# A Pilot Study of Aberration Detection Algorithms with Simulated Data

**Hwa-Gan Chang[1] PhD, Jian-Hua Chen [1] MD, MsPH, Dennis G. Cochrane[2] MD, John R. Allegra[2] MD, PhD, Howard Burkom[3]PhD, Jerome I Tokars[4] MD, Perry F. Smith[1]MD**

[1]*New York State Department of Health, Albany, NY*

[2]*Emergency Medical Associates of New Jersey Research Foundation, New Jersey*

[3]*Applied Physics Laboratory, John Hopkins University, Laurel, MD*

[4] *Centers for Disease Control and Prevention, Atlanta, GA*

**Objectives:** To evaluate four algorithms with varying baseline periods and adjustment for day of week for anomaly detection in syndromic surveillance data.

**Background:** The New York State Department of Health (NYSDOH) Electronic Syndromic Surveillance system currently uses the Centers for Disease Control and Prevention (CDC) EARS algorithm (1) for detecting anomalies in daily emergency department visit data for 130 hospitals with 6 syndrome categories. Using a 7-day baseline and a 3 standard deviation cut off value routinely generates a large number of signals (alerts) for counts within ranges in the historical baseline. CDC's BioSense proposed a modified algorithm (W2) 7- day baseline stratified by weekday/weekend. (2) This study compares the signals generated with inserted simulated outbreak data for different baseline periods and for the 7-day baseline period with and without adjustments for weekends.

**Methods:** Daily chief complaint (CC) data from two New York hospital emergency departments (EDs) during the period April 13, 2004 through December 31, 2005 were classified into the gastrointestinal (GI) or respiratory syndromes through NY CC syndrome classifiers. For each hospital and syndrome, the syndrome counts were separately concatenated to create 10 years of daily count data (2 hospitals x 2 syndromes x 10 years). Additional daily counts of 1 to 10 representing simulated outbreaks (N=64) were inserted in these four time series, in the shape of a log-normal distribution, with lengths of 4, 6, or 12 days.(3) The first day of outbreak varied by day of week, with at least 45 days between outbreaks. The combined counts were analyzed with a 7-day baseline EARS algorithm (C2), and 7-, 14-, 28-day baseline W2 statistics. Sensitivity, percent of early detection (the outbreaks detected on or before peak day), the number of signals generated per 100 days, the positive predictive value (PPV) and specificity of signals were calculated.

**Results:** The GI syndrome had a mean daily count of 7.4 for weekdays and 7.3 for weekends, and the respiratory syndrome had mean daily syndrome counts of 11.3 and 13.4 respectively. Table 1 summarizes the mean of each statistic for the GI syndrome. The sensitivity, percent early detection and the number of signals trended lower whereas the PPV and specificity were higher for longer baseline period. Compared with the 7-day EARS C2, the 7-day W2 tended to have similar sensitivity and percent early detected with slightly fewer signals generated per 100 observation days and higher PPV. The respiratory syndrome showed similar trends.

**Conclusion**: Our results from this pilot study in two NYS hospitals suggested that the use of longer baseline periods for anomaly detection in syndromic surveillance data decrease the number of signals with only a small decrease in sensitivity and percent early detection. Adjustment for day of week with a 7-day baseline provided only a modest advantage. Further studies of data from more hospitals and different syndromes are needed to select the optimal detection algorithm.

**Table 1. Means of statistics for GI syndrome (baseline count mean=7.4, std dev=2.9)**

| ***Calculated by event*** | 7-day C2 | 7-day W2 | 14-day W2 | 28-day W2 |
|---|---|---|---|---|
| Sensitivity | 0.59 | 0.62 | 0.53 | 0.51 |
| % Early detection ***Calculated by days*** | 96% | 93% | 93% | 91% |
| Signals/100 days | 4.17 | 3.71 | 2.37 | 1.49 |
| PPV | 0.39 | 0.49 | 0.60 | 0.80 |
| Specificity | 0.97 | 0.98 | 0.99 | 1.00 |

**References**

1. Hutwagner L, Thompson W, Seeman GM, et al. The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS). J Urban Hlth 2003;80(2, suppl 1):i89--i96.

2. CDC. BioSense Real-Time Hospital Data User Guide, version 2.06 (September 2006), at http://0-www.cdc.gov.mill1.sjlibrary.org/biosense, 2007.

3. CR Goodall, S Halasz, JR Allegra, D Cochrane. A System for Simulation: Introducing Outbreaks into Time Series Data, Adv Dis Surveillance 2007; 2:199.