Methods for Detecting Reportable Disease Clusters in Space and Time
June 15, 2015, 5:45-6:30 pm — Sheraton Hotel, Clarendon Room
CSTE 2015 Annual Conference Roundtable

Facilitators:

- Sharon K. Greene, PhD, MPH (sgreene4@health.nyc.gov), Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, Queens, NY
- Lauren Torso, MPH (LTorso@achd.net), Allegheny County Health Department, Pittsburgh, PA

Objectives:

- Describe methods currently used by public health departments for reportable disease cluster detection.
- Identify strengths and limitations of implementing different approaches.
- Compile best practices for accounting for common data and analytic challenges.

2013 CSTE Epidemiology Capacity Assessment:

- "Does your state routinely use automated cluster detection software on reportable disease and laboratory finding data?" Fifteen states (29%) indicated 'yes.'[1]
- Tools include: EpiCenter, EARS, DESTEM (Disease Electronic Surveillance with Trigonometric Models), SAS

#	Issue or Challenge	Discussion points, based on NYC experience	
1	Which reportable diseases to analyze?	- Exclude diseases so rare that any clusters would be detected without automated analyses (e.g., tularemia and human rabies) and diseases with very long incubation periods (e.g., leprosy, Creutzfeldt-Jakob disease). - Include diseases for which prospective and timely cluster identification might result in public health action (e.g., campylobacteriosis, legionellosis, shigellosis, etc.) - Consider excluding *Salmonella* overall (too non-specific), but include *Salmonella* by serotype.	
2	Which analytic method?	**Refined Historical Limits Method**[2] (purely temporal)	**Prospective space-time permutation scan statistic in** SaTScan[3] (spatio-temporal)
3	Time period of interest for recent clusters	Four weeks	- Up to 30 days for most diseases - Up to 60 days for Shiga toxin-producing *E. coli*, hepatitis A, each *Salmonella* serotype, typhoid fever, and paratyphoid fever - Up to 180 days for legionellosis and listeriosis
4	Baseline time period for comparison	Cases diagnosed within 15 prior periods: same 4-week period, the preceding 4-week period, and the subsequent 4-week period during past 5 years	- 1 year for most diseases - 1.5 years for 60-day maximum temporal window - 5 years for 180-day maximum temporal window

#	Issue or Challenge	Refined Historical Limits Method	Prospective space-time permutation scan statistic in SaTScan
5	Generating a signal	Case counts in current period >2 standard deviations above baseline mean	Recurrence interval (length of follow-up required to expect one cluster at least as unusual as observed cluster by chance) ≥100 days
6	Accounting for seasonality	Baseline restricted to similar months	Automatically adjusted for purely temporal clusters (e.g., seasonal variation)
7	Analysis frequency	Weekly	Daily
8	Date of interest for analysis	Preferred hierarchy: onset date > diagnosis date > specimen collection date > report date > date event created in surveillance database	
9	Accounting for reported cases that are still pending investigation or confirmation	Include all reported cases in the analysis regardless of current status (e.g., confirmed, probable, suspected, pending, "not a case")	
10	Accounting for lags in data accrual	Repeat analysis for a given 4-week period over 4 consecutive weeks	- Generally none; assuming no spatial variation in data lags (have quality assurance procedures in place to detect reporting drop-outs). - Analyze *Salmonella* serotype at a two week delay to allow time for serotyping
11	Account for secular trends or past clusters in baseline data	Identify and remove any statistically significant linear trend in historical data	Automatically adjusted for temporal trends
12	Geographic aggregation of residential address for analysis	3 resolutions: citywide, borough (n=5), neighborhood (n=42)	- Census tract - Maximum spatial cluster size: 50% of cases
13	Presenting cluster summary information	- Detailed report for each signal, showing cluster summary information, maps of geocoded patient residences, case linelists - Compare signals across consecutive runs to flag new signals, and to flag new cases added to previous signals	
14	Automating analyses	- Write SAS code to automatically detect signals, generate output, place output in secured folder, and generate e-mails to appropriate staff for each disease - Dedicate a computer to routine automated analyses; set up to run SAS code at specified intervals using Microsoft Task Scheduler	

1. CSTE. 2013 national assessment of epidemiology capacity: Findings and recommendations. P. 39.
2. Levin-Rector A, Wilson EL, Fine AD, Greene SK. Refining historical limits method to improve disease cluster detection, New York City, New York, USA. Emerging Infectious Diseases, 2015; 21:265-272.
3. Kulldorff M, Heffernan R, Hartman J, Assunção RM, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Medicine, 2005; 2(3):e59.