

Syndrome and Outbreak Detection Using Chief-Complaint Data — Experience of the Real-Time Outbreak and Disease Surveillance Project

Michael M. Wagner,¹ J. Espino,¹ F-C. Tsui,¹ P. Gesteland,^{2,3} W. Chapman,¹ O. Ivanov,¹ A. Moore,^{1,4} W. Wong,⁴ J. Dowling,¹ J. Hutman¹
¹University of Pittsburgh, Pittsburgh, Pennsylvania; ²Intermountain Health Care Institute for Health Care Delivery Research, Salt Lake City, Utah;
³University of Utah, Salt Lake City, Utah; ⁴Carnegie Mellon University, Pittsburgh, Pennsylvania

Corresponding author: Michael M. Wagner, Real-Time Outbreak and Disease Laboratory, University of Pittsburgh, Suite 500, Cellomics Building, 500 Technology Drive, Pittsburgh, PA 15219. Telephone: 412-383-8137; Fax: 412-383-8135; E-mail: mmw@cbmi.pitt.edu.

Abstract

This paper summarizes the experience of the Real-Time Outbreak and Disease Surveillance (RODS) project in collecting and analyzing free-text emergency department (ED) chief complaints. The technical approach involves real-time transmission of chief-complaint data as Health Level 7 messages from hospitals to a regional data center, where a Bayesian text classifier assigns each chief complaint to one of eight syndrome categories. Time-series algorithms analyze the syndrome data and generate alerts. Authorized public health users review the syndrome data by using Internet interfaces with timelines and maps. Deployments in Pennsylvania, Utah, Atlantic City, and Ohio have demonstrated feasibility of real-time collection of chief complaints. Retrospective experiments that measured case-classification accuracy demonstrated that the Bayesian classifier can discriminate between different syndrome presentations. Retrospective experiments that measured outbreak-detection accuracy determined that the classifier's performance was adequate to support accurate and timely detection of seasonal disease outbreaks. Prospective evaluation revealed that a cluster of carbon monoxide exposures was detected by RODS within 4 hours of the presentation of the first case to an emergency department.

Introduction

In 1999, the Real-Time Outbreak and Disease Surveillance (RODS) project created a regional test bed in a large metropolitan area (population: 2.3 million persons) that had the characteristic of high sampling density (i.e., monitoring of >50% of the population for at least one type of data). The project then used this test bed to study detectability of outbreaks, especially detectability of cohort exposures (e.g., a citywide aerosolized *Bacillus anthracis* release) that have a narrow window of opportunity for mitigation and thus present a substantial surveillance challenge (1). After early studies of laboratory data (2) and *International Classification of Diseases, Ninth Revision* (ICD-9) coded chief complaints (3,4), later research focused on analysis of free-text chief complaints. This paper describes the experience of the RODS project in collecting and analyzing patient chief complaints.

Methods

The technical approach to Health Level 7 (HL7)-based data collection and chief-complaint processing has been described previously (5-9). Briefly, when a patient registers for care at an ED, a triage nurse or registration clerk enters the patient's reason for visit (known as a chief complaint) into a

registration system. This step is part of the normal workflow in multiple U.S. hospitals (10). The registration system transmits chief-complaint data in the form of HL7 messages (5) to an HL7 message router in the hospital, which can de-identify these messages and transmit them via the Internet to a health department in real time.

At the health department, a naïve Bayesian classifier (9) encodes each chief complaint into one of eight mutually exclusive and exhaustive syndromic categories (respiratory, gastrointestinal, botulinic, constitutional, neurologic, rash, hemorrhagic, and none of the above). RODS software then aggregates the data into daily counts by syndrome and residential zip code for analysis by time-series algorithms and display on interfaces using timelines and maps.

Validation

A goal of the project has been to test whether early detection of outbreaks can be achieved through statistical analysis of chief-complaint data (or other routinely collected data). Although chief complaints are insufficient for accurate diagnosis of an individual patient, the hypothesis is that they contain sufficient information so that, when aggregated into daily population counts and analyzed by using spatio-temporal algorithms, early detection of an abnormally high number of

persons who have contracted a respiratory or other illness is possible.

Case-Detection Accuracy

The research team conducted numerous experiments to test this hypothesis. The first type of experiment measured the information content of chief complaints for syndrome categorization by measuring the sensitivity and specificity with which patients with different syndromes can be detected from their chief complaints alone (Table). Each experiment tested the ability of a classifier program to accurately assign a syndrome to a patient on the basis of the chief complaint alone (in certain experiments, the patient data were ICD-9-coded ED diagnoses). For example, one experiment measured the accuracy of the Bayesian text classifier for respiratory syndrome in comparison with a manual determination made by the Utah Department of Health during the 2002 Winter Olympic

Games. In that experiment, the Bayesian respiratory classifier detected 52% of affected patients, with a specificity of 89%.

The experiments demonstrated that chief-complaint data contains information about the syndromic presentation and that a naïve Bayesian classifier can extract that information. For certain syndromes of interest to terrorism preparedness, the sensitivity of classification is approximately 0.5 (i.e., in the event of an outbreak causing respiratory complaints, 50% of affected patients examined at a monitored facility would be detected).

Outbreak Detection

As expected, the case-detection experiments demonstrate that the specificity of case classification from chief complaints is <100%, meaning that daily counts of patients with respiratory syndrome would contain noise attributable to falsely classified nonrespiratory patients. Therefore, a second type of

TABLE. Performance of Bayesian and other classifiers in detecting selected syndromes

Classifier being tested	Standard cases for comparison	Sensitivity		Specificity		Positive likelihood ratio (LR+)	
		No.	(95% CI)*	No.	(95% CI)	No.	(95% CI)
Respiratory syndrome							
Chief complaint Bayesian classifier (CoCo) respiratory†	Utah Department of Health (UDOH) respiratory with fever§	0.52	(0.51–0.54)	0.89	(0.89–0.90)	5.0	(4.74–5.22)
CoCo respiratory (11)	Human review	0.77	(0.59–0.88)	0.90	(0.88–0.92)	7.9	(5.8–10.8)
CoCo respiratory†	Utah ICD-9¶	0.60	(0.59–0.62)	0.94	(0.94–0.95)	10.5	(9.90–11.05)
CoCo respiratory with fever (11)	Human review	0.22	(0.06–0.55)	0.99	(0.98–0.99)	24.5	(5.7–105.3)
ICD-9 respiratory (4)	Human review	0.44	(0.29–0.61)	0.97	(0.96–0.98)	15.6	(8.6–28.1)
Gastrointestinal (GI) syndrome							
CoCo GI†	UDOH gastroenteritis without blood	0.71	(0.69–0.74)	0.90	(0.90–0.90)	7.1	(6.80–7.51)
CoCo acute infectious GI (12)	Human review	0.63	(0.35–0.85)	0.94	(0.92–0.96)	12.2	(8.3–18)
						0.82	(0.75–0.90)
ICD-9 acute infectious GI (12)	Human review	0.32	(0.14–0.54)	0.99	(0.98–0.99)	37.1	(16.2–85.3)
CoCo GI†	Utah ICD-9	0.74	(0.72–0.76)	0.92	(0.92–0.92)	9.5	(9.04–9.94)
CoCo GI with diarrhea (11)	Human review	0.12	(0.06–0.22)	0.99	(0.99–0.99)	81.1	(17.56–374.4)
CoCo GI with vomiting (11)	Human review	0.16	(0.11–0.24)	0.99	(0.99–0.99)	105	(24.85–444.33)
Neurologic/encephalitic syndrome							
CoCo neurologic†	UDOH meningitis/ encephalitis	0.47	(0.32–0.63)	0.93	(0.93–0.94)	7.1	(4.98–9.99)
CoCo neurologic†	Utah ICD-9	0.72	(0.69–0.76)	0.95	(0.94–0.95)	13.5	(12.57–14.41)
Rash							
CoCo rash†	UDOH febrile illness with rash§	0.50	(0.40–0.59)	0.99	(0.99–0.99)	55.6	(44.25–69.91)
CoCo rash†	Utah ICD-9	0.60	(0.52–0.67)	0.99	(0.99–0.99)	80.9	(67.43–97.07)
Botulinic syndrome							
CoCo botulinic†	UDOH botulism-like syndrome	0.17	(0.05–0.45)	0.998	(0.998–0.999)	104	(28.57–381.86)
CoCo botulinic†	Utah ICD-9	0.22	(0.13–0.36)	0.999	(0.998–0.999)	167	(89.07–312.90)
Fever							
Keyword fever (13)	Human review	0.61	(0.51–0.69)	1.0	(0.96–1.0)	**	—
Fever from emergency department report (13)	Human review	0.98	(0.94–0.99)	0.89	(0.82–0.94)	9.3	(5.3–16.2)

* CI = confidence interval.

† Source: Gesteland PH, unpublished results, August 4, 2003.

§ Classifier trained on less-specific training classifications than standard, which required documentation of fever in the patient record.

¶ International Classification of Diseases, Ninth Revision.

$$** LR+ = \frac{\text{sensitivity}}{1 - \text{specificity}} = \frac{0.61}{1 - 1} = \text{undefined}$$

experiment was needed to determine whether outbreaks would produce a sufficiently large spike to stand out from the background noise in the daily syndrome counts (and to determine how early any spikes would occur). In these outbreak-detection experiments, a time-series detection algorithm was run on 3 years of daily syndrome counts from metropolitan areas that had experienced annual winter outbreaks. The time of detection from daily syndrome counts was determined as the date the algorithm first signaled during the beginning of the seasonal outbreak and was compared with the time of detection from ICD-9-coded hospital diagnoses (14). For detection of three pediatric gastrointestinal outbreaks, detection occurred 29 days earlier (95% confidence interval [CI] = 4–53) with no false alarms. For pediatric and adult respiratory outbreaks, detection occurred 10 days earlier (95% CI = -15–35) and 11 days earlier (95% CI = -10–33), respectively, also with no false alarms.

Early Experience with Prospective Evaluation

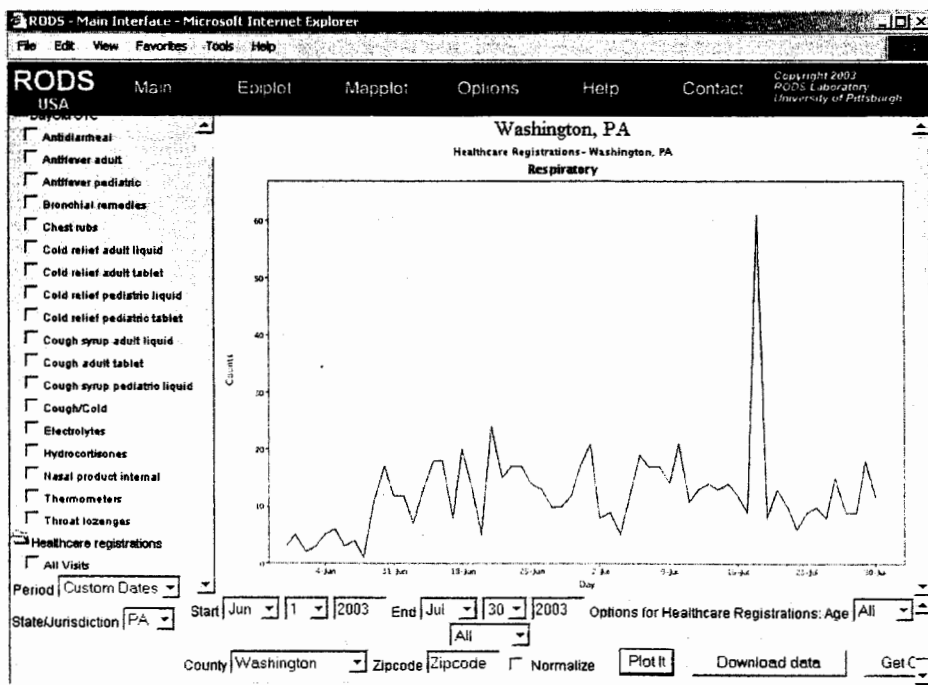
Retrospective studies cannot prove that, in field use, this type of system will lead to earlier detection than existing methods. For this reason, the project initiated a prospective evaluation.

The RODS test bed enables public health officials to examine timelines and maps whenever an outbreak occurs or whenever they receive alerts of anomalous syndrome activity. On Friday, July 18, 2003, an on-call epidemiologist received an alert regarding a spike in respiratory cases in a single county outside Pittsburgh (Figure). Normally, daily counts of respiratory cases numbered 10, but on that day they numbered 60. The epidemiologist logged onto the RODS interface, reviewed the verbatim chief complaints of affected patients, and discovered that all were related to carbon monoxide exposure from a faulty furnace. (Authorized public health users can access case studies of these and other outbreaks through the RODS interface by sending e-mail to nrdmaccounts@cbmi.pitt.edu).

Technology Dissemination

After rapid (6-week) deployment in February 2002 during the Winter Olympics, RODS had a proven model for building permanent, real-time, HL7-based data feeds of chief complaints from hospitals to public health agencies. Such feeds would have immediate surveillance use and could later be expanded to include transmission of data about microbiology results. However, because adoption of the RODS approach has been slower than expected, the project began to systematically identify and address barriers to dissemination. One barrier was the perception that such approaches are still unproven and would absorb public health resources through technology costs and false alarms (15,16). A second barrier was limited availability of software and lack of technical expertise. Accordingly, the University of Pittsburgh agreed to distribute the RODS system free of charge in 2002. Although this action resulted in hundreds of downloads of both the RODS system and the Bayesian parser, certain health depart-

FIGURE. Daily counts of respiratory cases — Washington County, Pennsylvania, June–July 2003



Source: Real-Time Outbreak and Disease Surveillance project.

* The June 2003 increase corresponds to new hospitals being added to the system.

† The sudden increase on July 18, 2003, was caused by 60 persons reporting to one emergency department within 4 hours for carbon monoxide exposure.

ments lack expertise in database administration, network administration, geographic information systems, HL7, and systems management. The RODS laboratory helped Utah and Pennsylvania avoid this barrier by hosting their surveillance operations. A cost model for this service was then developed, and the service was offered to other states, which led to implementation in Ohio and New Jersey. In addition, the RODS Open Source Project (<http://openrods.sourceforge.net>) was created in 2003 to catalyze the growth of a community of consultants to help health departments install and operate surveillance systems (17). In 2003, the University of Pittsburgh placed the RODS source code into the public domain under the GNU General Public License (18). Open-sourcing a project can facilitate technology dissemination because it directly addresses information technology managers' concerns about access to source code, code sustainability, customizability, and availability of expertise.

Status of RODS

RODS has operated continuously since 1999, connecting with 51 hospitals in Pennsylvania, 10 hospitals and 17 urgent care facilities in Utah, 12 hospitals in Ohio, and four hospitals in New Jersey. The system is also installed in Taiwan and Michigan.

Conclusions

Free-text chief-complaint data are useful in public health surveillance because they are widely available and can be obtained in real time for modest cost. Moreover, the HL7 technical infrastructure thus created can later be expanded to transmit other types of data. The technical expertise and cost to create and operate a real-time facility is substantial; therefore, sharing costs by using application service providers leads to cheaper and faster deployment.

Acknowledgments

This work was supported by Pennsylvania Department of Health Grant ME-107, the Defense Advanced Research Projects Agency, the Agency for Healthcare Research and Quality, and the National Library of Medicine. The authors also thank Jagan Dara, Feng Dong, William Hogan, Robert Olszewski, Hoah Su, and Virginia Dato.

References

1. Kaufmann A, Meltzer M, Schmid G. The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? *Emerg Infect Dis* 1997;3:83-94.
2. Panackal AA, M'ikanatha NM, Tsui F-C, et al. Automatic electronic laboratory-based reporting of notifiable infectious diseases. *Emerg Infect Dis* 2001;8:685-91.
3. Tsui F-C, Wagner MM, Dato V, Chang CC. Value of ICD-9 coded chief complaints for detection of epidemics. *Proc AMIA Symp* 2001:711-5.
4. Espino JU, Wagner MM. The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness. *Proc AMIA Symp* 2001:164-8.
5. Tsui F-C, 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.
6. Tsui F-C, Espino JU, Wagner MM, et al. Data, network, and application: technical description of the Utah RODS Winter Olympic biosurveillance system. *Proc AMIA Symp* 2002:815-9.
7. Gesteland PH, Gardner RM, Tsui F-C, et al. Automated syndromic surveillance for the 2002 Winter Olympics. *J Am Med Inform Assoc* 2003;10:547-54.
8. Gesteland PH, Wagner MM, Chapman WW, et al. Rapid deployment of an electronic disease surveillance system in the state of Utah for the 2002 Olympic Winter Games. *Proc AMIA Symp* 2002:285-9.
9. Olszewski R. Bayesian classification of triage diagnoses for the early detection of epidemics. In: Russell I, Haller S., eds. Recent advances in artificial intelligence: proceedings of Sixteenth International Florida Artificial Intelligence Research Society Conference. Menlo Park, CA: AAAI Press, 2003:412-6.
10. Travers D, Waller A, Haas S, Lober W, Beard C. Emergency department data for bioterrorism surveillance. *Proc AMIA Symp* 2003:664-8.
11. Chapman WW, Espino JU, Dowling JN, Wagner MM. Detection of multiple symptoms from chief complaints. Technical Report, CBMI Report Series, 2003.
12. 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.
13. Chapman WW, Dowling JN, Wagner MM. Fever detection from free-text clinical records for biosurveillance. Technical Report, CBMI Report Series, 2003.
14. Ivanov O, Gesteland PH, Hogan W, Tsui F-C, Wagner MM. Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints 2003. *Proc AMIA Symp* 2003:318-22.
15. Reingold A. If syndromic surveillance is the answer, what is the question? *Biosecur Bioterr* 2003;1:1-5.
16. Broome C, Pinner R, Sosin D, Treadwell T. On the threshold. *Am J Prev Med* 2002;23:229.
17. Espino JU, Wagner MM, Szczepaniak M, et al. Removing a barrier to computer-based outbreak and disease surveillance—The RODS Open Source Project. *MMWR* 2004;53(Suppl):32-9.
18. GNU General Public License. Available at <http://www.gnu.org/licenses/gpl.html>.



MMWRTM

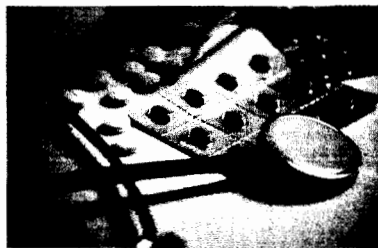
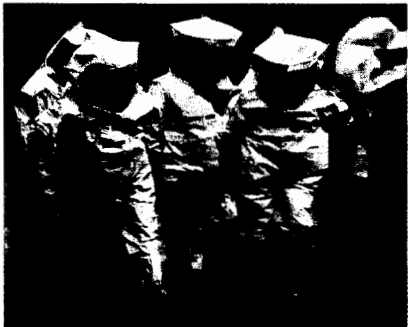
Morbidity and Mortality Weekly Report

Supplement

September 24, 2004 / Vol. 53

Syndromic Surveillance

Reports from a National Conference, 2003



DEPARTMENT OF HEALTH AND HUMAN SERVICES
CENTERS FOR DISEASE CONTROL AND PREVENTION