Free-Text Processing To Enhance Detection Of Acute Respiratory Infections

Sylvain DeLisle MD, MBA[1,2], Brett R. South MSc[3], Shobba Phansalkar[3], Trish M. Perl, MD, MSc[4], Adi Gundlapalli, MD, PhD[3], Matthew Samore MD[3]

[1]VA Maryland Health Care System, Baltimore, MD, [2]University of Maryland, Baltimore, MD, [3]University of Utah, Salt Lake City, UT, [4]Johns Hopkins Medical Institutions, Baltimore, MD

OBJECTIVE

We asked to what extent computerized processing of the full free-text clinical documentation could enhance syndrome detection compared to the sole use of structured data elements from a comprehensive electronic medical record (EMR).

METHODS

Using an explicit definition of acute respiratory infections (ARI) and CDC's definition of influenza-like illness (ILI), we manually reviewed all EMR entries within 24 hours of 15,377 randomly sampled outpatient encounters at two Veterans Administration medical systems. Uncovered ARI and ILI cases served as a reference target to develop automated case-detection algorithms (CDAs). We used logistic regression with backward elimination to select those structured parameters that significantly contributed to case detection. We then attempted to enhance those CDAs by pairing them with the result of two different text analysis strategies: 1) string searches for at least two non-negated case definition symptoms via an adapted NegEx agorithm; 2) natural language processing for those same symptoms via the native MedLEE software. CDAs that included both structured and free-text-derived data were compared to corresponding CDAs that used structured data alone for their statistical performance at detecting ARI and ILI cases.

RESULTS

Of the 22 structured clinical parameters considered, three contributed significantly to ARI and ILI case detection: ICD-9 diagnostic codes, a new prescription for cough remedies and measured elevations in body temperature. Statistical performance for CDAs that combined these structured parameters with or without text analyses are shown in the Table. Adding text analysis could increase ARI case detection sensitivity from 84 to 99% (compare Models 4 and 6). However, because of the low ARI incidence (1.8%), the accompanying drop in specificity translated into large declines in positive predictive value (PPV). CDAs that required satisfying both a query of structured EMR parameters as well as the NegEx algorithm yielded high PPVs while uniquely retaining sensitivities near 70% (Models 10 and 11). We could also construct CDAs targeting ILI that focused on maximizing sensitivity (Model 19) or PPV (Model 21). Benefits of free-text analysis persisted, albeit at a lower level, when we used a general-purpose text analysis software (MedLEE, see Models 7, 12, 20, 22) rather than the highly adapted NegEx algorithm.

CONCLUSIONS

Free-text processing of clinical notes brings information about disease symptoms that complements what is otherwise available as structured data in the EMR. Results from free-text analyses can be used to selectively enhance the statistical performance of CDAs that target acute respiratory infectious syndromes.