Earliest Event Detection: A Unique Approach Combining Syndromic Surveillance with Data Mining

Christopher S. Taylor, Bioterrorism Surveillance Program Manager, Division of Public Health Services, Department of Health and Human Services, State of New Hampshire, 29 Hazen Drive, Concord, NH 03301-6504, (603) 271-4987, ctaylor@dhhs.state.nh.us

Learning Objectives:
1. Understand the principles behind the use of syndromic surveillance and data mining.
2. Understand how New Hampshire's unique approach combining data mining with syndromic surveillance has enhanced disease surveillance efforts.
3. Describe the steps and code necessary to implement and enhance data mining.

Background: Syndromic surveillance is focused upon organizing data into categories to detect medium to large scale clusters of illness. Detection often requires that a critical threshold be surpassed [1]. Data mining searches through data to identify records containing keywords [2]. New Hampshire has combined data mining with syndromic surveillance since January 2003 to improve detection capacity.

Method: Data mining is conducted initially against free text data to allow for rapid detection. When ICD-9-CM diagnosis codes are available, the same data mining technique is used to identify confirmed cases of disease. Syndromic alerts are used to identify medium to large scale clusters of disease activity that may otherwise be missed.

MS Access queries and Oracle stored procedures are used in New Hampshire for both data mining, and calculating syndromic alerts. SAS, SPSS, MySQL, PostgreSQL or any of a number of readily available commercial and free software packages can be used to duplicate the approaches used in New Hampshire.

Results: Individual cases and small clusters of hemolytic uremic syndrome, encephalitis, pertussis, varicella, hepatitis and tuberculosis were detected through data mining of both free-text and ICD-9-CM data. Most of these cases would never have been identified by syndromic alerts.

Syndromic alerts have identified moderate to large clusters of pneumonia, respiratory, and gastrointestinal activity, as well as public health interventions.

Enhanced detection capacity has successfully reduced the risk of disease spread, and accelerated disease control interventions.

Conclusions: New Hampshire's unique approach of combining data mining with syndromic surveillance has greatly enhanced disease surveillance detection capacity. This concept can be widely adopted by public health departments across the country using readily available software. The most basic implementation can be rapidly implemented and yield immediate results.

References:
[1] Source: http://en.wikipedia.org/wiki/Syndromic_surveillance. Visited: 12 July 2007
[2] Source: http://en.wikipedia.org/wiki/Text_mining Visited: 12 July 2007