PUBLISHED BY THE INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Advances in Disease Surveillance

VOLUME 3

NUMBER 6

DATE 2007

ARTICLES

Can Chief Complaints Identify Patients with Febrile Syndromes?

Wendy W. Chapman and John N. Dowling

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA.

Received for publication May 26, 2006; accepted for publication May 3, 2007.

Background Syndromic surveillance systems often classify patients into syndromic categories based on emergency department (ED) chief complaints. There exists no standard set of syndromes for syndromic surveillance, and the available syndromic case definitions demonstrate substantial heterogeneity of findings constituting the definition. The use of fever in the definition of syndromic categories is arbitrary and unsystematic. We determined whether chief complaints accurately represent whether a patient has any of five febrile syndromes: febrile respiratory, febrile gastrointestinal, febrile rash, febrile neurological, or febrile hemorrhagic.

Methods We selected 1,557 patients admitted to the ED with discharge diagnoses potentially relevant to biosurveillance. We compared physician classification of the patients' chief complaints against criterion standard classifications from physician review of ED reports for five general syndromes (e.g., rash) and five febrile syndromes (e.g., febrile rash). We calculated sensitivity and specificity for general and febrile syndromes for the 1,557 cases.

Results Specificity for febrile and non-febrile syndromes was high. Sensitivity for the general syndromes ranged from 34 to 41%. Sensitivity for febrile syndromes ranged from 0 to 12%.

Conclusion Whereas chief complaints had modest sensitivity in predicting general syndromes correctly, they had poor sensitivity in predicting febrile syndromes. Respiratory, gastrointestinal, rash, neurological, and hemorrhagic syndromic case definitions for surveillance systems using chief complaints as input should not include fever.

chief complaints; disease outbreaks; fever; infection control; public health surveillance; surveillance; syndrome; syndromic surveillance

Abbreviations: ED, emergency department; GI, gastrointestinal illness; ICD-9, International Classification of Diseases, Ninth Revision; NLP, Natural language processing; NPV, negative predictive value; PPV, positive predictive value.

INTRODUCTION

Syndromic surveillance techniques often classify patients into syndromic categories based on emergency department (ED) chief complaints. 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, which suggests that an increase in the number of patients with a respiratory symptom

Correspondence to Wendy W. Chapman, Department of Biomedical Informatics, University of Pittsburgh, VALE M-183, 200 Meyran Avenue, Pittsburgh, PA 15260 (e-mail: chapman@cbmi.pitt.edu).

and fever may be more indicative of an infectious disease outbreak than an increase in the number of patients with respiratory syndromes without fever, for example. We do not know of any biosurveillance research supporting this assumption, but based on knowledge of infectious diseases, syndromic surveillance systems generally monitor febrile illness. However, the use of fever in the definition of syndromic categories is arbitrary and unsystematic. Some syndromic definitions include fever as one of a number of features that may occur in a syndromic category, some categorize fever as a separate syndrome, and some define febrile syndromes, requiring fever to be concurrently present with another presenting problem. We measured the predictive performance of manually classified chief complaints at identifying febrile syndromes.

Syndromic surveillance methods classify patients based on symptoms and signs that are available prior to a definitive diagnosis (1). Grouping cases into syndromes (e.g., respiratory syndrome) rather than into specific diagnoses (e.g., pneumonia) can provide earlier evidence of infection, because many diseases in their early phase have overlapping symptoms that may not initially alarm clinicians. For example, a retrospective analysis of a *Cryptosporidium* contamination in Milwaukee in 1993 that killed more than 100 people showed an unrecognized increase in the number of patients with diarrhea, weeks before the deaths (2).

Both manual (3, 4) and automated (5–16) syndromic surveillance systems have been developed; automated systems have been in use since 1999. A key element of a syndromic surveillance system's ability to detect outbreaks may be the syndromic definitions being monitored (17). Syndromic case definitions are broader than case definitions for the target diseases they encompass, but in the syndromic surveillance community there exists no standard set of syndromes to be monitored, and existing syndromic case definitions demonstrate substantial heterogeneity of findings constituting the definitions (18). For example, existing case definitions of respiratory syndrome vary depending on whether the definition includes upper respiratory symptoms, on how severe the respiratory illness must be to match the case definition, and on whether the patient must be febrile. Some syndromic surveillance systems classify patients into syndromic categories that explicitly include fever in their definitions (11, 19–22), while others do not (7, 9, 10, 13, 16, 17).

An additional consideration in designing a case definition is the need for timely detection of an outbreak (23). Near the end of a patient's visit to a healthcare facility, detailed information about the patient's clinical condition may be available from dictated reports and structured laboratory databases. However, when patients first present to a medical facility—which is the earliest point at which an outbreak could be detected from clinical data—only a small amount of information about their clinical conditions exists. Automated surveillance systems attempt to leverage pre-existing electronic ED data available upon admission, including date of admission, sex, age, address, coded admit diagnoses, and free-text triage chief complaint. Many automated syndromic surveillance systems use chief complaints to classify patients into syndromic case definitions (11, 18, 20, 21, 24, 25).

Our objective was to determine whether the information in chief complaints is sufficient for monitoring febrile syndromes by comparing manually classified chief complaints with the criterion standard of physician chart review for five febrile syndromes: febrile respiratory, febrile gastrointestinal, febrile rash, febrile neurological, and febrile hemorrhagic.

METHODS

We measured the performance of manual chief complaint classification into febrile syndromes by comparing chief complaint classification against criterion standard classification based on review of ED reports. The report is dictated by a physician after seeing the patient and details the patient's past history, chief complaint, physical findings, laboratory and radiology results and diagnoses. We also compared febrile syndromic classification (e.g., febrile respiratory) to general syndromic classification (e.g., respiratory), calculated what percentage of febrile syndromic cases could be identified by monitoring for general syndromes, and analyzed the content of chief complaints for febrile syndromic patients.

Setting

The study was conducted on data collected from the University of Pittsburgh Medical Center (UPMC) Presbyterian Hospital ED from December, 1990 to September, 2003. The ED at the UPMC Presbyterian Hospital admits approximately 40,000 adult patients a year, and patient visit data have been stored in the Medical Archival System (MARS) database since 1990, including free-text triage chief complaints, dictated and transcribed ED reports, and coded ICD-9 discharge diagnoses.

Selection of participants

Our goal was to select participants representing seven syndromes, including syndromes that occur infrequently. All patients with an electronic triage chief complaint and primary ICD-9 discharge diagnosis stored in MARS between December, 1990 and September, 2003 were eligible for this study, which was approved by the University of Pittsburgh's Institutional Review Board as exempt. From 527,228 eligible patients, we selected patients who potentially exhibited symptoms or signs consistent with the syndromes by filtering eligible patients according to their ICD-9 primary discharge diagnoses using a list of 831 unique ICD-9 codes. Table 1 shows examples of codes we used for each syndrome. From 96,818 potential participants, we selected a representative subset of approximately 200 patients per syndrome as described in (25). The resulting set included 1,557 cases for chart review by physicians who participated as criterion standard experts.

Criterion standard classification

Ten physicians board-certified in general internal medicine read ED reports for the 1,557 patients and applied case definitions described in (25) to classify patients into

Table 1. Examples of ICD-9 codes used for compiling a superset of positive patients

Syndrome (number of codes)	Examples of ICD-9 codes
Respiratory ($n = 287$)	022.1 (Pulmonary anthrax), 011 (Pulmonary tuberculosis), 480 (Viral pneumonia), 487 (Influenza), 163.3 Pneumocystosis, 033 (Whooping cough), 511 (Pleurisy), 786.05 (Shortness of breath)
Botulinic ($n = 60$)	005.1 (Botulism), 045.0 (Acute paralytic poliomyelitis, bulbar), 357 (Acute infective polyneuritis), 351.0 (Bell's palsy), 368.2 (Diplopia), 374.3 (Ptosis of eyelid), 787.2 (Dysphagia)
Gastrointestinal ($n = 119$)	003.0 (Salmonella gastroenteritis), 005 (Food poisoning, other bacterial), 006.1 (Infectious diarrhea), 007.4 (Cryptosporidiosis), 787.91 (Diarrhea)
Neurological ($n = 111$)	066.4 (West Nile fever), 331.81 (Reye's syndrome), 323 (Encephalitis, myelitis, and encephalomyelitis), 094.2 (Syphilitic meningitis), 320 (Bacterial meningitis), 780.01 (Coma), 784.0 (Headache)
Rash ($n = 99$)	022.0 (Cutaneous anthrax), 034.1 (Scarlet fever), 053 (Herpes zoster), 055 (Measles), 684 (Impetigo), 695.1 (Erythema multiforme)
Constitutional ($n = 66$)	002.0 (Typhoid fever), 075 (Infectious mononucleosis), 079.9 (Viral infection nos), 780.6 (Fever), 780.7 (Malaise and fatigue), 783.0 (Anorexia)
Hemorrhagic ($n = 89$)	065 (Arthropod hemorrhagic fever), 530.82 (Esophageal hemorrhage), 535.01 (Acute gastritis w/hemorrhage), 578.0 (Hematemesis), 599.7 (Hematuria)

syndromic categories (case definitions available at http://web.cbmi.pitt.edu/chapman/Syndromic_Case_Definitions.doc). They also determined whether the patient had a fever. Fever was defined as a measured temperature greater than 38.0˚C (100.4˚F) or a report of recent fever or chills. Every report was read by two physicians and by a third, if there were any disagreements on the report.

The resulting classifications by physicians allowed us to group patients into the following syndromic groups that have corresponding febrile syndromic groups: respiratory (congestion, shortness of breath, cough, etc.); gastrointestinal (nausea, vomiting, abdominal pain, etc.); rash (most rashes); hemorrhagic (bleeding from most sites); and neurological (non-psychiatric neurological symptoms, such as headache or seizure). To receive a positive classification for a febrile syndrome, the patient must have had a positive classification for both the syndrome and fever.

As described in (25), we measured agreement and reliability of the criterion standard classifications using methods described by Hripcsak (26). In this paper, we report the generalizability coefficients for the syndromes used in this study. A generalizability coefficient ranges from 0 to 1 and represents the reliability and reproducibility of the criterion standard. A coefficient above 0.70 is considered sufficient for most criterion standards.

Chief complaint classification

To avoid introducing errors in classification that may be caused by an automated chief complaint classifier, we manually classified chief complaints into syndromic categories. Author JND, who is a physician board-certified in internal medicine and infectious diseases with over 30 years of experience, read the chief complaint for each of the 1,557 patients and determined which of the syndromes was represented by the chief complaint and whether or not the complaint indicated that the patient had a fever. He applied the same syndromic case definitions as the physicians who generated the criterion standard classifications. However, he classified patients based only on their chief complaint (he was blinded to the ED reports), whereas the criterion standard physicians classified patients based on the ED report.

Outcome measures

The main outcome measure for our study was classification performance of the manually classified chief complaints when compared to criterion standard classification. We calculated the following outcome metrics and their 95% confidence intervals (27) for the febrile syndromes (e.g., febrile rash) and for the same syndromes without fever (e.g., rash):

- Sensitivity—the number of correct positive classifications divided by criterion standard positive classifications;
- Specificity—the number of correct negative classifications divided by criterion standard negative classifications;
- Positive predictive value (PPV)—the number of correct positive classifications divided by the total number of positive classifications;
- Negative predictive value (NPV)—the number of correct negative classifications divided by the total number of negative classifications.

Data analysis

We performed two data analyses. In the primary analysis, we addressed the question: What is the classification performance of chief complaints at identifying febrile syndromic cases when positive classification from the chief complaint requires both a syndromic and a febrile classification? For the primary data analysis we calculated the outcome measures by comparing manual classification from chief complaints into febrile and non-febrile syndromes against the analogous criterion standard classification.

A secondary analysis applied only to cases classified as positive by criterion standard for febrile syndromic definitions. The secondary analysis differed from the primary analysis in how chief complaints were classified—to be considered positive for a febrile syndrome, the chief complaint only needed to be positive for the syndrome (i.e., the chief complaint did not need to indicate fever). The secondary analysis addressed the question: What is the classification performance of chief complaints at identifying febrile syndromic cases when a positive case requires only a general syndromic classification? Table 2 illustrates the difference between the primary and secondary analyses.

The secondary analysis can help explain how well a syndromic surveillance system that does not include fever in its syndromic definitions can identify febrile syndromic cases. This analysis is important, because many syndromic surveillance systems only monitor for general syndromes and do not require fever in the chief complaint. Results of the secondary analysis reflect the sensitivity and precision (i.e., PPV) with which these systems will detect febrile syndromic cases. We expected sensitivity to be higher than was achieved in the primary analysis, because all syndromic cases are classified as febrile syndromic. We also expected PPV to be lower than in the primary analysis, because many of the cases classified as febrile syndromic would not actually have a fever.

To help explain the results of the secondary analysis, we examined what the chief complaints said for all cases classified by the criterion standard as positive for a febrile syndrome. We grouped positive cases into one of four mutually exclusive and exhaustive categories in which the chief complaint: 1) described a symptom consistent with the syndromic definition and described evidence of fever; 2) described a symptom consistent with the syndromic definition but not fever; 3) described evidence of fever but not of the syndrome; and 4) had no evidence of the syndrome or of fever.

Table 2. Difference between primary and secondary analyses

	Chief complaint classification	Criterion standard classification
Primary analysis	Syndrome + fever	Syndrome + fever
Secondary analysis	Syndrome	Syndrome + fever

RESULTS

The number of positive cases by criterion standard for the general syndromes ranged from 138 for rash to 633 for gastrointestinal (Table 3), with a mean of 455. The number of positive cases for febrile syndromes ranged from 34 for febrile rash to 196 for febrile respiratory, with a mean of 122. The criterion standard physicians agreed that fever was present in 315 (20%) of the 1,557 records. Table 3 shows the generalizability coefficients for three raters for febrile and non-febrile syndromes. All syndromes except for rash had a reliability coefficient higher than 0.70.

Primary data analysis

Table 4 shows the sensitivity, specificity, PPV, and NPV for chief complaint classification into syndromic categories

Table 3. Reliability of criterion standard classifications for febrile and non-febrile syndromes

Syndrome	Number of cases by criterion standard	Generalizability coefficient for three raters
Respiratory	607	0.87
Febrile respiratory	196	0.91
Gastrointestinal	633	0.89
Febrile gastrointestinal	181	0.94
Rash	138	0.63
Febrile rash	34	0.92
Hemorrhagic	328	0.86
Febrile hemorrhagic	52	0.89
Neurological	571	0.86
Febrile neurological	147	0.94

Table 4. Case classification performance of 1,557 manually encoded chief complaints for febrile and non-febrile syndromes. For a chief complaint to be classified as positive, the complaint required a symptom consistent with the syndromic definition and evidence of fever. For example, to be classified as febrile rash, the chief complaint required evidence of a rash and of fever. TP = true positives; FP = false positives; FN = false negatives. All outcome metrics are shown in percentages with 95% confidence intervals in brackets

Syndrome	No. of cases by criterion standard	No. of TP	No. of FP	No. of FN	Sensitivity	Specificity	PPV	NPV
Respiratory	607	231	18	376	38.1 [34–42]	98.1 [97–99]	92.8 [89–95]	71.3 [69–74]
Febrile respiratory	196	4	1	192	2.0 [0.8–5]	99.9 [99–100]	80.0 [38–96]	87.6 [86–89]
Gastrointestinal	633	215	20	418	34.0 [30–38]	97.8 [97–99]	91.5 [87–94]	68.4 [66–71]
Febrile gastrointestinal	181	12	0	169	6.6 [4–11]	100 [99–100]	100 [76–100]	89.1 [87–91]
Rash	138	53	17	85	38.4 [31–47]	98.8 [98–99]	75.7 [65–84]	94.3 [93–95]
Febrile rash	34	4	0	30	11.8 [5–27]	100 [99–100]	100 [51–100]	98.1 [97–99]
Hemorrhagic	328	134	14	194	40.9 [36–46]	98.9 [98–99]	90.5 [85–94]	86.2 [84–88]
Febrile hemorrhagic	52	0	0	52	0 [0–7]	100 [99–100]	*	96.6 [96–97]
Neurological	571	222	24	349	38.9 [35–43]	97.6 [96–98]	90.2 [86–93]	73.4 [71–76]
Febrile neurological	147	5	1	142	3.4 [2–8]	99.9 [99–100]	83.3 [44–97]	90.8 [89–92]

* Cannot calculate because of division by zero.

and febrile syndromic categories. Comparing chief complaint classification against the criterion standard for the syndromes (e.g., rash, neurological) yielded sensitivities ranging from 34 to 41% and specificities from 98 to 99%. Comparing chief complaint classification against criterion standard for febrile syndromes (e.g., febrile rash, febrile neurological) yielded sensitivities ranging from 0 to 12% and specificities near or at 100%. Classifying chief complaints based on presence of a syndromic symptom and fever identified true febrile syndromic cases with very low sensitivity. For febrile syndromic definitions, sensitivity was significantly lower than for the analogous non-febrile syndromic definitions, and NPV was significantly higher. Figure 1 plots the sensitivity of febrile and non-febrile syndromic classification from chief complaints.

Secondary analysis

Table 5 shows outcome metrics for febrile syndromic cases when chief complaints were classified based only on whether the complaint was consistent with the syndromic definition and not on the presence of fever. Sensitivity varied from 10% for febrile hemorrhagic to 37% for febrile respiratory. As expected, sensitivity increased for all syndromes when using a broader syndromic definition (i.e., not requiring the presence of fever in the chief complaint), and the differences were statistically significant for febrile respiratory, febrile gastrointestinal, and febrile neurological. PPV decreased significantly when using the broader definition, ranging between 3% for febrile hemorrhagic and 29% for febrile respiratory, with a median value of 22%.

In the secondary analysis, the case definition for chief complaint classification did not require fever in the chief complaint. Applying this case definition identified the majority of febrile syndromic cases but still missed some positive cases, because some of the positive cases had chief complaints that only indicated a fever without describing a symptom of the syndrome (e.g., “High temp”). Figure 2 shows that for all syndromes combined, only 5% of the chief complaints for positive patients described both a syndromic symptom and a fever, 23% described the syndrome without fever, and 19% described only fever. Over half of the chief complaints for positive cases described neither relevant syndromic complaints nor fever. Distribution of the four categories somewhat differed among the five febrile syndromes. The percentage of complaints describing neither the syndrome nor fever was highest for febrile hemorrhagic at 73%. The proportion of complaints describing only the syndrome was lowest for febrile hemorrhagic (5%) and highest for febrile respiratory (35%).

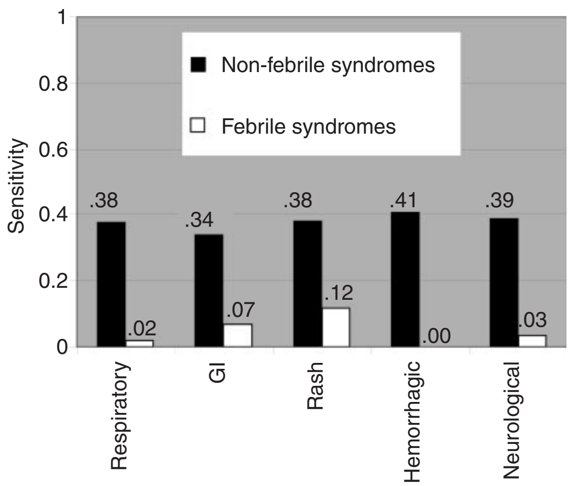

Figure 1. Sensitivity for febrile and non-febrile syndromic classification.

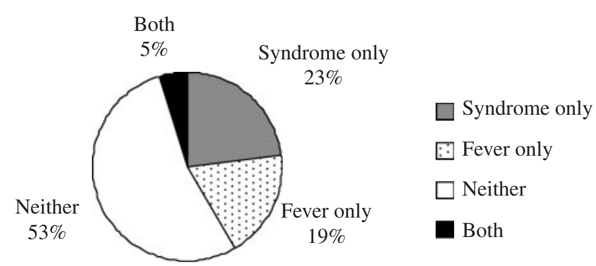

Figure 2. Content of chief complaints classified positive for febrile syndromes. “Syndrome only” indicates that the chief complaint only described symptoms consistent with the syndrome. “Fever only” indicates that the chief complaint described fever but did not describe symptoms consistent with the syndrome. “Neither” indicates that the chief complaint described neither the syndrome nor fever. “Both” is equivalent to sensitivity and indicates that the chief complaint described both the syndrome and fever.

Table 5. Case classification performance of 1,557 manually encoded chief complaints for febrile syndromes. For a chief complaint to be classified as positive, the complaint only required a symptom consistent with the general syndromic definition. For example, a case classified as Respiratory by the chief complaint would be considered positive for Febrile respiratory syndrome. TP = true positives; FP = false positives; FN = false negatives. All outcome metrics are shown in percentages with 95% confidence intervals in brackets

Syndrome	Number of cases by criterion standard	Num of TP	Num of FP	Num of FN	Sensitivity	Specificity	PPV	NPV
Febrile respiratory	196	72	177	124	36.7 [30–44]	87.0 [85–89]	28.9 [24–35]	90.5 [89–92]
Febrile gastrointestinal	181	52	183	129	28.7 [34–36]	86.7 [85–88]	22.1 [17–28]	90.2 [89–92]
Febrile rash	34	12	58	22	35.3 [22–52]	96.2 [95–97]	17.1 [10–28]	98.5 [98–99]
Febrile hemorrhagic	52	5	143	47	9.6 [4–21]	90.5 [89–92]	3.4 [2–8]	96.7 [96–98]
Febrile neurological	147	43	203	104	29.3 [23–37]	85.6 [84–87]	17.5 [13–23]	92.1 [91–93]

DISCUSSION

Chief complaint classification into febrile syndromes demonstrated poor sensitivity, ranging from 0 to 12% for all five febrile syndromes (Table 4). Based on this result, syndromic definitions should not require description of fever and a syndrome when using chief complaints as input.

Poor classification performance into febrile syndromes is consistent with the nature of triage chief complaints in several ways. First, on presentation to the ED, a patient may or may not know whether she has a fever. Second, she may have taken her temperature at home and may have self-treated with an anti-pyretic medication and so does not complain of fever. Third, clerks and nurses who enter chief complaints create short snippets of text and, in many cases, are allowed to enter only a limited number of characters. In an attempt to be brief, the nurses or clerks sometimes utilize creative abbreviations, summarize the complaints in their own words, and make decisions about the most important or relevant complaint in order to reduce the size of the textual entry (28–30). Therefore, some complaints conveyed by the patient may not be included in the entry.

If a febrile syndromic definition should not require description of fever and a syndrome when using chief complaints as input, what is the best method for detecting patients with febrile syndromes? There are a few options. First, we could monitor for broad syndromic definitions and hope to capture many of the febrile syndromic cases. This is the approach taken by many automated syndromic surveillance systems. In this study, sensitivity at detecting febrile syndromic cases, when monitoring for general syndromes, ranged from 34 to 41% with PPVs between 3 and 29%. Because we manually classified chief complaints into syndromic categories, the outcome measures we present are a ceiling level we would not expect an automated method to surpass. However, automatic chief complaint classifiers perform well when classifying chief complaints into syndromic categories (31–33). Our results indicate that monitoring for increases in the number of patients with general syndromes would capture over a third of the febrile syndromic cases. Figure 2 shows that about half of the positive febrile syndromic cases were potentially identifiable by chief complaint classification, because they had chief complaints describing either a syndrome, fever, or both. Of the potentially identifiable cases, about 57% could be identified by monitoring for broad syndromic definitions. The remaining 43% had chief complaints that only described fever.

Thus, a second option is to monitor chief complaints for fever—monitoring for fever would identify approximately 22% of the febrile syndromic cases in this study. Based on a previous study, a simple keyword search for fever identified 100% of study cases whose chief complaints described fever (34), so it is technically feasible to identify the febrile syndromic cases whose chief complaints only describe fever. However, it is not possible to classify patients with chief complaints only indicating a fever into syndromic categories such as febrile respiratory.

Monitoring for general syndromes and for fever would only capture about half of the febrile syndromic cases. The other half could not be identified by their chief complaints, because the chief complaints described symptoms not related to their true syndromic classification. Based on these results, chief complaint classification—regardless of the syndromic definition—cannot identify febrile syndromic cases with high sensitivity. Increasing sensitivity would require considering more detailed clinical information. After admission to the ED, a patient is examined by a physician, and the physician dictates an ED note detailing the patient's past history, chief complaint, physical findings, and diagnoses. At the University of Pittsburgh Presbyterian Hospital, ED reports are electronically available approximately 12 hours after the patient is admitted to the ED.

Extracting detailed clinical information from ED reports promises to increase sensitivity of detection of febrile syndromic cases. However, because ED reports are available later than chief complaints are, ED reports would not provide as timely a signal for an outbreak. Moreover, ED reports are more difficult than chief complaints to automatically process. In contrast to the short phrases of chief complaints, ED reports are complex narratives involving hundreds of sentences that display negation, uncertainty, and changes over time. Natural language processing (NLP) techniques have been successfully applied to narrative clinical reports for decision support (35–38) and quality assurance (39). To our knowledge, nobody has published evaluations comparing chief complaint classification of patients to classification from ED reports. However, we have shown that we can automatically identify febrile patients from ED reports with a sensitivity of 98% (34), and we are currently working on automated natural language processing methods for extracting respiratory-related information from ED reports for syndromic surveillance (40–43). Given the fact that input to most surveillance systems currently consists of only free-text chief complaints, the best way to monitor for febrile syndromic cases may be to initially classify patients into broad syndromic categories and into fever status based on their chief complaints, then to update the probability of having a febrile syndrome as ED reports become available.

Simple changes in triage data collection could potentially improve surveillance of febrile syndromes. For instance, some hospitals measure and record temperature during triage (e.g., L.D.S. Hospital in Salt Lake City, Utah and UNC Healthcare, Chapel Hill, North Carolina). Access to both a chief complaint and a measured temperature would probably improve our ability to accurately identify febrile syndromic patients. Structured chief complaint pick-lists could also help provide more complete triage data if the interface were designed to elicit more than one presenting problem and patients were asked specifically about fever in addition to other problems. Several sets of standardized chief complaints have been developed and implemented in EDs (44–46).

Future research for successful monitoring of febrile syndromic cases should include 1) designing algorithms that can combine evidence from general syndromes and from febrile patients and 2) developing and evaluating NLP techniques that can accurately extract relevant clinical information from text and can model the negation and temporal status of the conditions.

Our results apply to five febrile syndromic case definitions, but the results solicit the question of whether chief complaint classification would show poor sensitivity for any syndromic case definition that was narrowly defined. A case definition that requires multiple symptoms to be considered positive may be too narrowly defined when using chief complaints for classification. For example, a possible definition of Meningoencephalitic syndrome is stiff neck and headache, but it is unlikely a single chief complaint would describe both symptoms. A syndromic definition that required a single finding to be positive may also demonstrate poor sensitivity when using chief complaint classification. For instance, in looking for a GI (gastrointestinal illness) outbreak, a surveillance system may not lump all GI symptoms together but may attempt to monitor diarrhea separately from vomiting. In a previous study, we measured the sensitivity for diarrhea at 10% and for vomiting at 15%, which is three to four times lower than sensitivity for the broader GI syndromic definition (40%) (47). Chief complaints are variable in their quality and represent a limited version of a patient's problem when admitted to the ED. When defining syndromic definitions for chief complaint classification, the definitions should not be too narrow.

Limitations

We used ICD-9 discharge diagnoses to locate potentially positive cases to increase the prevalence of true syndromic cases in the test set. Therefore, our test set did not represent the natural prevalence of the syndromes. Because physician review is expensive, an advantage of artificially increasing the prevalence in a test set is decreasing the number of reports physicians need to read to find a reasonable number of positive cases. By including more positive cases, the confidence intervals of the outcome measures are reduced. Artificially increasing prevalence does not usually change measures of sensitivity and specificity but does affect measures of PPVs and NPVs. Therefore the PPV and NPV measures we report are not representative of a natural distribution of the syndromes.

Only a single physician classified the chief complaints into syndromes. Evaluations of NLP systems using physician judgment as a reference standard typically employ multiple physician judgments (42). It can be argued, however, that acting as a gold standard for this task requires project-specific expertise about syndromic definitions. The same physician who classified chief complaints also devised the syndrome definitions and trained the three physicians who read ED reports and classified patients into syndromic categories. This assured that the criterion standard physicians were all performing the same task.

Another limitation to our study is that sensitivities of chief complaint classification for general syndromes were lower than in other published studies. Many of the published studies on chief complaint case classification have used ICD-9 codes as the criterion standard (48, 49), whereas a few studies have used review of the ED report (50, 51). Sensitivities reported in this paper for many syndromes are lower than those for analogous studies using review of the ED report as criterion standard. Ivanov and colleagues (50) measured the sensitivity of a Bayesian chief complaint classifier against review of ED reports for acute infectious GI syndrome at 63%. Chang and colleagues (51) measured the sensitivity with which a keyword-based automated chief complaint classifier could identify several general syndromes. They demonstrated sensitivities of 71% for respiratory, 64% for GI, and 51% for Neurological syndromes, whereas sensitivities for manual classification into the same syndromes in this study were 38, 34, and 39%, respectively. In one syndrome (rash), we showed a higher sensitivity (38%) than that obtained by Chang (20%).

Both of these previous studies applied random selection to select cases. In the Ivanov study, there were only 22 positive cases, and the 95% confidence interval of sensitivity overlapped with that shown in this study. The Chang study did not list confidence intervals or the number of positive cases for each syndrome, so there is no way to determine whether our results were statistically different. Still, sensitivities in this study were lower than we anticipated.

Conclusion

An important element of understanding the effectiveness of syndromic surveillance is measuring the relationship between the data used as input and the case definition being monitored. Designing an appropriate case definition involves consideration of the tradeoff between sensitivity and specificity of detection of infected patients. On the one hand, a broad case definition inclusive of all signs, symptoms, and findings that may be experienced by an infected patient will be sensitive at detecting affected patients but will also detect uninfected patients with similar symptoms. On the other hand, a narrow case definition that only detects patients with verified evidence of a disease will almost never generate a false alarm but will have very low sensitivity and detection may occur too late in the process of the outbreak to effectively intervene (52). We found that sensitivity of chief complaint classification into five febrile symptoms was poor, suggesting that case definitions that use chief complaints as input should be more broadly defined.

ACKNOWLEDGMENTS

This work was funded by Defense Advanced Research Projects Agency (DARPA) Cooperative Agreement No. F30602-01-2-0550, Pennsylvania Department of Health grant number ME-01-737, AHRQ Grant No. 290-00-0009, and NLM K22 LM008301. These results were presented at the Syndromic Surveillance Conference in Seattle, Washington, September 2005.

REFERENCES

1. http://www.cdc.gov/epo/dphsi/phs/syndromic.htm. Accessed April 22, 2003.
2. Proctor ME, Blair KA, Davis JP. Surveillance data for waterborne illness detection: an assessment following a massive waterborne outbreak of *Cryptosporidium* infection. Epidemiol Infect 1998;120:43–54.

3. Kortepeter MG, Pavlin JA, Gaydos JC, Rowe JR, Kelley PW, Ludwig G, McKee KT Jr, Eitzen EM. Surveillance at US military installations for bioterrorist and emerging infectious disease threats. Mil Med 2000;165:ii–iii.
4. Moran GJ, Talan DA. Update on emerging infections: news from the Centers for Disease Control and Prevention. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. Ann Emerg Med 2003;41:414–8.
5. Lober WB, Karras BT, Wagner MM, Overhage JM, Davidson AJ, Fraser H, Trigg LJ, Mandl KD, Espino JU, Tsui FC. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc 2002;9:105–15.
6. Zelicoff A, Brillman J, Forslund DW, George JE, Zink S, Koenig S, et al. The rapid syndrome validation project (RSVP), a technical paper. Albuquerque, NM: Sandia National Laboratories, 2001.
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 Mak 2003;3:2.
9. Lazarus R, Kleinman K, Dashevsky I, Adams C, Kludt P, DeMaria A Jr, Platt R. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis 2002;8:753–60.
10. Matsui T, Takahashi H, Ohyama T, Tanaka T, Kaku K, Osaka K, Chijiwa K, Iwaki U, Okabe N. [An evaluation of syndromic surveillance for the G8 Summit in Miyazaki and Fukuoka, 2000]. Kansenshogaku Zasshi 2002;76:161–6.
11. 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;41:447–52.
12. Lewis MD, Pavlin JA, Mansfield JL, O'Brien S, Boomsma LG, Elbert Y, Kelley PW. Disease outbreak detection system using syndromic data in the greater Washington DC area. Am J Prev Med 2002;23:180–6.
13. 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.
14. Bravata DM, McDonald KM, Smith WM, Rydzak C, Szeto H, Buckeridge DL, Haberland C, Owens DK. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Intern Med 2004;140:910–22.
15. 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;10:399–408.
16. Gesteland PH, Gardner RM, Tsui FC, Espino JU, Rolfs RT, James BC, Chapman WW, Moore AW, Wagner MM. Automated syndromic surveillance for the 2002 Winter Olympics. J Am Med Inform Assoc 2003;10:547–54.
17. Reis BY, Mandl KD. Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. Ann Emerg Med 2004;44:235–41.
18. Graham J, Buckeridge D, Choy M, Musen M. Conceptual heterogeneity complicates automated syndromic surveillance for bioterrorism. Proc AMIA Annu Fall Symp 2002:1030.
19. Brillman J, Joyce E, Forslund D, et al. The bio-surveillance analysis, feedback, evaluation and response (B-SAFER) system. National Syndromic Surveillance Conference 2002. (Available from: http://www.nyam.org; http://www.nyam.org; accessed Aug 23, 2004).
20. Forslund DW, Joyce EL, Burr T, Picard R, Wokoun D, Umland E, Brillman JC, Froman P, Koster F. Setting standards for improved syndromic surveillance. IEEE Eng Med Biol Mag 2004;23:65–70.
21. Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, Sari J, Sniegoski C, Wojcik R, Pavlin J. A systems overview of the electronic surveillance system for the early notification of community-based epidemics (ESSENCE II). J Urban Health 2003;80(Suppl 1):i32–42.
22. Dembek Z, Myers M, Carley K, Hadler J. Connecticut hospital admissions syndromic data. J Urban Health 2003;80(Suppl 1): i121–22.
23. Wagner MM, Tsui FC, Espino JU, Dato VM, Sittig DF, Caruana RA, McGinnis LF, Deerfield DW, Druzdzel MJ, Fridsma DB. The emerging science of very early detection of disease outbreaks. J Public Health Manag Pract 2001;7: 51–9.
24. Wagner MM, Espino J, Tsui FC, Gesteland P, Chapman W, Ivanov O, Moore A, Wong W, Dowling J, Hutman J. Syndrome and outbreak detection using chief-complaint data—experience of the Real-Time Outbreak and Disease Surveillance project. MMWR Morb Mortal Wkly Rep 2004;53(Suppl): 28–31.
25. Chapman WW, Dowling JN, Wagner MM. Generating a reliable reference standard set for syndromic case classification. J Am Med Inform Assoc 2005;12:618–29.
26. Hripcsak G, Kuperman GJ, Friedman C, Heitjan DF. A reliability study for evaluating information extraction from radiology reports. J Am Med Inform Assoc 1999;6:143–50.
27. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 1927;22:209–212.
28. 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.
29. 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;36:260–70.
30. Shapiro AR. Taming variability in free text: Application to health surveillance. MMWR Morb Mortal Wkly Rep 2004;53(Suppl):95–100.
31. Olszewski RT. Bayesian classification of triage diagnoses for the early detection of epidemics. In: Proceedings of the 16th International FLAIRS Conference, St. Augustine, Florida, 2003. AAAI Press, 2003. pp. 412–416.
32. Chapman WW, Christensen LM, Wagner MM, Haug PJ, Ivanov O, Dowling JN, Olszewski RT. Classifying free-text triage chief complaints into syndromic categories with natural language processing. Artif Intell Med 2005;33:31–40.
33. Sniegoski CA. Automated syndromic classification of chief complaint records. Johns Hopkins APL Technical Digest 2004;25:68–75.
34. Chapman WW, Dowling JN, Wagner MM. Fever detection from free-text clinical records for biosurveillance. J Biomed Inform 2004;37:120–7.
35. Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. Proc AMIA Symp 1999:256–60.
36. 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;7:593–604.
37. Aronsky D, Fiszman M, Chapman WW, Haug PJ. Combining decision support methodologies to diagnose pneumonia. Proc AMIA Symp 2001:12–6.

38. Knirsch CA, Jain NL, Pablos-Mendez A, Friedman C, Hripcsak G. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system. Infect Control Hosp Epidemiol 1998;19: 94–100.
39. Fiszman M, Haug PJ, Frederick PR. Automatic extraction of PIOPED interpretations from ventilation/perfusion lung scan reports. Proc AMIA Symp 1998:860–4.
40. Chapman WW, Fiszman M, Dowling JN, Chapman BE, Rindflesch TC. Identifying respiratory findings in emergency department reports for biosurveillance using MetaMap. Medinfo 2004;11(Pt 1):487–91.
41. Chu D, Dowling JN, Chapman WW. Evaluating the effectiveness of four contextual features in classifying annotated clinical conditions in emergency department reports. AMIA Annu Symp Proc 2006:141–5.
42. Chu D. Clinical Feature Extraction from Emergency Department Reports for Biosurveillance [Master's Thesis]. Pittsburgh: University of Pittsburgh, 2007.
43. Chapman WW, Chu D, Dowling JN. ConText: An algorithm for identifying contextual features from clinical text. In: BioNLP Workshop of the Association for Computational Linguistics Prague, Czech Republic; 2007. (In press).
44. Aronsky D, Kendall D, Merkley K, James BC, Haug PJ. A comprehensive set of coded chief complaints for the emergency department. Acad Emerg Med 2001;8:980–9.
45. Barthell EN, Aronsky D, Cochrane DG, Cable G, Stair T. The frontlines of medicine project progress report: standardized communication of emergency department triage data for syndromic surveillance. Ann Emerg Med 2004;44:247–52.
46. Thompson DA, Eitel D, Fernandes CM, Pines JM, Amsterdam J, Davidson SJ. Coded chief complaints—automated analysis of free-text complaints. Acad Emerg Med 2006;13:774–82.
47. Chapman WW, Dowling JN, Espino JU, Wagner MM. Chief complaint detection of syndromic cases with broad and narrow case definitions. Technical Report, CBMI Report Series 2004.
48. Beitel AJ, Olson KL, Reis BY, Mandl KD. Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. Pediatr Emerg Care 2004;20:355–60.
49. Chapman WW, Dowling JN, Wagner MM. Classification of emergency department chief complaints into seven syndromes: a retrospective analysis of 527,228 patients. Ann Emerg Med 2005;46:445–455.
50. 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.
51. Chang HG, Cochrane DG, Tserenpuntsag B, Allegra JR, Smith PF. ICD9 as a surrogate for chart review in the validation of a chief complaint syndromic surveillance system. Advances in Disease Surveillance 2006;1:11.
52. Mocny M, Cochrane D, Allegra J, Nguyen T, Pavlin J, Rothman J, Heffernan R. Improving agreement between two algorithms for biosurveillance of respiratory disease in the emergency department—chief complaint and ICD-9 code. MMWR Morb Mortal Wkly Rep 2003;53(Suppl):153.