Correlation between Alerts Generated from Electronic Medical Record (EMR) Data Sources and Traditional Data Sources

Michael W. Thompson, Ph.D.
The Johns Hopkins University Applied Physics Laboratory

OBJECTIVE

The purpose of this study is to investigate the use of electronic medical record (EMR) data sources to improve the detection performance of a syndromic surveillance system. This analysis involves examining the temporal correlation between alerts generated from the EMR data sources and alerts generated from the more traditional data sources already being used by the surveillance system.

BACKGROUND

ESSENCE [1] is an electronic syndromic surveillance system that automatically processes data from multiple sources to assist public health officials in monitoring for disease outbreaks. Historically, these data sources have included *Call-Center Calls*, *Civilian Office Visits*, *Elderly Office Visits*, *Emergency Room Visits*, *Military Outpatient Visits*, and *Over-the-Counter Sales* from two different retail chains. The daily counts from each source are subtotaled by syndrome, including *Botulism-Like*, *Fever*, *Gastrointestinal*, *Hemorrhagic Illness*, *Localized Lesion*, *Lymphadenitis*, *Neurological*, *Rash*, *Respiratory*, and *Severe Illness or Death*. These time series of daily counts are processed by a detection algorithm that automatically signals an alert whenever an observed count is considered too large to have occurred by chance.

METHODS

In order to assess the potential advantage of using the EMR data sources in ESSENCE, the temporal correlation between the alerts generated from the EMR data sources and the alerts generated from the traditional ESSENCE data sources is examined.

The EMR dataset used in this study contains daily counts from five data sources from a large metropolitan area during a single recent calendar year. These sources are *Encounters*, *Encounters with Lab Orders*, *Encounters with Prescription Orders*, *Encounters with Lab and Prescription Orders*, and *Lab Orders*. The daily counts from each of these sources are subtotaled by the ten syndromes listed in the previous section.

The corresponding ESSENCE dataset contains daily counts from the seven data sources listed in the previous section, from the same metropolitan area, during the same time period, and subtotaled by the same ten syndromes.

Each time series of daily counts is processed by the ESSENCE detector, and the number of alerts occurring during a sliding seven-day window (the *alert count*) is calculated for each series.

For each syndrome, the Spearman rank correlation coefficient is calculated between each EMR alert-count series and each ESSENCE alert-count series. The p-value associated with each correlation coefficient is also calculated based on 999 Monte Carlo simulations involving random temporal permutations of the 365 daily alert flags in each series.

RESULTS

Of the 350 correlation coefficients calculated, fourteen are significant ($p \le 0.01$). Among these fourteen results, certain data sources and syndromes occur frequently. The EMR data sources that exhibit significant correlation over all ESSENCE data sources and syndromes are *Encounter*s and *Encounters with Prescription Orders*. The ESSENCE data sources that exhibit significant correlation over all EMR data sources and syndromes are *Over-the-Counter Sales* and *Civilian Office Visits*. The syndromes that exhibit significant correlation over all EMR and ESSENCE data sources are *Respiratory*, *Rash*, and *Fever*.

CONCLUSIONS

Alerts generated from the EMR data sources are significantly correlated with alerts generated from the traditional data sources used in ESSENCE. Because of this, it might be possible to improve the detection performance of a syndromic surveillance system, such as ESSENCE, by including EMR data sources.

REFERENCES

[1] J. Lombardo, H. Burkom, E. Elbert, S. Magruder, S. H. Lewis, W. Loschen, J. Sari, C. Sniegoski, R. Wojcik, and J. Pavlin, "A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II)," J. Urban Health **80**, Suppl. 1, i32–i42 (2003).