Detection Abilities of Several Commonly Used Algorithms as Determined by Simulation Analysis

Laura McDonald MSc, Glenn Guthrie MSc, Victoria Edge PhD, Jeff Aramini PhD

Foodborne Waterborne and Zoonotic Infections Division, Public Health Agency of Canada

Objective: To determine the sensitivity, specificity and days to detection of several commonly used algorithms in syndromic surveillance systems.

Background: The Public Health Agency of Canada (PHAC) is currently utilizing a syndromic surveillance prototype called the Canadian Early Warning System (CEWS). This system monitors several live data feeds, including emergency room (ER) chief complaint records from all seven local hospitals, Telehealth (24/7 nurse hotline) calls, and over-the-counter (OTC) drug sales from a number of the large chain drug stores. Data trends are analysed for aberrations as early indicators of outbreak events. Collaborators on this Winnipeg, Manitoba-based pilot include the Winnipeg Regional Health Authority and IBM Business Solutions. Algorithms currently in CEWS include the 3, 5 and 7-day moving averages, CUSUM and the CDC's EARS (1). We seek to investigate the performance of these algorithms in view of the fact that their detection ability may be dependent upon data source and/or the type of outbreak event.

Materials and Methods: First, three distinct outbreak scenarios were developed i) an intentional release of anthrax ii) a large outbreak of influenza and iii) a waterborne *E.coli* outbreak. Historical data for these events were used when available; otherwise estimates based on the expected event profile and counts were applied. Epidemiological curves for the ER, Telehealth, and OTC data streams were generated from these scenarios.

The second phase of this study involved generating expanded baseline data, incorporating the aforementioned outbreaks, and finally, testing the algorithms. One hundred years of baseline data were generated based on the trends and variability of two years of retrospective data for each data source. Outbreaks were then randomly introduced into the data at roughly one per year with at least 30 days interval passing between outbreaks (no overlaps). This type of simulation analysis allowed us to alter the parameters of the outbreak, including scale, duration, and variability. Tests were run in two ways: the threshold was either predetermined (i.e. 3 standard deviations from the mean), or was based on a set number of false positive alerts (1, 2, 3, 4, 6, or 12 per year).

Results: This study shows that the standard method of using a set threshold limit based on control chart methods or standard deviations may not be the optimal set point to maximize detection abilities (Table 1). The study also finds that tuning the threshold to the data for a given algorithm is more important than algorithm selection. By adjusting the thresholds for a given algorithm, the time to detection and false positive rate could be tuned to the point where the differences between the algorithms performance become minor.

Table 1. Results for a simulated *E. coli* outbreak incorporating 1750 cases into ER visit data for Winnipeg, Manitoba (pop. 700,000+) for the 7-day Moving Average algorithm.

Threshold[2]	0.7[3]	0.9[3]	1	1.3[3]	2	3
False pos/year[1]	6	3	1.6	1	0.0	0.0
Days to detec'n	6.7	7.2	7.5	7.7	8.9	9.9

1- Specificity was 100% (no false neg.) due to large outbreak scale
2- Number of Stan. Dev. to which the threshold is set
3- Threshold tuned to desired number of false pos.

Conclusion: Tuning algorithm thresholds to provide an acceptable number of false positives and days to detection can enhance the performance of a syndromic surveillance system for early event detection. The ability to customize algorithm parameters is essential when developing a syndromic surveillance system.

References:

1. Hutwagner LC, Thompson WW, Seeman GM, Treadwell T. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. Stat Med 2005 Feb 28;24(4):543-50.

2. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. Emerg Infect Dis 1997 Jul-1997 Sep 30;3(3):395-400.