

Predicting Levels of Influenza Incidence

Anna L. Buczak*, Liane Ramac-thomas, Erhan Guven, Yevgeniy Elbert, Steven Babin, Benjamin Baugher and Sheri Lewis

JHU APL, Laurel, MD, USA

Objective

Advanced techniques in data mining and integrating evidence from multiple sources are used to predict levels of influenza incidence several weeks in advance and display results on a map in order to help public health professionals prepare mitigation measures.

Introduction

Influenza epidemics occur seasonally but with spatiotemporal variations in peak incidence. Many modeling studies examine transmission dynamics [1], but relatively few have examined spatiotemporal prediction of future outbreaks [2]. Bootsma et al [3] examined past influenza epidemics and found that the timing of public health interventions strongly affected the morbidity and mortality. Being able to predict when and where high influenza incidence levels will occur before they happen would provide additional lead time for public health professionals to plan mitigation strategies. These predictions are especially valuable to them when the positive predictive value is high and subsequently false positives are infrequent.

Methods

This approach to predictive disease incidence modeling exploits the complicated relationships between disease outbreaks and measurable environmental, biological, and sociological variables. Data were obtained on variables found in previous studies (e.g., [4]) to be related to influenza: ambient temperature, humidity, precipitation, and date of first day of school (which varies annually and by jurisdiction). Epidemiological data included influenza-like illness (ILI) visits and pharmacy thermometer sales for the District of Columbia (DC) and 6 surrounding counties in Maryland and Virginia from the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE), as well as CDC-reported influenza-positive lab tests and pneumonia and influenza mortality. Data from 2003 to present were pre-processed to get one weekly value per jurisdiction. Fuzzy association data mining was applied to these data to extract rules used to build classifiers using the method described in [5]. Those classifiers that met certain performance criteria were then used to create a model for predicting future weekly ILI incidence levels four weeks in advance. A high incidence level was defined as equal to or greater than 1.5 standard deviations above the historical mean ILI incidence. The resulting model then made predictions for each jurisdiction, using weekly data not used for model development in order to avoid exaggerated values of performance. Also, the model only used input data that would actually be available on the date the prediction was made.

Results

Predictions of weekly ILI incidence levels made four weeks in advance were compared with ILI incidence from data not used in model development. Table 1 shows the positive and negative predictive values (PPV and NPV), as well as Sensitivity and Specificity, for predictions made using the test set for all counties and DC combined. High PPV was obtained, although a higher sensitivity is also desired.

Conclusions

Using available data to predict high incidence levels of influenza weeks in advance should facilitate public health interventions to mitigate the impact of the disease. High PPV is of principal importance as health officials may be unlikely to spend resources on mitigation efforts based on model predictions without evidence of accuracy on past outbreaks.

Model prediction of high influenza incidence 4 weeks in advance for all counties and DC combined.

PPV	NPV	Sensitivity	Specificity
0.941	0.935	0.410	0.997

Keywords

Influenza; outbreak prediction; data mining; data source integration

References

- [1] Coburn B, Wagner B, Blower S. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Medicine* 2009, 7:30.
- [2] Nishiura H. Real-time forecasting of an epidemic using a discrete time stochastic model: a case study of pandemic influenza (H1N1-2009). *Bio-Med Eng OnLine* 2011, 10:15.
- [3] Bootsma M, Ferguson N. The effect of public health measures on the 1918 influenza pandemic in US cities. *PNAS* 2007, 104:7588.
- [4] Tamerius J, Shaman J, Alonso W, Bloom-Feshbach K, Uejio C, Comrie A, Viboud C. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathogens* 2013, 9:e1003194.
- [5] Buczak A, Koshute P, Babin S, Feighner B, Lewis S. A data-driven epidemiological prediction method for dengue outbreaks using local and re-mote sensing data. *BMC Med. Inform. Decis. Making* 2012, 12:124.

*Anna L. Buczak

E-mail: anna.buczak@jhuapl.edu

