



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Fever detection from free-text clinical records for biosurveillance

Wendy W. Chapman,* John N. Dowling, and Michael M. Wagner

RODS Laboratory, Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA

Received 30 January 2004

Abstract

Automatic detection of cases of febrile illness may have potential for early detection of outbreaks of infectious disease either by identification of anomalous numbers of febrile illness or in concert with other information in diagnosing specific syndromes, such as febrile respiratory syndrome. At most institutions, febrile information is contained only 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 CoCo classified patients based on triage chief complaints; Keyword HP classified patients based on dictated emergency department reports. Keyword HP was the most sensitive (sensitivity 0.98, specificity 0.89), and Keyword CC was the most specific (sensitivity 0.61, specificity 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.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Disease outbreaks; Fever; Natural language processing; Computerized patient medical records; Public health surveillance; Surveillance; Infection control

1. Introduction

Many of the infectious diseases that represent threats to the public's health or have potential for bioterrorism produce a febrile response in affected individuals early in the course of illness. The ability to detect febrile illness in a community would be valuable in public health surveillance if surveillance could be performed routinely, with sufficient accuracy, and at low cost. For example, 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.

Some syndromic surveillance systems classify patients into syndromic categories [1–15] that include fever in their definitions [5,16–19]. Knowledge of the fever status of patients would be particularly helpful when combined with other information about their symptoms or clinical characteristics. For example, knowing the presence of

fever is crucial in determining whether a patient has Severe Acute Respiratory Syndrome (SARS) [20–22]. However, to our knowledge no one has measured the accuracy at which fever status can be determined automatically from routinely collected data.

Automatically determining if a patient has a fever from medical records is not straightforward. Very few clinical facilities encode temperature into computer readable format at the time the temperature is taken. In most instances, the only way to determine whether a patient has a fever is from the text written or dictated into the patient's medical record. 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.

Statistical and symbolic text processing techniques have been applied successfully to dictated clinical reports [23] for retrieval of relevant documents [24,25], classification of text into discrete categories [26–29], and extraction and encoding of detailed clinical information from text [30–34]. Applying text processing techniques to the field of biosurveillance is a fairly new area of research in medical informatics that has focused on

* Corresponding author. Fax: 1-413-383-8135.

E-mail address: chapman@cbmi.pitt.edu (W.W. Chapman).

processing free-text chief complaints recorded in the emergency department [26,35–39]. In this study, we evaluated the performance of three free-text processing algorithms to detect fever from triage chief complaints (Keyword CC and CoCo) and from the text of the patient’s emergency department report (Keyword HP).

2. Background

Below we describe one of the free-text processing algorithms that has already been developed for classifying patients into syndromic categories based on their chief complaints (CoCo) and the pre-existing negation algorithm that we apply in Keyword HP (NegEx).

2.1. CoCo

CoCo [26], currently used in the Real-time Outbreak and Disease Surveillance (RODS) system [40], is a naïve Bayesian text classification system that classifies chief complaints into one of eight different syndromic categories: chief complaints indicating upper or lower respiratory problems like congestion, shortness of breath, or cough are classified as respiratory; symptoms like nausea, vomiting, or abdominal pain are classified as gastrointestinal; any description of rash is classified as rash; ocular abnormalities and difficulty swallowing or speaking are classified as botulinic; bleeding from any site is classified as hemorrhagic; non-psychiatric neurological symptoms such as headache or seizure are classified as neurological; generalized complaints like fever, chills, or malaise are classified as constitutional; and clinical conditions not relevant to biosurveillance such as trauma and genitourinary disorders are classified as other. RODS monitors how frequently patients are classified into the syndromic categories and uses spatio-temporal algorithms to generate alerts when the observed number of patients classified into a particular syndromic category statistically exceeds the expected number [41].

CoCo is a naïve Bayesian classifier with a probabilistic model for every syndrome described above. A training set of 28,990 chief complaints manually classified by a physician was used to estimate the prior probability of syndromic classifications and the probabilities of unique words for each syndrome. Given a chief complaint, these probabilities are used to compute the posterior probability for each syndrome. The current implementation of CoCo classifies the chief complaint with the syndrome obtaining the highest posterior probability.

Given a chief complaint G consisting of a sequence of words w_1, w_2, \dots, w_n , the posterior probability of syndrome R , $P(R|G)$, can be expressed using Bayes’ rule and the expansion of G into words as

$$P(R|G) = \frac{P(R)P(w_1|R)P(w_2|w_1R) \cdots P(w_n|w_1 \cdots w_{n-1}R)}{\sum_R P(R)P(w_1|R)P(w_2|w_1R) \cdots P(w_n|w_1 \cdots w_{n-1}R)}$$

An approximation of $P(R|G)$ was computed by employing language models that made assumptions about the conditional independence of the words in a chief complaint [42]. A previous evaluation [26] showed that the unigram implementation of Bayes’ formula classified chief complaints into syndromes most accurately with the following areas under the ROC curves: botulism, 0.78; rash, 0.91; neurological, 0.92; hemorrhagic, 0.93; constitutional, 0.93; gastrointestinal, 0.95; other, 0.96; and respiratory, 0.96.

In this paper, one of the text processing methods classifies a patient as febrile if CoCo classifies the chief complaint as constitutional.

2.2. NegEx

One research study estimated that more than half of all findings described in dictated medical reports are negated [43]. Negation is not an issue in processing chief complaints, which are typically short, simple phrases [39]; however, in emergency department reports—which are the source of fever information for the Keyword HP algorithm—fever can be described as being present or absent in a patient. To account for negation in Keyword HP, we applied an algorithm called NegEx [44].

NegEx is a simple, regular-expression based algorithm whose input is a sentence with indexed findings and whose output is whether the indexed findings are explicitly negated in the text (e.g., “The patient denies chest pain”) or are mentioned as a hypothetical possibility (e.g., “Rule out pneumonia”). NegEx has two important components: regular expressions and a list of negation phrases. A complete description of NegEx, including a list of all negation phrases, can be found at <http://omega.cbmi.upmc.edu/~chapman/NegEx.html>.

NegEx uses two regular expressions that are triggered by three types of negation phrases.

Regular Expression 1: <negation phrase> * <indexed term>

Regular Expression 2: <indexed term> * <negation phrase>

The asterisk (*) represents five terms, which can be a single word or a UMLS phrase. Depending on the specific negation phrase in the sentence, an indexed term within the window of the regular expression may be marked as negated or possible. Three types of negation phrases are used by NegEx:

- (1) Pseudo-negation phrases—phrases that look like negation phrases but are not reliable indicators of pertinent negatives. If a pseudo-negation phrase is found, NegEx skips to the next negation phrase. NegEx’s current list of pseudo-negation phrases

includes 16 phrases such as “no increase,” “not cause,” and “gram negative.”

- (2) Pre-finding negation phrases—phrases that occur before the term they are negating. Pre-finding phrases are used in Regular Expression 1. NegEx currently applies 125 pre-finding negation phrases; however, seven of the pre-finding negation phrases (“no,” “denies,” “without,” “not,” “no evidence,” “with no,” and “negative for”) account for 90% of negations in most types of dictated reports [43]. In addition, NegEx uses 21 pre-finding negation phrases to indicate a conditional possibility (e.g., “rule out” and “r/o”).
- (3) Post-finding negation phrases—phrases that occur after the term they are negating. Post-finding phrases are used in Regular Expression 2. NegEx implements seven post-finding negation phrases (e.g., “free” and “are ruled out”) and 14 post-finding phrases to indicate conditional possibility (e.g., “did not rule out” and “is to be ruled out”).

NegEx’s algorithm works as follows:

- For each sentence, find all negation phrases.
- Go to the first negation phrase in the sentence (Neg1).
- If Neg1 is a pseudo-negation phrase, skip to the next negation phrase in the sentence.
- If Neg1 is a pre-finding negation phrase, define a window of six terms after Neg1; if Neg1 is a post-finding negation phrase, define a window of six terms before Neg1.
- Decide whether to decrease window size (relevant if another negation phrase or a conjunction like “but” is found within the window).
- Mark all indexed findings within the window as either negated (if negation phrase is a negating phrase) or possible (if negation phrase is a conditional possibility phrase).
- Repeat for all negation phrases in the sentence.
- Repeat for all sentences.

As an example, consider the following sentence: “He says he has **not vomited** but is *short of breath* with *fever*.” Indexed findings are italicized, negation phrases are in bold, and conjunctions that stop the scope of the negation phrase are underlined. All of the indexed findings in the sentence are eligible for negation based on Regular Expression 1 triggered by the negation phrase “not.” However, the presence of the word “but” within the window prevents *short of breath* and *fever* from being negated so that NegEx only negates *vomited*.

We incorporated NegEx within the Keyword HP algorithm to determine whether indexed instances of fever (described below) were negated.

3. Methods

We measured the classification accuracy of three fever detection algorithms on a test set consisting of 213 pa-

tients seen at the University of Pittsburgh Medical Center (UPMC) Presbyterian Hospital. We compared the detection performance of the three algorithms against physician classification of fever based on the patient’s ED report. In particular, we measured sensitivity, specificity, and likelihood ratio positive (LR+).

3.1. Test collection

The test and control cases were randomly selected from hospitalized patients seen during the period 02/01/02 to 12/31/02. One-half of the patients were drawn from patients with an ICD-9 hospital discharge diagnosis of 780.6 (fever). The other half from patients without an ICD-9 diagnosis of 780.6. All medical records were de-identified in accordance with the procedures approved by the UPMC Institutional Review Board. A physician board-certified in internal medicine and infectious diseases reviewed reports dictated from the emergency department to determine if the patients met our definition of febrile illness. We defined febrile illness as being present if there was either (1) a measured temperature $\geq 38.0^{\circ}\text{C}$ or (2) a description of recent fever or chills. The measured temperature could have been determined in the ED, by the patient, or at another institution such as a nursing home. The physician’s judgments about whether the patients were febrile based on the dictated ED reports comprised the gold standard answers for the test set. To better understand the how fever was described in ED reports, for patients that met the definition of febrile illness the physician also noted whether the patient had a measured temperature, at what location the temperature was measured, and who reported the fever.

3.2. Fever detection algorithms

We evaluated three free-text processing algorithms by comparing their classifications of febrile illness against classifications made by the gold standard physician based on manual review of ED reports.

3.2.1. Keyword HP algorithm

The Keyword HP algorithm was designed for this study to detect fever from history and physical exams dictated in the emergency department (ED). The algorithm accounts for contextual information about negation and hypothetical descriptions to eliminate false positive classifications. The logic of this algorithm is satisfied if either of two clauses is true:

1. (report contains a fever keyword AND the fever keyword is not negated AND the fever keyword is not in a hypothetical statement) OR
2. (report describes a measured temperature $\geq 38.0^{\circ}\text{C}$).

If either of the clauses is satisfied, the algorithm classifies the patient as febrile. We describe the logic for the two clauses below.

For clause 1 to be true, the report must contain a fever keyword. The set of fever keywords includes: *fever(s)*, *febrile*, *chill**, and *low(-)grade temp** where characters within parentheses are optional and asterisks indicate any character, including a white space. In this way, *fevers*, *chills*, and *low-grade temperature* would be considered fever keywords.

Clause 1 also requires the fever keyword not to be negated. To determine if a fever keyword was negated, the Keyword HP algorithm uses a regular-expression negation algorithm called NegEx [43,44], which is described in Section 2 of this paper. NegEx looks for dozens of negation phrases, such as *denies* or *no*, up to six terms before the fever keyword and for a few negation phrases, such as *free* or *unlikely*, up to six terms after the fever keyword. In the sentence “The patient **denies** any occurrence of *fever*” the keyword *fever* would be considered negated; therefore, keyword HP would not classify the patient as having febrile illness. If multiple fever keywords were found in a single sentence (e.g., *fever* and *chills*), and one of the fever keywords was negated, the other fever keywords were also negated, regardless of whether NegEx considered them negated or not.

The last requirement of clause 1 is that the fever keyword not occur in a hypothetical statement. To determine if a fever keyword was used in a hypothetical statement, the algorithm looks for descriptions of fever occurring in the future, as in “The patient should return for increased shortness of breath or fever.” If a fever keyword was preceded in the sentence by the word *if*, *return*, *should*, or *as needed* the keyword was considered to be used in a hypothetical statement, and the algorithm did not classify the patient as febrile. In addition, if the word *spotted* preceded a fever keyword, the algorithm did not consider it an instance of fever, in order to eliminate sentences hypothesizing that the patient might have (Rocky Mountain) spotted fever.

Clause 2 is true if the report describes a measured temperature. To determine the presence of a measured temperature in a report, we first located any occurrences of the word *temp**. If a number ranging inclusively from 38 to 44 or from 100.4 to 111.2 occurred within nine words after *temp**, the patient was considered to have a measured temperature meeting the definition of febrile illness. For example, in the sentence “Her *temperature* was measured at the nursing home as being 38.5 °C” the algorithm would classify the patient as febrile.

3.2.2. Keyword CC algorithm

The Keyword CC algorithm was designed to detect cases of febrile illness from chief complaints electronically entered on admission to the ED. If any of the fever keywords used by Keyword HP or the term *temp** appeared in the chief complaint, the patient was considered febrile. For example, a patient with the chief complaint “increased temperature” or “fever” would be classified

as febrile by Keyword CC. Because chief complaints are syntactically simple phrases describing a recent problem, Keyword CC did not use negation or hypothetical statement detection.

3.2.3. CoCo algorithm

We evaluated a second algorithm for detecting fever from chief complaints by applying CoCo (described in Section 2) to the patients’ chief complaints and determining whether the patients had a constitutional syndrome. Patients with chief complaints indicating a fever are currently classified by CoCo as having a constitutional syndrome as are patients who present with non-localized complaints typical of many illnesses in their early stages, such as malaise, lethargy, or generalized aches. We applied CoCo to the problem of fever detection to capture two possible scenarios. First, 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 fever. Second, because chief complaints are short phrases that are designed to represent the most pertinent symptoms rather than to give a complete description of a patient’s clinical condition, even when a patient reports fever to the triage nurse, a word indicating fever may not be included in the chief complaint. The CoCo algorithm represents a potentially more sensitive algorithm for detecting patients with a fever even though fever is not indicated in the chief complaint. For this study, any patient classified by CoCo as having a constitutional syndrome was considered febrile.

3.3. 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 below, where TP is the number of true positives, TN is true negatives, FP is false positives, and FN is false negatives.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

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

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

4. Results

Table 1 shows the accuracy of classification with 95% confidence intervals for the three algorithms when compared against gold standard classifications. The Keyword HP algorithm was the most sensitive algorithm.

Table 1
Performance of three fever detection algorithms

	Keyword HP	Keyword CC	CoCo
Sensitivity [95% CI]	0.98 (107/109) [0.94–0.995]	0.61 (66/109) [0.52–0.69]	0.57 (62/109) [0.48–0.66]
Specificity [95% CI]	0.89 (93/104) [0/82–0.94]	1.0 (104/104) [0.96–1]	0.95 (99/104) [0.89–0.98]
LR+ [95% CI]	9.28 [5.31–16.24]	^a	11.83 [4.95–28.26]

^a LR+ could not be calculated, because $LR+ = \frac{\text{sensitivity}}{1-\text{specificity}} = \frac{0.61}{1-1}$.

4.1. 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. Criteria by which the gold standard physician determined the patient was febrile were as follows. Of the 109 patients with fever, 96 (88%) had a measured temperature. In 80 (73%) of the patients with fever, 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.

4.2. Error analysis of fever detection algorithms

4.2.1. Keyword CC algorithm

All patients with fever indicated in the chief complaint were detected by Keyword CC, and all patients with a fever keyword in the chief complaint had a fever according to the gold standard classification, indicating that the Keyword CC algorithm is precise and specific. However, Keyword CC had a sensitivity of only 0.61. False negatives were due to chief complaints that did not explicitly indicate a fever. Thirteen of the false negatives included patients with constitutional chief complaints that were detected correctly by CoCo, as described below. However, the majority of the false negatives were due to chief complaints not generally associated with febrile illness, including *headache*, *tachypnea*, *sob*, *altered mental status*, *dehydration*, *leg swelling*, and *chest pain*. Five patients had chief complaints describing a disease or syndrome often associated with fever, such as *conjunctivitis*, *bacteremia*, and *flu like symptoms*, and four patients had chief complaints that instead of describing a clinical complaint described an evaluation or procedure for which the patient came to the ED.

4.2.2. CoCo algorithm

CoCo had slightly lower sensitivity and specificity than Keyword CC. Chief complaints for 16 febrile pa-

tients contained a fever or temperature keyword but were not detected by CoCo. The reason CoCo did not accurately classify these patients as having a constitutional syndrome—in spite of CoCo's being trained to classify chief complaints with fever 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 posterior probabilities for those syndromes were higher than the probabilities for constitutional syndrome.

Thirteen of the febrile patients were detected by CoCo 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, CoCo also generated five false positive classifications for patients with chief complaints of *viral infection*, *dizziness*, *muscle aches*, and *weakness*.

4.2.3. Keyword HP algorithm

Keyword HP generated two false negatives and 11 false positives. 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.

Four false positives were due to contradictions in the record 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°C. 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 describing an undocumented sign rather than a possible diagnosis.

5. Discussion

Keyword HP detected 98% of the febrile patients in the study, which was significantly better than the sensitivity of either of the detectors that analyzed chief complaints ($p < 0.05$), suggesting that dictated history and physical examinations contain better information than chief complaints for fever detection. Improving sensitivity would be difficult, because of the nature of the two false negatives. One was a mistake by the gold standard physician and the other was a vague description of fever (“felt warm”) that may generate false positives if added to the fever keyword list. Specificity may be somewhat improved with improvements in the negation and hypothetical situation identification. Four of the 11 false positives (36%) were due to mistakes by the Keyword HP algorithm. However, seven of Keyword HP’s 11 false positives (73%) were due to mistakes by the gold standard physician or ambiguous information in the ED report.

We hypothesized that ED reports would be a reliable resource for locating information about a patient’s fever status, because a patient’s temperature is almost always taken and recorded in the ED. Nevertheless, the error analysis revealed that some physicians still referenced 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%). Because 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).

The sensitivity of the algorithms detecting fever from chief complaints was higher than we expected, given the limited nature of triage chief complaints. Keyword CC performed with higher sensitivity than CoCo and had perfect specificity, indicating that patients whose chief complaints explicitly mention a fever actually had a fever according to the gold standard classification based on review of the ED report. If we were to classify patients as febrile if either Keyword CC or CoCo assigned a positive classification, sensitivity would increase to 72.5% (79/109) and specificity would be identical to that of CoCo at 95% (99/104). CoCo’s classification performance of fever from chief complaints is equal to or better than that reported for classification of patients based on their chief complaints for respiratory, gastrointestinal, neurological, rash, and botulinic syndromes [36] (P. Gesteleand, M.M. Wagner, R. Gardner, R. Rolfs, W.W. Chapman, O. Ivanov., in preparation).

This report only studies accuracy of classification. Other considerations affect the decision about which

type of input data and which classification algorithms are optimal for a surveillance application, including the availability and completeness of the data [24]. There are often tradeoffs to be considered. For example, 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; however, sensitivity of detection from chief complaints is lower than from ED reports.

A solution that represents the best of both worlds might be a biosurveillance system that initially monitors febrile illness from chief complaints with Keyword CC. Results of this study suggest that Keyword CC will not generate false alarms. As ED reports become available, Keyword HP could find cases not detected as febrile by Keyword CC and update the surveillance system with the more complete and sensitive detection provided from ED reports, potentially detecting smaller outbreaks.

The fever detection algorithms described in this paper should be generalizable to domains outside of bioterrorism surveillance. Fever is an important physical sign that manifests itself in naturally occurring infectious diseases and in other entities, such as collagen vascular, neoplastic, and inflammatory bowel diseases. Automatically monitoring whether patients are febrile could influence differential diagnosis, hospital epidemiology, and therapeutic choices. Because most institutions do not have coded information about fever status, automated fever detection must be obtained from textual records. Our results indicate that fever detection from textual medical records such as chief complaints and ED reports is feasible using fairly simple natural language processing technologies.

5.1. Limitations and future research

Because we enriched our test set with potentially febrile patients, we were not able to calculate a valid positive predictive value for any of the fever classification algorithms. Prevalence of fever in ED patients is low enough that a study relying on random selection of patients would require many more reports to be classified by the gold standard physician. However, a study with random selection would present a more realistic understanding of the prevalence of fever in the population and would give us better insight regarding the false alarm rate generated by the algorithms.

Our study involved a single university hospital in the city of Pittsburgh. A fuller understanding of the potential of biosurveillance for outbreaks of febrile illness from free-text clinical data on a regional or national level would require an expanded study that evaluated the fever detection algorithms on data from other hospitals and cities—particularly for Keyword HP, because

linguistic variation in reporting may exist across the United States.

6. Conclusion

We measured the ability of three 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 algorithms using information from the chief complaint were 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.

Acknowledgments

We acknowledge Zhongwei Lu for his programming assistance. This work was partially funded by Defense Advanced Research Projects Agency (DARPA) Cooperative Agreement No. F30602-01-2-0550; Pennsylvania Department of Health Grant No. ME-01-737; and AHRQ Grant No. 290-00-0009.

References

- [1] Available from: <http://www.health.pitt.edu/rods/>. Accessed April 16, 2003.
- [2] Available from: <http://www.geis.ha.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp>. Accessed April 16, 2003.
- [3] Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. *J Am Med Inform Assoc* 2002;9(2):105–15.
- [4] Available from: <http://secure.cirg.washington.edu/bt2001amia/index.htm>. Accessed April 16, 2003.
- [5] Zelicoff A, Brillman J, Forslund DW, et al. The rapid syndrome validation project (RSVP), a technical paper. Sandia National Laboratories; 2001.
- [6] Available from: <http://www.nytimes.com/2003/04/04/nyregion/04WARN.html?ex=1050552000&en=2c63a52eb80102bc&ei=5070>. Accessed April 16, 2003.
- [7] 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.
- [8] Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Making* 2003;3(1):2.
- [9] Lazarus R, Kleinman K, Dashevsky I, 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.
- [10] 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(1):9.
- [11] Matsui T, Takahashi H, Ohyama T, et al. An evaluation of syndromic surveillance for the G8 Summit in Miyazaki and Fukuoka, 2000. *Kansenshogaku Zasshi* 2002;76(3):161–6.
- [12] 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.
- [13] Gesteland PH, Wagner MM, Chapman WW, 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;3:285–9.
- [14] Tsui FC, Espino JU, Wagner MM, et al. Data, network, and application: technical description of the Utah RODS winter olympic biosurveillance system. *Proc AMIA Symp* 2002:815–9.
- [15] Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC area. *Am J Prev Med* 2002;23(3):180–6.
- [16] Recognition of illness associated with the intentional release of a biologic agent. *MMWR Morb Mortal Wkly Rep* 2001;50(41):893–7.
- [17] Bravata DM, McDonald K, Owens DK, 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.
- [18] Health CDOP. Bioterrorism: a public health threat. *CD Info* 2000;2(5).
- [19] Monterey County Health Department Disaster Plan. Monterey, CA: Monterey County Health Department; October 9, 2001. Available from: <http://www.co.monterey.ca.us/health/Publications/pdf/MCHDDisasterPlan.pdf>. Accessed December 9, 2003.
- [20] Available from: <http://www.who.int/csr/sars/casedefinition/en/>. Accessed April 12, 2003.
- [21] Available from: <http://www.cdc.gov/ncidod/sars/casedefinition.htm>. Accessed April 12, 2003.
- [22] Chen SY, Su CP, Ma MH, et al. Predictive model of diagnosing probable cases of severe acute respiratory syndrome in febrile patients with exposure risk. *Ann Emerg Med* 2004;43(1):1–5.
- [23] Friedman C, Hripesak G. Natural language processing and its future in medicine. *Acad Med* 1999;74(8):890–5.
- [24] Mendonca EA, Cimino JJ, Johnson SB. Using narrative reports to support a digital library. *Proc AMIA Symp* 2001:458–62.
- [25] Chapman WW, Cooper GF, Hanbury P, Chapman BE, Harrison LH, Wagner MM. Creating a text classifier to detect radiology reports describing mediastinal findings associated with inhalational anthrax and other disorders. *J Am Med Inform Assoc* 2003;10(5):494–503.
- [26] Olszewski RT. Bayesian classification of triage diagnoses for the early detection of epidemics. In: *Recent Advances in Artificial Intelligence: Proceedings of the 16th International FLAIRS Conference, 2003*. AAAI Press; 2003. p. 412–6.
- [27] Cho PS, Taira RK, Kangaroo H. Automatic section segmentation of medical reports. *Proc AMIA Annu Fall Symp* 2003.
- [28] Haug PJ, Christensen L, Gundersen M, Clemons B, Koehler S, Bauer K. A natural language parsing system for encoding admitting diagnoses. *Proc AMIA Annu Fall Symp* 1997:814–8.
- [29] Taira RK, Bui AA, Kangaroo H. Identification of patient name references within medical documents using semantic selectional restrictions. *Proc AMIA Symp* 2002:757–61.
- [30] Friedman C. A broad-coverage natural language processing system. *Proc AMIA Symp* 2000:270–4.
- [31] Baud RH, Lovis C, Ruch P, Rassinoux AM. A light knowledge model for linguistic applications. *Proc AMIA Symp* 2001:37–41.

- [32] Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc* 2000;7(6):593–604.
- [33] Hahn U, Romacker M, Schulz S. MEDSYNDIKATE—a natural language system for the extraction of medical information from findings reports. *Int J Med Inf* 2002;67(1–3):63–74.
- [34] Taira RK, Soderland SG. A statistical natural language processor for medical reports. *Proc AMIA Symp* 1999:970–4.
- [35] Chapman WW, Christensen L, Wagner MM, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing. *AI Med* 2004 [in press].
- [36] Ivanov O, Wagner MM, Chapman WW, Olszewski RT. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. *Proc AMIA Symp* 2002:345–9.
- [37] Ivanov O, Gesteland P, Hogan W, Mundorff MB, Wagner MM. Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. *Proc AMIA Annu Fall Symp* 2003:318–22.
- [38] Travers DA, Waller A, Haas SW, Lober WB, Beard C. Emergency department data for bioterrorism surveillance: electronic data availability, timeliness, sources and standards. *Proc AMIA Symp* 2003:664–8.
- [39] Travers DA, Haas SW. Using nurses' natural language entries to build a concept-oriented terminology for patients' chief complaints in the emergency department. *J Biomed Inform* 2003;36(4–5):260–70.
- [40] 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.
- [41] Wong W, Moore AW, Cooper G, Wagner M. Rule-based anomaly pattern detection for detecting disease outbreaks. In: *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI-02)*. 2002.
- [42] Manning CD, Schütze H. *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press; 1999.
- [43] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. Evaluation of negation phrases in narrative clinical reports. *Proc AMIA Symp* 2001:105–9.
- [44] 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.