

Disease Models for Event Prediction

Courtney D. Corley*¹ and Laura Pullum²

¹Pacific Northwest National Laboratory, Richland, WA, USA; ²Oak Ridge National Laboratory, Oak Ridge, TN, USA

Objective

The objective of this manuscript is to present a systematic review of biosurveillance models that operate on select agents and can forecast the occurrence of a disease event.

Introduction

One of the primary goals of this research was to characterize the viability of biosurveillance models to provide operationally relevant information to decision makers, in order to identify areas for future research. Two critical characteristics differentiate this work from other infectious disease modeling reviews [1,2]. First, we reviewed models that attempted to predict the disease event, not merely its transmission dynamics. Second, we considered models involving pathogens of concern as determined by the US National Select Agent Registry.

Background: A rich and diverse field of infectious disease modeling has emerged over the past 60 years and has advanced our understanding of population- and individual-level disease transmission dynamics, including risk factors, virulence and spatio-temporal patterns of disease spread. Recent modeling advances include biostatistical methods, and massive agent-based population, biophysical, ordinary differential equation, and ecological-niche models. Diverse data sources are being integrated into these models as well, such as demographics, remotely-sensed measurements and imaging, environmental measurements, and surrogate data such as news alerts and social media. Yet, there remains a gap in the sensitivity and specificity of these models not only in tracking infectious disease events but also predicting their occurrence.

Methods

We searched dozens of commercial and government databases and harvested Google search results for eligible models utilizing terms and phrases provided by public health analysts relating to biosurveillance, remote sensing, risk assessments, spatial epidemiology, and ecological niche-modeling. This returned 13,767 webpages and 12,152 citations. After de-duplication and removal of extraneous material, a core collection of 6,503 items was established, these publications and their abstracts are presented in a semantic wiki at <http://BioCat.pnnl.gov>. Next, PNNL's IN-SPIRE visual analytics software was used to cross-correlate these publications with the definition for a biosurveillance model. As a result, we systematically reviewed 44 papers, and the results are presented in this analysis.

Results

The models were classified as one or more of the following types: event forecast (9%), spatial (59%), ecological niche (64%), diagnos-

tic or clinical (14%), spread or response (20%), and reviews (7%). The distribution of transmission modes in the models was: direct contact (55%), vector-borne (34%), water- or soil-borne (16%), and non-specific (7%). The parameters (e.g., etiology, cultural) and data sources (e.g., remote sensing, NGO, epidemiological) for each model were recorded. A highlight of this review is the analysis of verification and validation procedures employed by (and reported for) each model, if any. All models were classified as either a) Verified or Validated (89%), or b) Not Verified or Validated (11%; which for the purposes of this review was considered a standalone category).

Conclusions

The verification and validation (V&V) of these models is discussed in detail. The vast majority of models studied were verified or validated in some form or another, which was a surprising observation made from this portion of the study. We subsequently focused on those models which were not verified or validated in an attempt to identify why this information was missing. One reason may be that the V&V was simply not reported upon within the paper reviewed for those models. A positive observation was the significant use of real epidemiological data to validate the models. Even though 'Validation using Spatially and Temporally Independent Data' was one of the smallest classification groups, validation through the use of actual data versus predicted data represented approximately 33% of these models. We close with initial recommended operational readiness level guidelines, based on established Technology Readiness Level definitions.

Keywords

Disease models; Event prediction; Operational readiness

References

- [1] Lloyd-Smith JO, George DB, Pepin KP, Pitzer VE, Pulliam JRC, Dobson A, Hudson PJ, and Grenfell BT. Epidemic Dynamics at the Human-Animal Interface. *Science* 4 December 2009: 326 (5958), 1362-1367. [DOI:10.1126/science.1177345]
- [2] Bravata DM, Sundaram V, McDonald KM, Smith WM, Szeto H, Schleinitz MD, et al. Detection and diagnostic decision support systems for bioterrorism response. *Emerg Infect Dis* [serial online] 2004 Jan [9-Sept 2012]. <http://wwwnc.cdc.gov/eid/article/10/1/03-0243.htm>

*Courtney D. Corley

E-mail: court@pnnl.gov

