OJPHI

ISDS
INTERNATIONAL SOCIETY
FOR DISEASE SURVEILLANCE

Contextualizing Data Streams for Infectious Disease Surveillance

Kirsten Taylor-McCabe*, Lauren Castro, Nicholas Generous, Kristen Margevicius, Mac Brown and Alina Deshpande

Los Alamos National Laboratory, Los Alamos, NM, USA

Objective

The goal of this project was to provide an approach and evaluation of data stream utility for integrated, global disease surveillance. This effort is part of a larger project which is developing tools to aid decision-makers with timely information to predict, prepare for, and mitigate the spread of disease.

Introduction

Los Alamos National Laboratory (LANL) was tasked with developing methods to determine the relevance of data streams for an integrated global biosurveillance system. We used a novel method of evaluating the effectiveness of data streams called the "surveillance window". We defined a surveillance window as the brief period of time when information gathered can be used to assist decision makers in effectively responding to an impending outbreak. Information obtained for data streams beyond this window is deemed to have limited use.

Methods

We used a stepwise approach to defining disease specific surveillance windows for nine representative historical case studies for FMD, Ebola, Influenza, Citrus Greening, Lassa fever, Malaria, Dengue, Cholera and E.coli. A timeline visualization of the progression of the disease linked with data stream utilization information was developed. These timelines were then used to determine the length of time between changes in "epidemiological state" that defined surveillance window durations. Surveillance windows were then identified and information for applicable data streams was collected for the duration of the outbreak. Traditional and non-traditional data streams were categorized into 16 groups and analyzed for each disease outbreak. A data stream was deemed useful if it was determined to be available within the defined surveillance window. In addition, evaluation of the ideal use case of the data streams was performed. In essence, if used more effectively could this data stream provide greater support to understanding, detection, warning or management of disease outbreaks or event situations?

Results

Using the surveillance window method, comparisons across all case studies could be made. Despite the diverse set of data, clear trends, could be seen in data stream categories that appeared consistently in the pre-rapid rise window as well as in later timeframes. These were ones that relied on specific, local and credible information. Data stream categories that were identified to have high utility for early detection and early warning regardless of disease were binned into three tiers of utility. The top tier consists of mostly traditional and more established surveillance methods such as Laboratory Records, ED/Hospital Records, Clinic/Health Provider, Official Records (Gov) and Personal Communication. The second tier contains more non-traditional and emerging data streams such as; Ambulance/EMT Records, News Aggregators, Social Media and Sales. The last tier consists of a mix of hard to access or less mined data streams such as Internet Search Queries, Financial Records, Employment/School Records, Help Lines, Police/Fire Department Records and Prediction Markets. Also, surveillance window durations and incubation periods of the disease were compared and shown to have little correlation. This suggests that the window of time available for early warning or detection of a specific disease is determined not only by the disease characteristics but also by operational characteristics, location, population and season, endemicity of disease, etc.

Conclusions

This study provides a simple, yet elegant methodology for which to ground truth known and emerging data streams for utility in integrated biosurveillance efforts. Evaluating data streams across multiple diseases provided a robust application of our method as well as allowed for the examination of the influence of disparate disease characteristics not only within human, but across species, on outbreak progression and surveillance window duration.

Surveillance window based evaluation of data streams during disease outbreaks helped identify data streams that are of significance for developing an effective biosurveillance system. Some data streams were identified to have high utility for early detection and early warning regardless of disease, while others were more disease and operations specific. This work also identified data streams currently not in use that could be exploited for faster outbreak detection. Key useful data streams that are underlying to all disease categories and thus important for integration into global biosurveillance programs will be presented here.

Keywords

datastream; global disease survelliance; surveillance window

*Kirsten Taylor-McCabe
E-mail: kjmccab@lanl.gov