

Spatio-Temporal Cluster Detection for Legionellosis using Multiple Patient Addresses

Eric R. Peterson* and Sharon K. Greene

Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, Queens, NY, USA

Objective

To improve timeliness and sensitivity of legionellosis cluster detection in New York City (NYC) by using all addresses available for each patient in one analysis.

Introduction

The Bureau of Communicable Disease (BCD) at the NYC Department of Health and Mental Hygiene performs daily automated analyses using SaTScan to detect spatio-temporal clusters for 37 reportable diseases.¹ Initially, we analyzed one address per patient, prioritizing home address if available. On September 25, 2015, a BCD investigator noticed two legionellosis cases with similar work addresses. A third case was identified in a nearby residential facility, and an investigation was initiated to identify a common exposure source. Four days later, after additional cases living nearby were reported, the SaTScan analysis detected a corresponding cluster. In response to this signaling delay, we implemented a multiple address (MA) analysis to improve upon single address (SA) analyses by using all location data available on possible exposure sites.²

Methods

Positive *Legionella* test results for NYC residents are reported to BCD with patient demographic and address data. BCD interviews all cases to elicit additional locations of potential exposure and enters the addresses into a disease surveillance database (Maven). Addresses are assigned X/Y coordinates in near real-time via integration with a geocoding webservice.

We used the prospective space-time permutation scan statistic in SaTScan,³ enabling the advanced input feature on the spatial neighbors tab to “include location ID in the scanning window if at least one set of coordinates is included.” This option considered a case as included in a given cluster if any of the case’s addresses were within the cluster. The case file included: unique case ID (as the location ID), number of cases, onset date, and day of week. The coordinate file included: case ID and X/Y coordinates for each address per case, resulting in one or more rows per case. We searched for alive clusters with a temporal range of 2 to 30 days and a maximum spatial size of 50% of observed cases. The study period was 1 year. Monte Carlo simulations (N=999) were used to determine statistical significance.

We mimicked prospective surveillance to determine when the September 2015 cluster would have been detected had this analysis been in place, by performing daily SA and MA analyses from September 21 (when the first outbreak-linked case was reported) to September 29 (when the initial SaTScan analysis signaled). Any cluster with a recurrence interval (RI) ≥ 100 days was summarized in a map and linelist. Prospective, automated analyses were launched in April 2016 and run daily using Microsoft Task Scheduler, SAS 9.4, and SaTScan 9.4.1. Signals through July 2016 were summarized.

Results

In mimicked prospective analysis, the SA and MA SaTScan analyses identified clusters of 13 and 11 cases, respectively, starting September 27, 2015. The MA cluster was more spatially focused (2.11 km vs. 5.42 km) and more unlikely to occur by chance alone (RI of 16,256 days vs. 8,758 days). In prospective analyses, a MA

cluster of 6 cases was identified on July 5, 2016 with a radius of 1.69 km (RI=100 days). On July 6, the MA cluster case count increased to 7 and maintained the same radius (RI=685 days), while a cluster of the same 7 cases was identified by the SA analysis with a larger radius (1.97 km) and lower RI (292 days). The RI for both clusters peaked on July 7 (MA: 2348 days, SA: 713 days).

Conclusions

In preliminary evaluation, the MA analysis facilitated cluster detection using non-residential possible exposure sites, such as workplaces. Timeliness was slightly improved, but the larger practical benefit was identifying more spatially focused clusters. Smaller clusters are useful for more precisely targeting legionellosis infection source identification and remediation activities, especially in urban environments with high population and building densities.

Keywords

Legionella; cluster detection; SaTScan

Acknowledgments

Martin Kulldorff suggested using multiple coordinates per location in SaTScan.

References

1. Greene SK, et al. Daily reportable disease spatiotemporal cluster detection, New York City, New York, USA, 2014–2015. *Emerg Infect Dis.* 2016;22(10).
2. Bull M, et al. The application of geographic information systems and spatial data during Legionnaires’ disease outbreak responses. *Euro surveillance.* 2012;17(49).
3. Kulldorff M, et al. A space-time permutation scan statistic for disease outbreak detection. *PLoS medicine.* 2005;2(3):e59.

*Eric R. Peterson

E-mail: epeterson@health.nyc.gov

