Accuracy of ICD-9–coded Chief Complaints and Diagnoses for the Detection of Acute Respiratory Illness

Jeremy U. Espino M.D. and Michael M. Wagner M.D., Ph.D.
Center for Biomedical Informatics, University of Pittsburgh and Biomedical Security Institute, Pittsburgh

ICD-9–coded chief complaints and diagnoses are a routinely collected source of data with potential for use in public health surveillance.

We constructed two detectors of acute respiratory illness: one based on ICD-9–coded chief complaints and one based on ICD-9–coded diagnoses. We measured the classification performance of these detectors against the human classification of cases based on review of emergency department reports. Using ICD-9–coded chief complaints, the sensitivity of detection of acute respiratory illness was 0.44 and its specificity was 0.97. The sensitivity and specificity using ICD-9–coded diagnoses were no different.

These properties of excellent specificity and moderate sensitivity, coupled with the earliness and electronic availability of such data, support the use of detectors based on ICD-9 coding of emergency department chief complaints in public health surveillance.

INTRODUCTION

It is a basic tenet of Public Health that, aside from prevention, rapid detection and response are the best means for reducing the morbidity and mortality from disease outbreaks—bioterrorist or naturally occurring.[1] The responsibility for detection falls upon astute clinicians and public health surveillance (PHS) systems.[2] When clinicians fail to notice or report outbreaks, PHS systems are the next line of defense.

A PHS system regularly collects data about a population to make assessments about the presence of disease outbreaks and disease spread. Because collection of surveillance data is expensive, and because such data need to be available in a timely manner, there is increasing interest in the use of routinely collected data such as grocery store purchases and emergency department (ED) visit data in PHS.[3] Studies have already indicated that routinely collected data, such as pharmacy records and ICD-9 data have good performance characteristics for detecting cases of tuberculosis and cardiovascular disease.[4, 5]

ICD-9–coded chief complaints and diagnoses from EDs, routinely collected for electronic insurance claims submission, are a timely alternative to other types of surveillance data such as mandatory case reporting and sentinel physician reporting. A major disadvantage of such non-electronic data sources is that they rely on slow methods of communication—mailed paper forms and faxes often collected on a weekly basis. In the UPMC Health System, EDs collect ICD-9–coded chief complaints and ICD-9–coded diagnoses for all patients. The chief complaint is recorded during patient registration and the diagnosis is recorded at the end of the visit. There is a delay of two seconds between the time an ICD-9–coded chief complaint or ICD-9–coded diagnosis is entered in ED workstations and the time it is available for further processing.

Many bioterrorist diseases such as inhalational anthrax, tularemia, and smallpox initially present with respiratory symptoms and then progress rapidly towards death. For example, in inhalational anthrax, treatment and prophylaxis must be initiated as soon as possible within 72 hours of inhalation to reduce mortality.[6] This rapid progression creates an urgent need for timely data. A detector that identifies patients at the stage of acute respiratory illness would be helpful for early detection of bioterrorism attacks involving diseases with this type of presentation.

In this study, we test hypotheses that detectors that use ICD-9–coded chief complaints or ICD-9–coded diagnoses from the ED as an input can identify cases of acute respiratory illness.

METHODS

Definition of Acute Respiratory Illness

We defined an acute respiratory illness as a patient with less than 5 days duration of the following findings: cough, shortness of breath, sputum production, abnormal pulmonary examination, or radiological evidence of pneumonia. We excluded cases explained by a working non-respiratory diagnosis.

Acute Respiratory Illness Detectors

To build detectors for acute respiratory illness, we reviewed ICD-9 codes and grouped selected codes into a respiratory illness class. In particular, we reviewed all unique ICD-9 codes that had been used to encode chief complaints in the EDs of nine UPMC Health System hospitals during the past three years. Two internists reviewed these codes and their code descriptions. The internists included a code in respiratory illness class if the ICD-9 code could be used for a patient presenting with respiratory symptoms of interest to public health officials.

We instructed the internists to include an ICD-9 code if they believed that, at least 5% of the time, the patient coded with that code would actually have an acute respiratory illness. The internists had to agree on the classification of a code before it was included into the respiratory illness class. We used the respiratory illness class—comprising 64 ICD-9 codes—to build two detectors of respiratory illness: Respiratory Illness Detector using ICD-9–coded Chief Complaints (RID-CC) and Respiratory Illness Detector using ICD-9–coded Diagnoses (RID-Dx). Table 1 shows the first 10 codes from an ICD-9 code sorting of the list.

Table 1. Partial list of the acute respiratory illness detector codes and their descriptions.

ICD-9	Description
769	Respiratory distress syndrome
770.9	Unspecified respiratory condition of fetus and newborn
786	Symptoms involving respiratory system and other chest symptoms
786.0	Dyspnea and respiratory abnormalities
786.00	Respiratory abnormality, unspecified
786.1	Stridor
786.2	Cough
786.4	Abnormal sputum
786.52	Painful respiration
786.7	Abnormal chest sounds

We stored the acute respiratory illness code table on the same computer system that functions as the clinical event monitor for the UPMC Health System.[7] For this evaluation, any patient with an ICD-9–coded chief complaint or ICD-9–coded diagnosis that matched a code in the table was labeled as a positive case by the respective detectors.

ICD-9 Encoding of Chief Complaints and Diagnoses in the ED

In the ED at the UPMC Health System, a triage nurse interviews each newly arriving patient and records the patient's chief complaint verbatim on a paper form. Prior to the patient being seen by the clinician, a registration clerk, using the paper form, encodes the written chief complaint as a single ICD-9 code.

After seeing the patient, the clinician in the ED records a diagnosis using a single ICD-9 code. The ICD-9–coded diagnosis may differ from the ICD-9–coded chief complaint. The ICD-9–coded chief complaint and diagnosis are available electronically in real-time from the medical center's Admission Discharge Transfer system.

Study Population

From 53,099 ED visits in the period July to October 2000, we selected a sample of 800 based on an ordering of the visits by social security number. We chose summer and fall because the incidence of respiratory illness is expected to be low and we wanted any bias due to prevalence of disease to be in direction opposite to the hypothesized effect. When multiple admissions were available for the same patient, we used only the most recent admission for that patient. This policy eliminated multiple patient visits for the same clinical problem. We excluded cases from analysis in which an ICD-9–coded chief complaint or an ICD-9–coded diagnosis was missing.

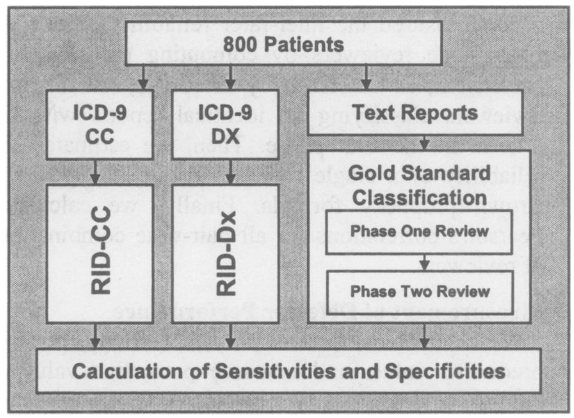

Figure 1. Evaluation Process. RID-CC - Respiratory Illness Detector using ICD-9–coded chief complaints. RID-Dx - Respiratory Illness Detector using ICD-9–coded diagnoses.

Creation of Gold Standard

We established a gold standard classification of the patients using a two-phase report review. (Figure 1) The reports were transcribed reports by attendings and residents available through the medical center's electronic medical record system.

In Phase One, four physicians reviewed a portion of the 800 ED reports in sets that overlapped by 50 reports (for the purpose of an inter-rater reliability study discussed below). Thus, each physician read 250 reports; flagging a report as positive if the patient met any of the following criteria: cough, sputum production, shortness of breath, signs of pneumonia on chest x-ray, abnormal physical exam of lungs, or a working diagnosis of a respiratory illness.

In Phase Two, two board-certified internists (one was also involved in creation of the ICD-9 sets) independently reviewed the positive reports identified in the first phase. The physicians flagged these reports as positive if they met the following criterion:

- Met Phase One criteria (Cough, sputum production, shortness of breath, pneumonia on chest x-ray, abnormal physical exam of lungs, or a working diagnosis of a respiratory illness)

  AND
- Duration of respiratory symptoms less than 5 days

  AND
- Absence of a non-respiratory working diagnosis that accounts for all the respiratory findings

The physicians met to resolve differing patient classifications. In both phases, the physicians were blinded to the ICD-9 encodings for the patients.

Inter-rater reliability

We measured the inter-rater reliability among the Phase One reviewers by computing the reliability coefficient, as described by Friedman,[8] for all four reviewers classifying 50 identical reports with the criteria in the first phase. Then, we estimated the reliability of a single reviewer using the Spearman Brown prophecy formula. Finally, we calculated Pearson's correlations for all pair-wise combinations of reviewers.

Measurement of Detector Performance

We calculated the sensitivity, specificity, positive predictive value, and negative predictive value of RID-CC and RID-Dx. We calculated 95% confidence intervals using the methods described by Fleiss.[9, 10]

One of the authors (JUE) performed an error analysis of the misclassifications made by RID-CC by reading the ED reports for the false negatives and false positives and comparing the transcribed attending diagnosis to the description of the ICD-9–coded chief complaint. If the ICD-9 code seemed appropriate for the patient, and the cause of the error was a result of the ICD-9 code being omitted from the Respiratory Illness Class, then we classified the error as "Correctable." Otherwise, we labeled the code as "Not Correctable."

To compare the relative timeliness of the two detectors, we calculated the time delay between the availability of an ICD-9–coded chief complaint and ICD-9–coded diagnosis for each ED visit.

RESULTS

We excluded 131 reports from the study because their ICD-9–coded chief complaint or diagnosis codes were not available. The majority of exclusions were due to a lack of chief complaint codes that occurred because of hospital direct admissions. Of 669 reports, 65 (10%) satisfied Phase One criteria. The total number of acute respiratory illness cases as determined by Phase Two review was 33, an incidence of 4.9%.

Table 2. Pearson's correlations for pair wise reviewers in the first phase of the report review.

Reviewer	A	B	C	D
A	-	0.759	0.759	0.806
B		-	0.883	0.941
C			-	0.941
D				-

The reliability coefficient for the four judges using Phase One criteria was 0.957. The estimated reliability for a single judge was 0.849. Pearson's correlations showed that the reviewers classified the reports in a similar manner (Table 2).

Table 3. Detector characteristics with 95% confidence intervals.

Measurement	RID-CC	RID-Dx
Sensitivity	0.44 ± 0.15	0.43 ± 0.15
Specificity	0.97 ± 0.01	0.97 ± 0.01
Pos. Predictive Value	0.44 ± 0.15	0.45 ± 0.16
Neg. Predictive Value	0.97 ± 0.01	0.97 ± 0.04

Table 3 shows the sensitivity, specificity, positive predictive value, negative predictive value and 95% confidence intervals for RID-CC and RID-Dx. As we will discuss, sensitivity and positive predictive value were lower than expected but specificity and negative predictive value were much higher than expected. The accuracy of RID-CC and RID-Dx were similar.

Table 4. Correctable and Not Correctable Errors

ICD-9	ICD-9–Coded Chief Complaint Description	Attending Diagnosis	Error
493.90	ASTHMA W/O STATUS ASTHM	Asthma	C
560.9	INTESTINAL OBSTRUCT NOS	COPD	NC
719.07	JOINT EFFUSION-ANKLE	R. Lung Effusion	NC
780.6	PYREXIA UNKNOWN ORIGIN	Pneumonia	NC
780.6	PYREXIA UNKNOWN ORIGIN	Acute Pylonephritis	NC
780.6	PYREXIA UNKNOWN ORIGIN	Fever Unknown Origin	NC
780.9	GENERAL SYMPTOMS NEC	R. Upper Lobe Pneumonia	NC
784.1	THROAT PAIN	Pharyngitis	NC
786.50	CHEST PAIN NOS	COPD, Pneumonia	NC
786.50	CHEST PAIN NOS	Pulmonary Embolus	NC
786.50	CHEST PAIN NOS	Pneumonia	NC
786.50	CHEST PAIN NOS	Pulmonary Embolus	NC
786.50	CHEST PAIN NOS	Pneumonia	NC
786.50	CHEST PAIN NOS	CHF	NC
786.50	CHEST PAIN NOS	Musculoskeletal Pain	NC
786.50	CHEST PAIN NOS	Asthma	NC
789.05	ABDOM PAIN, PERIUMBILIC	COPD, CHF	NC
999.1	AIR EMBOL COMP MED CARE	PCP Pneumonia	NC

Table 4 shows the results of the error analysis for the eighteen false negatives (respiratory patients missed by RID-CC). Even though it might be possible to improve the coding quality in EDs, we classified such errors as not correctable because the focus of the present research is on the potential of routinely collected data-with all of its noise and limitations-for public health surveillance.

The error analysis of the 17 false positives produced by RID-CC showed that six patients had respiratory symptoms/signs, but they were present for longer than four days and thus did not satisfy our case definition; five had no mention of respiratory symptoms/signs in the report, so these were instances in which the assigned ICD-9 code was incorrect; and six patients had a non-respiratory working diagnosis that explained the respiratory symptoms. None of these errors was felt to be correctable.

The mean time lag between availability of the ICD-9–coded chief complaints and diagnoses was 7.5 hours. The maximum time delay was 80.6 hours.

DISCUSSION

The measured sensitivity of 44% and the specificity of 97% were unexpected. Prior to this study, we expected that both RID-CC and RID-Dx would exhibit high sensitivity and at best moderate specificity and positive predictive value. The medical experts who created the detectors intended to make the detectors sensitive, because sensitivity is important in early detection of epidemics. So, whenever there was a doubt, the experts included codes even knowing that such inclusions would have made the detectors non-specific.

The sensitivity of ICD-9–coded chief complaints and diagnoses for detection of acute respiratory illness are relatively good, 0.44. Although missing half of the cases could affect the performance of a detection system, especially for small epidemics, it would probably be adequate to detect moderate or large-scale aerosolized respiratory bioterroristic releases with agents such as Bacillus anthracis, Francisella tularensis, and Yersinia pestis. It is an empirical question whether this level of sensitivity is sufficient to detect an outbreak of a more gradually developing disease such as Influenza. In another paper in these proceedings, Tsui et al show that during the 1999 and 2000 influenza seasons, a detection system using this set of ICD-9 codes and ED-coded chief complaints detects increases in the incidence of respiratory cases during influenza outbreaks.[11] Brinsfield reports similar findings using their set of ICD-9 codes.[3]

Higher sensitivity is of course desirable in a detection system. In particular, detection systems with high sensitivity have value for ruling out bioterrorist attack or natural disease outbreaks. There were 252 bioterrorism threats in the United States from the January of 1997 to May 1999[12] and many of these were anthrax hoaxes that were taken seriously, sometimes resulting in decontamination and other measures.[12] A detector with high sensitivity, showing no activity during a hoax, would be useful to prevent unnecessary quarantine, decontamination and fear.

Our respiratory class of ICD-9 codes was surprisingly specific for acute (<5 days) respiratory illness. This result was unexpected since neither the ICD-9 codes, nor the detector itself had any information about the duration of symptoms. The fact that there were so few errors of this type (6 out of 17 false positives for RID-CC) suggests that either patients rarely came to the ED for chronic respiratory illness during the study period, or the unlikely possibility that such patients are differentially coded by the ICD-9 coders in some way.

In considering the relative values of RID-CC and RID-Dx, we note that ICD-9–coded diagnoses offer no advantages in positive predictive value and specificity. However, because the diagnosis codes are significantly delayed (on average by 7.5 hours), it is clearly the case that detection systems should focus on the chief complaint data, when it is available.

These results were obtained in a single institution, and moreover it is likely that there are coding differences from institution to institution. Therefore, additional studies are needed to determine the generalizability of our findings. One question is whether our particular set of respiratory ICD-9 codes will produce the same classification performance if used elsewhere.

We also do not know if our ICD-9 respiratory detector uses the optimal set of ICD-9 codes. We could add, for example, the ICD-9 codes of the false negative cases in Table 4 to the detectors, which would improve sensitivity, albeit probably at a cost to specificity. Alternatively, we could use machine-learning methods to identify optimal sets, given assumptions about where such a system should be operating on an ROC curve.

Most of the errors were not correctable and our error analysis results mirror the results of others when a set of ICD-9 codes is used to determine the incidence of a disease or condition.[13] The majority of false negative errors are caused by human coding error and are unlikely to be corrected without a massive expenditure. Therefore, withstanding a "better" set of ICD-9 codes, the measurements we obtained are an accurate estimate of the potential of ICD-9–coded chief complaints and diagnoses to detect acute respiratory illness.

It is important to note that this study measured specificity, as well as positive predictive value. Typically, in disease surveillance research, only the sensitivity and positive predictive value of a detector is established, which limits the applicability of the results to the specific surveyed population (unless the disease prevalence is somehow also known). In the present research, the specificity results may be useful to other researchers wishing to incorporate such a system into larger models of detection and response under epidemic conditions.

In this study, we examined the performance characteristics of a method to detect acute respiratory illness in patients presenting to EDs that is based on the idea of a respiratory class of ICD-9 codes. A generalization of this idea is the concept of ICD-9-based syndromic detection. For example, we have also defined "diarrheal," "botulinic," "encephalitic," "flu-like," and "rash" sets to cover common prodromal presentations of epidemic diseases. Syndromic (or prodromic) detection is of current interest because it can provide earlier information than laboratory reporting (cultures often take 72 hours for microbial identification), and for diseases for which no laboratory tests exist, there may be no other alternative for automatic disease surveillance.[14]

ACKNOWLEDGEMENTS

This work was supported in part by grants GO8 LM06625-01, and T15 LM/DE07059 from the National Library of Medicine; contract no. 290-00-0009 from the Agency for Healthcare Research and Quality; and a cooperative agreement U90/CCU318753-01 from the Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC.

REFERENCES

1. Inglesby TV, O'Toole T, Henderson DA. Preventing the use of biological weapons: improving response should prevention fail. *Clin Infect Dis.* 2000;30(6):926-929.
2. Pesik N, Keim M, Sampson TR. Do US emergency medicine residency programs provide adequate training for bioterrorism? *Ann Emerg Med.* 1999;34(2):173-176.
3. Brinsfield KH, Gunn JE, Barry MA, McKenna V, Dyer KS, Sulis C. Using Volume-based Surveillance for an Outbreak Early Warning System. *Acad Emerg Med.* 2001;8(5):492.
4. Yokoe DS, Subramanyan GS, Nardell E, Sharnprapai S, McCray E, Platt R. Supplementing tuberculosis surveillance with automated data from health maintenance organizations. *Emerg Infect Dis.* 1999;5(6):779-787.
5. Assaf AR, Lapane KL, McKenney JL, McKinlay S, Carleton RA. Coronary heart disease surveillance: field application of an epidemiologic algorithm. *J Clin Epidemiol.* 2000;53(4):419-426.
6. Kaufmann AF, Meltzer MI, Schmid GP. The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? *Emerg Infect Dis.* 1997;3(2):83-94.
7. Wagner MM, Pankaskie M, Hogan W, et al. Clinical event monitoring at the University of Pittsburgh. *Proc AMIA Annu Fall Symp.* 1997:188-192.
8. Friedman C, Wyatt J. *Evaluation methods in medical informatics.* New York: Springer; 1997.
9. Fleiss JL. *Statistical methods for rates and proportions.* 2d ed. New York: Wiley; 1981.
10. Cameron D, Jones IG. John Snow, the broad street pump and modern epidemiology. *Int J Epidemiol.* 1983;12(4):393-396.
11. Tsui F-C, Wagner MM, Dato V, Chang C-CH. Value of ICD-9-Coded Chief Complaints for Detection of Epidemics. Paper presented at: Submitted to AMIA Annual Symposium, 2001; Washington, DC.
12. Cole LA. Risks of publicity about bioterrorism: anthrax hoaxes and hype. *Am J Infect Control.* 1999;27(6):470-473.
13. MacIntyre CR, Ackland MJ, Chandraraj EJ, Pilla JE. Accuracy of ICD-9-CM codes in hospital morbidity data, Victoria: implications for public health research. *Aust N Z J Public Health.* 1997;21(5):477-482.
14. Pavlin JA. Epidemiology of bioterrorism. *Emerg Infect Dis.* 1999;5(4):528-530.