

Comparison of Two Major Emergency Department-Based Free-Text Chief-Complaint Coding Systems

Christina A. Mikosz,¹ J. Silva,¹ S. Black,¹ G. Gibbs,¹ I. Cardenas²

¹Rush University Medical Center, Chicago, Illinois; ²Chicago Department of Public Health, Chicago, Illinois

Corresponding author: Julio Silva, Rush University Medical Center, 1653 W. Congress Parkway, Chicago, IL, 60612. Telephone: 312-942-7802; Fax: 312-942-4021; E-mail: Julio_Silva@rush.edu.

Abstract

Introduction: Emergency departments (EDs) using free-text chief-complaint data for syndromic surveillance face a unique challenge because a complaint might be described and coded in multiple ways.

Objective: Two major ED-based free-text chief-complaint coding systems were compared for agreement between free-text interpretation and syndrome coding.

Methods: Chief-complaint data from 21,736 patients at an urban ED were processed through both the New York City Department of Health and Mental Hygiene (DOHMH) syndrome coding system as modified by the Chicago Department of Public Health and the Real-Time Outbreak Detection System Complaint Coder (CoCo, version 2.1, University of Pittsburgh). To account for differences in each system's specified syndromes, relevant syndromes from the DOHMH system were collapsed into the corresponding CoCo categories so that a descriptive comparison could be made. DOHMH classifications were combined to match existing CoCo categories as follows: 1) vomit+diarrhea = Gastrointestinal; 2) cold+respiratory+asthma = Respiratory; 3) fevflu = Constitutional; 4) rash = Rash; 5) sepsis+other = Other; 6) unknown = Unknown.

Results: Overall agreement between DOHMH and CoCo syndrome coding was optimal (0.614 kappa). However, agreement between individual syndromes varied substantially. Rash and Respiratory had the highest agreement (0.711 and 0.594 kappa, respectively). Other and Constitutional had an intermediate level of agreement (0.453 and 0.419 kappa, respectively), but less than optimal agreement was identified for Gastrointestinal and Unknown (0.270 and 0.002 kappa, respectively).

Conclusions: Although this analysis revealed optimal overall agreement between the two systems evaluated, substantial differences in classification schemes existed, highlighting the need for a consensus regarding chief-complaint classification.

Introduction

Syndromic surveillance has emerged as a novel approach to early disease detection. Both the public and private health sectors are exploring different approaches to disease-outbreak detection using real-time, automated syndromic surveillance systems (1–5). These systems are composed of a series of distinct steps that work collectively to shorten the time necessary to detect an aberrant pattern in clinical activity, potentially indicating a disease outbreak. The flow of information in such systems begins with the collection of patient chief-complaint data, often in free-text form, by triage staff in an emergency department (ED) or outpatient clinic. Then, these complaints are coded into specific broadly defined syndromes for epidemiologic surveillance. Next, syndrome counts for a predetermined period are compared with baseline data from a previous interval. Finally, any suspicious anomalies in syndrome trends detected during the analytic phase are investigated.

One of the first and most important steps in syndromic data processing is the classification of free-text chief complaints into syndromes. In Chicago, Illinois, a major university

teaching hospital and the Chicago Department of Public Health (CDPH) are working to implement automated, real-time syndromic surveillance. However, each institution is using a different free-text chief-complaint coding system. Processing of chief-complaint data collected as free text poses a unique challenge. One solution to the problem is to use software specifically designed to evaluate the patient's chief complaint and then assign it a syndrome category. Different computerized algorithms, or complaint coders, are trained to prioritize and code symptoms differently. As a result, depending on what algorithm is in place in a given clinical setting, a syndrome profile for a group of patients in a certain span of time might vary considerably, not only skewing the potential accuracy of patient data tracking within the hospital but also affecting public health surveillance efforts on a broader geographic scale.

Methods

For this study, two major ED-based free-text chief-complaint coding algorithms were tested. One system was

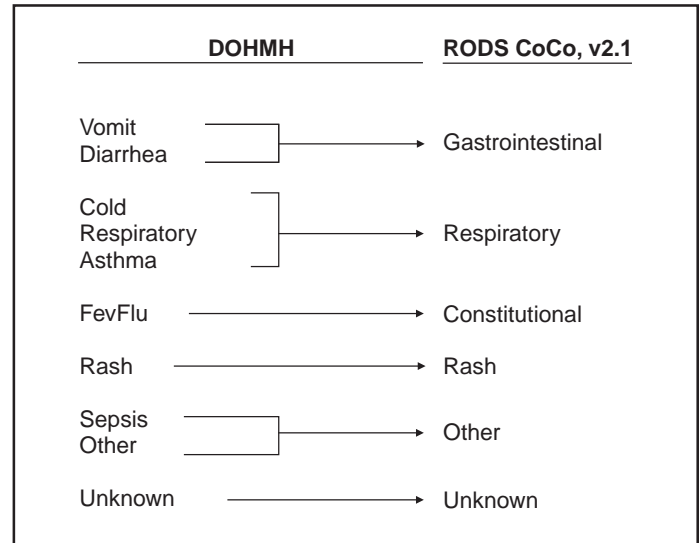
version 2.1 of the Complaint Coder (CoCo) developed by the Real-Time Outbreak Detection System (RODS) laboratory at the Center for Biomedical Informatics, University of Pittsburgh (5). This Bayesian classifier codes symptoms into syndromes on the basis of probability (i.e., the chances that a given symptom or group of symptoms will fall into a certain syndrome grouping), and then the syndrome code with the highest computed probability is assigned. These probabilities are determined from a default probability file included as a part of the CoCo software; this file was derived from 28,990 complaint strings collected from a single ED that were each manually coded by a physician into a syndrome category. Although CoCo has the capability to be retrained by using patient data obtained locally, for this study, the included default file was used. The second system was the complaint classifier algorithm developed and implemented by the New York Department of Health and Mental Hygiene (DOHMH) after the events of September 11, 2001 (6). This system codes complaints into syndromes on the basis of keywords, for which the algorithm searches, to assign a particular syndrome. The basis for choosing these specific keywords was data previously collected from New York City area EDs. In Chicago, CDPH has obtained the DOHMH algorithm and uses certain syndrome modules for routine public health surveillance. Both of these complaint-classifying algorithms require that free-text chief-complaint data be preprocessed into a preferred format to be read correctly (i.e., all text must be in lowercase, with all punctuation removed, for CoCo to process the text); in contrast, the DOHMH system requires data to be in uppercase, but punctuation does not need to be removed or altered.

Data for this study included all chief complaints collected during January–June 2002 at a Chicago ED where all chief-complaint data are logged in a free-text manner, for a total of 21,736 free-text complaint strings. All complaints were preprocessed for each of the two coding algorithms; only case and punctuation were altered as necessary. Spelling or grammatical errors were not corrected and instead left in place. These complaints were then processed by each coding algorithm separately and compared for agreement, by using the kappa statistic; all statistical analysis was implemented by using SPSS 10.0 software (7).

Results

CoCo's syndromes are more broadly defined and distinct from one another, whereas the syndromes of the DOHMH coding algorithm are of a more specific nature with a certain level of overlap (e.g. not just *Gastrointestinal*, as in CoCo, but *Vomit* and *Diarrhea* in particular, to specify upper and lower gastrointestinal symptoms, respectively) (Figure 1). Because

FIGURE 1. Scheme used to collapse New York City Department of Health and Mental Hygiene (DOHMH) categories, as modified by the Chicago Department of Public Health (CDPH), into the Real-Time Outbreak Detection System (RODS), Complaint Coder (CoCo) Version 2.1 (Center for Biomedical Informatics, University of Pittsburgh)



of apparent differences in syndrome specificity between the two coding algorithms, to make a descriptive comparison between the two systems, the syndrome categories of the DOHMH coder were collapsed to more accurately match the wider scope of the CoCo syndromes. This scheme was based on the types of chief complaints that were classified into each syndrome in both systems, as follows:

- Any symptom coded as *Vomit* or *Diarrhea* by the DOHMH algorithm was renamed *Gastrointestinal* to match CoCo.
- Any symptom classified as *Cold*, *Respiratory*, or *Asthma* by the DOHMH coder was renamed *Respiratory*.
- *FevFlu* in the DOHMH system was renamed *Constitutional*.
- Any symptom coded as *Sepsis* or *Other* by the DOHMH algorithm was renamed *Other*.
- Symptoms coded as *Rash* were left as is and were not combined with any other syndromes.

Three syndromes existed, *Hemorrhagic*, *Botulinic*, and *Neurologic*, into which CoCo classifies symptoms for which no exact equivalent exists in the DOHMH algorithm, as used by CDPH. These are syndromes that, although extremely narrow in their scope, have considerable relevance in surveillance for biologic terrorism agents. However, because no direct comparison could be made for the current analysis, any chief complaints coded as *Hemorrhagic*, *Botulinic*, or *Neurologic* were removed from the study.

Of the specific syndromes into which CoCo classifies chief complaints, *Respiratory* was the most frequently represented, at 14.0% (Figure 2). *Constitutional* and *Gastrointestinal* were roughly equivalent in their representation, at 8.7% and 10.2%, respectively. *Rash* was present at the same frequency as the *Unknown* category, at 2.4%. *Unknown* represents a catch-all category into which CoCo places symptoms it is not trained to handle. Symptoms commonly reported in an ED setting, yet not fitting into any of the four tracked syndromes (*Respiratory*, *Constitutional*, *Gastrointestinal*, and *Rash*) represented the largest group of all, the *Other* category, at 62.3%.

For the DOHMH system, *Respiratory* was also the most frequently coded of the tracked syndromes (11.3%), and *Constitutional* and *Gastrointestinal* were similar in representation (5.4% and 3.7%, respectively) (Figure 2). A total of 77.8% of the chief complaints in the data set were classified not into any of the tracked syndrome categories but as *Other*, a larger proportion than the 62.3% coded as *Other* by CoCo. The DOHMH coding algorithm can classify chief-complaint data into substantially variable syndromes, and the syndromes into which the study data were categorized were only those syndromes within the DOHMH coder that CDPH uses on a daily basis for routine public health surveillance. Had CDPH been actively using every possible syndrome into which the DOHMH algorithm is trained to code data, possibly all of the chief complaints coded as *Other* would instead have been coded into a distinct syndrome category. The *Unknown*

category prevalence was approximately 0, indicating that the DOHMH coder was able to recognize and categorize virtually all of the chief complaints.

This study used the kappa statistic to assess the agreement in syndrome classification between both coding algorithms (i.e., the chances that a given chief complaint was coded as the same syndrome by both algorithms were analyzed). On a scale of zero to one, with one representing complete overall agreement between both algorithms, the kappa statistic was calculated to be 0.614. This represents a substantial level of overall agreement between the two coding systems (8). This value was also statistically significant (standard error = 0.05; 95% confidence interval = 0.604–0.624; total = 145.866 [$p < 0.0005$]). However, when the kappa statistic was calculated for each syndrome, the results varied substantially. The *Rash* syndrome had the highest level of agreement, with a kappa of 0.711. Examination of the data confirms this level of optimal agreement, because the majority of free-text chief complaints coded as *Rash* by both algorithms — representing 66.8% of the complaints coded as *Rash* by the DOHMH system and 55.4% by CoCo — was simply the word *rash*. Any coding algorithm trained to classify symptoms into a *Rash* category would be capable of correctly classifying a complaint of *rash*. The *Respiratory* syndrome had the next highest level of agreement, with a kappa of 0.594. A sample of the complaints used to define the respiratory and gastrointestinal categories of the different coding systems is provided (Table). The majority of free-text strings coded as *Respiratory* by both algorithms were common respiratory complaints (e.g., *shortness of breath*, *asthma attack*, and *difficulty breathing*). One notable difference was that a complaint of *dib* was recognized by the DOHMH algorithm as an abbreviation for *difficulty in breathing* and subsequently coded as *Respiratory*. In contrast, CoCo coded all 106 complaints of *dib* as *Unknown*. In fact, *dib* represented the largest proportion of the symptoms coded as *Unknown* by CoCo. An even lower level of agreement was identified within the *Constitutional* syndrome (kappa statistic = 0.419). A complaint string of *fever* was the most common symptom coded as *Constitutional* by both systems; however, beyond this single common symptom, distinct differences in complaints coded as *Constitutional* existed between the two algorithms.

The *Gastrointestinal* syndrome had the lowest level of agreement of all four tracked syndromes between the two coding algorithms, with a kappa of only 0.270. A key contributor to this low agreement is the handling of a free-text complaint of *abdominal pain* (and all abbreviations indicative of pain in the abdomen [e.g., *abd pain*]). CoCo codes a complaint of *abdominal pain* as *Gastrointestinal*; in fact, *abdominal pain* (and

FIGURE 2. Emergency department free-text chief complaint data as processed by Real-Time Outbreak Detection System (RODS), Complaint Coder (CoCo) Version 2.1 (Center for Biomedical Informatics, University of Pittsburgh) and the New York City Department of Health and Mental Hygiene (DOHMH) system, as modified by the Chicago Department of Public Health

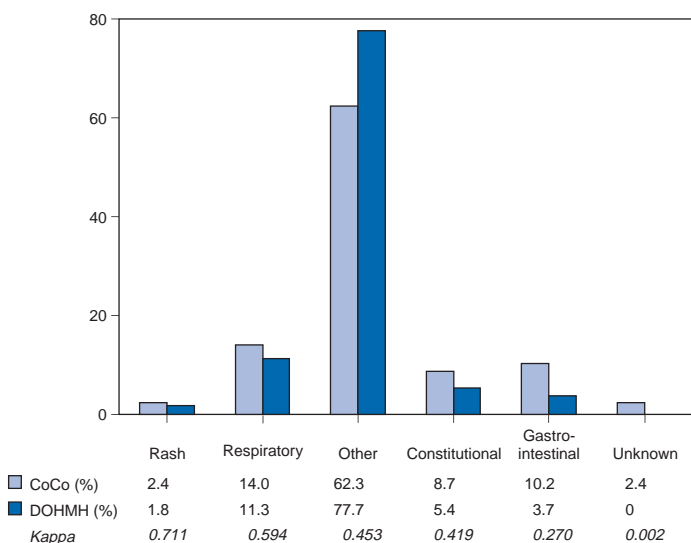


TABLE. Sample comparison of syndrome category definitions between the New York City Department of Health and Mental Hygiene (DOHMH) algorithm, as modified by the Chicago Department of Public Health (CDPH), and the Real-Time Outbreak Detection System (RODS), Complaint Coder (CoCo) Version 2.1 (Center for Biomedical Informatics, University of Pittsburgh)

Category	DOHMH/CDPH	RODS CoCo*
Gastrointestinal	<i>Diarrhea</i> : diar, dair, diah, enteri, gastroent, 558; IF stool, bowel, bm AND loose, watery, or liquid <i>Vomit</i> : throwing up, threw up, food pois, vom, vmt, n/v, 787	pain or cramps anywhere in the abdomen, nausea, vomiting, diarrhea, and abdominal distension or swelling
Respiratory	<i>Cold</i> : stuffi, stuffy, sneez, nasal, nase1 AND cong, conj, drip, disch, runn, cong, runn, nose, congested, congestion, cold <i>Respiratory</i> : pneumon, gasp, SOB, pulmon, monia, infiltr, croup, bronch, hypox, 786.2, 786.0, 480, 481, 482, 483, 465, 466, 484, 485, 486, pleur, dyspn, coug, couh, breat, beath, dib, d i b, d.i.b., brathing, diff dr, uri, uri/, uri; u r i, u.r.i, sob, s o b, s.o.b. <i>Asthma</i> : asth, asmtha, ashtma, astma, asyhma, whez, azth, az, airway, whee, wheel, 490, 491, 492, 493, COPD, c.o.p.d.	problems of the nose (coryza) and throat (pharyngitis), as well as the lungs; examples of <i>Respiratory</i> include congestion, sore throat, tonsillitis, sinusitis, cold symptoms, bronchitis, cough, shortness of breath, asthma, chronic obstructive pulmonary disease, and pneumonia; the presence of both cold and flu symptoms is <i>Respiratory</i> and not <i>Constitutional</i>

* For CoCo to process, all text must be in lowercase.

its similar spellings or abbreviations) accounted for 33.7% of the complaint strings coded by CoCo as *Gastrointestinal*, considerably more than any other complaint. However, the DOHMH algorithm is not trained to code a complaint of *abdominal pain* as *Gastrointestinal* and instead codes it as *Other*. The symptom that represented the largest proportion of DOHMH's *Gastrointestinal* syndrome (26.7%) was a complaint of *vomiting*, which was the second-most frequent complaint within CoCo's *Gastrointestinal* category (at 10.5%). Additionally, examination of the data coded as *Gastrointestinal* by each algorithm provides insight into the hierarchy of syndromes within each system. When a single complaint string included multiple symptoms, a decision had to be made by each algorithm regarding which syndrome is most important for surveillance purposes, because each algorithm is trained to settle on a single syndrome and not allow for multiple codings of a single complaint string. For example, a complaint string of *vomiting blood* was coded by CoCo as *Hemorrhagic*, indicating that CoCo is trained to weigh the presence of *blood* within the complaint as more important than *vomiting*, which is otherwise coded as *Gastrointestinal*. In contrast, the DOHMH algorithm codes a complaint of *vomiting blood* as *Gastrointestinal*. Another example is a complaint string of *rash fever*. The DOHMH algorithm codes such a complaint as *Constitutional*, meaning that it has been trained to consider the keyword *fever* as more important than the keyword *rash*, whereas CoCo codes *rash fever* as *Rash*, demonstrating that even in the presence of fever, CoCo considers a rash to be the more important finding.

Conclusions

This study's findings demonstrate the substantial variability that exists between these two chief-complaint coding systems. Whereas the overall agreement for coding of the data set was satisfactory, agreement between individual syndrome classifications ranged from substantial to unsatisfactory. These differences are not necessarily a problem of accuracy or performance, but rather a result of the choices made in designing the coding systems. When relying on automated classification of chief-complaint strings, public health officials need to be aware of the symptom hierarchy within systems because this prioritization will result in changes in syndrome classification prevalence.

The programs in this study allow for individual modification of the algorithms that classify each complaint, and changing each program is possible so that the user can obtain approximately complete concordance for the syndromes of choice. However, this highlights a more substantial problem: what are the syndrome categories that surveillance systems should be monitoring? A recent literature review revealed multiple syndrome categories under surveillance in different programs throughout the country (1–5). These categories ranged from such individual syndrome categories as *respiratory* or *gastrointestinal* to such groups of syndromes as *rash with fever* or *upper/lower respiratory infection with fever*.

No set standards exist regarding which syndrome classifications should be regularly monitored. The ultimate goal of surveillance should be early detection of disease outbreaks, either natural or as a result of biologic terrorism. Although surveillance systems should remain flexible to adapt to local public health needs, national consensus is required to define which syndromes should be monitored as well as what chief complaints accurately define these syndrome categories. After agree-

ment is reached, efforts can focus on refining systems' ability to perform automated real-time syndromic surveillance accurately.

This study had certain limitations. First, the DOHMH system already has a newer version available that might change the outcome of coding complaints. Second, neither system was designed to be specific to Chicago, where the chief complaints were made; regional differences in demographics, language, and culture might affect the coding. Third, this study did not examine the validity or efficacy of the two surveillance systems. Further studies are needed to evaluate the diverse approaches available for automated chief-complaint classification in ED-based syndromic surveillance.

References

1. 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.
2. Begier EM, Sockwell D, Branch LM, et al. The national capitol region's emergency department syndromic surveillance system: do complaint and discharge diagnosis yield different results? *Emerg Infect Dis* 2003;9:393–6.
3. Tsui F-C, Espino JU, Wagner MW, et al. Data, network, and application: technical description of the Utah RODS Winter Olympic biosurveillance system. *Proc AMIA Symp* 2002;815–9.
4. Komatsu K. Syndromic surveillance for the 2001 World Series in Arizona. Presented at the National Syndromic Surveillance Conference, September 24, 2002, New York.
5. Tsui F-C, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MW. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc* 2003;10:399–408.
6. CDC. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. *MMWR* 2002;51(Special Issue):13–5.
7. SPSS, Inc. SPSS for Windows,® version 10.0.0 [Software]. Chicago, IL: SPSS, Inc., 1999. Available at <http://www.spss.com>.
8. Landis J, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.