Adaptation of the NegEx algorithm to Veterans Affairs electronic text notes for detection of Influenza-like illness (ILI)

Brett R. South, MS¹ Shobha Phansalkar, RPh, MS¹, Ashwin Deepak Swaminathan, BS¹, Sylvain Delisle, MD, MBA², Trish Perl, MD, MS³, Matthew H Samore, MD¹

¹VA Salt Lake City Health Care System and the Department of Medicine, University of Utah, School of Medicine, Salt Lake City, UT, ²VA Maryland Health Care System and University of Maryland, School of Medicine, Baltimore, Maryland, ³The Johns Hopkins Hospital, Baltimore, Maryland

Abstract

We tested automated detection of influenza-like illness (ILI) from free-text clinician notes extracted from the VA electronic medical record using a simple negation algorithm coupled with string matching for relevant ILI symptoms mapped to UMLS concepts. Additionally, we documented negation problems encountered and adaptations made to the negation algorithm when applied to the full texts of a diverse array of VA note types and templates.

Methods

A randomised sample of 15,377 patient encounters from the VA Maryland Health Care system (VAMHCS) and VA Salt Lake City Health Care system (VASLCHCS) were selected from emergency and selected outpatient clinics during the study period 10/01/03 to 3/31/04. ILI cases were identified by chart review using an explicit definition of ILI based on CDC criteria. ILI concepts from the case definition were mapped to a standard vocabulary using the UMLS Metathesaurus® search tool. The full texts of a total of 76,500 notes associated with the sampled encounters were processed by a negation algorithm called NegEx¹ to identify pertinent negations. Presence of any two non-negated UMLS concepts per note was interpreted to denote ILI. Initially, the publicly available NegEx algorithm was adapted to the VA note structures without modifying the negation logic. We then sought to refine the logic to improve both sensitivity and specificity, via an iterative process. Initial modifications targeted false negatives and subsequent changes addressed the false positives. With each cycle, 10-15% of notes were systematically reviewed to identify problems that were amenable to rule modification. Thus, revisions to NegEx specific to VA notes were produced. Iterations were terminated when further changes were either not feasible or beneficial. The accuracy of the unmodified (“out-of-the-box”) version of NegEx was compared to the final, VA customized version using chart review as the reference standard. Comparisons between predictive values derived from the final optimized version and case detection based on specific document source were also estimated.

Results

Chart review identified 280 cases of ILI. Predictive values derived from case detection based on the unmodified, publicly available version of NegEx in terms of sensitivity, specificity, PPV and area under the ROC were: (85%, 88%, 12%, 87%), compared with the optimized version of NegEx: (88%, 93%, 18%, 91%). By tailoring the NegEx algorithm to VA note documents, the number of false positive cases was reduced by half between the initial “out of the box” application of NegEx and the final optimized version. Additionally, case detection models incorporating the entire note corpus yielded better performance over case detection models based on chief complaint or emergency department notes alone. Only 15 percent of all documents in the note corpus contained a chief complaint field, and only 6 percent of notes were related to emergency department encounters. Predictive values for detection models using notes containing only chief complaint (51%, 96%, 21%, 74%) or emergency department notes (53%, 98%, 36%, 76%) had lower sensitivities, higher specificities, but lower areas under the ROC than the final optimized version of NegEx.

Conclusion

Previous studies in syndromic surveillance have focused on extraction of clinical signals from two primary textual sources, namely the chief complaint field and emergency department encounters. Our study showed that superior ILI case detection models can be achieved when the customized version of NegEx is applied in combination with the entire note corpus, and all textual data, regardless of document source.

References