Automated Syndromic Classification of Chief Complaint Records

Carol A. Sniegoski

Syndromic surveillance, a medical surveillance approach that bins data into broadly defined syndrome groups, has drawn increasing interest in recent years for the early detection of disease outbreaks for both public health and bioterrorism defense. Emergency department chief complaint records are an attractive data source for syndromic surveillance owing to their timeliness and ready availability in electronic form. As part of the ESSENCE prototype biosurveillance system, APL has developed an automated process for classifying chief complaint data into syndrome groups, thus increasing the feasibility of chief complaint–based surveillance by avoiding laborious manual classification. Preliminary studies indicate positive results in classifying daily chief complaint data feeds from multiple hospital sources.

INTRODUCTION

The intent of biosurveillance is to detect a disease outbreak at the earliest possible time. Traditionally, detection has relied primarily on the observations of astute clinicians or laboratory-confirmed diagnoses. These approaches, however, may miss the early period of an outbreak, as several days may elapse before affected individuals seek medical treatment from traditional health care providers. In addition, terrorist biological warfare attacks may cause symptoms that, in early stages, mimic flu or other traditional disease outbreaks. The early detection of a biological warfare attack could result in significant savings of lives and resources owing to earlier intervention and containment of an outbreak.

Recent biosurveillance methods explore two innovations. The first is syndromic surveillance. Instead of focusing on relatively late-arriving, highly specific disease indicators such as diagnoses or laboratory reports, broad syndrome groups are defined that group both specific and non-specific indicators. Longitudinal behaviors of data counts in each group are analyzed to obtain early indication of outbreaks. The second innovation in biosurveillance is the integration of traditional and non-traditional health information. The latter encompasses information that either is considered outside the health domain or is not normally available in a timely manner. Non-traditional information includes over-the-counter pharmacy sales, school absenteeism, animal health data, outpatient medical visit insurance claims data, and emergency room chief complaint data.

A biosurveillance prototype system exploring these methods, ESSENCE (Electronic Surveillance System for the Early Notification of Community-based Epidemics), has been under development over the past several years as a joint project between APL and the Department of Defense's Global Emerging Infections System (DoD-GEIS) program.[1] The system merges data from both the military and civilian sectors to form a more complete understanding of regional health. Currently, DoD-GEIS operates a prototype version for the early detection of infectious disease outbreaks at military treatment facilities. Civilian data sets are being collected from various sites across the country, including the National Capital Area (Washington, D.C., Maryland, and Virginia), to demonstrate how these data can be integrated for early disease outbreak detection.

OVERVIEW OF THE CHIEF COMPLAINT CLASSIFICATION TASK

Chief complaint records are an attractive data source for surveillance because they are relatively timely and easy to obtain. The chief complaint, which describes the patient's stated reason for his or her visit to the emergency room, is recorded as part of a standard hospital procedure by the triage nurse upon the patient's arrival. Although some hospitals keep paper records, computer-based electronic chief complaint records are more common. Thus, the chief complaint is a natural product of the medical system, can (usually) be obtained in electronic form, and is available immediately upon admission to the emergency room.

Unlike most types of non-traditional data explored in ESSENCE, there is no standardized coding system for chief complaint data. Every type of over-the-counter pharmacy product, for example, is identified by a unique 10-digit NDC (National Drug Code) number that can be used to assign it to a syndrome group for surveillance. Similarly, every distinct medical claims diagnosis is described by a unique ICD-9 code (the World Health Organization's International Classification of Diseases, version 9). The chief complaint, however, is a free-form text field. Classifying chief complaint into syndrome groups is thus a text classification problem.

Although other syndromic classification systems exist,[2,3] the discussion in this article refers to the system developed at Maryland's Department of Hygiene and Mental Health (DHMH) that consists of eight groups (Table 1): Death (unexplained sudden death), Gastrointestinal (non-specific abdominal and gastrointestinal complaints), Neurological (complaints related to stroke, vision problems, paralysis, etc.), Rash (skin rash of unknown cause), Respiratory (lower respiratory or breathing problems), Sepsis (sepsis or septicemia), Unspecified Infection (non-specific fever or flu-like symptoms), and Other, which includes records not assigned to another group. Records not belonging in the Other category are referred to here as "records of interest." Each syndrome group is defined by a listing of individual symptoms and diseases to be included in that group. Exclusionary symptoms, such as alcohol use or trauma, and constraints on features, such as patient age, may also apply. A chief complaint may belong to more than one syndrome group, either because it contains one complaint belonging to multiple categories or because it includes multiple distinct complaints. Under this classification system, the percentage of records falling in each syndrome group ranges from approximately 12% for the most common syndromes to less than 0.1% for the rarest (Fig. 1).

The correspondence between chief complaint and other medical data sources has been examined elsewhere. Studies include the association between chief

Table 1. Syndrome classification system developed at the Maryland Department of Hygiene and Mental Health.

Syndrome	Example chief complaint data
Death	Cardiac Arrest POSS HEART ATTACK PCP REFERRED
Gastr	Coughing/Vomiting Fever, Vom R/o Meningitis R/o Gi Bleed Ab Pain
Neuro	Fever, Vom R/o Meningitis Change Mental Status, Noteating
Rash	Spots/welts All Over Body Exposed To Chickenpox Rash cough
Respi	Back pain/sob SORE THROAT, CONGESTION, TROUBLE BREA PNEUMONIA
Sepsi	R/o Sepsis VOMITING FEVER POSS SEPSIS
Unspe	High Fever Weak Noteating Or Drinking Chills Cold Sweats
Other	INJURY R HAND MVC PAIN JAW NECK FELL AMBO

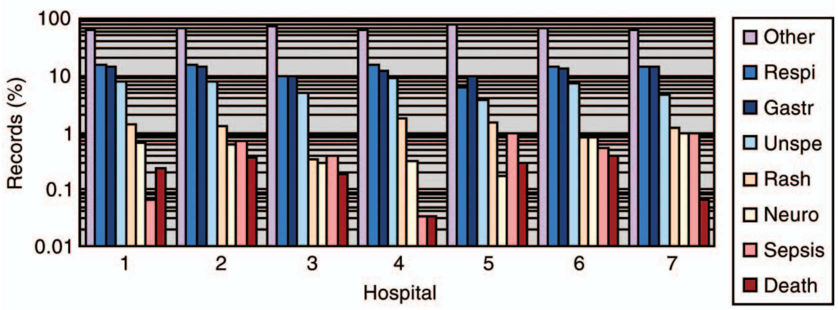


Figure 1. Syndrome distributions by hospital.

The labor-intensive nature of manual chief complaint categorization has led to attempts to automate the process. The simplest may be NYCDOH's SAS-based coding system, which uses simple keyword and phrase-based pattern matching.[9] (No accuracy estimates are available.) The University of Pittsburgh's Real-time Outbreak and Disease Surveillance (RODS) system[10] uses a Bayesian classifier and reports areas under the ROC (receiver-operating characteristic) curve from 0.80 to 0.97, depending on the syndrome.[11]

complaint–based syndromes and ICD-9–based syndromes[4,5] and the association between chief complaint and true disease as determined by physician review of patient discharge records.[6,7] Although these are important aspects of surveillance, the current discussion focuses on classifying chief complaint records by syndrome group.

Some syndromic surveillance efforts other than ESSENCE[2,8] categorize chief complaint data by hand. Manual categorization can be performed either onsite at individual hospitals or offsite at central facilities such as public health departments. Onsite categorization has the drawback of increasing already overburdened emergency department staffs' workloads and may be marked by low or sporadic compliance. For example, the New York City Department of Health (NYCDOH) instituted short-term syndromic surveillance at 15 city hospital emergency departments after the events of 9/11. Clinicians were asked to complete paper forms classifying patients into syndrome groups. Despite the round-the-clock presence of health department personnel acting as prompters, compliance hovered near 60%.[9]

On the other hand, Sandia National Laboratory's Rapid Syndromic Surveillance Project (RSVP), which requires emergency department clinicians to use a Web-based interface to record syndrome data for each patient, has unofficially reported surprisingly high compliance levels of nearly 90%. Given an approach such as RSVP's, with its estimated data entry time of 1 min per patient, however, onsite manual coding would represent some 3 man-hours of labor per day for a typical Washington, D.C., metropolitan area hospital's emergency department patient load of approximately 200 patients daily. Offsite coding reduces the compliance problem since hospitals merely have to transfer raw data logs, but it requires comparable amounts of labor. The Maryland DHMH spends an estimated 4.5 man-hours per day manually coding chief complaint logs faxed in from area hospitals (personal communication, David Blythe, State Epidemiologist, Maryland DHMH).

APL'S APPROACH TO CHIEF COMPLAINT CLASSIFICATION

Processing Electronic Records

Chief complaint data are characterized by brevity, irregular punctuation, misspellings, numerous abbreviations, unreliable syntax, and variable word choices. APL's approach to chief complaint classification attempts to either cope with or exploit each of these features.

Electronic records are processed in three sequential steps:

1. Normalize the text.
2. Establish memberships in lower-level syndrome categories using weighted keyword matching.
3. Build higher-level categories out of lower-level ones.

The first step removes punctuation and digits and expands abbreviations. Although punctuation can yield clues about distinguishing multiple symptoms within a record, its use in chief complaint classification is so idiosyncratic that it is simply removed. Digits, too, may contribute to meaning, but the number of uninterpretable numeric codes appearing in chief complaint data is large enough that they too are removed.

Abbreviations appear extensively in chief complaint data. A review of a day's worth of chief complaint logs from area hospitals shows that nearly half of all records contain one or more abbreviations (Fig. 2). Of those records classified in non-Other syndrome groups, over a third contain abbreviations critical to correct

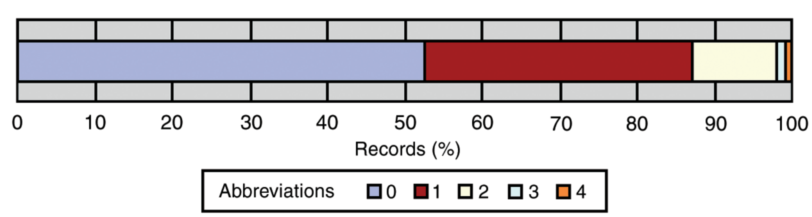


Figure 2. Percentage of chief complaint records containing abbreviations for all syndrome groups. Data were obtained from a review of 1 day's worth of records (1017 total, including Other).

classification (Fig. 3). When abbreviations do appear in records of interest, they are critical to correct classification 82% of the time. Thus, handling abbreviations appropriately is necessary for accurate classification.

Although numerous compilations of medical abbreviations exist, they unfortunately cannot be applied in a straightforward manner to interpret abbreviations in chief complaints. Abbreviation usage varies among hospitals and even within hospitals (Fig. 4). Many abbreviations have multiple definitions that need to be distinguished by context in practice (Fig. 5). Of the 770 abbreviations in APL's application that are compiled into an internal dictionary to expand abbreviations into their longer forms, 167 have had to be defined contextually.

In the second major processing step for classifying chief complaint, memberships in lower-level syndrome categories are determined for each record. These categories generally represent a single symptom or a related cluster of symptoms. Weighted keyword matching is used, an approach that resembles the well-established vector cosine method for determining relevance among documents. Chief complaint categorization is a document filtering process: individual chief complaint records act as the documents being filtered, and syndrome category descriptions act as filters. If a chief complaint category is deemed sufficiently relevant to a syndrome category, the chief complaint is said to belong to that syndrome category. The keywords and weights have been compiled manually, although in principle they could be determined statistically if given adequately representative training data.

The weighted keyword approach makes two simplifying assumptions: (1) it treats all words in a record as equally important, and (2) it ignores syntax and word order. Simplifying the classification task in this manner exploits the brevity and non-repetitious nature of chief complaint documents, which average between two and four words long across all syndromes (Fig. 6), although it can be seen that different hospitals trend longer or shorter (Fig. 7). A different approach might be required for classifying longer, more structured medical documents.

For many syndromes, the majority of instances appear as canonical phrases—for example, "abdominal pain" or an equivalent abbreviation such as "abd pn" for the abdominal pain symptom. These phrases could conceivably be classified using simpler unweighted

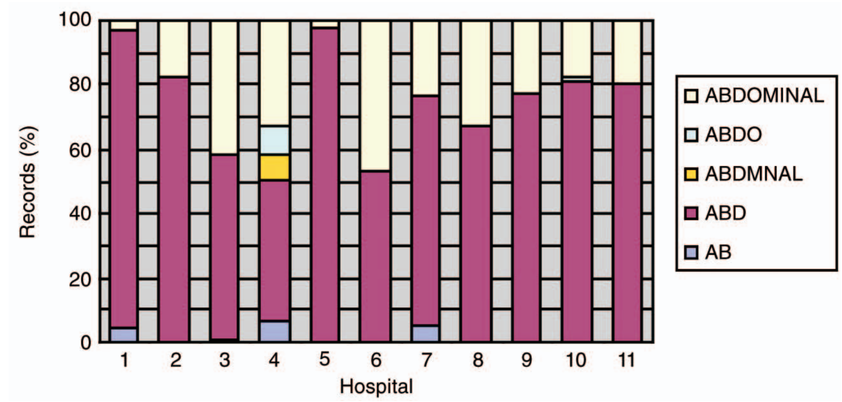


Figure 4. Abbreviations for "abdominal" by hospital.

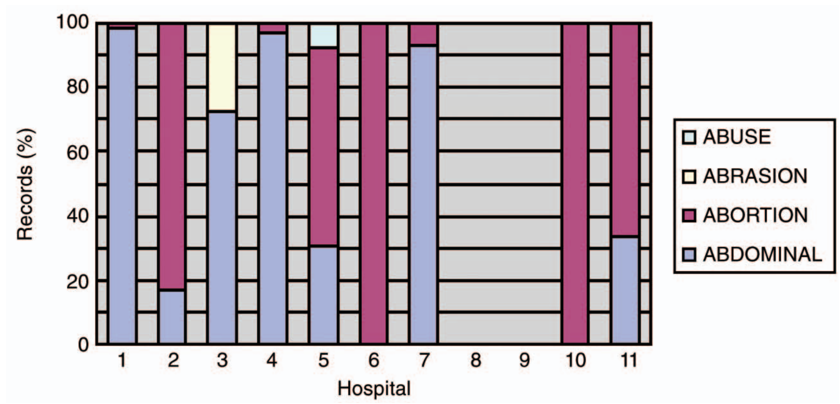


Figure 5. Meanings for the "AB" abbreviation by hospital.

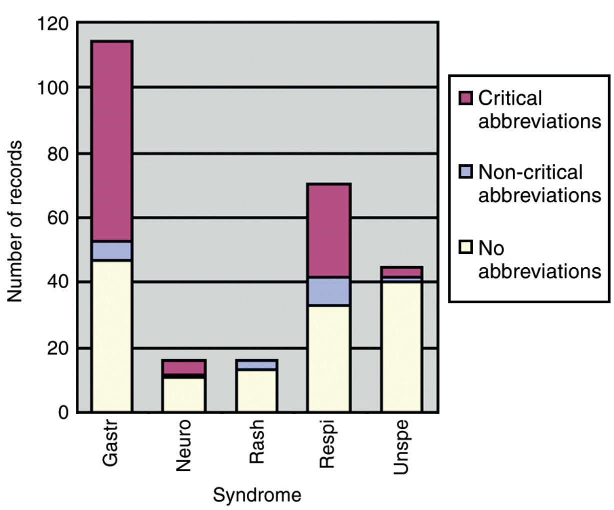


Figure 3. Number of critical and non-critical abbreviations per chief complaint record for non-Other syndrome groups. A critical abbreviation is one whose correct interpretation is necessary for correct classification of the record in which it appears. Data were obtained from a review of 1 day's worth of records (1017 total, including Other).

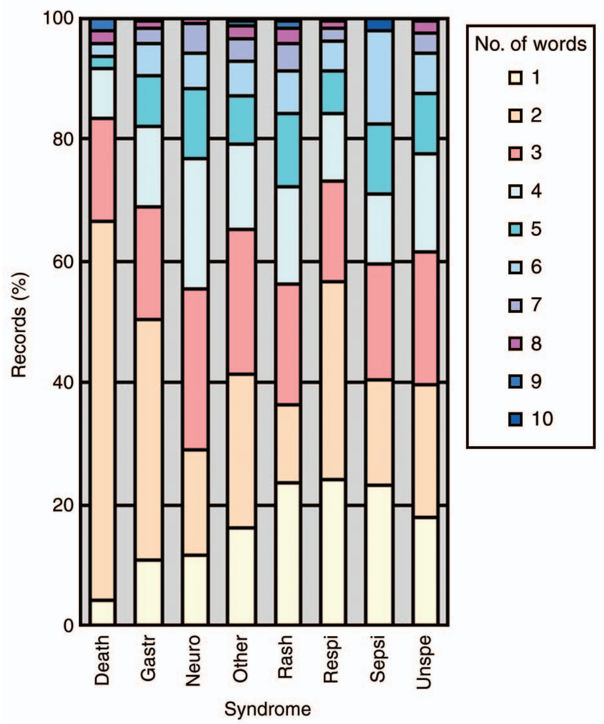


Figure 6. Length of chief complaint by syndrome for all hospitals.

keyword or phrase-based matching. However, a significant minority of instances of abdominal pain appear with no consistent word order (non-standard syntax), e.g.,

Abdominal/left Side Pains
PAIN ABD RADIAT TO LT BACK SIDE
PAIN ON R LOWER ABDOMEN
Sharp Pain in Abd, nauseated, diarrhea

with non-standard word usage, e.g.,

RUQ/RLQ pan, Fever
STOMACH HURTS, DIARRHEA, VOMIT
FEVER STOMACH & HEAD ACHE
"tummy Hurt"

and with no reliable word proximities that a simple matching algorithm could detect (Fig. 8). These cases are best handled using weighted keywords.

Keyword matching must be able to cope with spelling errors. Although less numerous than abbreviations overall, misspellings (Figs. 9 and 10) merit attention, as some terms of interest are frequent sources of error (Fig. 11). The issue is addressed using a variation of the Levenstein distance metric, or edit distance. The similarity between two strings is measured by the minimum number of single-character insertions, deletions, substitutions, or inversions needed to make them equal. An edit distance of 1 appears to catch the majority of misspellings (as illustrated for the term "pneumonia" in Fig. 12). This approach, however, misses most word concatenations or truncations. Truncations are particularly common, probably due to the size of the chief complaint text field. Commonly occurring truncations can be handled by treating them as known abbreviations.

For the third major processing step in chief complaint classification, lower-level syndrome categories are used as building blocks to construct higher-level categories of interest. The application's current configuration uses 4 to 20 lower-level categories to build each of the 7 upper-level categories of Death, Gastrointestinal, Neurological, Rash, Respiratory, Sepsis, and Unspecified Infection. Options supported for constructing upper-level categories out of lower ones include logical OR, logical AND, constraints, exclusions, multiple levels, and

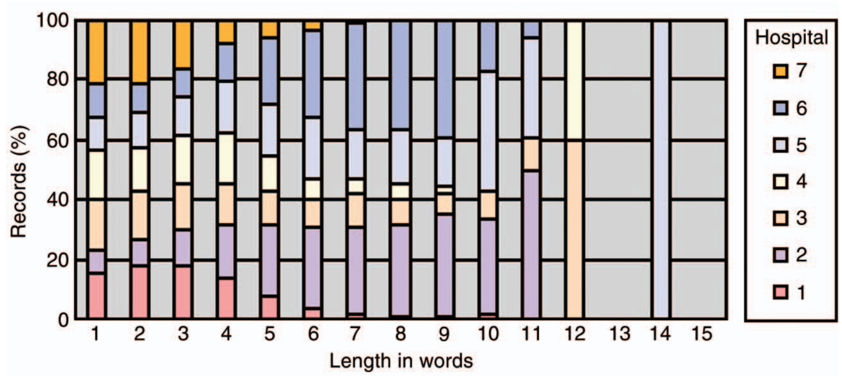


Figure 7. Percentage of chief complaints of given lengths by hospital for all syndromes.

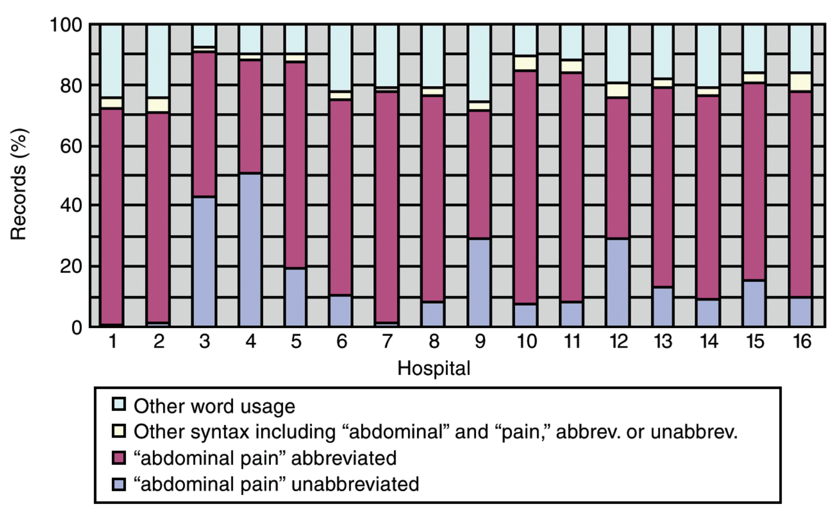


Figure 8. Expressions of "abdominal pain" in chief complaint by hospital.

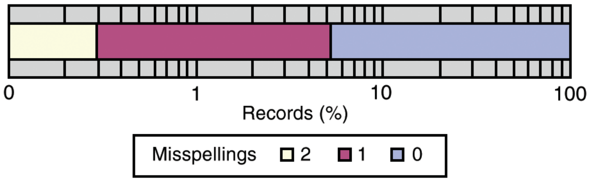


Figure 9. Percentage of chief complaint records containing misspellings. Data were obtained from a review of 1 day's worth of records (1017 total, including Other).

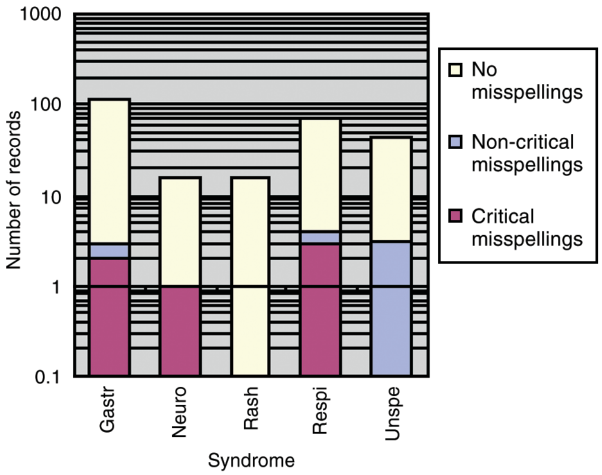


Figure 10. Number of critical and non-critical misspellings per chief complaint record for non-Other syndrome groups. A critical misspelling is one whose correct interpretation is necessary for correct classification of the record in which it appears. Data were obtained from a review of 1 day's worth of records (1017 total, including Other).

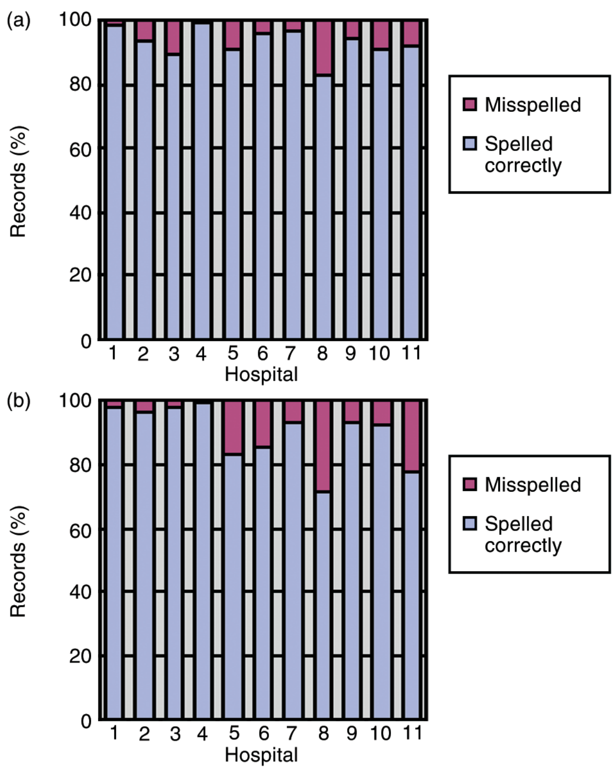


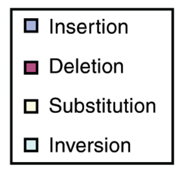
Figure 11. Percentage of misspellings for (a) "diarrhea" and (b) "pneumonia" by hospital.

category reuse (Fig. 13). These options were chosen to reflect the way public health practitioners appear to naturally formulate syndrome groups.

Relationships among syndrome categories are flexibly specified in a configuration table. This allows easier accommodation of requirements to reinterpret or redefine categories. If appropriate lower-level building blocks are already established, redefining upper-level ones is a simple matter of editing a table. This makes it possible to quickly develop new syndrome groups for short-term surveillance. During last year's SARS outbreak, for example, a new SARS syndrome category was built out of preexisting lower-level categories and deployed for surveillance in less than a day (Fig. 14).

Effectiveness

Although the overall accuracy of the system is difficult to establish without the extensive record-by-record manual coding that it was designed to avoid, it is subject to periodic review by developers and ongoing inspection by users in the field. Feedback from these review processes forms the basis of ongoing incremental improvements to the system. Figure 15 shows overall accuracy



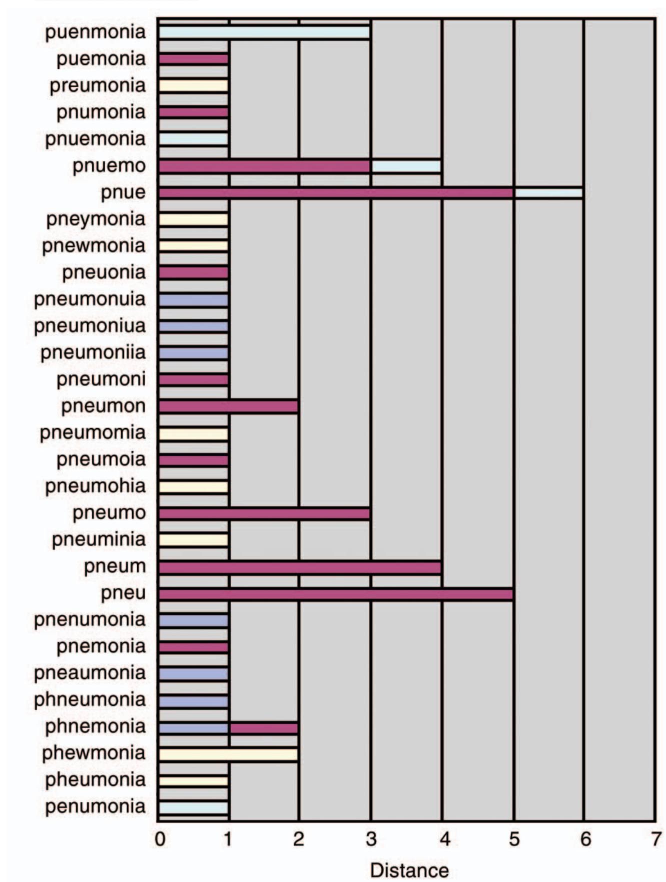


Figure 12. Effectiveness of Levenstein distance metric (edit distance) for recognizing observed misspellings of "pneumonia." A distance of 1 or less indicates a recognized misspelling.

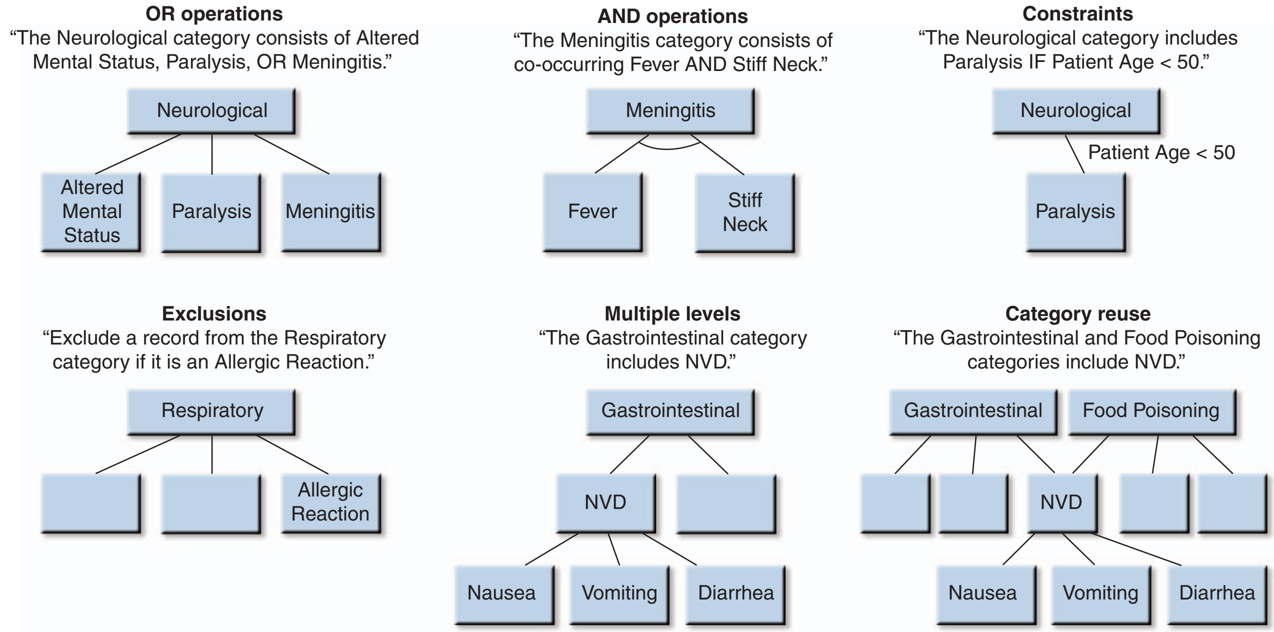


Figure 13. Supported operations for building syndrome groups out of symptoms.

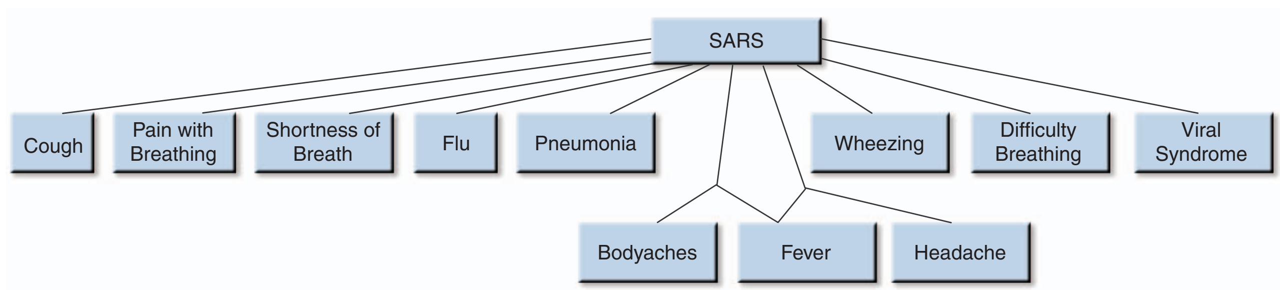


Figure 14. The SARS syndromic surveillance category was quickly implemented for short-term surveillance.

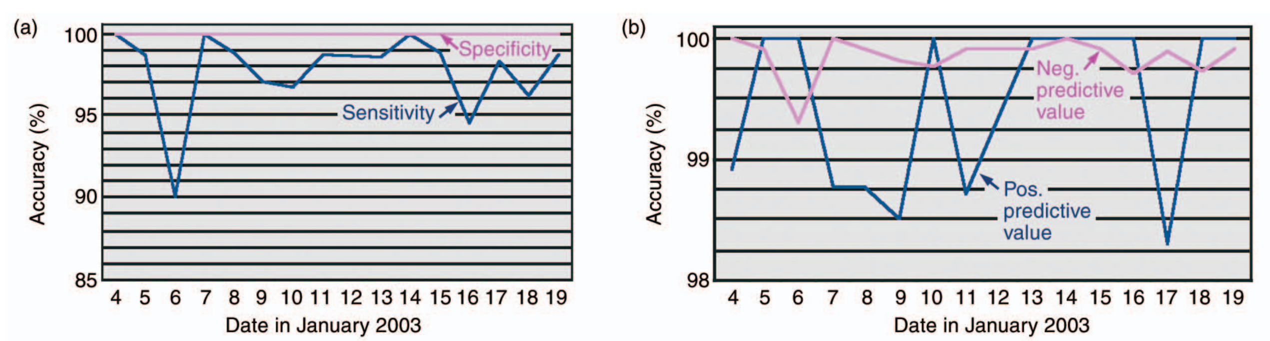


Figure 15. Overall sensitivity/specificity (a) and positive and negative predictive values (b) for automated categorization of chief complaint data from an individual hospital over a 2-week period as determined by manual review of chief complaint records.

rates across all syndromes of interest (e.g., excluding the default Other syndrome) as determined by manual review of chief complaint records for an individual hospital during a 2-week period. Analysis of the remaining errors in the system suggests that they are largely due to unrecognized abbreviations, unrecognized truncated words, and either unrecognized or misrecognized misspellings. The dip in sensitivity down to 90% in Fig. 15a, for example, was caused primarily by the introduction of a previously unseen abbreviation affecting the Gastrointestinal syndrome.

FUTURE DIRECTION

Future efforts should include larger-scale performance accuracy tests based on manual classification of relatively substantial sets of records. Manual classification would preferably be performed by more than one person in order to allow measurement of the inevitable inter-coder error. Given a test set extensive enough to support trained classification methods, such as Bayesian classification, comparisons between the current approach and more traditional statistical classification methods could be performed.

Further analysis of inter-hospital differences in chief complaint data is also warranted. Numerous discrepancies are evident in features such as style of expression, vocabulary choice, abbreviation use, and spelling error distributions that either demonstrably or potentially affect classification accuracy. How effectively a system configured for one set of hospitals performs on another set is important to establish because of its relevance to the ongoing expansion of the ESSENCE biosurveillance system to additional hospitals and geographic regions.

REFERENCES

[1] Lombardo, J., Burkom, H., Elbert, E., Magruder, S., Lewis, S. H., et al., "A Systems Overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II)," *J. Urban Health* **80**(2 Suppl 1), 32–42 (Jun 2003).

[2] *Bioterrorism Preparedness and Response: Use of Information Technologies and Decision Support Systems*, Evidence Report/Technology Assessment No. 59, Agency for Healthcare Research and Quality, Rockville, MD (Jul 2002).

[3] Graham, J., Buckeridge, D., Choy, M., and Musen, M., "Conceptual Heterogeneity Complicates Automated Syndromic Surveillance for Bioterrorism," in *Proc. Am. Med. Informatics Assoc. (AMIA) Symp.*, p. 1030 (9–13 Nov 2002).

[4] Mocny, M., Cochrane, D. G., Allegra, J. R., Nguyen, T., Heffernan, et al., "A Comparison of Two Methods of Biosurveillance of Respiratory Disease in the Emergency Department: Chief Complaint vs. ICD9 Diagnosis Code," *Acad. Emerg. Med.* **10**(5), 513 (2003).

[5] Begier, E. M., Sockwell, D., Branch, L. M., Davies-Cole, J. O., Jones, L. H., et al., "The National Capital Region's Emergency Department Syndromic Surveillance System: Do Chief Complaint and Discharge Diagnosis Yield Different Results?" *Emerg. Infect. Dis.* **9**(3), 393–396 (Mar 2003).

[6] Ivanov, O., Wagner, M. M., Chapman, W. W., and Olszewski, R. T., "Accuracy of Three Classifiers of Acute Gastrointestinal Syndrome for Syndromic Surveillance," in *Proc. Am. Med. Inform. Assoc. (AMIA) Symp.*, pp. 45–49 (9–13 Nov 2002).

[7] Espino, J. U., and Wagner, M. M., "Accuracy of ICD-9-coded Chief Complaints and Diagnoses for the Detection of Acute Respiratory Illness," in *Proc. Am. Med. Inform. Assoc. (AMIA) Symp.*, pp. 164–168 (9–13 Nov 2002).

[8] Bravata, D. M., Vandana, S., McDonald, K. M., Smith, W. M., Szeto, H., et al., "Evaluating Detection and Diagnostic Decision Support Systems for Bioterrorism Response," *Emerg. Infect. Dis.* **10**(1), 100–108 (Jan 2004).

[9] Das, D., Weiss, D., Mostashari, F., Treadwell, T., McQuiston, J., et al., "Enhanced Drop-in Syndromic Surveillance in New York City Following September 11, 2001," *J. Urban Health* **80**(2 Suppl 1), 76–88 (Jun 2003).

[10] Real-time Outbreak and Disease Surveillance (RODS) Laboratory Web site, http://www.health.pitt.edu/rods/sw/default.htm.

[11] Olszewski, R. T., "Bayesian Classification of Triage Diagnoses for the Early Detection of Epidemics," in *Recent Advances in Artificial Intelligence: Proc. 16th Int. FLAIRS Conf.*, pp. 412–416 (11–15 May 2003).

THE AUTHOR

CAROL A. SNIEGOSKI is a member of the Senior Professional Staff in the Special Applications Branch of the National Security Technology Department. She holds a B.A. in linguistics, a B.S. in computer science from the University of Maryland at College Park, and an M.S. in computer science from The Johns Hopkins University. Ms. Sniegoski has pursued additional graduate courses in bioinformatics, applied mathematics, and epidemiology at the Bloomberg School of Public Health and Whiting School of Engineering at Johns Hopkins. At APL she initially worked on command and control systems for the Air Defense Systems Department before moving to her current group to work on biosurveillance. Her e-mail address is carol.sniegoski@jhuapl.edu.