

Early effect of validation efforts of Massachusetts syndromic surveillance data

Mark Bova* and Rosa Ergas

Massachusetts Department of Public Health, Boston, MA, USA

Objective

To develop a detailed data validation strategy for facilities sending emergency department data to the Massachusetts Syndromic Surveillance program and to evaluate the validation strategy by comparing data quality metrics before and after implementation of the strategy.

Introduction

As a participant in the National Syndromic Surveillance Program (NSSP), the Massachusetts Department of Public Health (MDPH) has worked closely with our statewide Health Information Exchange (HIE) and National Syndromic Surveillance Program (NSSP) technical staff to collect and transmit emergency department (ED) data from eligible hospitals (EHs) to the NSSP. Our goal is to ensure complete and accurate data using a multi-step process beginning with pre-production data and continuing after EHs are sending live data to production.

Methods

We used an iterative process to establish a framework for monitoring data quality during onboarding of EHs into our syndromic surveillance system and kept notes of the process.

To evaluate the framework, we compared data received during the month of January 2016 to the most recent full month of data (June 2016) to describe the following primary data quality metrics and their change over time: total and daily average of message and visit volume; percent of visits with a chief complaint or diagnosis code received in the NSSP dataset; and percentage of visits with a chief complaint/diagnosis code received within a specified time of admission to the ED.

Results

The strategies for validation we found effective included examination of pre-production test HL7 messages and the execution of R scripts for validation of live data in the staging and production environments. Both the staging and production validations are performed at the individual message level as well as the aggregated visit level, and included measures of completeness for required fields (Chief Complaint, Diagnosis Codes, Discharge Dispositions), timeliness, examples of text fields (Chief Complaint and Triage Notes), and demographic information. We required EHs to pass validation in the staging environment before granting access to send data to the production environment.

From January to June 2016, the number of EHs sending data to the production environment increased from 44 to 48, and the number of messages and visits captured in the production environment increased substantially (see Table 1). The percentage of visits with a chief complaint remained consistently high (>99%); however the percentage of visits with a chief complaint within three hours of admission decreased during the study period. Both the overall percentage of visits with a diagnosis code and the percentage of visits with a diagnosis code within 24 hours of admission increased.

Conclusions

From January to June 2016, Massachusetts syndromic surveillance data improved in the percentage of visits with diagnosis codes and the time from admission to first diagnosis code. This was achieved while the volume of data coming into the system increased. The timeliness of chief complaints decreased slightly during the study period, which may be due to the inclusion of several new facilities that are unable to send real-time data. Even with the improvements in the timeliness of the diagnosis code field, and the subsequent decrease in the timeliness of the chief complaint field, chief complaints remained a more timely option for syndromic surveillance. Pre-production and ongoing data quality assurance activities are crucial to ensure meaningful data are acquired for secondary analyses. We found that reviewing test HL7 messages and staging data, daily monitoring of production data for key factors such as message volume and percent of visits with a diagnosis code, and monthly full validation in the production environment were and will continue to be essential to ensure ongoing data integrity.

Table 1: ED Data in the Production Environment

Time period	January 2016	June 2016	Percent Change
Number of EDs	44	48	+9.1%
Messages (Daily Mean)	1,457,988 (47,031.7)	1,954,261 (65,142)	+38.5%
Visits (Daily Mean)	176,246 (5,685.4)	190,682 (6,356.1)	+11.8%
Visits with Chief Complaint	99.32%	99.45%	+0.13%
Visits with Chief Complaint within 3 hours	35.9%	30.2%	-5.7%
Visits with Diagnosis Code	72.96%	78.04%	+5.08%
Visits with Diagnosis Code within 24 hours	27.3%	42.3%	+15%

Keywords

Syndromic Surveillance; Data Quality; Validation; R Studio

*Mark Bova

E-mail: mark.bova@state.ma.us

