OJPHI

ISDS 2015 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Alert-Enabled Application Integrating Data Quality Monitoring for Multiple Sources

Harold Gil* and Nicholas L. Michaud

Marion County Public Health, Indianapolis, IN, USA

Objective

An interactive, point-and-click application was developed to facilitate the routine assessment of known data quality factors that compromise the integrity and timeliness of data sets used at the Marion County Public Health Department (MCPHD). The code (and associated documentation) for this application is being made available for other surveillance practitioners to adopt.

Introduction

Data sets from disparate sources widely vary in the number and type of factors which most hamper integrity and timeliness of the data. To maintain high quality data, data sets must be regularly assessed, particularly for those vulnerabilities that each is especially prone to due to the methods involved in collecting the data. For surveillance practitioners charged with monitoring data from multiple data sources, keeping track of the issues that each data set is susceptible to, and quickly identifying any inconsistencies or deviations from normal trends, may be a challenge. An application that can track all those issues, and trigger alerts when patterns diverge from what is expected, could help to enhance the efficiency and effectiveness of the surveillance efforts.

Methods

MCPHD data sets that have experienced recurrent and well-defined issues were identified and subsequently prioritized based on the impact to public health surveillance and response efforts. Our electronic lab report data, with its pattern of consistent delays and missing positive cases, was selected as the highest priority. Originally the intention was for our solution to be utilized by personnel whom are not expected to be literate in a programming language; therefore, the potential solution needed to be user-friendly and easy-to-use.

R is a statistical computing application, known for its versatility and ability to create powerful visualizations. Shiny is an R package that facilitates the creation of interactive, easy-to-use point-and-click applications. For end users, Shiny applications eliminate the need to be familiar with the R programming language and retain the ability to harness R's analytic tools. We looked to R and its Shiny package extension as a candidate solution.

Results

The R-Shiny application was developed and is presently in use at MCPHD. The application in its current state allows the user to select 1) the time frame for analysis, 2) the sliding window size used by the analysis as the number of days, and 3) the alert threshold as the number of standard deviations. Upon selection of any of these parameters, the program will automatically recalculate alerts for the selected time frame and plots them as red dots on the disease-specific trend lines. A dynamic table is also displayed in which the user can sort variables (date, disease name, count, alert status).

Conclusions

In the past, MCPHD has had a significant period of time go by without realizing that certain electronic lab reports were not being received. The alert-enabled R-Shiny application that was developed in this project allows us to check report trends on a daily basis in order to confirm whether the expected number of reports is being submitted. Further metrics and visualizations will be added to the application in order to monitor other aspects of the data that have been problematic.

We are next looking to integrate two other MCPHD data sets (HIV and Vital Records) into the application. Based on our initial success, we are also beginning to create an online library of code in order to facilitate the adoption of this R-Shiny tool.

Figure 1

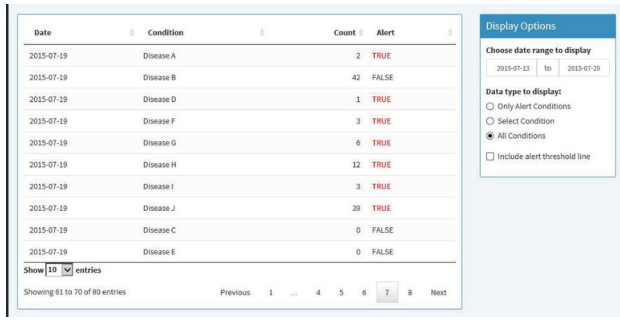

Figure 2

Keywords

Data Quality; R; Shiny

***Harold Gil**
E-mail: hgil@marionhealth.org