Advanced Visualization and Analysis of Data Quality for Syndromic Surveillance Systems

Brian E. Dixon*[1, 2, 3], Chen Wen[2], Tony French[2], Jennifer Williams[2] and Shaun J. Grannis[4, 2]

[1]Epidemiology, Indiana University Fairbanks School of Public Health, Indianapolis, IN, USA; [2]Regenstrief Insitute, Indianapolis, IN, USA; [3]Roudebush VA Medical Center, Health Services Research & Development Service, Indianapolis, IN, USA; [4]Indiana University School of Medicine, Indianapolis, IN, USA

Objective

To extend an open source analytics and visualization platform for measuring the quality of electronic health data transmitted to syndromic surveillance systems.

Introduction

Effective clinical and public health practice in the twenty-first century requires access to data from an increasing array of information systems. However, the quality of data in these systems can be poor or "unfit for use." Therefore measuring and monitoring data quality is an essential activity for clinical and public health professionals as well as researchers[1]. Current methods for examining data quality largely rely on manual queries and processes conducted by epidemiologists. Better, automated tools for examining data quality are desired by the surveillance community.

Methods

Using the existing, open-source platform Atlas developed by the Observational Health Data Sciences and Informatics collaborative (OHDSI; www.ohdsi.org), we added new functionality to measure and visualize the quality of data electronically reported from disparate information systems. Our extensions focused on analysis of data reported electronically to public health agencies for disease surveillance. Specifically, we created methods for examining the completeness and timeliness of data reported as well as the information entropy of the data within syndromic surveillance messages sent from emergency department information systems.

Results

To date we transformed 111 million syndromic surveillance message segments pertaining to 16.4 million emergency department encounters representing 6 million patients into the OHDSI common data model. We further measured completeness, timeliness and entropy of the syndromic surveillance data. In Figure-1, the OHDSI tool Atlas summarizes the analysis of data completeness for key fields in over one million syndromic surveillance messages sent to Indiana's health department in 2014. Completeness is reported by age category (e.g., 0-10, 20-30, 60+). Gender is generally complete, but both race and ethnicity fields are often complete for less than half of the patients in the cohort. These results suggest areas for improvement with respect to data quality that could be actionable by the syndromic surveillance coordinator at the state health department.

Conclusions

Our project remains a work-in-progress. While functions that assess completeness, timeliness and entropy are complete, there may be other functions important to public health that need to be developed. We are currently soliciting feedback from syndromic surveillance stakeholders to gather ideas for what other functions would be useful to epidemiologists. Suggestions could be developed into functions over the next year. We are further working with the OHDSI collaborative to distribute the Atlas enhancements to other platforms, including the National Syndromic Surveillance Platform (NSSP). Our goal is to enable epidemiologists to quickly analyze data quality at scale.

Data completeness for key syndromic surveillance fields summarized across 1 million emergency department messages transmitted in 2014.

Keywords

Data quality; Syndromic surveillance; Health information exchange

References

1. Dixon BE, Rosenman M, Xia Y, Grannis SJ. A vision for the systematic monitoring and improvement of the quality of electronic health data. Studies in health technology and informatics. 2013;192:884-8.

***Brian E. Dixon**
E-mail: bedixon@regenstrief.org