



Alternative Cluster Detection Methods and Communications

Whole Genome Sequencing
Live Learning Series - Session 3

Friday, June 12th

Introduction



Kelly Orejuela, MPH

- Epidemiologist/Epi-Lab Liaison
- Foodborne and Enteric Diseases Program
- Tennessee Department of Health

Kelly.Orejuela@tn.gov

Learning Objectives

Alternative Cluster Detection Tools	SaTScan and other scan statistics tools are used to detect and monitor emerging disease clusters
	Identify situations where non-WGS surveillance tools are useful
Communicating WGS Results	Use and interpret standard phrasing for communicating WGS results
	Summarize complex laboratory testing and genetic comparisons during outbreak investigations for a general audience

Scan Statistics to Support Enteric Disease Surveillance

Sharon K. Greene, PhD, MPH
Epidemiology and Analysis Team Lead
Bureau of Communicable Disease
New York City Department of Health and Mental Hygiene

June 12, 2026
Tennessee Food Safety CoE WGS for Cluster Detection and
Investigation Live Learning Series

Space-time scan statistic

- A signal detection/data-mining method for space-time data (e.g., reportable diseases)



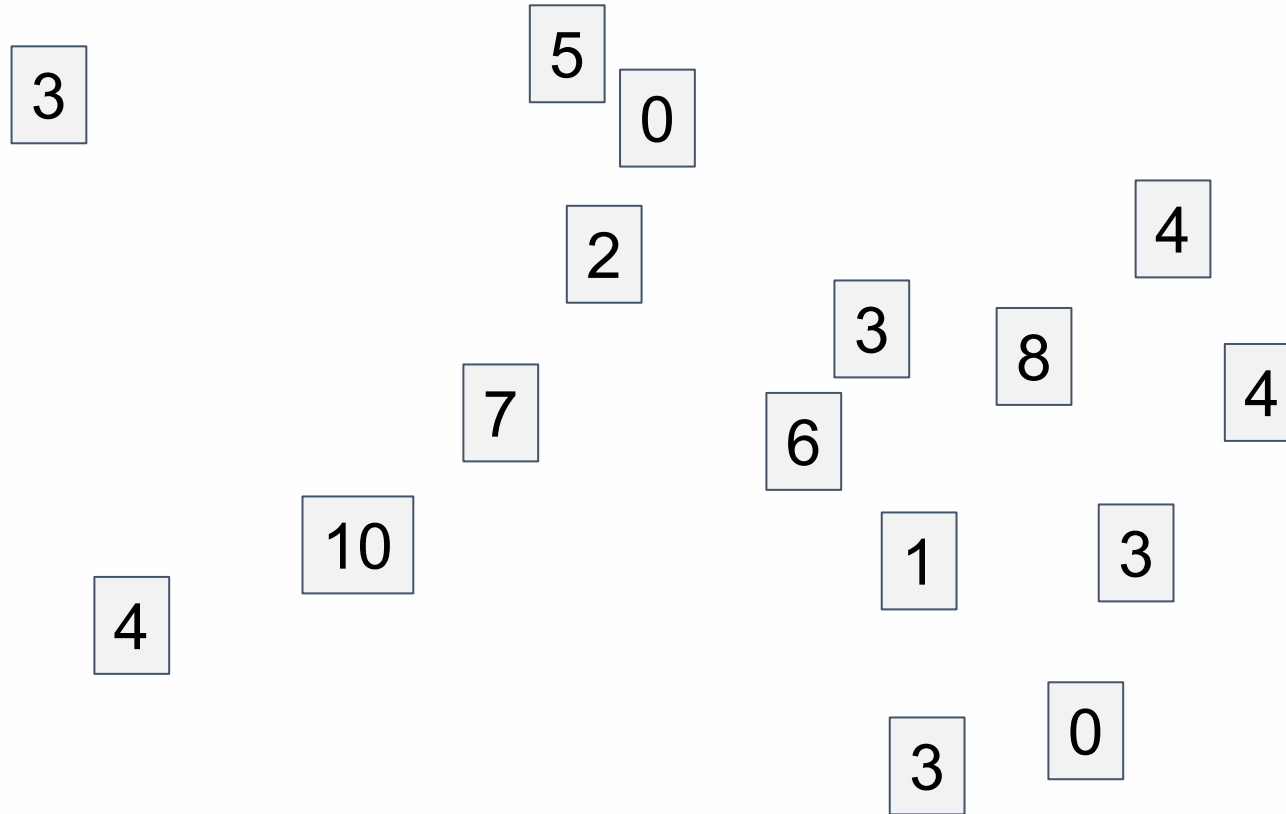
- SaTScan™ software

- Free download: www.satscan.org
- Open-source code: github.com/scanstatistics/satscan
- We worked with SaTScan software developer to build in new features for scanning along a network, automating analyses, and tools to support cluster interpretation and investigation

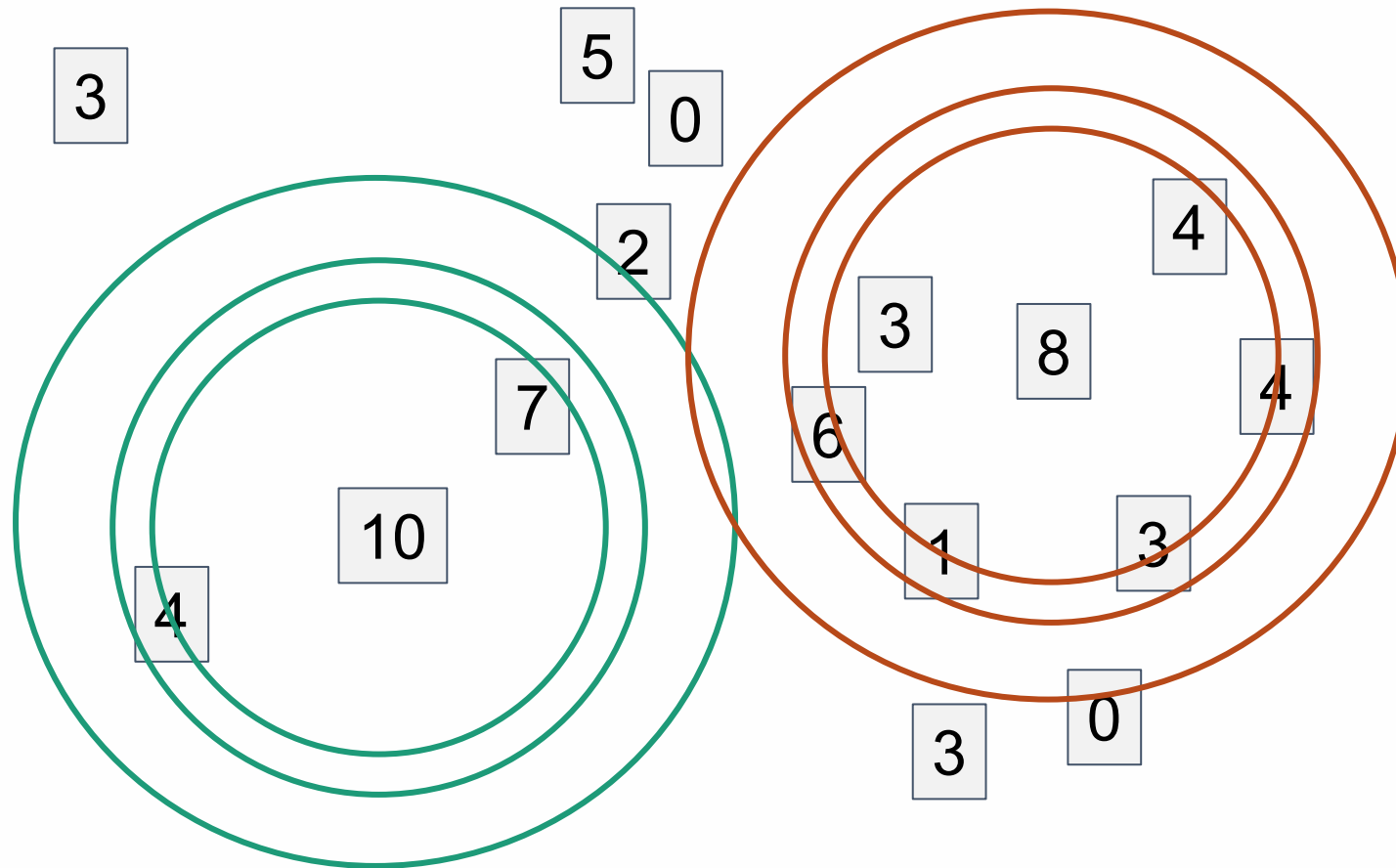
New!

- Video tutorials: www.satscan.org/videotutorials.html

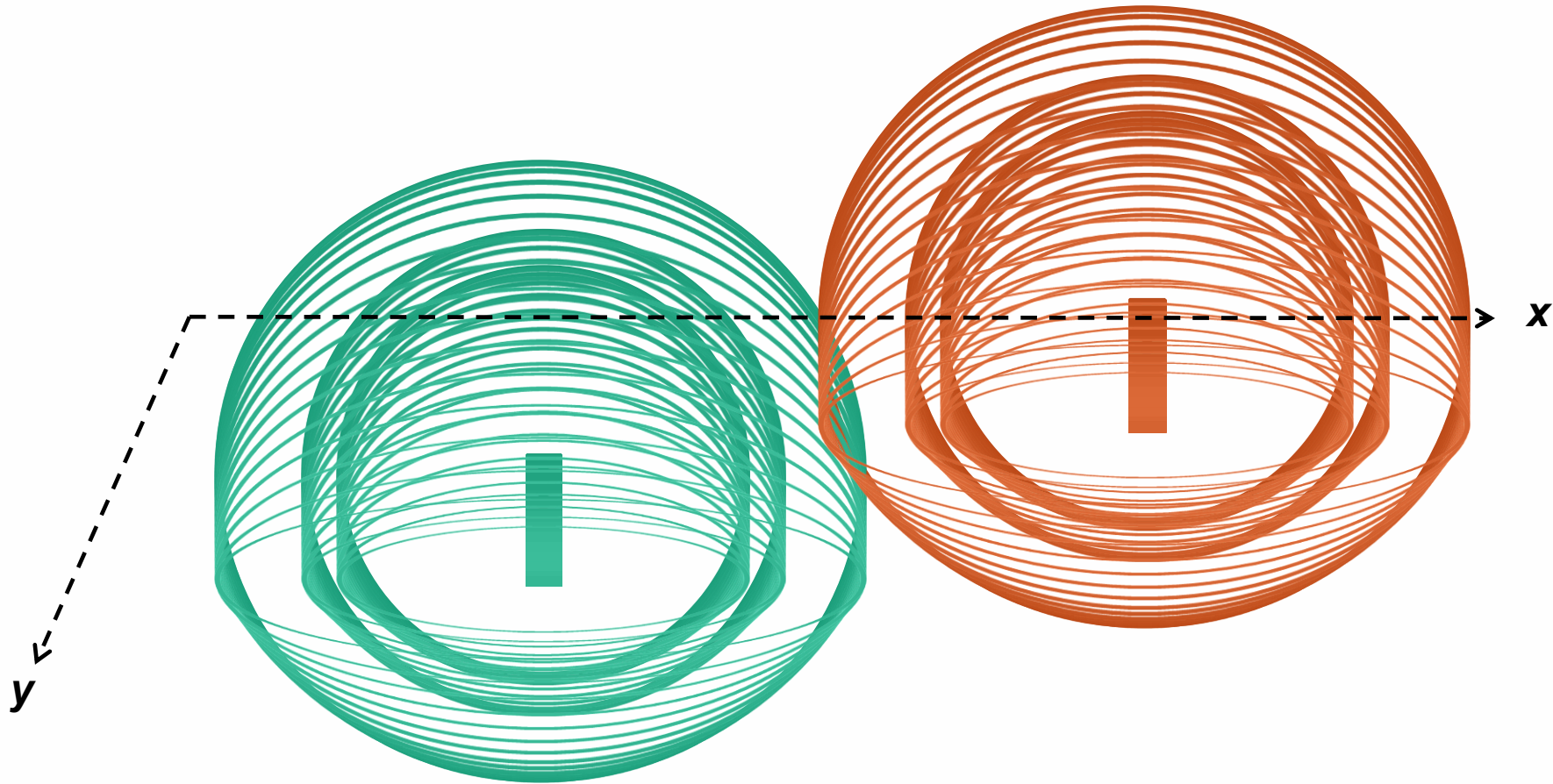
Scan Statistic



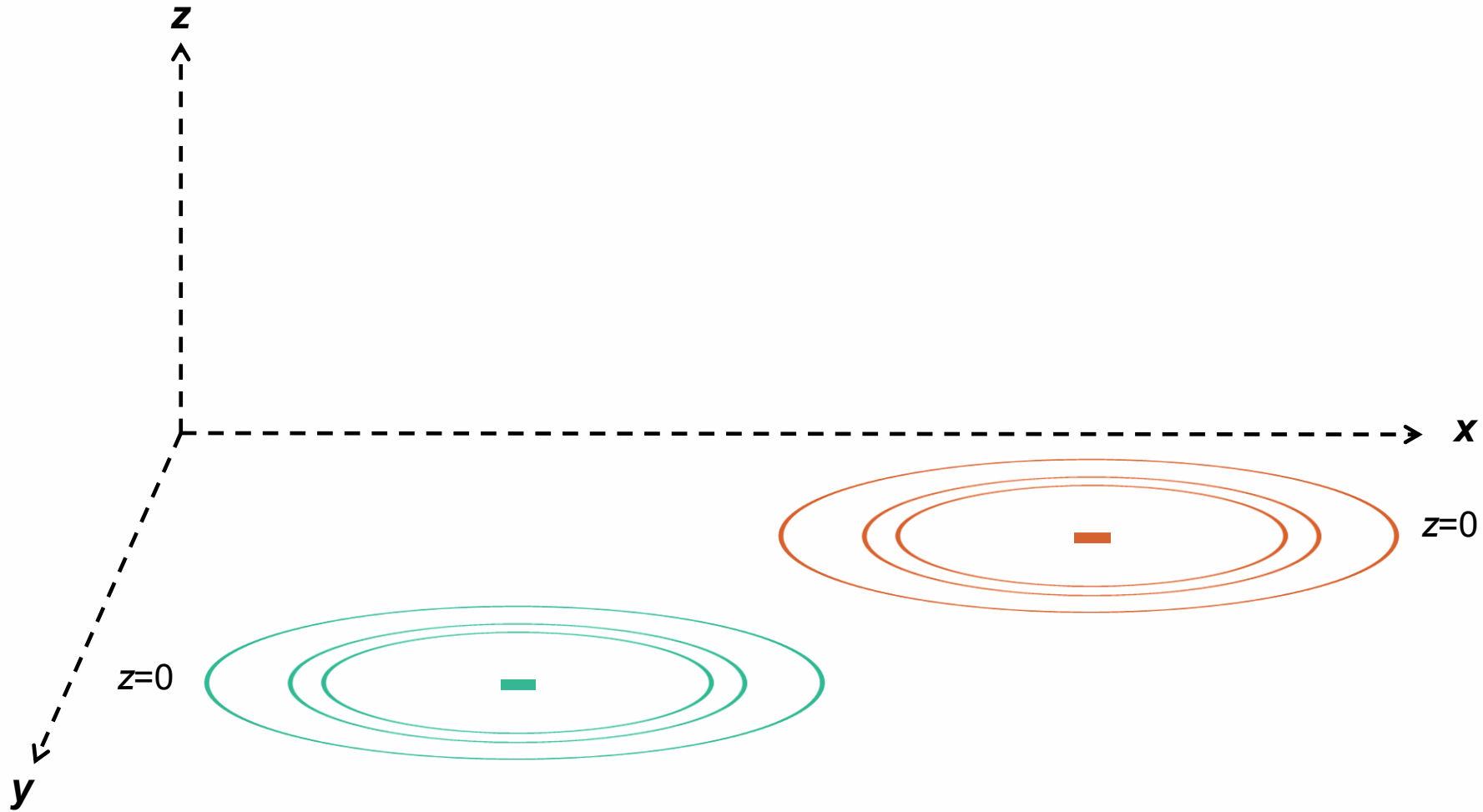
Scan Statistic



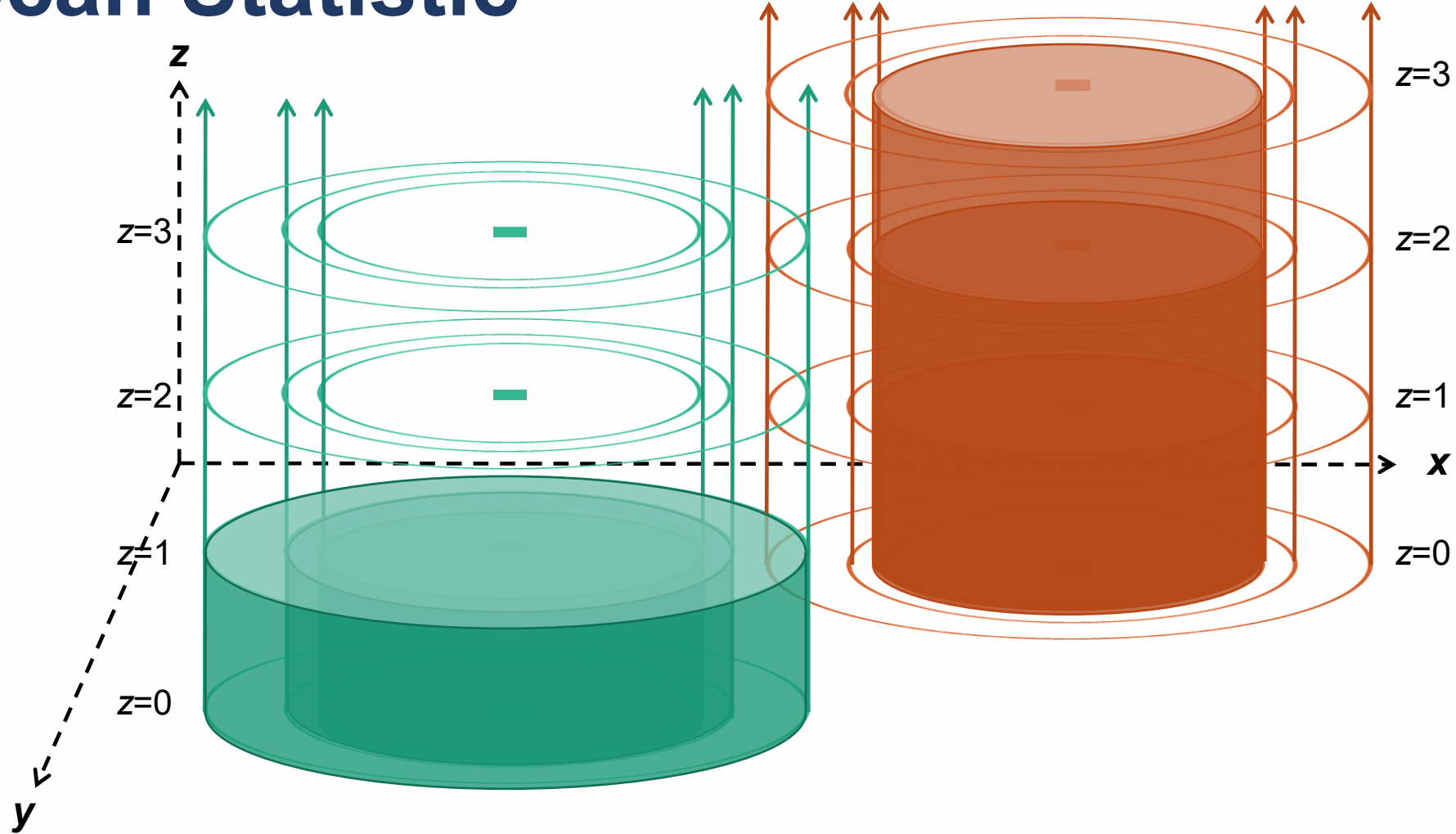
Scan Statistic



Scan Statistic



Scan Statistic

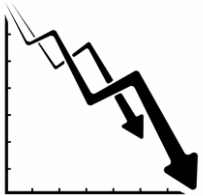




Detect spatiotemporal clusters in reportable diseases



Detect emerging clusters in genomic surveillance data



Detect drop-offs in electronic laboratory reporting



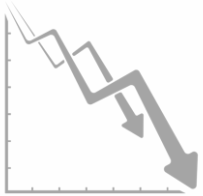
Detect increases in ICD-10 diagnosis codes for ED visits



Detect spatiotemporal clusters in reportable diseases



Detect emerging clusters in genomic surveillance data



Detect drop-offs in electronic laboratory reporting



Detect increases in ICD-10 diagnosis codes for ED visits

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

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

Julia Latash, MPH¹*; Sharon K. Greene, PhD¹*; Faina Stavinsky, MS²; Sandy Li³; Jennifer A. McConnell, MS³; John Novak, PhD³; Teresa Rozza³; Jing Wu, PhD³; Enoma Omoregie, PhD³; Lan Li, MPH¹; Eric R. Peterson, MPH¹; Bruce Gutelius, MD¹; Vasudha Reddy, MPH¹ ([View author](#))

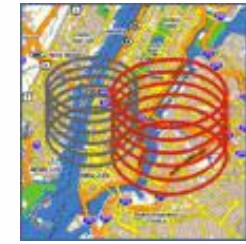
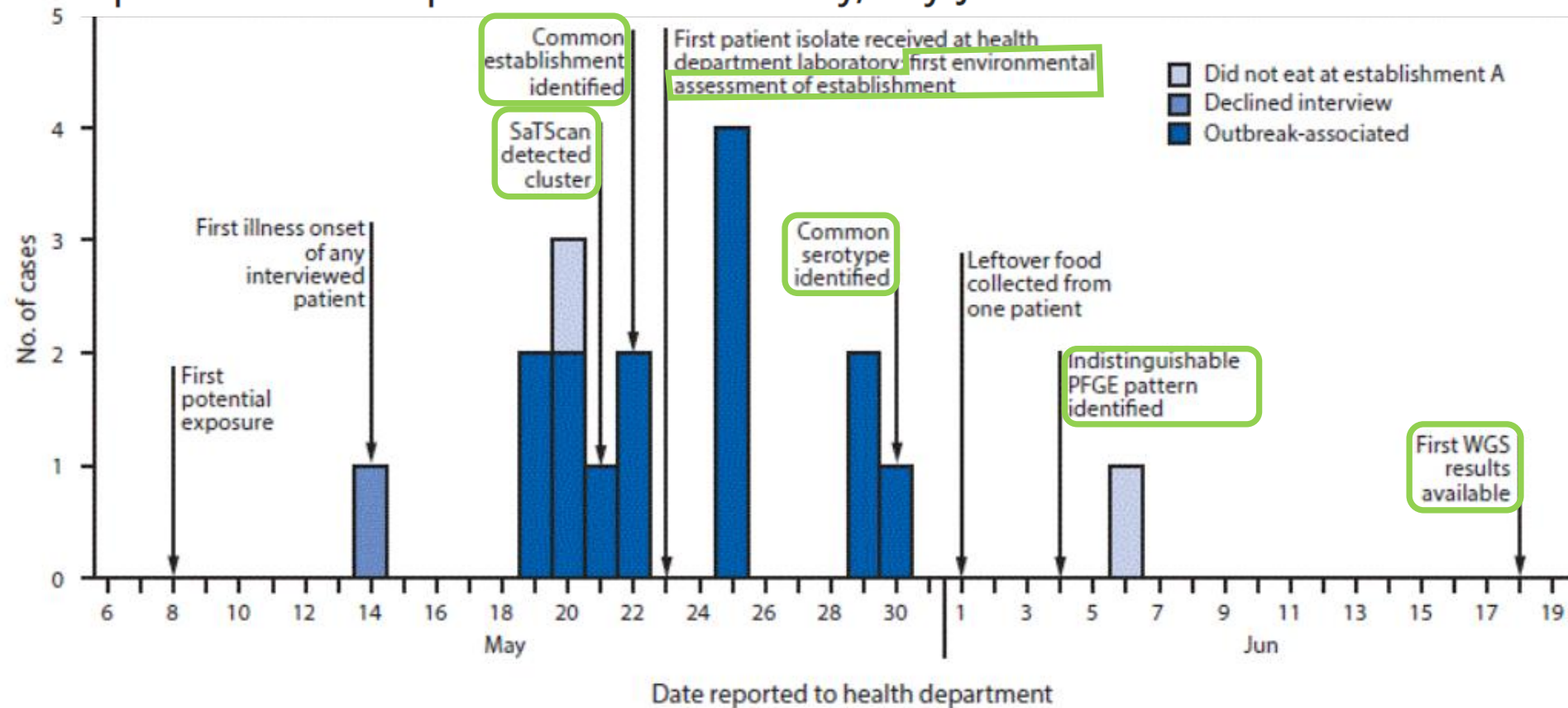


FIGURE. Cases of salmonellosis (N = 17) included in a SaTScan* spatiotemporal cluster, by date reported to health department — New York City, May–June 2019



DONATE

THE CITY

WEDNESDAY, OCTOBER 11, 2023 | REPORTING FOR NEW YORKERS



[ABOUT US](#) ▾ [BOROUGHs](#) ▾ [TOPICS](#) ▾ [CORONAVIRUS TRACKER](#) [FAQNYC](#) [NEWSLETTERS](#) ▾ [HELP US REPORT](#) ▾

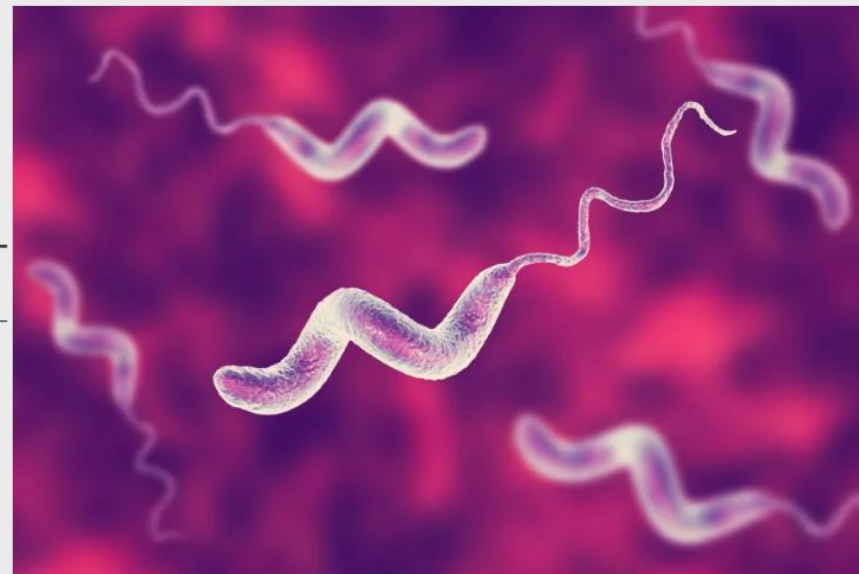


HEALTH

Health Department Probing Possible Foodborne Illness Outbreak in Brooklyn

Campylobacter is a gastrointestinal bug that can come from eating raw or undercooked poultry. About 50 cases have been reported so far to city health authorities.

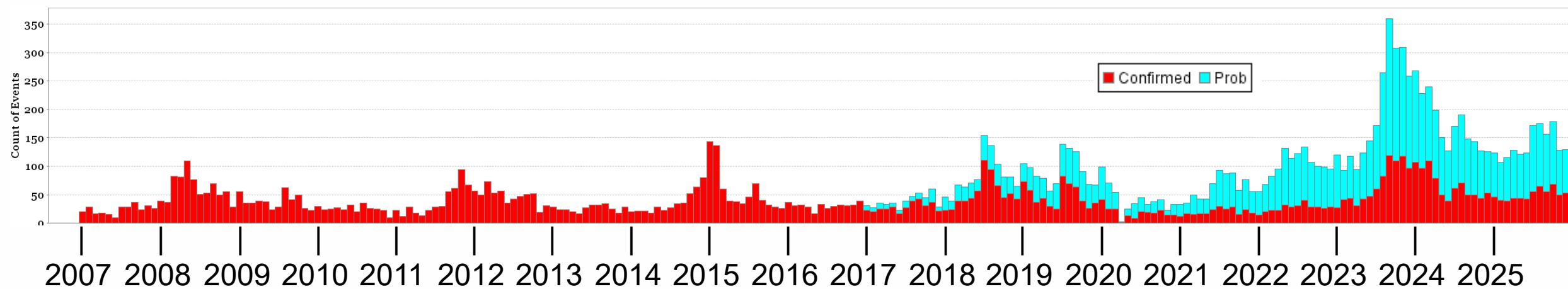
BY REUVEN BLAU | RBLAU@THECITY.NYC | OCT 26, 2022, 12:38PM EDT



The Campylobacter can be found in undercooked poultry. | Illustration/Kateryna Kon/Shutterstock

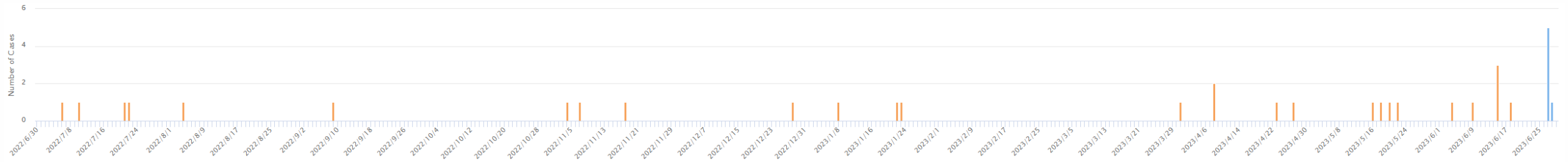
- Campylobacteriosis is not routinely investigated
- SaTScan facilitated both outbreak detection and daily outbreak monitoring
 - Supported investigators to select which cases were considered potentially outbreak-related and assigned for investigation

Monthly shigellosis cases among NYC residents, 2007–2025

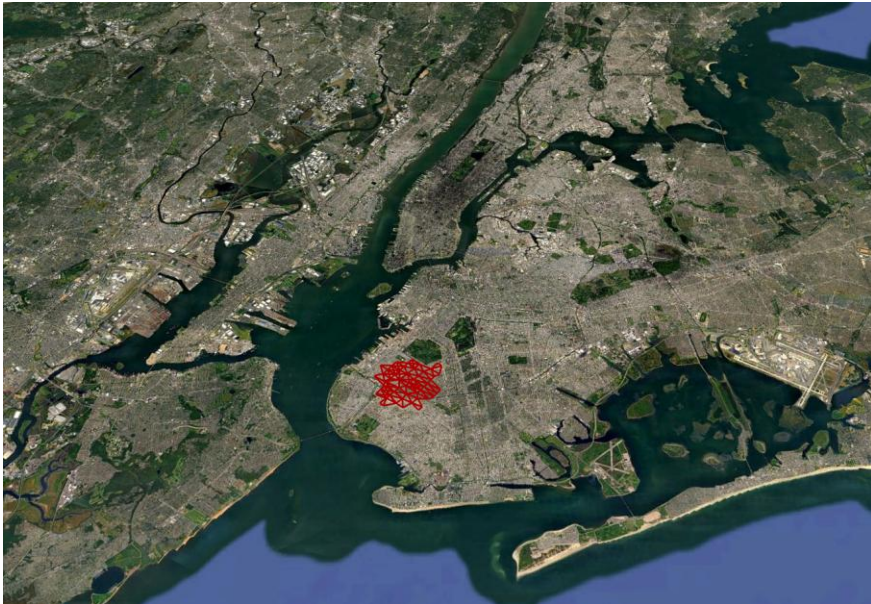


Unpublished data as of 2/13/26, NYC Health Department. Not for further distribution.

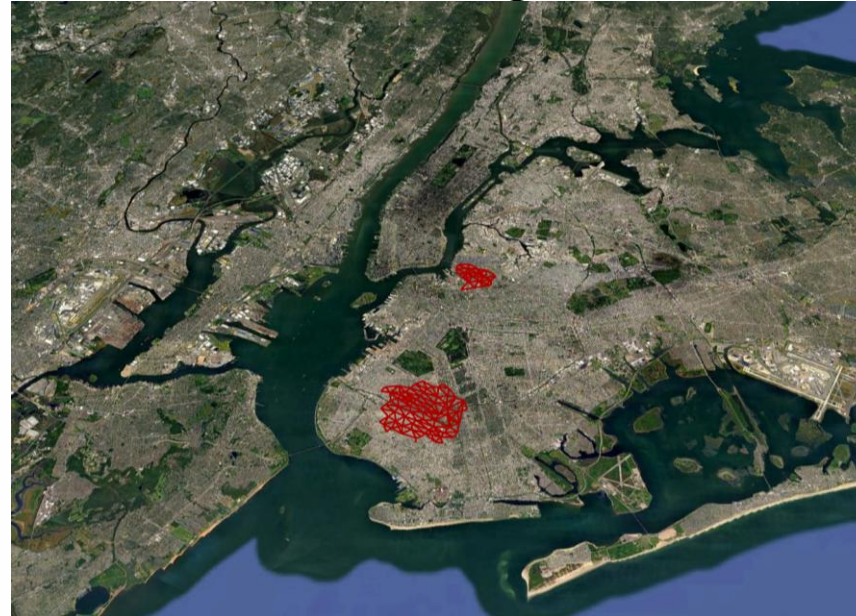
Monitoring Recurrent Shigellosis Outbreaks



Initial cluster detection, 6/30/23

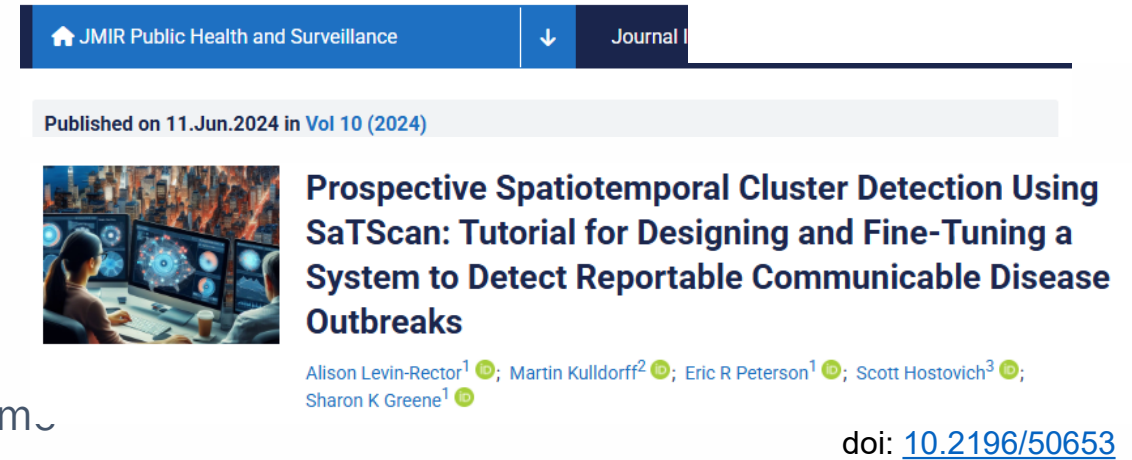


Cluster monitoring, 10/15/23



Implementation details are complex, nuanced

- Data quality assurance for input data
- Parameter settings
- System finetuning
 - When difficulty detecting:
 - geographically small outbreaks or outbreaks at the boundaries of geographic units
 - outbreaks affecting a particular age group
 - outbreaks where people are exposed far from their homes
 - outbreaks because of missing data
 - new outbreaks in areas with prior outbreaks
 - outbreaks affecting people residing in areas not connected in the network file
 - For consistently finding large, uninteresting clusters
 - For too many signals, where clusters are driven by:
 - duplicate events or incorrect patient addresses
 - changing nature of input data, such as recent adoption of culture-independent diagnostic testing
 - reports that do not represent true illnesses, such as reports of patients with only negative laboratory test results
 - within-household transmission so do not represent community transmission
- Not covered by any existing academic curriculum or workforce training program





SaTScan for Cluster Detection of Enteric Pathogens

How do we use SaTScan in Tennessee?



TN uses SaTScan to find potential clusters of *Salmonella* and STEC



FED epidemiologists run a SAS code to pull the data from NBS (NEDSS Based System)



Spreadsheet and other required data are set up in SaTScan prior to first run



SaTScan analyzes the input data and produces various file types, including a map, to show detected clusters

How do we use SaTScan?

Three epidemiologists run SaTScan and switch every week

Clusters of interest/statistically significant are investigated

- Use CRF data
- Info is documented in an Excel spreadsheet and saved to folder
- Updates are shared during weekly Cluster Meeting
- If a potential outbreak is detected, info gets escalated to Outbreak Team the same day

Examples of Detected Clusters: STEC

STEC was the first pathogen to be analyzed with SaTScan for cluster detection

- Investigated 3 clusters in 2024
 - Range of 3 – 5 cases in each cluster (July – September)
- Investigated 5 clusters in 2025
 - Range of 3 -5 cases in each cluster (February – November)

Examples of Detected Clusters: *Salmonella*

- *Salmonella* was implemented in 2025
 - Investigated 6 clusters in 2025
 - Range of 5-14 cases in each cluster (June – September)
 - 1 outbreak (2508KYDT)

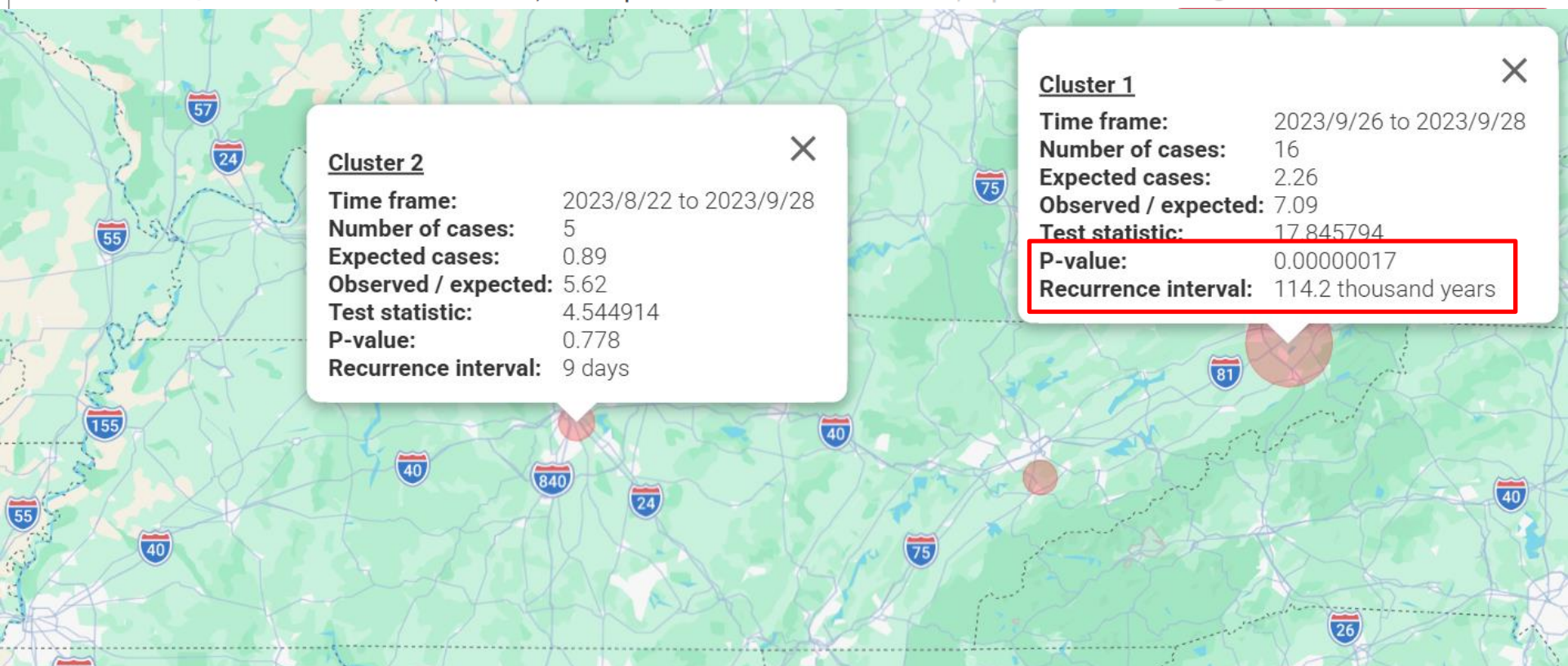
Cluster	Hierarchy	New Individ	Onset	Collection Date	Case Status	Hospitalized	Notes
1	Primary	New	Unk LTFU	8/12/2025	Probable	Unk	Not in LIMS
1	Primary	New	8/3/2025	8/13/2025	Probable	Yes	Case N Sam's Club; 2% milk, strawberries; chicken from Resort A
1	Primary	New	Unk LTFU	8/13/2025	Probable	Unk	Lab Not in LIMS
1	Primary	New	8/10/2025	8/10/2025	Probable	No	Case K Walmart or Kroger (chicken, chicken fries, popcorn chicken, eggs, American cheese, watermelon, apples, bananas, cucumbers, Jif peanut butter, frozen pizza); Butchered beef and sausage from unknown location, venison processed at home, eggs from grandmother's chickens; Resort A ; salad-lettuce, croutons, dressing, carrots, red cabbage, tomatoes, cheese, egg
1	Primary	New	7/31/2025	8/18/2025	Probable	No	Case N Walmart (chicken breasts Perdue, ground beef, 2% milk, Chobani flip) Chicken fingers at waterpark and restaurant R
			8/4/2025	8/9/2025	Confirmed	No	Case N Quick Stop in Paris, TN (Krispy Krunchy chicken legs); Resort A : (catfish, chicken strips, cole slaw, white beans) Case N Kroger (chicken, Lan-O-Lakes turkey deli or Kroger turkey deli, beef and burgers, ham, <i>Starkist Tuna</i> , eggs, American cheese, feta cheese, cheddar, mixed greens salad, arugula)

Examples of SaTScan Cluster Detection Viz



Cluster #1

● Observed in cluster area (red=cluster) — Expected in cluster area — Observed / expected in cluster area ● Observed outside cluster area



Pros and Limitations and Considerations



Pros

- Quick and easy to run once set up
- Identify clusters/ outbreaks faster vs. WGS
- Captures CIDT-only cases
- Helps prioritize case interviews and investigations

Limitations and Considerations



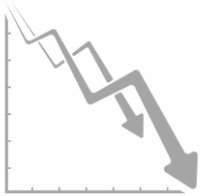
- Not intuitive at first
- Results depend on parameter settings
- Data quality affects detection
- Large clusters can be time-consuming to review



Detect spatiotemporal clusters in reportable diseases



Detect emerging clusters in genomic surveillance data



Detect drop-offs in electronic laboratory reporting



Detect increases in ICD-10 diagnosis codes for ED visits

Detecting salmonellosis clusters



GUIDELINES FOR FOODBORNE DISEASE OUTBREAK RESPONSE



Foodborne Illness Surveillance and Outbreak Detection



Several “rules of thumb” based on the number of allele differences have been developed to help define a cluster by WGS.

- These rules vary by pathogen and mode of transmission.
- Generally, PulseNet uses a definition of at least **3** cases within a **60-day** window with **0–10** allele differences, where at least **2** of the cases differ by **5** or fewer alleles, for *Salmonella* and STEC.
 - PHLs may consider a narrower definition (such as **0–5** alleles) to reduce the number of clusters that need to be investigated and to focus investigation resources.
- Similar to PFGE, there can be **common** sequence types or **rare** sequence types, which should be considered during cluster investigations.
- In addition, if the outbreak occurs over a long period or is **zoonotic**, **more** allele differences are detected than in an outbreak representing a **point source** contamination event.
- When an outbreak source is contaminated with multiple diverse sequence types, known as a polyclonal outbreak, sequence data may be used to identify multiple independent clusters, which can then be used to identify the polyclonal outbreak.
 - One strategy is to use a **narrow** cluster definition to identify clusters. That strategy will reduce the number of misclassified cases and will increase the measure of association
 - Once an outbreak is identified, the cluster definition can be **expanded** to identify additional cases that were missed because of the initial stringent cluster definition.



Tree-temporal scan statistic

- A signal detection/data-mining method
 - Used by CDC, FDA, and academic scientists for pharmacovigilance
 - Scan electronic health data grouped into hierarchical tree structures (e.g., ICD-10-CM diagnoses, allele codes for WGS data)
- 
- The logo for TreeScan, featuring the word "Tree" in blue and "Scan" in red, with a circular graphic of red and blue dots forming a spiral pattern behind the text.
- TreeScan™ software
 - Free download: www.treescan.org
 - Open-source code: github.com/scanstatistics/treescan
 - We worked with TreeScan software developer to accommodate our use case for prospective disease surveillance



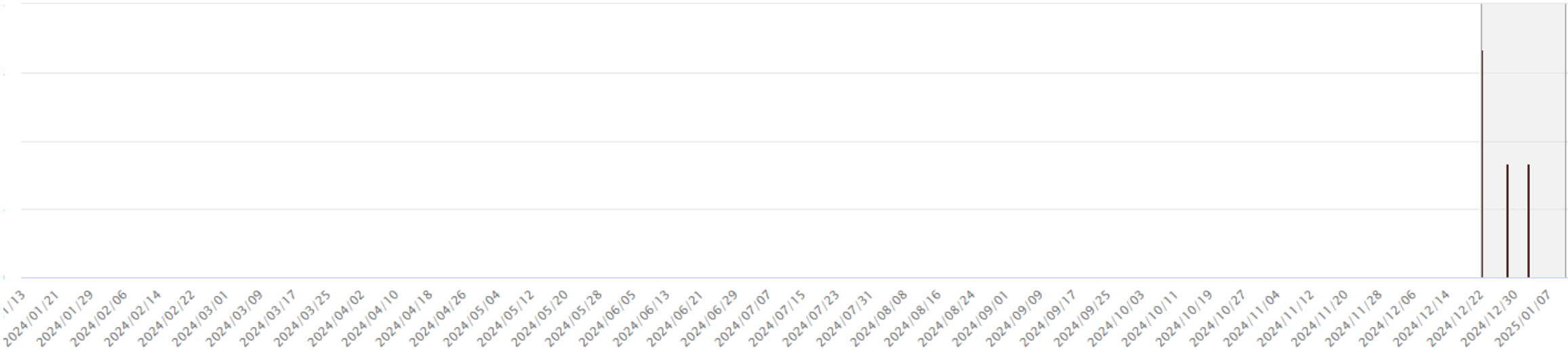
Detect emerging clusters in genomic surveillance data



(Specimen
collection date)

Node ID	Tree Level	Node Cases	Time Window	Cases in Window	Expected	Relative Risk	Excess Cases	Test Statistic	P-value	Recurrence Interval
SAL.Enteritidis.6.6.6.6x.8x.69011	8	4	2024/12/22 to 2025/01/11	4	0.16	infinity	4.00	8.986963	0.00274	7.0 years

SAL.Enteritidis.6.6.6.6x.8x.69011





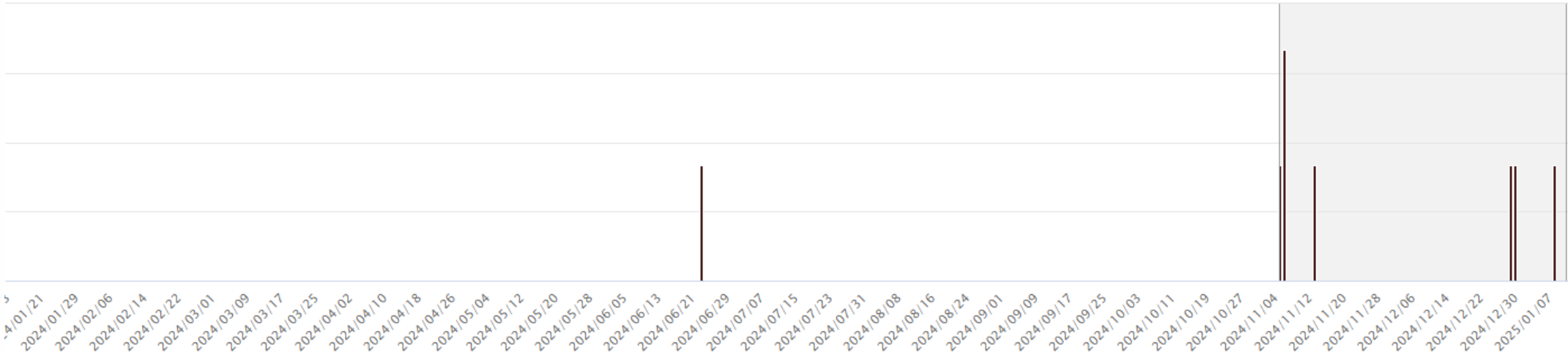
Detect emerging clusters in genomic surveillance data



(Specimen
collection date)

Node ID	Tree Level	Node Cases	Time Window	Cases in Window	Expected	Relative Risk	Excess Cases	Test Statistic	P-value	Recurrence Interval
SAL.Enteritidis.6.6.6.6x.8x.69011	8	4	2024/12/22 to 2025/01/11	4	0.16	infinity	4.00	8.986963	0.00274	7.0 years
SAL.Agona.78.1960.2992.10396.12707	7	8	2024/11/05 to 2025/01/11	7	1.06	47.83	6.85	7.278872	0.03140	224 days

SAL.Agona.78.1960.2992.10396.12707





Detect emerging clusters in genomic surveillance data



(Upload date)

Node ID	Tree Level	Node Cases	Time Window	Cases in Window	Expected	Relative Risk	Excess Cases	Test Statistic	P-value	Recurrence Interval
SAL.Newport.88.111.129.147x.194632.308139	8	3	2025/08/12 to 2025/08/19	3	0.11	infinity	3.00	7.066348	0.01353	1.4 years

- Initiated an investigation
- Noticed a related patient in another jurisdiction
- Patients dined at the same restaurant and reported beef consumption
- Case-reinterviews, an environmental assessment, additional case-finding

Original article

Applying prospective tree-temporal scan statistics to genomic surveillance data to detect emerging SARS-CoV-2 variants and salmonellosis clusters in New York City

Sharon K. Greene^{1,*} , Julia Latash¹ , Eric R. Peterson¹, Alison Levin-Rector¹ , Elizabeth Luoma¹, Jade C. Wang¹, Kevin Bernard¹, Aaron Olsen¹ , Lan Li¹, HaeNa Waechter¹, Aria Mattias¹, Rebecca Rohrer¹ , Martin Kulldorff²

¹Division of Disease Control, New York City Department of Health and Mental Hygiene, Long Island City, NY, United States

²Independent Biostatistician, Ashford, CT, United States

International Journal of Epidemiology, 2025, 54(4), dyaf084
<https://doi.org/10.1093/ije/dyaf084>
Commentary



Commentary

Commentary: An example of translational science with tree-temporal scan statistics—building methodological innovations into public health practice

Shirley V. Wang^{1,*} 

¹Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

This work illustrates a shift in how we might think about pathogen surveillance. Rather than relying on fixed rules, public health may benefit from flexible frameworks that can adapt to changing biology. Tree-temporal scan statistics, as implemented here, provide such a framework. However, they are neither a panacea nor a replacement for expert review. Rather, they provide a structured way to prioritize attention and resources.

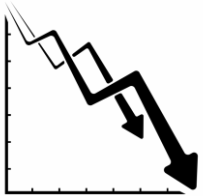
Greene et al. offer a template for translational science—taking a generalizable method, validating it through prospective use in parallel with existing systems, and then embedding it within the operational realities of a public health agency. This is a model worth replicating.



Detect spatiotemporal clusters in reportable diseases



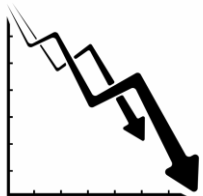
Detect emerging clusters in genomic surveillance data



Detect drop-offs in electronic laboratory reporting



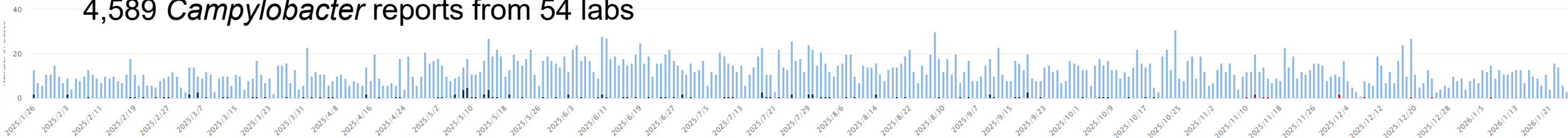
Detect increases in ICD-10 diagnosis codes for ED visits



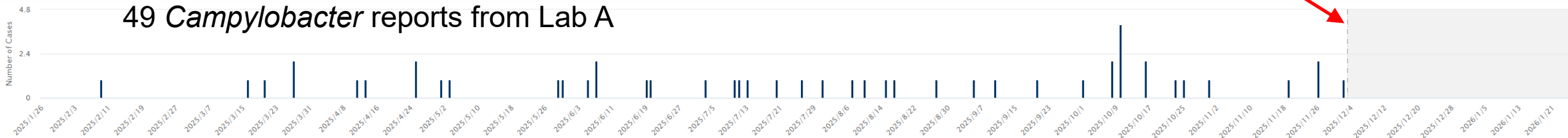
Detect drop-offs in electronic laboratory reporting



4,589 *Campylobacter* reports from 54 labs



49 *Campylobacter* reports from Lab A



Unpublished data as of 1/27/26, NYC Health Department. Not for further distribution.

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



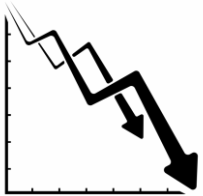
PRESIDENTIAL PRIORITIES AWARD



Detect spatiotemporal clusters in reportable diseases



Detect emerging clusters in genomic surveillance data



Detect drop-offs in electronic laboratory reporting



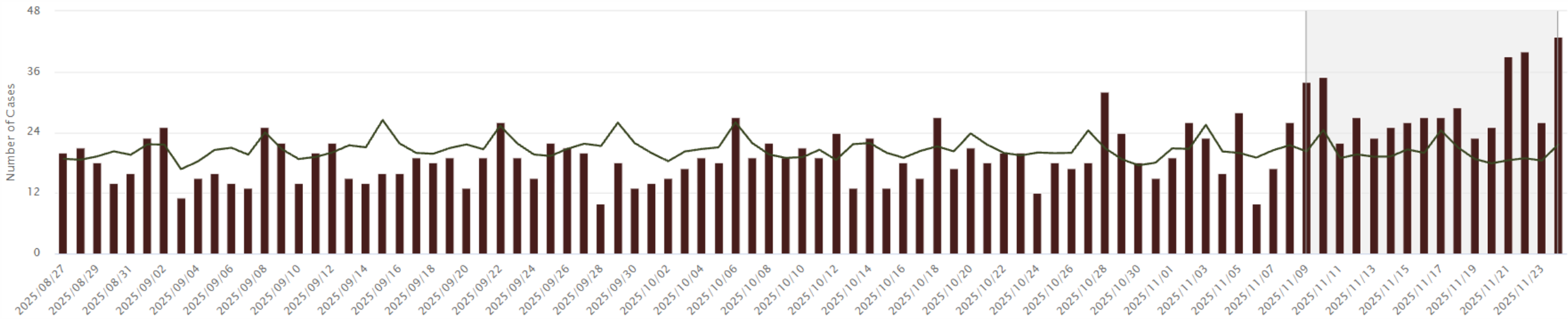
Detect increases in ICD-10 diagnosis codes for ED visits

Goal: detect emerging health issues that might not be captured by reportable disease or syndromic surveillance

- Use TreeScan to monitor for increases in ED visits overall or in hospital admissions for any of thousands of ICD-10-CM diagnoses and related diagnosis groupings, over varying time frames, while adjusting for the multiplicity of overlapping codes and time frames evaluated

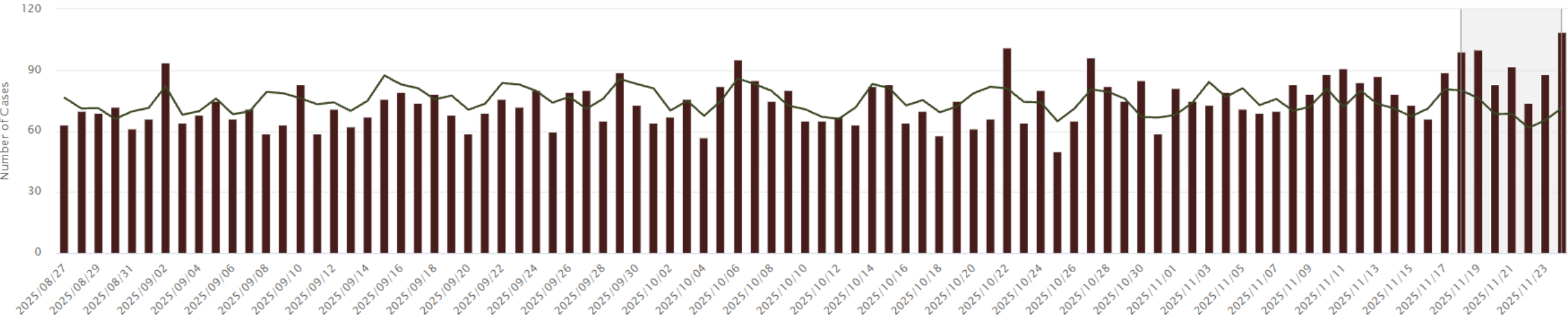
A08.4: Viral intestinal infection, unspecified

Node ID	Node Name	Tree Level	Node Cases	Time Window	Cases in Window	Expected	Relative Risk	Excess Cases	Test Statistic	P-value	Recurrence Interval
2-A08.4	2-Viral intestinal infection, unspecified	5	1851	2025/11/09 to 2025/11/24	471	321.45	1.65	184.70	30.392492	0.00001	274 years

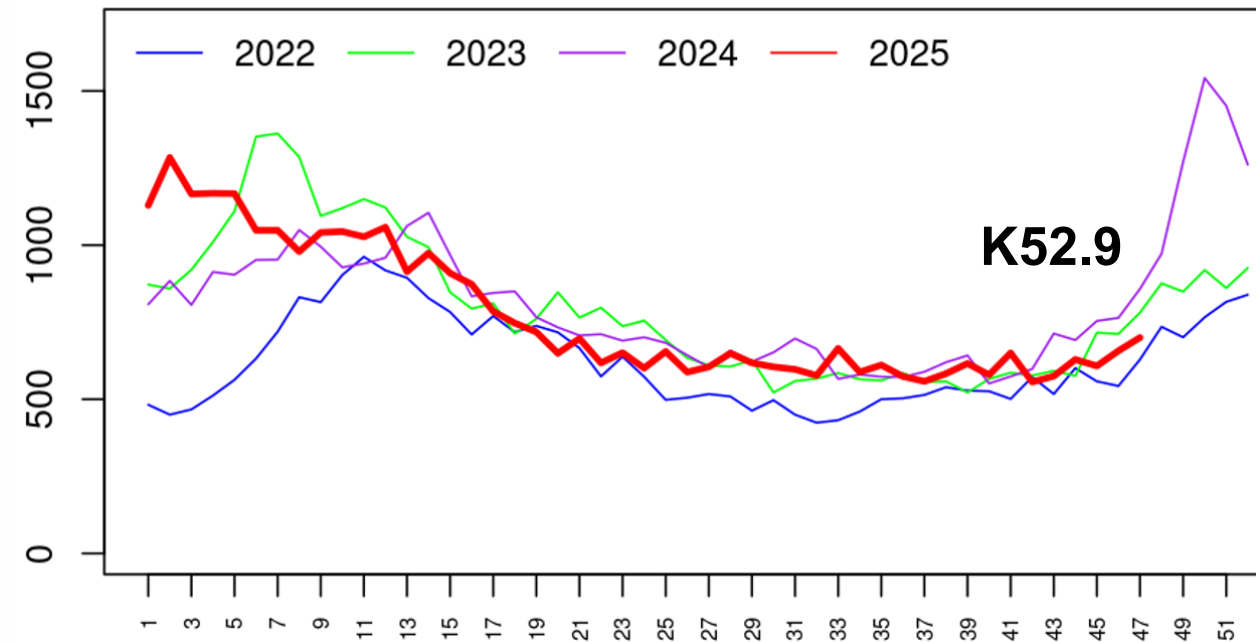
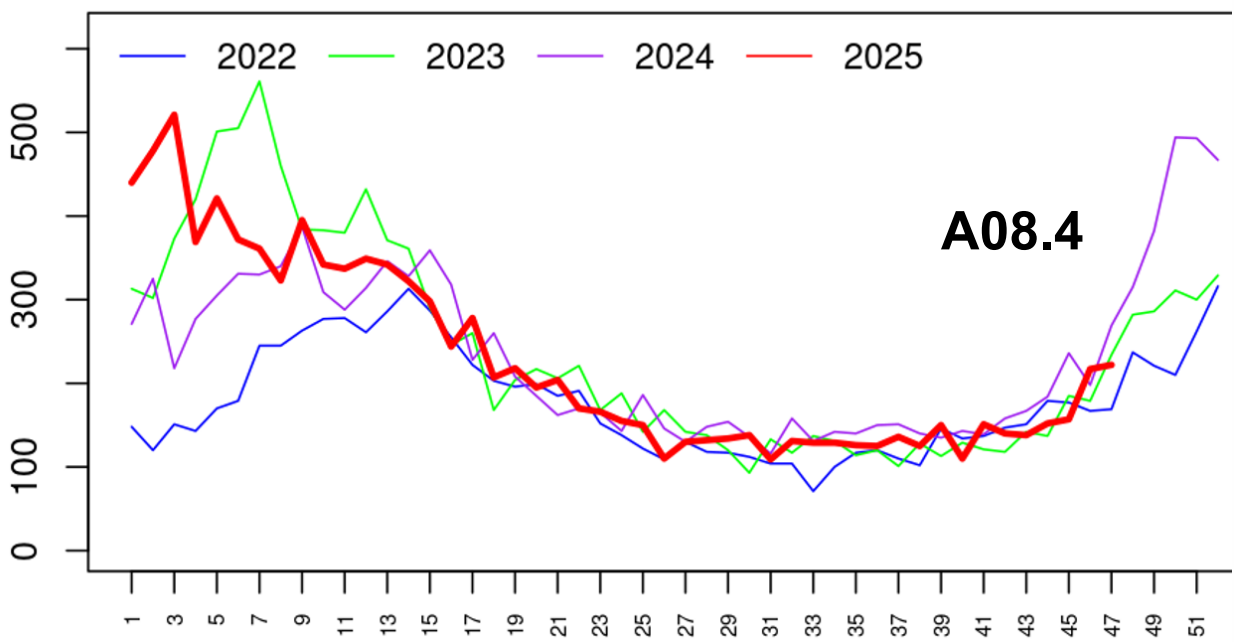


K52.9: Noninfective gastroenteritis and colitis, unspecified

Node ID	Node Name	Tree Level	Node Cases	Time Window	Cases in Window	Expected	Relative Risk	Excess Cases	Test Statistic	P-value	Recurrence Interval
2-K52.9	2-Noninfective gastroenteritis and colitis, unspecified	5	6717	2025/11/18 to 2025/11/24	645	491.19	1.35	167.19	21.913558	0.00001	274 years



Both increases were within typical seasonal bounds



Asyndromic surveillance of New York City emergency department diagnoses with the tree-temporal scan statistic

Journal of Public Health Research

Vol. 15(2): 1–16

© The Author(s) 2026

DOI: 10.1177/22799036261457753

journals.sagepub.com/home/phj



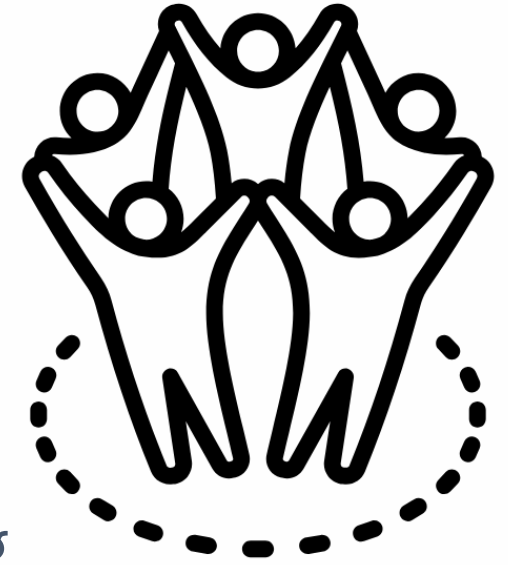
Sharon K. Greene, PhD, MPH¹ , Alison Levin-Rector, MPH¹, Martin Kulldorff, PhD², and Ramona Lall, PhD¹ 



See also: epiengage.org/projects/treescan/

Would health departments like more support with scan statistics for enteric disease surveillance?

- Provide data, methodological, epidemiological, and educational support to HD staff
- Produce training modules on data sources, statistical techniques, software, and public health practice
- Develop community of practice groups for sharing best practices, successes, and challenges
- Provide 1:1 coaching for implementation, problem solving, and interpretation
- Increase biostatistical support and enhance software to meet needs and use cases of additional HDs





ESSENCE for Clusters and Outbreaks

What is ESSENCE?

- ESSENCE is the Electronic Surveillance System for the Early Notification of Community-based Epidemics

ESSENCE - NSSP Home

Version 1.21

Date	Description
15Jul21	Notice to users: Data loaded to the Staging BioSense platform should not be considered permanently retained. Data older than 90 days may be deleted from the system to free up system resources.
15Jul21	Data files transmitted to the Staging system for Onboarding purposes are not retained. Site Administrators will be notified in advance of any planned data storage maintenance which results in data being deleted earlier than 90 days.
14Jul21	Note to Users: You are logged into the Staging ESSENCE Application
14Jul20	Staging system maintenance occurs the third Tuesday of every month from 6:00 AM to 10:00 AM

This Information is for Authorized use only.

Your ability to access this information is granted with the expectation and understanding that you will comply with and not violate privacy information policies. This is a private system and is only to be used by authorized users. By continuing, the user is stating that they are the indicated user.

ESSENCE PROTOTYPE DISCLAIMER

- ESSENCE receives near real-time data from emergency department (ED) chief complaints and urgent care visit notes
 - The system classifies the data into syndromic categories, including a “GI syndrome” (e.g., diarrhea, vomiting)

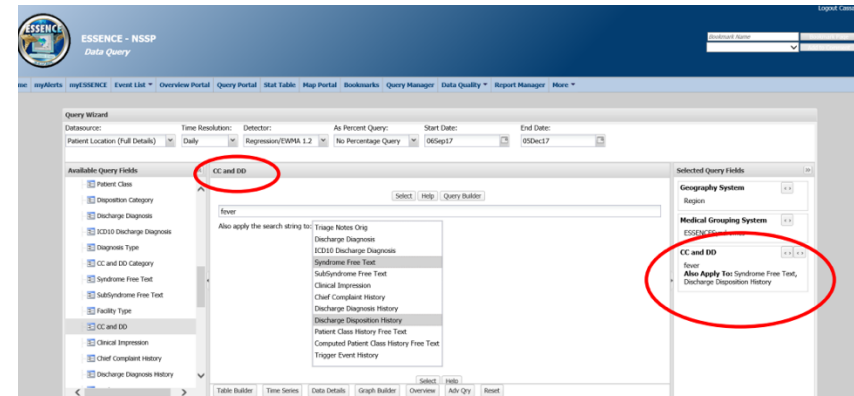
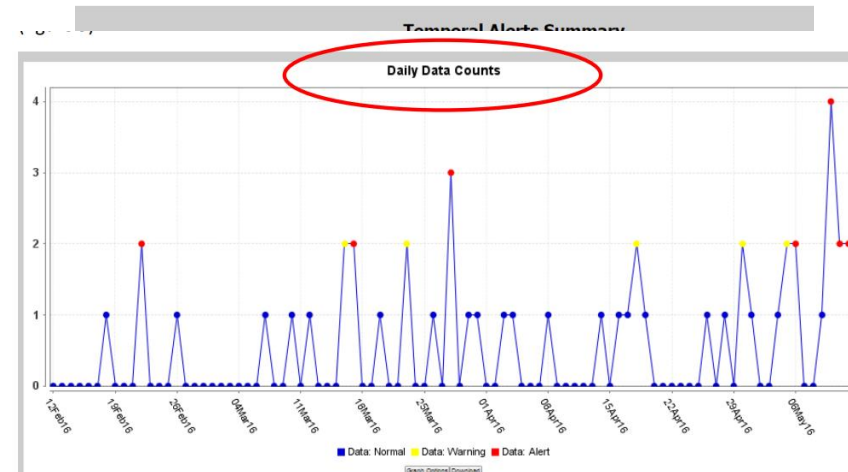
ESSENCE as an additional tool

There is a GI syndrome in ESSENCE

- Useful to monitor for any obvious increases in GI visits
- Can be viewed based on region/county in state
- May generate a lot of noise

Users can create customized queries to monitor more specific outcomes or exposures using:

- chief complaint text
- discharge diagnoses
- triage notes
- other data fields



ESSENCE as an additional tool



Can detect illness clusters before specimens are collected



May capture outbreaks not found via WGS

- Viral outbreaks
- Toxin-related outbreaks
- Outbreaks with quick recovery (no testing)

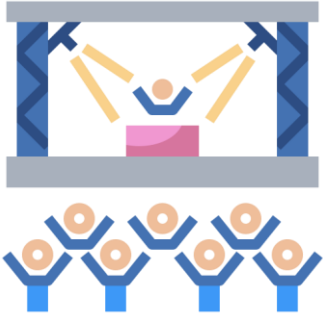


Helps identify unusual increases of GI-related visits by county, zip code, or specific hospitals



Support known outbreak investigations by assessing size of outbreak, geographic spread, if activity is back to baseline

Outbreaks and ESSENCE



Music Festival in Summer 2024

ESSENCE helped find additional cases and corroborate reports of illness mentioned via the complaint system and social media

Church Thanksgiving Potluck Fall 2025

Emergency Preparedness colleagues alerted the FED Team of four ED visits in West Tennessee reporting GI illness after attending a church Thanksgiving potluck and uncovered an outbreak!



What will ESSENCE help with in 2026?

- Exploring a specific HUS query
- Testing a “Stomach Bug” query
- Creating a process to triage or alert region or county

Knowledge Check

What type of data does ESSENCE receive?

- A. Deidentified data including chief complaint, diagnosis codes, patient characteristics, and location
- B. Patient name, DOB, address, and contact information
- C. Chief complaint data only



Communicating Results

Communicating WGS Results

Poll Question:

When communicating WGS results, which challenge do you encounter most often or anticipate encountering?

- a. Explaining technical terms
- b. Managing uncertainty
- c. Delays in lab results
- d. Tailoring messages to different audiences

Communicating WGS Results

Why is it challenging?

Whole genome sequencing (WGS) provides powerful lab evidence, but it can be difficult to explain to non-technical audiences



Common Challenges:

- Scientific jargon
- Interpreting uncertainty
- Different audiences and needs
- Delay or lack of testing and results



Communicating WGS Results

WGS CAN tell us	WGS (alone) CANNOT tell us
Whether illnesses may be related	Exactly where someone got sick
Which cases belong in an investigation	Which specific meal caused illness
Which cases are likely unrelated	Who infected whom
If isolates are highly similar	The complete outbreak story

Communication Tips

- Use plain language
- Set expectations for turnaround times
- Provide regular updates
- WGS is one piece of evidence
- Investigations may evolve

Challenge Scenario #1

Sequence data suggests relatedness, but Epi/EH data is missing or does not align

Situation:

- Twelve *Salmonella* Enteritidis cases spread in multiple counties are highly related by WGS in your state
- Interviews do not identify a common exposure
- Environmental samples from a potential source are highly related to the clinical cases

How would you share a high-level update with leadership at this point in the investigation?

Knowledge Check

When WGS findings and epidemiologic data do not fully align, what is the best communication approach?

- A. Focus only on WGS results because they are the strongest piece of evidence
- B. Focus only on the epidemiologic findings because WGS did not identify a source
- C. Explain what each piece of evidence suggests, acknowledge uncertainty, and describe next steps
- D. Delay communication until all evidence is aligned

Challenge Scenario #2

No genetic relatedness with clinical and environmental samples, but epidemiological evidence is strong

Situation:

- Twenty-three *Salmonella* Newport cases report eating at the same Mexican-styled restaurant. No environmental samples have been linked to the clinical cases and food testing results are still pending. Stakeholders want definitive answers.

Knowledge Check

No/Unclear WGS Linkage with clinical and environmental isolates

The restaurant manager asks about the environmental sampling and food testing results. What is the BEST way to describe the current situation?

- A. The environmental isolates were 14 alleles and 16 SNPs different from the outbreak strain. The food samples are still being cultured and will be sequenced if anything grows on the media.
- B. The environmental samples were not as closely related genetically to the clinical cases. The food samples are still pending and will be shared as early as Tuesday of next week.
- C. We cannot share an update at this time because all testing has not been completed yet.
- D. Environmental isolates were unrelated to the cases and food testing is pending.

WGS Communication Best Practices



- Explain what is known and unknown
- Avoid overstating certainty



- Describe how findings may evolve
- Set expectations with your audiences



- Emphasize WGS is one piece of the investigation

Questions?