COACTION

Can we identify 'bellwether' states with respect to syphilis incidence?

Mauricio Monsalve*, Sean Tolentino, Sriram Pemmaraju and Philip Polgreen

Computational Epidemiology Group, The University of Iowa, Iowa City, IA, USA

Objective

(i) To forecast syphilis cases per state in the United States to support early containment of outbreaks. (ii) For each state, to determine which states are most correlated, to find 'bellwether' states to inform surveillance efforts. (iii) To determine a small collection of states whose syphilis incidence patterns are most closely correlated with all the states.

Introduction

The time series of syphilis cases has been studied at the country and state level at the yearly basis (1, 2), and it has been found that syphilis has a periodicity of approximately 10 years (2). However, to inform prevention efforts, it is important to understand the short-term dynamics of disease activity.

Methods

We used data from the MMWR. It contains weekly syphilis counts per state. We consider the time period from 1995 to 2009. We removed week 53 when present, due to inconsistencies in reporting. We considered 53 locations: the 50 states plus Puerto Rico, and the cities of New York City and Washington DC. To predict disease activity in each state, we constructed a series of linear lagged regression models that used several states as covariates. To benchmark our models, we constructed a basic ARIMA model with one autocorrelation term. All the models were constructed to forecast 4 weeks in advance. Prediction at week t was performed by fitting the models using all past data prior to week t–4. To identify bellwether states, we proceeded as follows. First, we repeatedly fitted 2-covariate models to forecast each state and obtained the top 5 most frequent bellwether states for each state. Then, we obtained the most frequent bellwether states from the above lists.

Results

We found that forecasting states using less than 10 states as covariates is better than using more or the state itself as covariate (ARIMA), as shown in Fig. 1. An example of out of sample prediction is shown in Fig. 2, for New York City.

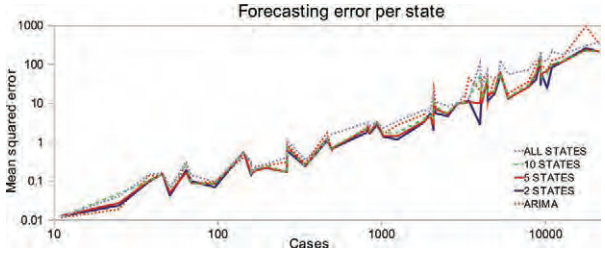

Fig. 1. Mean squared error in out of sample forecasting per number of covariates used. Plotted in log-log scale. States were sorted by their historical number of cases. Solid lines illustrate the methods with the smallest errors (using 2 and 5 states).

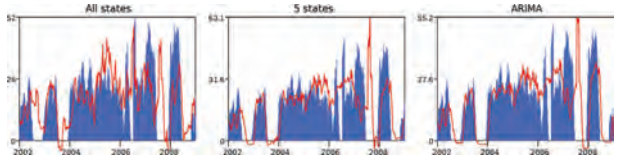

Fig. 2. Forecasting New York City one month in advance.

We also found that the 10 most frequent states in models with two covariates are California, Virginia, Florida, New York City (treated as state), Alabama, Ohio, Tennessee, North Carolina, New Hampshire and New Mexico. The first 5 are covariates of 40 states, and the amount increases to 50 when adding the later.

Conclusions

Using several states as covariates in models seem to improve their forecasting power. This suggests that these models 'learn' the dynamics of syphilis between different states. In addition, we have identified the existence of specific bellwether states. By using these bellwether states, it is possible to forecast syphilis cases in almost all the states in the country.

Several limitations undermine the quality of the predictions. First, cases are counted at reporting time instead of acquisition time (3). Second, the MMWR file supposedly contains cumulative numbers within a year, but this is not always true. Third, some states exhibit strong yearly periodicity, which seems to be due to patterns in disease reporting.

Keywords

Syphilis time series; forecasting; disease surveillance

References

1. Zaidi AA, Schnell DJ, Reynolds GH. Time series analysis of syphilis surveillance data. Stat Med. 1989;8:353–62; discussion 363.
2. Grassly NC, Fraser C, Garnet GP. Host immunity and synchronized epidemics of syphilis across the United States. Nature. 2005;433:417–21.
3. Hook EW 3rd, Richey CM, Leone P, Bolan G, Spalding C, Henry K, et al. Delayed presentation to clinics for sexually transmitted diseases by symptomatic patients. Sex Transm Dis. 1997;24:443–8.

***Mauricio Monsalve**
E-mail: mauricio-monsalve@uiowa.edu

Emerging Health Threats Journal 2011. © 2011 M. Monsalve et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Emerging Health Threats Journal 2011, **4**: 11134 - DOI: 10.3402/ehtj.v4i0.11134