Fever Detection from Free-text Clinical Records for Biosurveillance

Wendy W. Chapman, PhD
John N. Dowling, MD, MS
Michael M. Wagner, MD, PhD

RODS Laboratory
Center for Biomedical Informatics
University of Pittsburgh

Abstract

Surveillance of febrile illness would be useful for public health and bioterrorism outbreak detection. At most institutions, febrile information is only contained in free-text clinical records. We compared the sensitivity and specificity of three fever detection algorithms for detecting fever from free-text. Keyword-CC and Syndrome-CC classified patients based on their triage chief complaint; Keyword-ED classified patients based on their dictated emergency department report. Keyword-ED was the most sensitive (sens 0.98; spec 0.89), and Keyword-CC was the most specific (sens 0.61, spec 1.0). Because chief complaints are available sooner than emergency department reports, we suggest a combined application that classifies patients based on their chief complaint followed by classification based on their emergency department report, once the report becomes available.

Introduction

Many of the diseases that need to be identified for public health and bioterrorism detection produce fever. Therefore, monitoring febrile illness upon admission to health care facilities may be useful for surveillance. Detection of an increase in the number or percentage of patients with fever compared to a baseline number could alert public health officials to an outbreak of a known or new disease or to a terroristic threat. Knowledge of the fever status of patients would be particularly helpful when combined with syndromic data. For example, an outbreak distinguished as febrile gastroenteritis would be easier to characterize and track than a simple increase in the number of cases of gastroenteritis. Syndromic surveillance systems classify patients into syndromic categories [1-20] that often include fever in their case definitions [5, 21-24]; however, to our knowledge no one has measured the ability to automatically determine whether a patient has a fever.

Determining if a patient has a fever is not straightforward. Some institutions may have a coded record of temperature upon admission, but many do not. In most instances, the only way to determine whether a patient has a fever is from the text of the patient’s medical records. Patient records include complaints, symptoms, vital signs, physical examination, radiological findings, microbiology results, prescribed medications, and diagnoses. However, much of the clinical information is locked in free-text reports and must be automatically extracted to be useful for a real-time surveillance system. In this study, we evaluate the performance of three fever detection algorithms.
algorithms, two that detect fever from the triage chief complaint and one that detects fever from the text of the patient’s emergency department (ED) report.

Methods

Case Selection

We measured the sensitivity, specificity, and likelihood ratio positive (LR+) of three fever detection algorithms on a test set comprised of 213 patients seen at the University of Pittsburgh Medical Center (UPMC) Presbyterian Hospital from 02/01/02 to 12/31/02. One-half of the patients were randomly selected from patients with an ICD-9 discharge diagnosis of 780.6 (fever). The other half were randomly selected from patients discharged during a day at the midpoint of the study time period without an ICD-9 diagnosis of 780.6.

Case Definition. We compared the detection performance of the three algorithms against physician classification of fever based on the patient’s ED report. Fever was defined as a measured temperature \( \geq 38^\circ C \) (100.4\(^\circ F\)) or a description of recent fever or chills. The reported temperature could have been measured in the ED, by the patient, or at another institution such as a nursing home or other hospital.

A physician board-certified in internal medicine and infectious diseases reviewed the ED reports of the 213 patients. For every report the physician noted whether the definition of fever was met. For patients with a fever, the physician also noted whether the patient had a measured temperature, where the temperature was measured, and who reported the fever.

Fever Detection Algorithms

1. Detection of fever from ED reports. To detect fever in ED reports, we developed a keyword-based algorithm that accounts for contextual information about negation and hypothetical descriptions (Keyword-ED). The algorithm (Figure 1) detects two types of fever description in the reports, including a reported fever and a measured temperature meeting the definition of fever described above. If either of the fever descriptions exists in the report, the patient is considered febrile.

Fever is true IF:
A. non-negated fever keyword exists AND fever keyword not in a hypothetical statement
OR
B. measured temperature \( \geq 38^\circ C \)

Figure 1. Keyword-ED algorithm.

Fever keywords include \textit{fever(s)}, \textit{febrile}, \textit{chill*}, and \textit{low(-)grade temp*} where characters within parentheses are optional and asterisks indicate any character, including a white space. In this way, \textit{fevers, chills}, and \textit{low-grade temperature} would be considered fever keywords. Any fever keywords that were not negated and were not in a hypothetical statement indicated the presence of a fever in the patient.
We used a regular-expression negation algorithm called NegEx [25, 26] to determine if a fever keyword was negated. NegEx looks for negation phrases, such as *denies* or *no*, up to six terms before the fever keyword and for negation phrases, such as *free* or *unlikely* up to six terms after the fever keyword. A term can be a single word (e.g., *fever*) or a phrase from the Unified Medical Language System Metathesaurus (e.g., *chest pain*). If a negation phrase exists within six terms of a fever keyword, the fever keyword is considered to be negated. For example, in the sentence “The patient denies any occurrence of fever” the keyword *fever* would be considered negated and therefore not an indication that the patient was febrile.

ED reports often describe symptoms that may occur in the future, as in “The patient should return for increased shortness of breath or fever.” If a fever keyword was preceded by the words *if*, *return*, *should*, *spotted*, or *as needed* the keyword was not considered an indication that the patient was febrile.

A measured temperature was detected if a number was identified no more than nine words following the temperature keyword *temp*. Any measured temperature greater than or equal to 38 degrees Celsius was considered an indication the patient was febrile. For example, in the sentence “Her temp was measured at the nursing home as being 38.5C” the patient would be considered febrile.

2. Detection of fever from chief complaints. We implemented two fever detection algorithms based on the patient’s triage chief complaint electronically entered on admission to the ED. The first detector from chief complaints (Keyword-CC) employed the fever and temperature keywords used by the ED report fever detection algorithm. If any of the fever or temperature keywords appeared in the chief complaint, the patient was considered febrile. Because chief complaints are syntactically simple phrases describing a current problem, we did not assess whether the phrase was negated or in a hypothetical statement.

The second detector from chief complaints was an attempt to be more sensitive at detecting patients who may have a fever but may not have measured their temperature before admission to the ED. This detector (Syndromic-CC) is a syndromic detector called CoCo [27] that is currently used in the Real-time Outbreak and Disease Surveillance (RODS) system [28]. RODS monitors seven different syndromes including respiratory, gastrointestinal, rash, neurological, hemorrhagic, botulinc, and constitutional. The constitutional syndrome is a classification for capturing patients who present with non-localized complaints typical of many illnesses in their early stages. Any patient complaining of fever, malaise, lethargy, or generalized aches would be classified as having a constitutional syndrome. It is possible that some febrile patients presenting to the ED do not yet complain of fever but are experiencing other constitutional symptoms that typically occur with a febrile illness. Any patient classified by CoCo as having a constitutional syndrome was considered febrile.

**Outcome Measures**

We calculated the sensitivity, specificity, and LR+ of the fever detection algorithms compared against the gold standard classifications made by the physician as shown in Figure 2.
Sensitivity = $\frac{TP}{TP + FN}$

Specificity = $\frac{TN}{TN + FP}$

LR+ = $\frac{Sensitivity}{1 - Specificity}$

Figure 2. Outcome measures where TP is the number of true positives, TN is true negatives, FP is false positives, and FN is false negatives.

Results

In Table 1 we show the outcome measures and their respective 95% confidence intervals of the three automatic fever detection algorithms when compared against gold standard classifications. The Keyword-ED algorithm was more sensitive and had higher overall accuracy than either of the chief complaint algorithms.

| Table 1. Performance of Three Fever Detection Algorithms |
|--------------------------------------------------------|
| **Keyword-ED** | **Keyword-CC** | **Syndromic-CC** |
| Sensitivity    | 0.98 (107/109) | 0.61 (66/109) | 0.57 (62/109) |
|                | [0.94-0.995]   | [0.52-0.69]   | [0.48-0.66]   |
| Specificity    | 0.89 (93/104)  | 1.0 (104/104) | 0.95 (99/104) |
|                | [0.82-0.94]    | [0.96-1]      | [0.89-0.98]   |
| LR+           | 9.28 [5.31-16.24] | *  | 11.83 [4.95-28.26] |

Error Analysis

Detection of Fever from Chief Complaints. Twenty-six febrile patients were undetected by both Keyword-CC and Syndromic-CC. The majority of the chief complaints for the 26 patients were not related to febrile illness but described other complaints, including headache, tachypnea, sob, altered mental status, dehydration, leg swelling, and chest pain. Five of the 26 patients had chief complaints describing a disease or syndrome that is often associated with fever, such as conjunctivitis, bacteremia, and flu like symptoms. Four of the patients had chief complaints that instead of describing a clinical complaint described an evaluation or procedure for which the patient came to the ED.

Thirteen of the febrile patients were detected by Syndrome-CC but not by Keyword-CC. All of these patients had chief complaints indicating a constitutional illness that did not specifically mention fever, such as sepsis, viral infection and weakness. However, Syndrome-CC also generated five false positive classifications for patients with chief complaints of viral infection, dizziness, muscle aches, and weakness.
Chief complaints for sixteen febrile patients contained a fever or temperature keyword but were not detected by Syndrome-CC. The reason CoCo did not accurately classify these patients as having a constitutional syndrome — in spite of training CoCo to classify chief complaints with fever keywords as constitutional — involves the current method CoCo uses for determining the best syndromic classification when multiple classifications exist. Currently, CoCo selects the single syndromic classification with the highest probability. Thus, chief complaints such as rash/fever, nausea/vomiting/fever, or fever and headaches were classified as rash, gastrointestinal, and neurological, respectively, because the probabilities for those syndromes were higher than the probabilities for constitutional syndrome.

Detection of Fever from ED Reports

Keyword-ED generated 11 false positives and 2 false negatives. Four false positives were due to contradictions in the report between report of fever and measured temperature. For example, one patient was described as febrile for the last week, but his measured temperature in the ED was 37.6 degrees. Three false positives were due to NegEx errors in which fever keywords were not properly negated and one was due to not identifying a fever keyword as a hypothetical statement (“We have given her instructions on what to watch out for, including … fever, chills, distention …”). Two false positive were due to a conflict between the resident’s and the attending’s notes, and in one instance “fever of unknown origin” was interpreted by the gold standard physician as a sign rather than a possible diagnosis.

One of the false negatives was due to the vague description of fever “he felt warm.” The other false negative was an error on the part of the expert physician, who classified an afebrile patient as febrile.

Description of Fever in Data Set. Prevalence of fever in the data set was 51% (109/213). Only nine (4%) of the 213 reports contained no information about temperature or fever. The criteria by which it was determined that the patient was febrile were as follows. Of the 109 patients with fever, 96 (88%) had a measured temperature. In 80 of the patients with fever (73%), the temperature was measured and found to be elevated in the ED, whereas in 16 (15%) instances, the temperature had been taken by the patient or at an institution where the patient had been previously. In 13 (12%) of 109 patients with fever, the fever or chills was self-reported or a report of fever came from another institution.

Discussion

Keyword-ED detected 98% of the febrile patients in the study and was statistically significantly more sensitive than either of the detectors from chief complaints (p < 0.05). Keyword-ED’s specificity of 89% can potentially be increased by improving negation and hypothetical situation identification.

We hypothesized that ED reports would be a fairly good resource for locating information about a patient’s fever status, because HCFA guidelines for ED reporting require that vital signs be measured and recorded [**need ref]. Nevertheless, some physicians referred the reader to the nursing notes for details about vital signs, and some failed to report anything about fever. Still, the majority of the ED reports in our sample contained fever information (96%). However, our sample was enriched with patients having a discharge diagnosis of fever, and none of the patients without information about febrile status in the ED reports came from the enriched sample. A
more accurate estimate of the proportion of ED reports without a description of febrile status can be calculated from the non-enriched portion of the sample at 8.4% (9/107).

Keyword-CC was fairly sensitive (61%), given the limited nature of triage chief complaints, and performed with perfect specificity. All patients detected as febrile by Keyword-CC actually had a fever according to the gold standard classification.

Considerations for selecting the optimal surveillance methods for biosurveillance of febrile illness include the completeness of the input data, the performance of the fever detectors, and the timeliness [29] with which the input data is electronically available. A chief complaint is available immediately upon admission to an emergency facility, whereas an ED report is not available until the report is dictated by the physician, manually transcribed, and stored on the hospital information system. Time lags between admission of the patient to the ED and electronic availability of the dictated report vary by institution.

Our results suggest that monitoring febrile illness may be best accomplished with an application that combines the output from chief complaints with that from ED reports. Initially, a biosurveillance system could monitor febrile illness from chief complaints with Keyword-CC. Based on results of our study, cases detected by Keyword-CC will be specific and will not generate false alarms. As ED reports become available, Keyword-ED could process cases not detected as febrile by Keyword-CC and update the surveillance system with the more complete and sensitive detection provided from ED reports.

Limitations and Future Work

Our study involved a single hospital in the city of Pittsburgh. Biosurveillance of febrile illness on a regional or national level would require an expanded study that evaluates the fever detection methods on data from other hospitals and cities – particularly for Keyword-ED, because linguistic variation in reporting may exist across the United States.

Conclusion

We measured the ability of three febrile detection algorithms to detect patients with a fever from free-text medical records. Two of the algorithms used triage chief complaints to classify the patients, and a third used the information described in the dictated ED report. The algorithm using information from the ED report was the most sensitive, whereas the algorithm using information from the chief complaint was the most specific. A surveillance application incorporating fever detection from chief complaints – which are the earliest electronic clinical data available in an emergency care facility – followed by detection from ED reports as they become available may provide an effective method for surveillance of febrile illness.
References

1. http://www.health.pitt.edu/rods/. Accessed April 16, 2003.
2. http://www.geis.ha.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp. Accessed April 16, 2003.
3. Lober WB, Karras BT, Wagner MM, Overhage JM, Davidson AJ, Fraser H, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc 2002;9(2):105-15.
4. https://secure.cirg.washington.edu/bt2001amia/index.htm. Accessed April 16, 2003.
5. Zelicoff A, Brillman J, Forslund DW, George JE, Zink S, Koenig S, et al. The rapid syndrome validation project (RSPV), a technical paper. Sandia National Laboratories.
6. http://www.nyam.org/events/syndromicconference/2002/posterpdf/buckeridge_poster.pdf. Accessed April 16, 2003.
7. http://www.nyam.org/events/syndromicconference/2002/posterpdf/brillman_poster.jpg. Accessed April 16, 2003.
8. http://www.nyam.org/events/syndromicconference/2002/posterpdf/foldy_poster.pdf. Accessed April 16, 2003.
9. http://www.chip.org/research/biosurv_projects.htm. Accessed April 16, 2003.
10. http://www.nytimes.com/2003/04/04/nyregion/04WARN.html?ex=1050552000&en=2c63a52eb80102bc&ei=5070. Accessed April 16, 2003.
11. Brinsfield K, Gunn J, Barry M, McKenna V, Dyer K, Sulis C. Using volume-based surveillance for an outbreak early warning system. Acad Emerg Med 2001;8:492.
12. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak 2003;3(1):2.
13. Lazarus R, Kleinman K,Dashevsky I, Adams C, Kludt P, DeMaria A, Jr., et al. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis 2002;8(8):753-60.
14. Lazarus R, Kleinman KP, Dashevsky I, DeMaria A, Platt R. Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. BMC Public Health 2001;1:9.
15. Matsu T, Takahashi H, Ohyama T, Tanaka T, Kaku K, Osaka K, et al. [An evaluation of syndromic surveillance for the G8 Summit in Miyazaki and Fukuoka, 2000]. Kansenshogaku Zasshi 2002;76(3):161-6.
16. http://www.nyam.org/events/syndromicconference/2002/posterpdf/chapman.pdf. Accessed April 16, 2003.
17. Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients’ chief complaints: An opportunity for bioterrorism and influenza surveillance. Ann Emerg Med 2003;41(4):447-52.
18. Gesteland PH, Wagner MM, Chapman WW, Espino JU, Tsui F, Gardner RM, et al. Rapid deployment of an electronic disease surveillance system in the state of Utah for the 2002 Olympic Winter Games. Proc AMIA Symp 2002:285-9.
19. Tsui FC, Espino JU, Wagner MM, Gesteland P, Ivanov O, Olszewski R, et al. Data, Network, and Application: Technical Description of the Utah RODS Winter Olympic Biosurveillance System. Proc AMIA Symp 2002:815-9.
20. Lewis MD, Pavlin JA, Mansfield Jl, O’Brien S, Boomsma LG, Elbert Y, et al. Disease outbreak detection system using syndromic data in the greater Washington DC area. Am J Prev Med 2002;23(3):180-6.
21. Recognition of illness associated with the intentional release of a biological agent. MMWR Morb Mortal Wkly Rep 2001;50(41):893-7.
22. Bravata DM, McDonald K, Owens DK, Smith W, Rydzak C, Szeto H, et al. Bioterrorism: Information technologies and decision support systems. Evidence Report/Technology Assessment No. 5 (Prepared by UCSF-Stanford Evidence-based Practice Center under Contract No. 290-97-0013). Rockville, MD: Agency for Healthcare Research and Quality. 2002.
23. Bioterrorism: a public health threat. CD Info 2000;2(5).
24. Monterey County Health Department Disaster Plan. Monterey, CA: Monterey County Health Department. October 9, 2001. http://www.co.monterey.ca.us/health/Publications/pdf/MCHDDisasterPlan.pdf. Accessed Dec. 9, 2003.
25. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001;34(5):301-10.
26. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. Evaluation of negation phrases in narrative clinical reports. Proc AMIA Symp 2001:105-9.
27. Olszewski RT. Bayesian classification of triage diagnoses for the early detection of epidemics. In: Recent Advances in Artificial Intelligence: Proceedings of the Sixteenth International FLAIRS Conference; 2003: AAAI Press; 2003. p. 412-416.
28. Tsui FC, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. J Am Med Inform Assoc 2003;10(5):399-408.
29. Wagner MM, Tsui FC, Espino JU, Dato VM, Sittig DF, Caruana RA, et al. The emerging science of very early detection of disease outbreaks. J Public Health Manag Pract 2001;7(6):51-9.