Linking Public Health and Healthcare Data for Syndromic Surveillance

Z. Mnatsakanyan, PhD[1], P. Cutchis, MD, MS[1], S. Babin, MD, PhD[1], L. Hung, MPH[1], D. Mollura, MD[2], *J. Lombardo, MS[1]

[1]The Johns Hopkins University Applied Physics Laboratory
[2]Johns Hopkins University School of Medicine

INTRODUCTION

A number of syndromic surveillance systems include tools that quickly identify potentially large disease outbreak events [1–4]. However, the high false-positive rate continues to be a problem [4–6] in all of these systems. Our earlier work has showed that multi-source information fusion can improve specificity of the syndromic surveillance systems[7]. However, an anomalous health event that presents as only a few cases may remain undetected because the chief complaint data does not contain enough details. New linked data sources need to be used to enhance detection capabilities. The focus of this project examining the incorporation of laboratory, prescription medications and radiology data linked to the patient encounter within syndromic surveillance systems. These data source linkings may enhance the sensitivity of syndromic surveillance.

METHOD

Selected linked data sources, such as laboratory requests and results, radiology requests, and prescription medications were associated with the individual patient encounter number. Our approach included multi-step information processing. First linked data was processed on the individual patient level and then a severity score was assigned for each patient encounter. The Bayesian Network (BN) technique was chosen to estimate severity scores. Severity scores should be calculated daily. Second, temporal anomaly detection algorithms detected anomalies within scored patients counts for each of the age groups such as infants (0-4 years), children (5-17), adults (18-64), and elderly (65+). Third, detected anomalies fused with the population level BN to detect outbreak.

RESULTS

Analysis of the data showed correlations among different data sources. Females and adults in the 45-54 year age group showed the highest frequency among the prescription and the patient encounter data source. In the laboratory dataset, the 65+ year age group showed the highest frequency. A consistent seasonal trend, as well as a day-of-week effect, were also observed in the prescription, encounter, and laboratory data sources. Laboratory-confirmed flu data and influenza-like illnesses (ILI) were obtained from the CDC Influenza website http://www.cdc.gov/flu/ (accessed on 12/18/06), to make comparisons with prescription and patient encounter data. Results from the prescription data source showed that penicillin and cold/cough medication corresponded well with the flu epidemic curve. Similarly in the patient encounter dataset, same day office visits corresponded well with the flu epidemic curve.

CONCLUSION

Preliminary analysis showed that syndromic groups can be made more specific for diseases of interest by utilizing the information available in links. Monitoring only cases that have such linkings or that have an increased severity index should remove much noisy data that are not related to the health events of the user's interest. In this study, we showed that use of linked data improves detection of events with small numbers of cases and may provide earlier detection of large outbreaks while the numbers are still small.

REFERENCES

1. Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. *J Biomed Inform.* 2005;**38**:99–113. doi: 10.1016/j.jbi.2004.11.007.

2. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med.* 2005;**2**:e59. doi: 10.1371/journal.pmed.0020059.

3. Wong W-K, Moore A, Cooper G, Wagner M. Rule-based anomaly pattern detection for detecting disease outbreaks. In: AAAI-02; Edmonton, Alberta: AAAI Press/The MIT Press;2002:217--23.

4. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. PNAS 2003;100:1961--5.

5. Marsden-Haug, Nicola, V. Foster, S. Hakre, A. Anderson, J. Pavlin, "Evaluation of Joint Services Installation Pilot Project and BioNet Syndromic Surveillance Systems --- United States," 2004: *MMWR* August 26, 2005 / 54(Suppl);194.

6. Bravata DM, McDonald KM, Smith WM, et al. A critical evaluation of existing surveillance systems for illnesses and syndromes potentially related to bioterrorism. *Ann Intern Med 2004*;140:910--22.

7. Z. Mnatsakanyan, S. Murphy, R. Ashar, H. Burkom: Hybrid Probabilistic Modeling and Automated Data Fusion for Biosurveillance Applications. PHIN 2006 Atlanta, GA.