Epi Evident: Biosurveillance to Monitor, Compare, and Forecast Disease Case Counts

Natalie Tomaszewski*[1], Meeshu Agnihotri[3], Huiwen Cheng[4], Ashutosh Bhadke[5], Michael Henry[2] and Lauren E. Charles[2]

[1]University of Southern California, Santa Ana, CA, USA; [2]Pacific Northwest National Laboratory, Richland, WA, USA; [3]Georgia Tech University, Atlanta, GA, USA; [4]University of Indiana, Bloomington, IN, USA; [5]University of Arizona, Tucson, AZ, USA

Objective

Epi Evident is a web based application built to empower public health analysts by providing a platform that improves monitoring, comparing, and forecasting case counts and period prevalence of notifiable diseases for any scale jurisdiction at regional, country, or global-level. This proof of concept application development addresses improving visualization, access, situational awareness, and prediction of disease behavior.

Introduction

The Epi Evident application was designed for clear and comprehensive visualization for monitoring, comparing, and forecasting notifiable diseases simultaneously across chosen countries. Epi Evident addresses the taxing analytical evaluation of how diseases behave differently across countries. This application provides a user-friendly platform with easily interpretable analytics which allows analysts to conduct biosurveillance with minimal user tasks. Developed at the Pacific Northwest National Laboratory (PNNL), Epi Evident utilizes time-series disease case count data from the Biosurveillance Ecosystem (BSVE) application Epi Archive (1). This diverse data source is filtered through the flexible Epi Evident workflow for forecast model building designed to integrate any entering combination of country and disease. The application aims to quickly inform analysts of anomalies in disease & location specific behavior and aid in evidence based decision making to help control or prevent disease outbreaks.

Methods

A workflow was constructed to define the best disease forecast model for each location based on an adjustable method approach. The differences in disease behavior across countries was achieved through a React/Python application with a user-friendly output for monitoring and comparing different combinations.

The forecast model building workflow consisted of three major steps to determine the best fit model for a given disease-country pair: data type, model type, and model comparison & selection. Testing various disease-country combinations allowed for direct evaluation of the workflow efficiency, flexibility, and criteria for determining the best fit model. Data type was characterized as either seasonal, cyclic, or sporadic. Depending on data type, a specific time series forecasting model was applied. In general, seasonal or cyclic data required either an Auto-Regression Integrated Moving Average (ARIMA) model or a Seasonal Auto-Regression Integrated Moving Average (SARIMA) model while sporadic datasets employed a Poisson model. Several model candidates for a single country and disease combination were then compared to determine which was the best fit model. ARIMA and SARIMA model selection criteria included their respective order significance, residual diagnostics, and lowest possible combination of Akaike Information Criterion and Root Mean Square Error (RMSE) values. Poisson model selection criteria involved Poisson or negative binomial distribution and event probability, lag dependency of immediate past events or seasonality, and lowest possible RMSE. To enhance the user's monitoring and comparisons across multiple countries and diseases, each forecasted case counts supplied a corresponding period prevalence. This period prevalence was calculated by dividing the case counts by the population in the selected country and timeframe. Population records were obtained through the public World Health Organization database (2).

Results

A variety of visualization tools on Epi Evident allows convenient interpretation on behaviors of diseases spanning multiple countries simultaneously (Figure 1). Countries, diseases, and timeframe are selected and displayed within a matrix alongside with their corresponding forecasts for case counts and period prevalence. By providing this full representation, users can easily interpret and anticipate disease behavior while monitoring, comparing, and forecasting case counts and period prevalence across multiple countries. For future work, the Epi Evident workflow can be scaled to accommodate any disease-country combination with automated model selection to allow easier and more efficient biosurveillance.

Conclusions

Epi Evident empowers analysts to visualize, monitor, compare, and forecast disease case counts and period prevalence across countries. Epi Evident exemplifies how filtering diverse data through a flexible workflow can be scalable to output distinctive models for any given country and disease combination. Thus, providing accurate forecasting and enhanced situational awareness throughout the globe. Implementing this application's methodology helps enhance and expand biosurveillance efficacy for multiple diseases across multiple countries simultaneously.
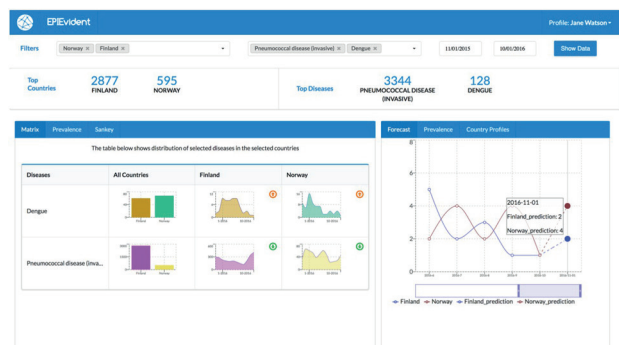

Figure 1. Epi Evident workbench with a toolbar (top) for country, disease, and timeframe selection, descriptive statistics, and matrix display of disease-country pairs (left) and corresponding forecasts (right).

Keywords

Biosurveillance; notifiable diseases; public health; forecast

References

1. Generous Nicholas, Fairchild Geoffrey, Khalsa Hari, Tasseff Byron, Arnold James. Epi Archive: An automated data collection of notifiable disease data. Online Journal of Public Health Informatics. 2017. 9(1):e37

2. http://apps.who.int/gho/data/view.main.POP2040?lang=en Accessed: 6/20/2017

*Natalie Tomaszewski
E-mail: ntomasze@usc.edu