OJPHI

ISDS 2014 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Validation of New Jersey Emergency Department (ED) Registration Data in BioSense 2.0

Feiyan Chen[1], Elizabeth Kostial[2], Teresa Hamby[1] and Stella Tsai*[1]

[1]New Jersey Department of Health, Trenton, NJ, USA; [2]Health Monitoring Systems, Inc, Pittsburgh, PA, USA

Objective

To assess and validate New Jersey's ED registration data feed from EpiCenter to BioSense 2.0.

Introduction

BioSense 2.0, a redesigned national syndromic surveillance system, provides users with timely regional and national data classified into disease syndromes, with views of health outcomes and trends for use in situational awareness. As of July 2014, there are 60 jurisdictions nationwide feeding data into BioSense 2.0. In New Jersey, the state's syndromic surveillance system, EpiCenter, receives registration data from 75 of NJ's 80 acute care and satellite emergency departments. EpiCenter is a system developed by Health Monitoring Systems, Inc. (HMS) that incorporates statistical management and analytical techniques to process health-related data in real time. To participate in BioSense 2.0, New Jersey worked with HMS to connect existing data to BioSense. In May, 2013, HMS established a single data feed of New Jersey's facility data to BioSense 2.0. This transfer from HMS servers occurs twice daily via SFTP. The average daily visit volume in the transfer is around 10,000 records. This data validation project was initiated by the New Jersey Department of Health (NJDOH) in 2013 to assure that the registration records are delivered successfully to BioSense 2.0.

Methods

For this assessment, NJDOH searches and exports weekly ED visit counts by facility and date from EpiCenter using built-in export functions and from BioSense 2.0 via SQL query scripts and then compares them using the data validation tools developed by SAS. Figure1 shows the procedure of data validation.

A summary report of this comparison is generated by SAS ODS, which includes total number of facilities feeding data, the list of problematic facilities not reporting data, and unmatched visit counts by facility and date between EpiCenter and BioSense 2.0. This SAS tool imports, cleans and manipulates these two data sets exported from EpiCenter and BioSense 2.0 using the merge and subgroup functions, and applies the SQL procedure to obtain the timeframe of the validation in the final report. A SAS macro automatically generates a directory to store the report file.

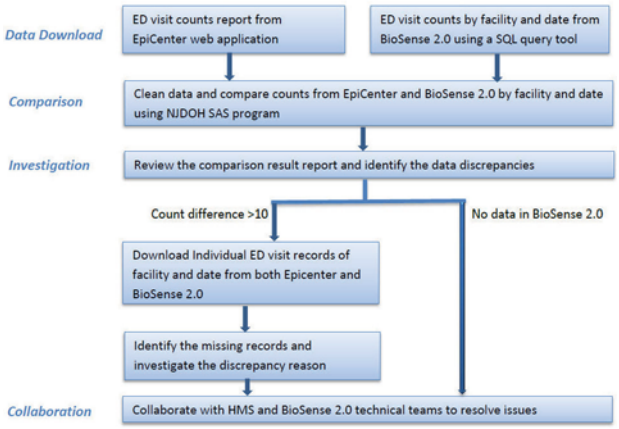
Results

Using the validation tools described, project staff investigate, identify and resolve the data issues. For data discrepancies (count difference > 10) for a specific facility on a specific date, the missing records are identified to compare the individual ED visit records exported from EpiCenter to those from BioSense 2.0. A common reason for discrepancy is programmatic. EpiCenter is able to process messages without a visit number while BioSense cannot do so. Other reasons for discrepancies found during validation include: missing data in BioSense 2.0 due to processing issues upon receipt, messages from newly added facilities where IDs have not yet been fully processed by HMS, NJDOH, and BioSense, an identifier is changed by a facility in the message that is then unrecognized, facilities delay sending records for some period after visit date. In addition, a facility may stop sending data due to system maintenance or upgrade resulting in gaps in data. Where applicable for these discrepancies, HMS resends the missing data to BioSense 2.0.

Conclusions

The data validation tools and procedures used by NJDOH are useful to assess the ED registration data sent to the back end of BioSense 2.0. Most missing data and discrepancies can be detected by these tools. NJDOH and HMS will continue to improve the tools to validate the data between the back end and the front end of BioSense 2.0 based on the knowledge of how data is processed in BioSense 2.0. In addition, work on developing data quality tools to evaluate and report the data status will continue.

Figure 1: Data Validation Process



Keywords

Data validation; Data quality, Data assessment; SAS, SQL; EpiCenter, BioSense; New Jersey

*Stella Tsai
E-mail: stella.tsai@doh.state.nj.us

ISDS Annual Conference Proceedings 2014. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.