

# Evaluation of Emergency Medical Text Processor For Pre-Processing Chief Complaint Data for Syndromic Surveillance

Debbie Travers<sup>1</sup>, PhD, RN, Aaron Kipp<sup>1</sup>, BS,  
Jennifer MacFarquhar<sup>2</sup>, BSN, Anna Waller<sup>1</sup>, ScD

Departments of Emergency Medicine<sup>1</sup>, and Medicine<sup>2</sup>, University of North Carolina

## OBJECTIVE

The goal of this project is to compare automated syndromic surveillance queries using raw chief complaints to those pre-processed with the Emergency Medical Text Processor (EMT-P) system.

## BACKGROUND

The North Carolina Bioterrorism and Emerging Infection Prevention System (NC BEIPS) is being developed as an early detection system for public health surveillance in North Carolina [1]. In 2004, the NC BEIPS data warehouse included 813,536 emergency department (ED) visits. The ED records are queried daily with 6 syndrome reports, using chief complaint (CC) and, when available, triage note (TN) and temperature data. The automated syndrome queries are written in standard query language (SQL) and search raw CC and TN text for keywords (e.g., *myalgia*, *fever*), including synonyms and abbreviations (e.g., *muscle aches*, *temp. fvr*). The Emergency Medical Text Processor (EMT-P) cleans ED CCs and maps them to standard concepts from the Unified Medical Language System (UMLS) [2]. Cleaning modules address free text CC variability such as acronyms, truncations and punctuation. The standardized CCs are then available for a variety of primary clinical and secondary uses. EMT-P has been validated for general clinical purposes, but not specifically for syndromic surveillance. EMT-P was recently implemented at NC BEIPS and is used to clean all ED CCs on a daily basis.

## METHODS

This study analyzed a subset of the 1,000 ED visits which were previously sampled from the 2004 NC BEIPS data warehouse. 500 visits were selected that did not meet any of the original automated syndrome queries, and another 500 that met at least one of the original queries. The subset for this study consisted of 717 ED visits, 215 of which met the original query for influenza-like illness (ILI), and 502 that did not. All ED records were individually reviewed by two clinical experts who determined whether the visit met the ILI clinical case definition. Any discrepancies were resolved by a third expert.

The original automated ILI query was revised to search the CC field for EMT-P-generated standard concept unique identifiers (CUI) for ILI symptoms (e.g., C0015967 for *fever*). All visits were processed

twice, once using the original query on the raw CCs and once using the revised query on the CUIs. The clinical decisions were used as the gold standard, and compared to the output of both automated ILI queries. SAS version 8 (Cary, NC) was used to generate kappa statistics ( $\kappa$ ), sensitivity (Se), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV).

## RESULTS

The two clinical experts agreed on 93% of the ILI decisions ( $\kappa=0.82$ ). The final gold standard set included 191 (27%) ILI cases and 526 (73%) non-ILI cases. The statistics for the raw and cleaned CC data were identical: Se: 92%, Sp: 92%, PPV: 81%, NPV: 97%.

## CONCLUSIONS

Given that electronic ED CC data are widely available and more timely than diagnosis data, CC is commonly used for early detection. However, syndromic surveillance is hampered by the variability of free text CC data [3]. Pre-processing offers a potential solution for this problem, and we found that EMT-P performed as well as a conventional natural language query for ILI detection. Natural language queries are labor-intensive and complex to develop and maintain compared to CUI-based queries of processed CC data. Since EMT-P was not developed specifically for syndromic surveillance, performance may improve with customization. Future research will evaluate EMT-P for pre-processing of pre-hospital CC data. In light of the fact that no standard terminology exists for ED CC, there is an ongoing need for pre-processors such as EMT-P to support applications like syndromic surveillance.

## REFERENCES

- [1] Ising, A., McLamb, J., Waller, A., Li, M., Falls, D. (2005, May). North Carolina Bioterrorism and Emerging Infection Prevention System: The syndromic surveillance component of the NC PHIN. Demonstration at the 2005 PHIN Conference, Atlanta, GA.
- [2] Travers, D.A. & Haas, S.W. (2004). Evaluation of Emergency Medical Text Processor, a system for cleaning chief complaint text data. *Academic Emergency Medicine* 11(11): 1170-1176.
- [3] Shapiro, A. (2004). Taming variability in free text: Application to health surveillance. *MMWR* 53S: 95-100.

Thanks to the following contributors: Meichun Li, Dennis Falls and Megan Davie and funder: NC Division of Public Health.

Further Information:  
Debbie Travers, [dtravers@med.unc.edu](mailto:dtravers@med.unc.edu)