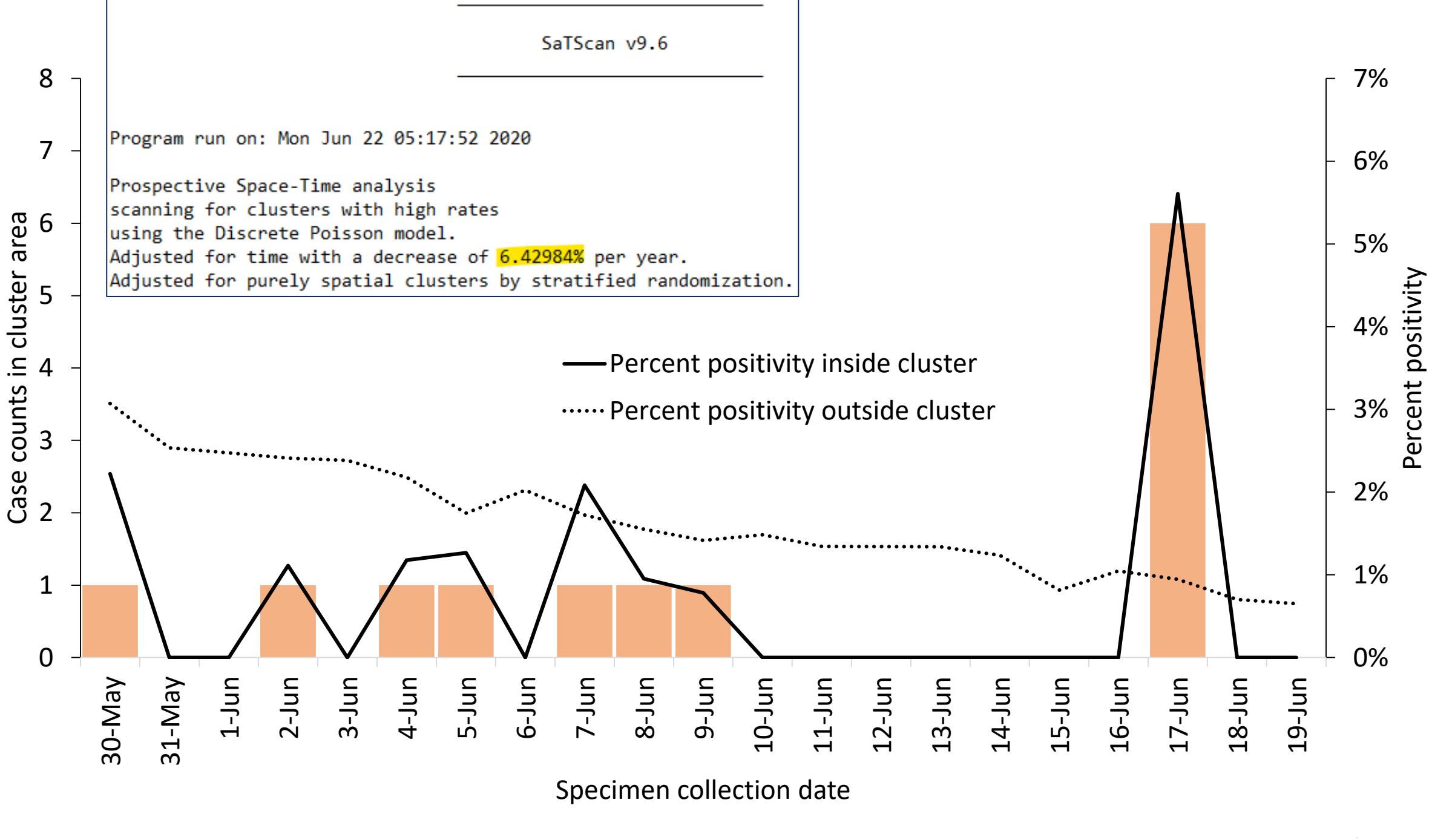
Radius (km)	Observed cases	Relative risk	Recurrence interval (days)	% positivity inside cluster	% positivity outside cluster
0.6	6	4.0	365	2.2	0.8

- Prompted investigation identifying a gathering with inadequate social distancing where viral transmission likely occurred
- Response included testing, contact tracing, community engagement, and health education


Case counts in cluster area
0
1
2
3
4
5
6
7
8
30-May
31-May
1-Jun
2-Jun
3-Jun
4-Jun
5-Jun
6-Jun
7-Jun
8-Jun
9-Jun
10-Jun
11-Jun
12-Jun
13-Jun
14-Jun
15-Jun
16-Jun
17-Jun
18-Jun
19-Jun
Specimen collection date


Case counts in cluster area
Percent positivity
Percent positivity inside cluster
Specimen collection date
0
1
2
3
4
5
6
7
8
0%
1%
2%
3%
4%
5%
6%
7%
30-May
31-May
1-Jun
2-Jun
3-Jun
4-Jun
5-Jun
6-Jun
7-Jun
8-Jun
9-Jun
10-Jun
11-Jun
12-Jun
13-Jun
14-Jun
15-Jun
16-Jun
17-Jun
18-Jun
19-Jun


Case counts in cluster area
Percent positivity
Percent positivity inside cluster
Cluster period
Specimen collection date
0
1
2
3
4
5
6
7
8
0%
1%
2%
3%
4%
5%
6%
7%
30-May
31-May
1-Jun
2-Jun
3-Jun
4-Jun
5-Jun
6-Jun
7-Jun
8-Jun
9-Jun
10-Jun
11-Jun
12-Jun
13-Jun
14-Jun
15-Jun
16-Jun
17-Jun
18-Jun
19-Jun


SaTScan v9.6
Program run on: Mon Jun 22 05:17:52 2020
Prospective Space-Time analysis
scanning for clusters with high rates
using the Discrete Poisson model.
Adjusted for time with a decrease of 6.42984% per year.
Adjusted for purely spatial clusters by stratified randomization.
Case counts in cluster area
Percent positivity
Percent positivity inside cluster
Percent positivity outside cluster
Specimen collection date
0
1
2
3
4
5
6
7
8
0%
1%
2%
3%
4%
5%
6%
7%
30-May
31-May
1-Jun
2-Jun
3-Jun
4-Jun
5-Jun
6-Jun
7-Jun
8-Jun
9-Jun
10-Jun
11-Jun
12-Jun
13-Jun
14-Jun
15-Jun
16-Jun
17-Jun
18-Jun
19-Jun

Cluster Interpretation

- Prioritize clusters with:
  - High percent positivity
  - High relative risk
  - High recurrence interval
    - Areas with many persons tested can have highly significant clusters of only minor increases in percent positivity
- If searching for explicit epi-links:
  - Consider secondary clusters with their overlapping primary cluster
  - Track evolving clusters

Automation

- Generate input files
- Invoke SaTScan in batch mode
- Output files to secured folders
- Define signals (e.g., clusters with recurrence interval ≥100 days), and generate patient linelist, visualizations, and investigator notification email

- NYC DOHMH, using SAS:
  https://github.com/CityOfNewYork/communicable-disease-surveillance-nycdohmh

- Cook County Dept of Public Health, using R:
  https://github.com/kb230557/rsatscan-cluster-detection-covid

Planned SaTScan software enhancements for COVID-19

- Prioritize short emerging clusters
- Accommodate irregular geographies
- Detect clusters along a distance matrix/network
- Support weekly prospective analysis using data with daily resolution
- Temporal and spatial adjustments for Bernoulli model
- Descriptive and analytic outbreak investigation tools
- Software interface for prospective surveillance
- Automated outbreak alert notices