# Systematic comparison of algorithms used in syndromic surveillance

**Mike Jackson, MPH[1,2]; Atar Baer, PhD[1]; Ian Painter, PhD[3]; Jeff Duchin MD[1,2]**

*[1]Public Health: Seattle & King County; [2]Department of Epidemiology, University of Washington; [3]Foundation for Healthcare Quality*

## OBJECTIVE

We conducted a large simulation study to evaluate the detection properties of 6 different algorithms across a range of outbreak characteristics.

## BACKGROUND

Varied approaches have been used by syndromic surveillance systems for aberration detection. However, the performance of these methods has been evaluated only across a small range of epidemic characteristics.

## METHODS

We based our evaluation on the approach recommended by Mandl et al [1]. We used the actual daily counts of four different syndromes from 9 emergency departments in Seattle & King County in 2004 as baselines. We selected syndromes having average daily counts of 2, 10, 35 and 50 visits per day. We then simulated epidemics by creating all possible combinations of 3 different parameters: temporal distribution, outbreak size, and outbreak duration. We used 9 different distributions (4 simple mathematical functions and 5 chosen from "epi-curves" of historical outbreaks), 10 different outbreak sizes (from 5 to 50 cases per day), and 8 different durations (from 1 to 32 days). We added the simulated counts to the actual baseline counts, with the epidemic starting on January 2nd, 2004. For each day of 2004, we then calculated each of 6 algorithms at each of 8 false positive rates. We repeated this process starting the epidemic on every other day of 2004, for a total of 183 repeats per epidemic per baseline. We calculated the sensitivity of detecting the outbreak on any given day, overall probability of detecting the epidemic, and the timeliness of detection for each of 6 algorithms: 3 CUSUM algorithms currently used by CDC-EARS[2]; 2 Exponential Weighted Moving Average (EWMA) algorithms used by the ESSENCE system [3], with smoothing constants of 0.4 and 0.9; and a generalized linear model (GLM) [4] similar to that used by BioSense. Algorithm performance was examined overall and stratified by outbreak distribution, duration, and size and by baseline syndrome.

## RESULTS

Overall, timeliness was similar for all algorithms, with time to detection ranging between 2 and 4 days after the start of the outbreak. When averaged across all outbreak parameters, the GLM had the best sensitivity and outbreak detection probability. Setting the false positive rate at 1 per month, the GLM detected 64% of outbreak days, and 30% of outbreaks overall. Stratified by baseline levels, distribution, and outbreak size, however, the relative performance of the algorithms varied greatly. In general, the GLM performed better than the other algorithms. However, the EWMAs outperformed the other algorithms when the number of false positives was 1 per month or greater, the average daily count of the syndrome was small (2 per day), and the outbreak lasted 4 or more days. In addition, the CUSUMs had the best performance when the average daily count was high (50 per day) and the outbreak size was small (5 per day).

Table 1: Overall outbreak detection and timeliness of 6 algorithms at one false positive per month.

| Algorithm | % outbreak-days detected | % outbreaks detected | Mean first day detected |
|---|---|---|---|
| CUSUM1 | 18.8% | 56.5% | 3.79 |
| CUSUM2 | 22.8% | 53.9% | 3.78 |
| CUSUM3 | 22.1% | 51.8% | 3.84 |
| EWMA (0.4) | 28.9% | 56.1% | 3.71 |
| EWMA (0.9) | 27.7% | 60.6% | 3.70 |
| GLM | 30.2% | 64.5% | 3.61 |

## CONCLUSIONS

No single algorithm outperformed the others across all outbreak parameters. To summarize the performance of the algorithms across outbreak distribution, duration, size, baseline levels and false positive rates, we created an Algorithm Performance Matrix (APM). The APM allows syndromic surveillance users to: (1) determine which algorithm performs best overall across various outbreak parameters; (2) determine the sensitivity of outbreak detection at a set baseline and false positive rate; and (3) estimate the minimum outbreak size needed in order to achieve a desired level of sensitivity.

## REFERENCES

[1] Mandl K, Reis B, and Cassa C, Measuring Outbreak-Detection Performance by Using Controlled Feature Set Simulations. MMWR 2004:53(Suppl):130-136.

[2] Hutwagner L, Thompson W, Seeman M, and Treadwell T, The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS). J Urban Health. 2003:80(2 Suppl 1):i89-96

[3] Burkom H, Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance. Johns Hopkins Apl Technical Digest 2003:24(4):335-342

[4] Kleinman K, Lazarus R, and Platt R, A Generalized Linear Mixed Models Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism. Am J Epidemiol 2004:159(3):217-224

Further Information:
Mike Jackson, Michael.Jackson@metrokc.gov