Using the Emergency Medical Text Processor to Standardize Chief Complaints in Boston's Syndromic Surveillance System

Michael Donovan, Julia E. Gunn, RN, MPH
Boston Public Health Commission

OBJECTIVE

To describe the implementation and baseline evaluation of Emergency Medical Text Processor (EMT-P) developed by the University of North Carolina.

BACKGROUND

Abbreviation, misspellings, and site specific terminology may misclassify chief complaints syndromes. EMT-P is system that cleans emergency department chief complaints and returns standard terms. However, little information is available on the implementation of EMT-P in a syndromic surveillance system.

METHODS

The Boston Public Health Commission's (BPHC) syndromic surveillance system receives daily, de-identified patient-visit data from 10 hospital emergency departments in the city of Boston. Data is received every 24 hours, and the chief complaint for each visit is categorized into non-exclusive syndrome groups. Syndromes are stratified by age groups, zip codes, hospital, and gender, and analyzed for statistical aberrations in the number of daily visits using the Early Aberration Reporting System (EARSv3).

EMT-P is a natural language processing program that corrects non-standard chief complaint terms and then maps cleaned chief complaints to standardized terms from the Unified Medical Language System (UMLS) Metathesaurus. Known issues with EMT-P include one to many return for a chief complaint with more than one concept in UMLS and stray chief complaint segments (c/c cough returned as cough and c).

A system upgrade to EARS v4.5 provided the opportunity deploy EMT-P to maximize the standardization and processing of chief complaints. A business process analysis was used to develop the project plan and identify the information flow and procedures for the use of the EMT-P system. A random sample of 49,792 visits was selected to evaluate EMT-P.

RESULTS

BPHC syndromic system is running EMT-P on a windows 2003 server with a SQL Server 2005 backend database. In order to use s same database for the LVG/UMLS datasets, changes in the configuration file were made to the JDBC connector for EMT-P and UMLS import/install processes. Because BPHC opted to use the 2007AA UMLS data set, changes to the structure of the data were required. The string fields in the 2007 UMLS data are 65k bytes compared to 255 (text vs tinytext) in the older version. In addition, a set of clustered and non-clustered indexes was created to improve the query performance.

The EMT-P system is structured to run daily as a SQL server job. The package runs parallel and serial tasks and is divided into a set of work units. Tasks include chief complaint pre-process, EMT-P process and chief complaint post-process. The EMT-P component was modified with addition of a new directory which functions as a staging area for the process. The pre-processing is used to standardize local chief complaint terms that are not recognized by EMT-P system (pna = pneumonia). A correction process for visits with a one to many return was developed that includes concatenating the returned chief complaints back into a single chief complaint field and deduplicating terms appearing more than once. Post processing corrects for errors in the return. For example, the term "od" is always returned as overdose. A chief complaint of "od keratitis" is returned as "overdose keratitis" instead of "right eye keratitis". Post processing is used to correct the error with "overdose kerititis" returned as "eye keratitis".

There were 20,129 unique chief complaints reported for 49,792 visits. EMT-P without the one to many correction returned 61,725 records. EMT-P with the BPHC modifications returned 49,792 visits with a 27% (n=14, 619) decrease in unique chief complaints. There were 1,125 visits with "fever" reported as the chief complaint. EMT-P returned 1,209 visits with "fever" by standardizing terms such as "fver", fevr, "temp" and "fever`". The EMT-P return for concept terms related to chest pain included "chest pain", "pain chest", and "chest arm pain".

CONCLUSIONS

Standardizing chief complaint without changing hospital practices can improve data accuracy and system performance. EMT-P decreases the need to capture of every misspelling, acronym, and regional variant. However, the development of a syndrome will need to consider the various concept terms that are associated with a syndrome. Although free text chief complaint classification can never be 100% accurate, the EMT-P process can facilitate the development of a syndrome in response to a developing public health emergency.

REFERENCES

Further Information:
EMT-P: http://nursing.unc.edu/research/current/emtp/