ISDS 2018 Conference Abstracts

Free-Text Mining to Improve Syndrome Definition Matching Across Emergency Departments

Kristin Arkin*[1, 2]

[1]Centers for Disease Control and Prevention, Atlanta, GA, USA; [2]Idaho Division of Public Health, Boise, ID, USA

Objective

We sought to use free text mining tools to improve emergency department (ED) chief complaint and discharge diagnosis data syndrome definition matching across facilities with differing robustness of data in the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) application in Idaho's syndromic surveillance system.

Introduction

Standard syndrome definitions for ED visits in ESSENCE rely on chief complaints. Visits with more words in the chief complaint field are more likely to match syndrome definitions. While using ESSENCE, we observed geographic differences in chief complaint length, apparently related to differences in electronic health record (EHR) systems, which resulted in disparate syndrome matching across Idaho regions. We hypothesized that chief complaint and diagnosis code co-occurrence among ED visits to facilities with long chief complaints could help identify terms that would improve syndrome match among facilities with short chief complaints.

Methods

The ESSENCE-defined influenza-like illness (ILI) chief complaint syndrome was used as the base syndrome for this analysis. Syndrome-matched visits were defined as visits that match the syndrome definition.

We assessed chief complaints and diagnosis code co-occurrence of syndrome-matched visits using the RCRAN TidyText package and developed a bigram network from normalized, concatenated chief complaint and diagnosis code (CCDD) fields and normalized diagnosis code (DD) fields per previously described methodologies.[1] Common connections were defined by a natural break in frequency of pair occurrence for CCDD pairs (30 occurrences) and DD pairs (5 occurrences).

The ESSENCE syndrome was revised by adding relevant bigram network clusters and logic operators. We compared time series of the percent of ED visits matched to the ESSENCE syndrome with those matched to the revised syndrome. We stratified the time series by facilities grouped by short (average < 4 words, "Group A") and long (average ≥ 4 words, "Group B") chief complaint fields (Figure 1). Influenza season start was defined as two consecutive weeks above baseline, or the 95% upper confidence limit of percent syndrome-matched visits outside of the CDC ILI surveillance season. Season trends and influenza-related deaths in Idaho residents were compared.

Results

During August 1, 2016 through July 31, 2017, 1,587 (1.17%) of 135,789 ED visits matched the ESSENCE syndrome. Bigram networks of CCDD fields produced clusters already included by the ESSENCE syndrome. The bigram network of DD fields (Figure 2) produced six clusters. The revised syndrome definition included the ESSENCE syndrome, 3 single DD terms, and 3 two DD terms combined. The start of influenza season was identified as the same week for both ILI syndrome definitions (ESSENCE baseline 0.70%; revised baseline 2.21%). The ESSENCE syndrome indicated the season peaked during Morbidity and Mortality Weekly Report (MMWR) week 2017-05 with the season ending MMWR week 2017-14. The revised syndrome indicated 2017-20 as the season end. Multiple peaks seen with the revised syndrome during MMWR weeks 2017-02, 2017-05, and 2017-10 mirrored peaks in influenza-related deaths during MMWR weeks 2017-03, 2017-06, and 2017-11.

ILI season onset was five weeks earlier with the revised syndrome compared with the ESSENCE syndrome in Group A facilities, but remained the same in Group B. The annual percentage of ED visits related to ILI was more uniform between facility groups under the revised syndrome than the ESSENCE syndrome. Unlike the trend seen with the ESSENCE syndrome, the revised syndrome shows low-level ILI activity in both groups year-round.

Conclusions

In Idaho, dramatic differences in ED visit chief complaint word counts were seen between facilities; bigram networks were found to be an important tool to identify diagnosis codes and logical operators that built more inclusive syndrome definitions when added to an existing chief complaint syndrome. Bigram networks may aid understanding the relationship between chief complaints and diagnosis codes in syndrome-matched visits.

Use of trade names and commercial sources is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service, or the U.S. Department of Health and Human Services.

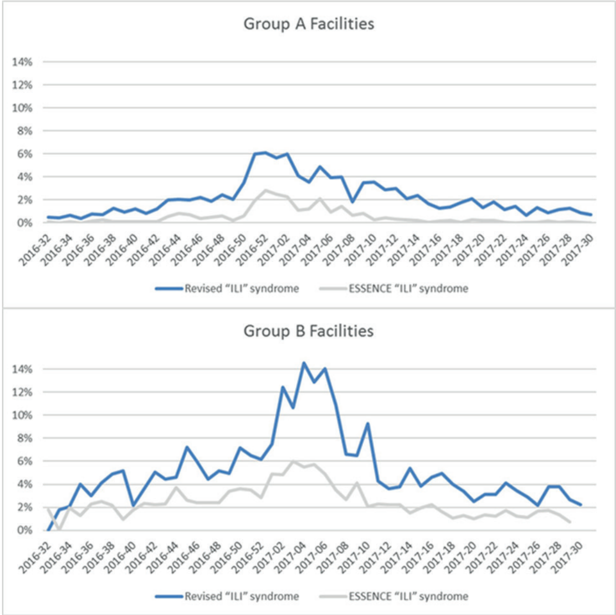

Figure 1. Percent of influenza-like illness-related emergency department visits by MMWR week for the original ESSENCE syndrome (grey) and revised syndrome (blue) grouped by facilities with short (top) and long (bottom) chief complaint fields.

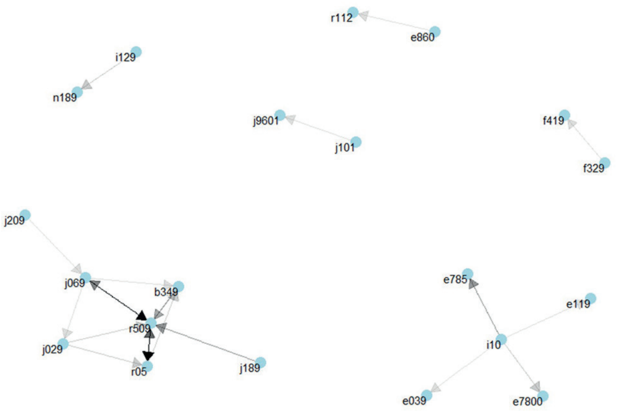

Figure 2. A bigram network displaying common diagnosis code pairs for emergency department visits matched to the ESSENCE influenza-like illness syndrome.

Keywords

Syndromic; Syndrome Definition; Free-text Mining; ESSENCE; ILI

References

1. Silge, J., Robinson, D. (2017). “Text Mining with R”. O’Reilly.

***Kristin Arkin**
E-mail: Kristinaarkin@gmail.com