Emerging Health Threats Journal 2011, **4**:s17. **doi:** 10.3134/ehtj.10.017

www.eht-journal.org

ABSTRACT

Incorporating seasonality and other long-term trends improves surveillance for acute respiratory infections

H Zheng[1,2], WH Woodall[3], and S DeLisle[1,2]

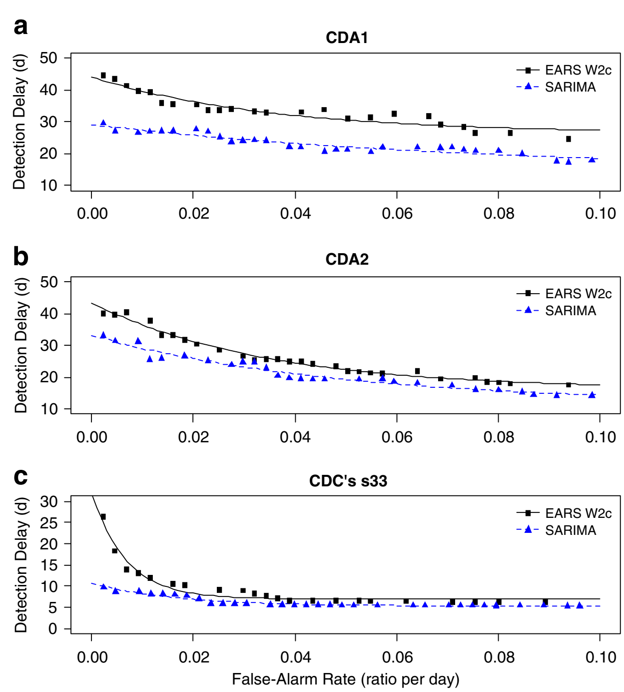
[1] *Veterans Administration Maryland Health Care System, Baltimore, MD, USA;* [2] *Department of Medicine, University of Maryland, Baltimore, MD, USA; and* [3] *Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*
E-mail: sdelisle@umaryland.edu

Introduction

As the electronic medical record (EMR) market matures, long-term time series of EMR-based surveillance data are becoming available. In this work, we hypothesized that statistical aberrancy-detection methods that incorporate seasonality and other long-term data trends reduce the time required to discover an influenza outbreak compared with methods that only consider the most recent past.

Methods

Authentic background time series of daily case counts were created by applying either of two case detection algorithms (CDAs) to EMR entries related to outpatient encounters at the Baltimore VA. The CDAs targeted acute respiratory infections (ARIs), and had the following composition and performance, compared with a manual review of 15,377 records:[1] (1) CDA1, a grouping of ICD-9 diagnostic codes similar to that used by the BioSense surveillance system (sensitivity 63%, positive predictive value (PPV) 16%; (2) CDA2, an optimized grouping of ICD-9 codes combined with the results of a computerized free-text analysis of whole-clinical notes (sensitivity 69%, PPV 54%). We used an age-structured metapopulation influenza epidemic model for Baltimore to inject factitious influenza cases into these backgrounds. From the time of this injection, aberrancy-detection statistics were applied each successive day on paired background + injection vs background-only time series. Each injection-prospective-surveillance cycle was repeated 52 times, each time shifting the injection to a different week of the study period (2003-04). We applied the same study scheme to CDC's s33 synthetic background and injection data sets (http://www.bt.cdc.gov//surveillance/ears/datasets.asp, accessed 20 February 2010). We computed two whole-system benchmarks: (1) the 'Detection Delay', the average time from injection to the first true-positive signal, defined as a statistical alarm originating in the background + injection data set but not present in the background-only data set; (2) the 'False-Alarm Rate' (FAR), defined as the number of unique false-alarms originating in the background-only data set during the study year, divided by 365 days. Detection Delay–FAR pairs were determined empirically over broad ranges of alert thresholds. We compared two aberrancy-detection approaches: (1) CDC's early aberration reporting system (EARS) W2c,[2] which makes predictions using 4 weeks of past data; (2) background-specific seasonal autoregressive integrated moving average (SARIMA) models,[3] which used 8 years of historical data. These methods were implemented for weekdays and for weekend/holidays time series.



Results

The figure shows activity monitoring operating characteristic (AMOC) curves for epidemic Detection Delay (*y*-axis, in days) as a function of FAR (*x*-axis). Results using EARS W2c (squares) are compared with those using SARIMA (triangles) for CDA1 (panel 'a'), CDA2 (panel 'b') and for CDC's s33 synthetic data set (panel 'c'). Note that at any given FAR within a practical range (0–10%), the SARIMA methods yield lower detection delays than EARS W2c.

Conclusions

Forecasting approaches that incorporate long-term data trends may improve the performance of surveillance systems, at least for diseases that exhibit strong seasonality.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

1 DeLisle S, South B, Anthony JA, Kalp E, Gundlapallli A, Curriero FC, *et al*. Combining free text and structured electronic medical record entries to detect acute respiratory infections. *PLoS ONE* 2010;**5:** e13377.
2 Tokars JI, Burkom H, Xing J, English R, Bloom S, Cox K, *et al*. Enhancing time-series detection algorithms for automated biosurveillance. *Emerg Infect Dis* 2009;**15:**533–9.
3 Williamson GD, Weatherby Hudson G. A monitoring system for detecting aberrations in public health surveillance reports. *Stat Med* 1999;**18:**3283–98.