COACTION PUBLISHING

The exploration of various methods for *Shigella* outbreak detection

Xin Jin*, Jian-Hua Chen, Charlene Weng, Bryon Backenson, Dina Hoefer and Hwa-Gan Chang

New York State Department of Health, Albany, NY, USA

Objective

To explore the possibility of using statistical methods to detect Shigella outbreaks, assess the effectiveness of the methods to signal real outbreaks, provide manageable information for follow-up activities and avoid unnecessary surveillance work.

Introduction

Shigella remains highly infectious in the United States, and rapid detection of Shigella outbreaks is crucial for disease control and timely public health actions. The New York State Department of Health (NYSDOH) implemented a Communicable Disease Electronic Surveillance System (CDESS) for local health departments (LHDs) to collect clinical and laboratory testing information and supplement epidemiologic information for the patients from New York State, excluding New York City, with infectious diseases. The CDESS includes reported cases that are involved in outbreaks and which constituted the base for identifying any outbreak. The selection of a fitted outbreak detection method would play a critical role in enhancing disease surveillance.

Methods

Weekly case numbers were obtained from CDESS and counted patients with Shigella who had diagnosis or specimen collection dated between January 1, 2006, and December 31, 2010. Six statistical models were applied to the weekly case numbers in generating signals to identify outbreaks, and signals were compared to the actual outbreak to evaluate their detection powers. Outbreak-related cases from CDESS were removed for the modeling purpose except for the cumulative sum-related methods, which used all cases. The sensitivity (SE), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV) were calculated to evaluate the performance of each method.

General Linear method (GL)

$Yt = a + \Sigma bi\ ct, i, i = 1 \ldots 52$, where Yt is the expected number of cases in week t, ct,*i* is the dummy value which equals 1 if the week of the year for Yt is the same as *i*, else it equals 0.

Poisson method (PO)

It applies the same statistical procedure as GL except for the assumption that the case numbers follow Poisson distribution.

Time Series method (TS)

$Yt = a + bt + c1\sin(2\pi t/52) + c2\sin(4\pi t/52) + c3\sin(6\pi t/52) + d1\cos(2\pi t/52) + d2\cos(4\pi t/52) + d3\cos(6\pi t/52) + at$, where Yt is the expected number of cases in week t, and at is the random error.

A signal was generated when the case number exceeded the 95% confidence limit for the prediction value from the above three methods.

Cumulative Sum method (CuSum)

A signal was created when the case number exceeded the baseline mean, i.e., mean of previous two weeks, plus three standard deviations.

Historical Limit method (HL)

Similar procedures applied as CuSum except that data for the prior 8 weeks of the last year were used as the baseline.

Negative Binomial CuSum method (NBC)

Prior 8 weeks of data excluding current week were used to calculate the baseline mean and variance, which derived the NBC parameter. A signal occurred when the parameter exceeded the threshold value.

For the purpose of evaluations, an outbreak week was defined as any week that had over two outbreak-related cases during the study period.

Results

Fourteen outbreak weeks were identified to evaluate the detection ability of the six methods. The table below summarizes the measures of each method.

Model	*SE*	*SP*	*PPV*	*NPV*	*Total No signals*
GL	50%	98%	63%	97%	11
PO	93%	82%	23%	99%	57
TS	50%	97%	50%	97%	14
CuSum	7%	99%	33%	95%	3
HL	64%	90%	27%	98%	33
NBC	43%	80%	11%	96%	56

The SPs did not vary much across six methods while the SE of the PO method was higher than the rest. The PPV ranged from 11% to 63%, and the NPV did not vary greatly. The total numbers of signals generated from the PO and NBC methods were higher than the rest.

Conclusions

Among the above six methods, the PO method had the ability to detect a high percentage of true outbreaks. However, the high number of signals and the relatively low PPV indicated the limitations of the PO method. Other information such as geographical clusters should be considered in determining further public health investigations as needed.

***Xin Jin**
E-mail: xxj01@health.state.ny.us

Citation: Emerging Health Threats Journal 2011, **4**: 11056 - DOI: 10.3402/ehtj.v4i0.11056
(page number not for citation purpose)