Evaluation of Syndrome Algorithms for Detecting Pneumonia Emergency Department Visits

Priscilla Wong* and Hilary Parton

New York City Department of Health and Mental Hygiene, Long Island City, NY, USA

Objective

To validate and improve the syndromic algorithm used to describe pneumonia emergency department (ED) visit trends in New York City (NYC).

Introduction

The NYC Department of Health and Mental Hygiene (DOHMH) uses ED syndromic surveillance to monitor near real-time trends in pneumonia visits. The original pneumonia algorithm was developed based on ED chief complaints, and more recently was modified following a legionella outbreak in NYC. In 2016, syndromic data was matched to New York State all payer database (SPARCS) for 2010 through 2015. We leveraged this matched dataset to validate ED visits identified by our pneumonia algorithm and suggest improvements. An effective algorithm for tracking trends in pneumonia could provide critical information to inform and facilitate public health decision-making.

Methods

The DOHMH syndromic surveillance system includes daily ED data from 53 NYC hospitals. Most syndrome algorithms rely solely on chief complaint, which has historically been reported more consistently than discharge diagnosis. For this analysis, the validation dataset was restricted to matched visits with consistent age (plus or minus two years) and sex between the syndromic and SPARCS datasets.

The original pneumonia algorithm used a basic text search function to identify any mention of ICD-9-CM and ICD-10-CM diagnosis codes indicating pneumonia or key words "PNEUMON" or "MONIA" within the chief complaint only. The updated algorithm additionally searches the chief complaint for any mention of key words specific to legionella ("LEGIONA", "LEGIONN", "LEGIONE") and also searches for pneumonia ICD codes in the discharge diagnosis field. Syndrome sensitivity and positive predictive value (PPV) were evaluated by comparing visits identified by each algorithm to visits identified by billing diagnosis codes. A true SPARCS pneumonia ED visit was defined to contain an admitting or principal diagnosis billing code indicating pneumonia.

Alternate algorithms were created using regular expressions, which allowed for more flexible and accurate pattern matching. The algorithms were further revised based on additional inclusion and exclusion key words identified using the validation dataset.

Results

Between 2010 and 2015, there were a total of 204,101 true pneumonia visits based on the SPARCS billing records. Evaluation of the original algorithm found a sensitivity of 15.6% (31,771/204,101) and a PPV of 55.6% (31,771/57,180). Over the same time period, syndromic surveillance identified a total of 127,560 pneumonia visits using the updated algorithm; 86,590 of the 127,560 syndromic cases identified were determined to be a true visit based on the billing diagnosis codes, resulting in an algorithm sensitivity of 42.4% and PPV of 67.9%. Of the 127,560 cases identified by the updated algorithm, 19 cases were classified as a pneumonia visit solely due to the presence of legionella key words in the chief complaint. Regular expression usage as opposed to the basic text search on the updated algorithm found similar sensitivity (42.4%, 86,561/204,101) and PPV (68.0%, 86,561/127,238).

Among all true pneumonia visits with a non-blank discharge diagnosis field, 65.3% (68,001/104,223) had mention of a pneumonia diagnosis code. Use of the discharge diagnosis code field in addition to the chief complaint found the algorithm to be almost three times more sensitive and 1.2 times greater in PPV than an algorithm without use of discharge diagnosis. Seasonal trends captured with and without use of discharge diagnosis were both similar to the true pneumonia trend indicated by SPARCS.

Among the 117,540 pneumonia cases missed by the updated algorithm, 58.6% had fewer than three words in the chief complaint. With the most popular key words among the missed cases being highly non-specific (i.e., "fever", "cough", "pain"), inclusion of these key words in addition to regular expression and discharge diagnosis field usage elevated algorithm sensitivity at a severe cost to the PPV. Including "fever" in the list of pneumonia key words resulted in a sensitivity of 56.5% (115,280/204,101) and a PPV of 9.0% (115,280/1,282,342), while addition of the key word combination "fever" and "cough" led to a sensitivity of 46.7% (95,264/204,101) and a PPV of 29.8% (95,264/319,876).

As we were unable to identify novel key word indicators that were good markers for pneumonia events, regular expression search functionality was the best improvement to the pneumonia syndrome algorithm. This revised, new algorithm maintained sensitivity (42.4%, 86,561/204,101) and provided slight improvements to PPV (68.0%, 86,561/127,219).

However, performance of the updated algorithm varied across age groups. The algorithm was most effective in identifying younger cases (43.9% sensitivity, 80.4% PPV for those 17 years and younger), while it performed the worst among those 65 years and older (39.6% sensitivity, 58.7% PPV).

Conclusions

Based on our evaluation of the pneumonia syndromic surveillance algorithm, we found that search of the discharge diagnosis field greatly improved algorithm sensitivity and PPV and usage of regular expressions increased PPV slightly. Including additional words possibly indicating pneumonia did not substantially improve sensitivity or PPV. However, integration of the ED chief complaint triage notes which are not currently utilized could further enhance the effectiveness of the pneumonia syndrome algorithm and better characterize daily pneumonia trends in NYC.

Monthly count of true pneumonia ED visits identified by pneumonia algorithm, NYC, 2010-2015
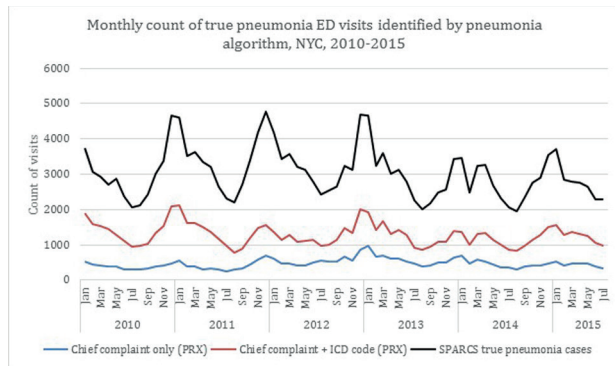
Keywords

pneumonia; syndromic surveillance; validation

*Priscilla Wong

E-mail: pwong4@health.nyc.gov