Validation of Emergency Department and Outpatient Data Using ILI syndrome classifiers

Jenna Iberg Johnson*[2] and Komal Brown[1]

[1]LPHI, New Orleans, LA, USA; [2]Office of Public Health, New Orleans, LA, USA

Objective

The goal of this analysis is to compare the results of influenza-like-illness (ILI) text and International Classification of Diseases (ICD) code classifiers applied to the Louisiana Office of Public Health's (OPH) syndromic surveillance data reported by New Orleans area emergency departments and the Greater New Orleans Health Information Exchange's (GNOHIE) data reported by New Orleans area outpatient clinics.

Introduction

The Louisiana Office of Public Health conducts ED syndromic surveillance using the Louisiana Early Event Detection System (LEEDS). Using outpatient data for syndromic surveillance is a relatively new concept, brought about due to the increasing use of EHRs and HIEs making such data readily available. Previously, there has been limited means of syndrome classification validation for the LEEDS data and the GNOHIE data has not been studied widely as a population sample, so this analysis and comparison is valuable on both fronts. Due to differences in the types of data (ADT messages from EDs and CCD from outpatient clinics), as well as different patient populations and site visit capability, the percentages of patients classified as ILI between data sets are unequal. The main focus of this analysis is determining whether the ILI classifiers applied to both data sets detect similar syndrome trends.

Each indicator used in the study represents the percentage of total patients seen that week who are classified as ILI cases. The study period covered the 13-14 influenza season, CDC week 1340 through 1420 (9/29/2013-5/17/2014). Two ILI classifiers were applied to both the GNOHIE and LEEDS data:the first classifier consisted of ICD-9 influenza codes and the second classifier consisted of keywords applied to encounter notes(GNOHIE data) and chief complaint, admit reason and diagnoses (LEEDS data). A graph of the data, below, shows the four data sets.

Methods

In order to compare data trends, the first step of analysis is basic correlation coefficients, for overall data comparison. Then lagged cross-correlation was applied to examine if the two data sources recognize rises and falls in syndrome trends during the same time period. The EARS C2 method was applied to examine simultaneous alerts between data sets. Lastly, the trend of the differences between the data sets was analyzed by scaling the GNOHIE data to have the same mean and standard deviation as the LEEDS data, and creating control charts of the difference data.

Results

The correlation coefficients for the GNOHIE data compared with the LEEDS data, are shown in the table below.

EARS C2 method generated alarms for weeks 1349 through 1401 for LEEDS ICD9, GNOHIE ICD9, and LEEDS text classifiers and for weeks 1350 through 1401 for GNOHIE text classifier.

Using lagged cross correlation, evidence that the two data sets correspond to one another by time of fluctuation would be a maximized correlation value at t=0. Both data sets displayed maximum correlation at t=0.

A control chart was created of the value by week of LEEDS percentage of ILI minus the scaled value of GNOHIE percentage of ILI, for both data sets, where the GNOHIE mean and standard deviation were adjusted to match the LEEDS data. For the ICD-9 data, only two observations at weeks 1351 and 1401 were below and above the 3 standard deviation limits, respectively. For text data, one outlier existed below the limit at week 1420.

Conclusions

The GNOHIE data and LEEDS data show similar trends in timing and severity of seasonal ILI fluctuations. The correlation coefficients indicated strong correlations, the EARS C2 method and lagged cross correlation show a consistency in timeliness of aberration detection, and the control charts maintained a moderate level of statistical control. Overall this indicates that the GNOHIE clinical outpatient data is a good candidate to supplement the LEEDS emergency department syndromic surveillance data. This is important not just for ILI surveillance but for surveillance of other syndromes with a lower incidence rate that may need larger sample sizes for aberration detection. As well, clinical data is becoming more readily available with emergent health information technologies, so outpatient encounter data will be widely available for study in the coming years.

Data Type	Pearson's Coefficient	Spearman's Rho
ICD-9 Code	0.912	0.856
Text Filter	0.754	0.632
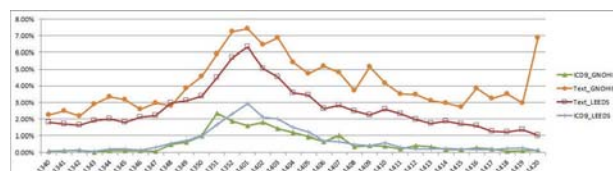
Keywords

influenza-like illness; surveillance; outpatient data

*Jenna Iberg Johnson
E-mail: jenna.ibergjohnson@la.gov