MMWR™

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

The *MMWR* series of publications is published by the Epidemiology Program Office, Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services, Atlanta, GA 30333.

SUGGESTED CITATION

General: Centers for Disease Control and Prevention. Syndromic Surveillance: Reports from a National Conference, 2003. MMWR 2004:53(Suppl).

Specific: [Author(s)]. [Title of particular article]. In: Syndromic Surveillance: Reports from a National Conference, 2003. MMWR 2004;53 (Suppl):[inclusive page numbers].

Notice

This supplement is being published as a courtesy to the *MMWR* readership. These papers represent the views of the authors, and their publication does not imply endorsement by CDC.

CONTENTS

CONTENTS (*Continued*)

Preface

Effective public health surveillance is essential for detecting and responding to emerging public health threats, including terrorism and emerging infectious diseases. New surveillance methods are being developed and tested to improve the timeliness and completeness of detection of disease outbreaks. One promising set of approaches is syndromic surveillance, in which information about health events that precede a firm clinical diagnosis is captured early and rapidly from existing, usually electronic, data sources, and analyzed frequently to detect signals that might indicate an outbreak requiring investigation.

To provide a forum for scientists and practitioners to report on progress in developing and evaluating syndromic surveillance systems, the New York City Department of Health and Mental Hygiene, the New York Academy of Medicine, and CDC convened the second annual National Syndromic Surveillance Conference in New York City during October 23–24, 2003. The conference, supported by the Alfred P. Sloan Foundation, was attended by more than 460 public health practitioners and researchers, who had the opportunity to hear 41 oral presentations and view 50 poster presentations.

The original papers and posters for this conference were chosen by a scientific program committee after a review of submitted abstracts. Senior researchers in the field were also invited to address key concerns in surveillance for early detection of outbreaks. All participants who presented papers or posters at either the conference or at a preconference workshop were invited to submit manuscripts based on their presentations for publication in this *Morbidity and Mortality Weekly Report Supplement*. Each manuscript was then reviewed by at least two peer reviewers and final publication decisions were made by an editorial committee. Many of the articles are considerably different from the material originally presented at the conference. Certain authors updated their findings, and others were asked to revise their papers into descriptions of syndromic surveillance systems. Other presenters chose to submit only abstracts. Papers are presented here in the following order: system descriptions, research methods, evaluation, and public health practice.

In addition to these reports, other resources on syndromic surveillance are available. The proceedings of the 2002 National Syndromic Surveillance Conference were published in the *Journal of Urban Health* (accessible at http://jurban.oupjournals.org/content/suppl_1/index.shtml). In May 2004, a revised *Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks* was published (*MMWR* 2004;53[No. RR-5]). An annotated bibliography of published papers and other Internet-accessible materials has been developed and is maintained monthly on a CDC website (http://www.cdc.gov/epo/dphsi/syndromic/index.htm). An Internet-based forum (http://syndromic.forum.cdc.gov) was established for discussion of topics related to syndromic surveillance and was used to distribute answers to audience questions raised at the conference. A related forum (http://surveval.forum.cdc.gov) has been maintained for discussion of topics related to surveillance system evaluation. Finally, the website of the Annual Syndromic Surveillance Conferences (http://www.syndromic.org) includes links to recent news and scientific articles about syndromic surveillance, oral and poster presentations and workshop materials from past conferences, and notices of upcoming conferences. The third National Syndromic Surveillance Conference is planned for November 3–4, 2004, in Boston, Massachusetts.

The editorial committee acknowledges the work of the scientific planning committee: Dennis Cochrane, Christine Hahn, Patrick Kelley, Martin Kulldorff, John Loonsk, David Madigan, Richard Platt, and Don Weiss. The committee is also grateful for the support and efforts of the following staff members in conducting this conference and developing this *Supplement*: Alan Fleischman, Irv Gertner, and Jessica Hartman, New York Academy of Medicine; Rick Heffernan, New York Department of Health and Mental Hygiene; and Alan Davis, Division of Public Health Surveillance and Informatics, Epidemiology Program Office, CDC; Valerie Kokor, Division of International Health, Epidemiology Program Office; and Stephanie Malloy, Jeffrey Sokolow, and Malbea LaPete, *MMWR*, Epidemiology Program Office, CDC. Special thanks are given to JoEllen DeThomasis, Division of Applied Public Health Training and Division of Public Health Surveillance and Informatics, Epidemiology Program Office, CDC, who coordinated the preparation of these reports.

— *The Editorial Committee*

Peer Reviewers

The following persons served as peer reviewers for these proceedings: Raymond D. Aller, M.D.; Los Angeles County Public Health; Janet M. Blair, Ph.D., Division of Applied Public Health Training, Epidemiology Program Office, CDC; David L. Buckeridge, M.D., Stanford University; Palo Alto Veterans Healthcare; James W. Buehler, M.D., Rollins School of Public Health, Emory University; Howard S. Burkom, Ph.D., Johns Hopkins University Applied Physics Laboratory; Man-huei Chang, M.P.H., Division of Public Health Surveillance and Informatics, Epidemiology Program Office, CDC; Dennis G. Cochrane, M.D., Emergency Medical Associates of New Jersey Research Foundation; Morristown Memorial Hospital Residency in Emergency Medicine; Andrew G. Dean, M.D., EpiInformatics.com; Rosaline Dhara, M.A., M.P.H.; Division of Public Health Surveillance and Informatics, Epidemiology Program Office, CDC; Jeremy U. Espino, M.D.; University of Pittsburgh; Laura J. Fehrs, M.D., Division of Applied Public Health Training, Epidemiology Program Office, CDC; Kathleen F. Gensheimer, M.D., Maine Department of Human Services; Joe Fred Gonzalez, Jr., M.S.; Office of the Director, National Center for Health Statistics, CDC; Ivan J. Gotham, Ph.D., New York State Department of Health; University at Albany, School of Public Health; Timothy A. Green, Ph.D., Division of HIV, STD, and TB Prevention/Surveillance and Epidemiology, National Center for HIV, STD, and TB Prevention, CDC; Jesse Greenblatt, M.D., New Hampshire Department of Health and Human Services; Joel R. Greenspan, M.D., Quintiles Transnational; Samuel L. Groseclose, D.V.M.; Division of Sexually Transmitted Disease Prevention, National Center for HIV, STD, and TB Prevention, CDC; Christine G. Hahn, M.D., Idaho Department of Health and Welfare; Jessica J. Hartman, M.S., New York City Department of Health and Mental Hygiene; New York Academy of Medicine; Richard T. Heffernan, M.P.H., New York City Department of Health and Mental Hygiene; Kelly J. Henning, M.D., New York City Department of Health and Mental Hygiene; Patrick W. Kelley, M.D., Dr.P.H., The National Academies; Ken Kleinman, Sc.D., Harvard Medical School and Harvard Pilgrim Health Care; Melvin Kohn, M.D., Oregon Department of Human Services; Denise Koo, M.D., Division of Applied Public Health Training, Epidemiology Program Office, CDC; Martin Kulldorff, Ph.D., Harvard Medical School and Harvard Pilgrim Health Care; Harvey B. Lipman, Ph.D., Division of Laboratory Systems, Public Health Practice Program Office, CDC; William B. Lober, M.D., University of Washington; David Madigan, Ph.D., Rutgers University; Julie M. Magri, M.D., Division of Applied Public Health Training, Epidemiology Program Office, CDC; Martin Mendelson, M.D., Ph.D., Division of Public Health Surveillance and Informatics, Epidemiology Program Office, CDC; Nkuchia M. M'ikanatha, Dr.P.H., Pennsylvania Department of Health; Ronald L. Moolenaar, M.D., Division of Applied Public Health Training, Epidemiology Program Office, CDC; Roy Gibson Parrish, II, M.D., formerly of CDC; Julie A. Pavlin, M.D., Walter Reed Army Institute of Research; Carol A. Pertowski, M.D., Office of the Director, Epidemiology Program Office, CDC; Stephen P. Pickard, M.D., Division of Applied Public Health Training, Epidemiology Program Office, CDC; North Dakota Department of Health; Robert W. Pinner, M.D., Office of the Director, National Center of Infectious Diseases, CDC; Richard Platt, M.D., Harvard Medical School and Harvard Pilgrim Health Care; Daniel A. Pollock, M.D., Office of the Director, CDC; Julie A. Rawlings, M.P.H., Texas Department of Health; Janise Richards, Ph.D., Division of Public Health Surveillance and Informatics, Epidemiology Program Office, CDC; Robert T. Rolfs, M.D., Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion, CDC and Utah Department of Health; Henry Rolka, M.P.S., M.S., Office of the Director, CDC; Patricia Schumacher, M.S., R.D., Division of Public Health Surveillance and Informatics, Epidemiology Program Office, CDC; David William Siegrist, M.A., M.S.M., Potomac Institute for Policy Studies; Maya R. Sternberg, Ph.D., Division of Sexually Transmitted Disease Prevention, National Center for HIV, STD, and TB Prevention, CDC; Donna F. Stroup, Ph.D., Office of the Director, National Center for Chronic Disease Prevention and Health Promotion, CDC; Stephen B. Thacker, M.D., Director, Epidemiology Program Office; Van T. Tong, M.P.H., Division of Applied Public Health Training, Epidemiology Program Office, CDC; David Walker, M.P.H., Office of the Director, CDC; Sylvan Wallenstein, Ph.D., Mount Sinai Medical Center; Don Weiss, M.D., New York City Department of Health and Mental Hygiene; Scott F. Wetterhall, M.D., DeKalb County Board of Health; Christopher W. Woods, M.D., Duke University Medical Center; and Akbar A. Zaidi, Ph.D., Division of Sexually Transmitted Disease Prevention, National Center for HIV, STD, and TB Prevention, CDC.

Overview of Syndromic Surveillance

What is Syndromic Surveillance?

Kelly J. Henning
New York City Department of Health and Mental Hygiene, New York, New York

Corresponding author: Kelly J. Henning, New York City Department of Health and Mental Hygiene, 125 Worth Street, CN-6, New York, NY 10013. Telephone: 212-788-0234; Fax: 212-788-4473; E-mail: khenning@health.nyc.gov.

Abstract

Innovative electronic surveillance systems are being developed to improve early detection of outbreaks attributable to biologic terrorism or other causes. A review of the rationale, goals, definitions, and realistic expectations for these surveillance systems is a crucial first step toward establishing a framework for further research and development in this area. This commentary provides such a review for current syndromic surveillance systems.

Syndromic surveillance has been used for early detection of outbreaks, to follow the size, spread, and tempo of outbreaks, to monitor disease trends, and to provide reassurance that an outbreak has not occurred. Syndromic surveillance systems seek to use existing health data in real time to provide immediate analysis and feedback to those charged with investigation and follow-up of potential outbreaks. Optimal syndrome definitions for continuous monitoring and specific data sources best suited to outbreak surveillance for specific diseases have not been determined. Broadly applicable signal-detection methodologies and response protocols that would maximize detection while preserving scant resources are being sought.

Stakeholders need to understand the advantages and limitations of syndromic surveillance systems. Syndromic surveillance systems might enhance collaboration among public health agencies, health-care providers, information-system professionals, academic investigators, and industry. However, syndromic surveillance does not replace traditional public health surveillance, nor does it substitute for direct physician reporting of unusual or suspect cases of public health importance.

Introduction

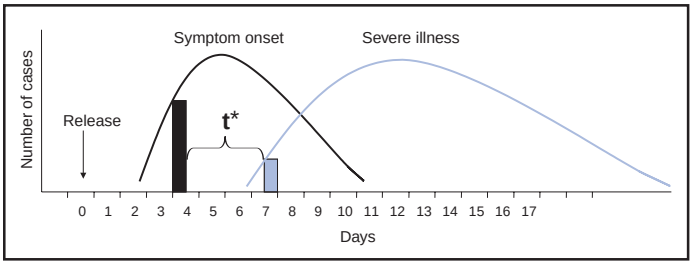
The desire to expand and improve upon traditional methods of public health surveillance is not new. Even before the 2001 terrorist attacks on the United States and the subsequent anthrax outbreak, public health officials had begun to enhance detection of emerging infections and illnesses caused by biologic agents. A primary objective of a 1998 CDC plan was to develop programs for early detection and investigation of outbreaks (*1*). CDC's 2000 strategic plan for biologic and chemical preparedness called for early detection by integrating terrorism preparedness into existing systems and developing "new mechanisms for detecting, evaluating, and reporting suspicious events" (*2*). Although the need for innovative surveillance techniques had already been identified, the anthrax outbreak after *Bacillus anthracis* spores were released through the mail in 2001 (*3*) accelerated the implementation of syndromic surveillance systems across the United States. An overview of the location and scope of the earliest systems implemented before and after fall 2001 has been published (*4*).

Goals and Rationale

Although syndromic surveillance was developed for early detection of a large-scale release of a biologic agent, current surveillance goals reach beyond terrorism preparedness. Medical-provider reporting remains critical for identifying unusual disease clusters or sentinel cases. Nevertheless, syndromic surveillance might help determine the size, spread, and tempo of an outbreak after it is detected (*5*), or provide reassurance that a large-scale outbreak is not occurring, particularly in times of enhanced surveillance (e.g., during a high-profile event). Finally, syndromic surveillance is beginning to be used to monitor disease trends, which is increasingly possible as longitudinal data are obtained and syndrome definitions refined.

The fundamental objective of syndromic surveillance is to identify illness clusters early, before diagnoses are confirmed and reported to public health agencies, and to mobilize a rapid response, thereby reducing morbidity and mortality. Epidemic curves for persons with earliest symptom onset and those with severe illness can be depicted graphically (Figure). The time between symptom onset for an increasing number of cases caused by deliberate release of a biologic agent and subsequent patient visits to a health-care facility resulting in a definitive diagnosis is represented by *t*. Syndromic surveillance aims to identify a threshold number of early symptomatic cases, allowing detection of an outbreak *t* days earlier than would conventional reporting of confirmed cases. The ability of syndromic surveillance to detect outbreaks earlier than con-

FIGURE. Syndromic surveillance — rationale for early detection



* t = time between detection by syndromic (prediagnostic) surveillance and detection by traditional (diagnosis-based) surveillance.

ventional surveillance methods depends on such factors as the size of the outbreak, the population dispersion of those affected, the data sources and syndrome definitions used, the criteria for investigating threshold alerts, and the health-care provider's ability to detect and report unusual cases (*6*). CDC's framework for evaluating public health surveillance systems for early detection of outbreaks should be useful for comparing syndromic surveillance across jurisdictions and for evaluating system performance (*7*).

Specific definitions for syndromic surveillance are lacking, and the name itself is imprecise. Certain programs monitor surrogate data sources (e.g., over-the-counter prescription sales or school absenteeism), not specific disease syndromes. Meanwhile, certain well-defined disease or clinical syndromes (e.g., hemolytic uremic syndrome or Kawasaki's syndrome) are not included in syndrome definitions, often leading to confusion about what "syndromic" surveillance actually monitors. Diverse names used to describe public health surveillance systems for early outbreak detection include

- early warning systems (*8,9*);
- prodrome surveillance (*10*);
- outbreak detection systems (*11*);
- information system-based sentinel surveillance (*12*);
- biosurveillance systems (*13–15*);
- health indicator surveillance (*16*); and
- symptom-based surveillance (*17*).

However, *syndromic surveillance* is the term that has persisted.

In defining syndromic surveillance, certain authors have emphasized the importance of monitoring the frequency of illnesses with a specific set of clinical features (*18*), a definition that does not account for nonclinical data sources. Others have emphasized the importance of prediagnostic data to estimate a community's health status, particularly by relying on outpatient visits (*19*). Inherent in the use of existing electronic data to describe prediagnostic health indicators is the central role of timeliness in the analysis, detection, and investigation of alerts. Perhaps the most comprehensive definition to date, and likely the one to be broadly adopted, is provided by CDC's evaluation framework, which describes syndromic surveillance as "an investigational approach where health department staff, assisted by automated data acquisition and generation of statistical alerts, monitor disease indicators in real-time or near real-time to detect outbreaks of disease earlier than would otherwise be possible with traditional public health methods" (*7*).

Syndromic surveillance systems vary by their planned duration and their manner of acquiring data (Table). Short-duration, event-based systems are usually used to provide enhanced surveillance around a discrete event (e.g., the Olympic Games or a national political convention) (*20,23*). Historically, these short-term syndromic surveillance projects, sometimes termed *drop-in surveillance*, have required medical providers or others to collect nonroutine information (*20*). More recent event-based surveillance systems have relied on rapid implementation of electronically transferred data (*23*). Manual data entry, which occurred after September 11, 2001, in 15 New York City emergency departments (EDs), is difficult to sustain (*21*). Using pre-existing health data for syndromic surveillance offers immediate accessibility and poses limited burden to providers and health-care institutions.

Categorizing symptoms and diagnoses into syndromes is a fundamental component of syndromic surveillance systems that use clinical data sets. Although the majority of investigators have devised broad categories aimed at early detection of biologic terrorism, validation of syndrome definitions is only beginning. Respiratory, gastrointestinal, rash, neurologic and sepsis syndromes have been monitored consistently (*19,22*). Because numerous ED and outpatient settings have *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) data available electronically, ICD-9-CM codes have been used to categorize syndromes. To facilitate comparability between surveillance systems, a CDC working group published lists of candidate syndrome groups based on ICD-9-CM codes (*27*). The usefulness of ICD-9-CM codes compared with other data streams, particularly with regard to the data's timeliness, requires evaluation by each surveillance program.

Syndromic surveillance focuses on the early symptom (prodrome) period before clinical or laboratory confirmation of a particular disease and uses both clinical and alternative data sources (Box). Strictly defined, syndromic surveillance gathers information about patients' symptoms (e.g., cough, fever, or shortness of breath) during the early phases of illness. However, in practice, certain syndromic surveillance systems collect surrogate data indicating early illness (e.g., school or work

TABLE. Types of syndromic surveillance — selected characteristics, advantages, and disadvantages

Surveillance type	Selected characteristics	Advantages	Disadvantages
Event-based surveillance			
Drop-in (*20,21*)	Active Defined duration Emergency departments (EDs) Large clinics	Develop relationships with ED staff and infection-control professionals Transportable to various sites	Labor-intensive Not sustainable Not scalable
Sustained surveillance			
Manual (*22*)	Active and passive Fax-based reporting ED triage staff typically log and tally sheets	Develop relationships with hospital staff Easy to initiate Detailed information obtainable	Labor-intensive Difficult to maintain 24 hours, 7 days/week Not sustainable
Electronic (*8,19,23,24*)	Passive Automated transfer of hospital (usually ED triage or diagnosis) or outpatient data Use of data collected for other purposes Data mining of large collections or from multiple sources	Can be scalable Requires minimal or no provider input Data available continuously Data are standardized	Need programming and informatics expertise Confidentiality issues
Novel modes of collection (*25*)	Passive Hand-held or touch-screen devices	Easy to use; rapid provider feedback; can post alerts and information	Requires provider input Not sustainable
Novel data sources (*26*)	Active and passive Medical examiner data Unexplained death or severe illness data	Clearly defined syndrome Can be supplemented with laboratory data	Not an early warning Unclear whether it can be rapidly and broadly expanded

absenteeism data or veterinary data such as unexpected avian deaths or other potential precursors of human illness). Alternative data sources have potential problems, including a presumed low specificity for syndromes of interest, high probability of influence by factors unrelated to personal health (e.g., weather or holidays), and difficulty in retracing data aberrations to individual patients. Despite these qualifiers, the optimal system might be one that integrates data from multiple sources, potentially increasing investigators' confidence in the relevance of an alert from any single data source.

Analytic Methods for Signal Detection

The analytic challenge in using syndromic surveillance for outbreak detection is to identify a signal corresponding to an outbreak or cluster amid substantial "background noise" in the data. Syndromic surveillance systems use an array of aberration-detection methods to identify increases in syndromes above predetermined thresholds. However, signal-detection methods have not yet been standardized. Temporal and spatio-temporal methods have been used to assess day-to-day and day and place variability of data from an expected baseline (*27,28*).

BOX. Potential data sources for syndromic surveillance

Clinical data sources
Emergency department (ED) or clinic total patient volume
Total hospital or intensive-care–unit admissions from ED
ED triage log of chief complaints
ED visit outcome (diagnosis)
Ambulatory-care clinic/HMO outcome (diagnosis)
Emergency medical system (911) call type
Provider hotline volume, chief complaint
Poison control center calls
Unexplained deaths
Medical examiner case volume, syndromes
Insurance claims or billing data
Clinical laboratory or radiology ordering volume

Alternative data sources
School absenteeism
Work absenteeism
Over-the-counter medication sales
Health-care provider database searches
Volume of Internet-based health inquiries by the public
Internet-based illness reporting
Animal illnesses or deaths

Response Protocols

Response protocols for investigating syndromic surveillance alerts are under development by multiple programs. Obstacles to effective, efficient follow-up include the difficulty of predicting how well the syndromes themselves correlate with target diseases under surveillance; the extremely low positive predictive value of any given signal based on the high level of system sensitivity; and investigators' relative lack of experience with syndromic surveillance under real-world conditions (*30*).

Programmatic requirements for effective signal response (e.g., documented procedures; staff with appropriate expertise; 24-hour/day, 7-day/week analysis and response; and plans for information dissemination) are complex. Certain circumstances surrounding an alert might prompt rapider investigation, including clustering of cases by location; severe symptoms; unexplained deaths; sudden, substantial case numbers; simultaneous alerts from multiple data sources; or restriction of an alert to a particular population (e.g., age group or sex) (*31*). Diagnostic confirmation is a paramount step in investigating alerts, particularly given the nonspecific nature of certain syndrome categories. Developing protocols to address alerts from data sources in which individual cases are unidentifiable (e.g., over-the-counter medication sales) is particularly challenging.

Perspectives and Challenges

Distinguishing those points on which multiple investigators agree from those that are less well-delineated might be helpful in defining realistic expectations for syndromic surveillance. Investigators usually agree on the following:

- Syndromic surveillance is being used in numerous states and localities to detect a potential large-scale biologic attack.
- Pre-existing electronic health data will likely become increasingly available, thereby enhancing system development.
- Syndromic surveillance does not replace traditional public health surveillance.
- Syndromic surveillance is unlikely to detect an individual case of a particular illness.
- Syndromic surveillance cannot replace the critical contribution of physicians in early detection and reporting of unusual diseases and events.

Although syndromic surveillance's ability to detect a terrorism-related outbreak earlier than traditional surveillance remains unknown, it will likely be useful for defining the scope of an outbreak, providing reassurance that a large-scale outbreak has not occurred, and conducting surveillance of noninfectious health problems (e.g., monitoring nicotine replacement therapy sales following tobacco-tax increases). However, integral components of syndromic surveillance require additional research and evaluation, including the following:

- defining optimal data sources;
- evaluating appropriate syndromic definitions;
- standardizing signal-detection methods;
- developing minimally acceptable response protocols;
- clarifying the use of simulation data sets to test systems; and
- advancing the debate regarding resource commitment for syndromic versus traditional surveillance.

On a broader policy level, defining the role of academic partners in bridging any potential analytic gaps, defining the role and scope of a national syndromic data repository, and developing policy for integrating laboratory testing and laboratory information systems with syndromic surveillance are on the horizon.

Acknowledgments

Farzad Mostashari, Don Weiss, Rick Heffernan, and other members of the New York City Department of Health and Mental Hygiene syndromic surveillance team provided data and program information.

References

1. CDC. Preventing emerging infectious diseases: a strategy for the 21st century. Atlanta, GA: US Department of Health and Human Services, 1998. Available at http://www.cdc.gov/ncidod/emergplan/index.htm.
2. CDC. Biological and chemical terrorism: strategic plan for preparedness and response: recommendations of the CDC strategic planning workgroup. MMWR 2000;49(No. RR-4):1–14.
3. CDC. Update: investigation of anthrax associated with intentional exposure and interim public health guidelines, October 2001. MMWR 2001;50:889–93.
4. Henning KJ. Syndromic surveillance. In: Smolinski MS, Hamburg MA, Lederberg J (eds.). Microbial threats to health: emergence, detection, and response. Washington, DC: National Academies Press, 2003;309–50.
5. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. J Urban Health 2003;80(2 Suppl 1):i1–7.
6. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis 2002;9:1197–1204.
7. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No.RR-5).
8. Wagner MM, Tsui F-C, Espino JU, et al. The emerging science of very early detection of disease outbreaks. J Public Health Manag Pract 2001;7:50–8.
9. Brinsfield KH, Gunn JE, Barry MA, McKenna V, Syer KS, Sulis C. Using volume-based surveillance for an outbreak early warning system [Abstract]. Acad Emerg Med 2001;8:492.

10. Mostashari F, Karpati A. Towards a theoretical (and practical) framework for prodromic surveillance [Abstract]. International Conference on Emerging Infectious Diseases, Atlanta, GA, March 24–27, 2002.
11. Stern L, Lightfoot D. Automated outbreak detection: a quantitative retrospective analysis. Epidemiol Infect 1999;122:103–10.
12. Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information systems-based surveillance. J Am Med Inform Assoc 2002;9:105–15.
13. Cochrane DH, Allegra JR, Rothman J. Comparison of physician's choice of charting template to ICD-9 codes for biosurveillance using an emergency department electronic medical records database [Abstract]. Acad Emerg Med 2003;10:525.
14. Olson KL, Mandl K. Geocoding patient addresses for biosurveillance [Abstract]. Proc AMIA Symp 2002;1119.
15. Mocny M, Cochrane DG, Allegra JR, et al. A comparison of two methods of biosurveillance of respiratory disease in the emergency department: chief complaint vs. ICD-9 diagnosis code [Abstract]. Acad Emerg Med 2003;10:513.
16. Pavlin JA, Mostashari F, Kortepter MG, et al. Innovative surveillance methods for rapid detection of disease outbreaks and bioterrorism: results of an interagency workshop on health indicator surveillance. Am J Public Health 2003;93:1230–5.
17. Osaka K, Takahashi H, Ohyama T. Testing a symptom-based surveillance system at high-profile gatherings as a preparatory measure of bioterrorism. Epidemiol Infect 2002;129:429–34.
18. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecur Bioterror 2003;1:1–5.
19. US Department of Defense. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE). Silver Spring, MD: US Department of Defense, Global Emerging Infections Surveillance and Response System. Available at http://www.geis.ha.osd.mil/geis/surveillanceactivities/essense/essence.asp.
20. County of Los Angeles, Department of Health Services, Acute Communicable Disease Control. Special studies report 2000: Democratic National Convention—bioterrorism syndromic surveillance. Los Angeles: County of Los Angeles Department of Health Services, 2000. Available at http://www.lapublichealth.org/acd/reports/spclrpts/spcrpt00/demonatconvtn00.pdf.
21. CDC. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. MMWR 2002; 51(Special Issue):13–15.
22. Begier EM, Sockwell D, Branch LM, et al. The National Capitol Region's emergency department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? Emerg Infect Dis 2003;9:393–6.
23. 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;285–289.
24. Lazarus R, Kleinman K, Sashevesky I, et al. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis 2002;8:753–60.
25. Zelicoff A, Brillman J, Forslund DW, et al. The rapid syndrome validation project. Proc AMIA Symp 2001;771–5.
26. Kluger MD, Sofair AN, Heye CJ, Meek JI, Sodhi RK, Hadler JL. Retrospective validation of a surveillance system from unexplained illness and death: New Haven County, Connecticut. Am J Public Health 2001; 91:1214–9.
27. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Atlanta, GA: US Department of Health and Human Services, CDC. Available at http://www.bt.cdc.gov/surveillance/syndromedef/index.asp.
28. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response early aberration reporting system (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96.
29. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. J Urban Health 2003;80(2 Suppl 1):i76–88.
30. Duchin JS. Epidemiological response to syndromic surveillance signals. J Urban Health 2003;80(2 Suppl 1):i115–6.
31. Pavlin JA. Investigation of disease outbreaks detected by syndromic surveillance systems. J Urban Health 2003;80(2 Suppl 1):i107–4.

Linking Better Surveillance to Better Outcomes

Seth L. Foldy
City of Milwaukee Health Department, Milwaukee, Wisconsin, and Medical College of Wisconsin, Milwaukee, Wisconsin

Corresponding author: Seth L. Foldy, 3061 N. Marietta Avenue, Milwaukee, WI 53211. Telephone: 414-906-0036; Fax: 414-906-9667; E-mail: sfoldy@sbcglobal.net.

Abstract

Syndromic surveillance aspires to achieve rapid outbreak detection and response, but stand-alone systems not integrated into local business processes might fail to offer better health outcomes. To describe how surveillance can most directly serve action, the author presents a model of local public health work as a series of outcome-driven business processes consisting of information input, information processing, actions, and outcomes. This report derives lessons for improving each of these elements from public health emergencies occurring in Milwaukee, Wisconsin. Lessons for improving input include 1) creatively mining internal or readily accessible information; 2) integrating information flow into routine business practices before an emergency; 3) reusing information in multiple business processes, and ensuring that information-management systems enable such recycling; 4) fostering relationships with information providers by reducing burdens and meeting their needs; and 5) using agile tools to focus surveillance on pressing problems. Lessons for better processing include 1) combining diverse information in well-organized visual displays ("surveillance dashboards"); 2) creating alerts that warn of unusual patterns; 3) using Internet tools to view and share information on demand; 4) using diverse expertise to interpret complex information; 5) assembling surveillance so as to be scalable (from local to global); and 6) ensuring sufficient environmental, laboratory, and clinical capacity for rapid confirmation and response. Lessons for linking surveillance to more efficient action include 1) building surveillance directly into response plans; 2) feeding surveillance information directly into response systems; and 3) employing those information and communications systems used in daily practice to the greatest extent possible. Using surveillance information systematically in outcome-driven business processes can improve emergency response while building day-to-day organizational effectiveness.

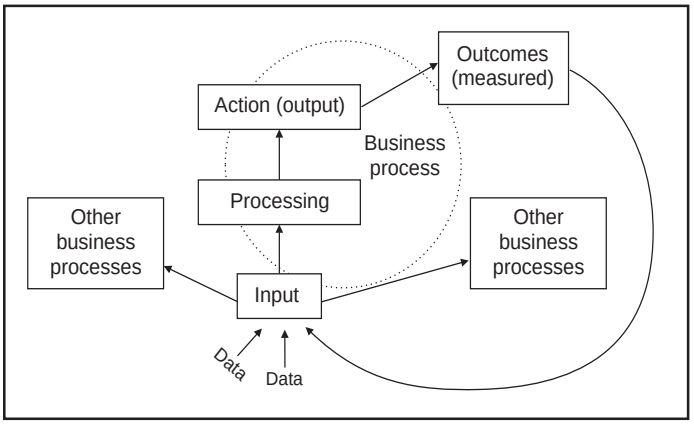
Introduction

Public health surveillance has been defined as "the ongoing collection, analysis, interpretation, and dissemination of data regarding a health-related event for use in public health action to reduce morbidity and mortality and to improve health" (*1*). The primary goal of surveillance is to support action. Because surveillance of established diagnoses might be too slow or insensitive to initiate timely countermeasures, the threats of biologic terrorism and emerging infections (e.g., severe acute respiratory syndrome [SARS]) have spurred interest in syndromic surveillance of near real-time illness indicators (e.g., chief complaints, laboratory test orders, and absenteeism). In addition to its new relevance for homeland security, syndromic surveillance or case management has been used to track influenza, polio, and sexually transmitted diseases for which laboratory confirmation is impractical (*2–4*).

Excellent criteria have been proposed for determining whether syndromic surveillance systems provide reliable, useful information to decision-makers. (*5*). Different considerations are required to determine whether a system facilitates rapid, effective action, whether a system can be sustained, and whether it will be used in an actual emergency. Answers depend on how well surveillance is integrated into the day-to-day work of local public health agencies (LPHAs). Local professionals are best situated to validate a suspected threat (by rapid assessment of local health-care, environmental, and laboratory information); define the evolving direction of the threat and who is at risk (by interpreting local information on place, time, occupation, and environmental conditions); notify and mobilize the most immediately affected parties; and offer timely, locally relevant risk communications. State and federal resources can help but cannot supplant local knowledge and relationships.

LPHAs are typically small but complex organizations working simultaneously on multiple desired community outcomes (e.g., improvements in infant nutrition, food safety, tobacco use, elder quality-of-life, or communicable disease). Work toward each outcome can be viewed as a series of business processes (Figure 1) in which information *input* (e.g., a referral, an inspection, a survey, a client assessment, or a disease report) is *processed* to reach an action decision. *Actions* (e.g., issuing a WIC coupon, a sanitation order, a citation for tobacco sales to minors, or an isolation order; conducting a home visit; or writing a prescription) aim to improve a population outcome. A community that tracks outcomes (e.g., teenage smoking rates) quantitatively also uses this information as

FIGURE 1. The work of public health represented as a series of business processes in which information inputs* are translated to actions toward a desired set of outcomes†



* One piece of information can serve as input for multiple business processes (e.g., a report of a death might prompt a communicable disease investigation, a death certificate, and collection of a death certificate processing fee).

† The business-process model of local public health work was developed by Stephen Downs, Seth Foldy, Peter Kitch, Patrick O'Carroll, and David Ross at a meeting on information modeling for public health practice sponsored by the Robert Wood Johnson Foundation, Denver, Colorado, October 13, 2004.

input, creating a feedback loop to adjust the type, quality, or quantity of actions.

An efficient organization will apply one piece of information to multiple business processes. For example, a patient address received in a disease report can be used to dispatch an investigator, locate household contacts, issue an isolation order, and map an outbreak.

This idealized, informatic view of public health emphasizes the importance of considering how information is most effectively converted into action. Too often, information *collection* is emphasized over information *use*. Poorly processed information produces information glut and unread reports. Particularly when all staff are responding to an emergency, surveillance information must feed multiple action processes simultaneously (e.g., case finding, specimen collection, laboratory reporting, outbreak characterization [person-place-time patterns], isolation and quarantine, environmental surety, and risk communications).

Various public health emergencies helped the Milwaukee (Wisconsin) Health Department (MHD) learn to integrate surveillance into well-organized business processes serving both emergency and everyday functions. This report describes these experiences and summarizes lessons that can help improve syndromic surveillance systems.

Linking Surveillance to Outcomes — Local Experience

In 1993, approximately 400,000 Milwaukee-area residents were sickened by a waterborne outbreak of cryptosporidiosis, a then-emerging disease for which reporting was not mandated and testing rarely performed. Although drinking water turbidity levels increased 10 days before reported onset of symptoms, the outbreak was recognized only after shortages of diarrhea medications and enteric culture media were reported (*6*). At the time, different agencies held information (e.g., on water turbidity, customer complaints, and employee or student absenteeism) that, viewed together, might have alerted authorities to the outbreak earlier.

After the outbreak, MHD initiated surveillance of water quality, pharmacy sales, and diarrhea in nursing facilities. Fourteen LPHAs in Milwaukee County established a single disease-reporting site (SurvNet) to simplify reporting, improve outbreak recognition, and increase communication and feedback regarding public health trends to clinicians and laboratories (data reporters). An interagency task force was formed to monitor and improve water quality and to compile and interpret all available information when concerns arose. MHD upgraded clinical and environmental microbiology capabilities and established fax-broadcast and Internet-based communication with laboratories, physicians, infection-control practitioners, and emergency departments (EDs). Debriefings held after each outbreak identified needed changes in policy or procedures. MHD adopted a community outcome goal of 20 reportable enteric infections/100,000 residents (adapted from *Healthy People 2010* goal 10-1) (*7*).

These improvements helped speed effective response in 2000 when a nurse contacted SurvNet about four children from three health jurisdictions who had suspected but unconfirmed *Escherichia coli* O157:H7 infection. Patient interviews identified a restaurant, which was rapidly inspected and closed. Broadcasts to clinicians and laboratories provided diagnostic, treatment, and prevention advice and resulted in rapid identification of additional cases. Evidence from rapidly performed epidemiologic, environmental, and laboratory investigations demonstrated conclusively that processing of contaminated whole-beef cuts could cause sustained disease transmission in restaurants, which helped change national food policy (*8*). In this instance, one telephone call arising from clinical suspicion triggered rapid action and comprehensive investigation and contributed to health-policy change. Each such success increases interest and confidence in public health surveillance among clinicians and other reporters of public health information.

The health effects of a 1995 Milwaukee heat wave were recognized belatedly after the medical examiner was overwhelmed by investigations of heat-related deaths (*9*). An extreme-heat plan was created, under which action is triggered by an environmental signal (weather forecast) and further accelerated if heat illness is observed by emergency medical services (EMS) or the medical examiner (*10*). MHD uses communications tools developed for outbreaks to alert multiple human service agencies to take planned action to protect those at greatest risk (*9,11*). The plan is updated annually and available continuously online. Heat-adjusted morbidity and mortality were reduced by 50% during a 1999 heat wave, compared with 1995 (*12*).

In 1999, local hospitals established EMSystem,® a regional emergency medicine Internet (REMI) application that enables EDs to communicate when they must divert ambulances. When too many EDs simultaneously signal diversion, the paramedic system overrides diversion, generating an e-mail/text-page alert. In January 2000, REMI data were used to track influenza-related ED congestion, and a Health Care Capacity Alert Committee was formed (including public health, EMS, medical, and hospital representatives) to issue recommendations to ease ED crowding (*13*). In fall 2000, an unusual volume of diversion-override text pages alerted MHD to severe ED congestion, months before influenza season. Review of REMI data indicated that congestion was primarily attributable to inpatient bed shortages. Committee recommendations to adjust vacation leave, facilitate timely discharge, and control elective admissions were followed in 2 days by a rapid decline in ED diversions.

REMI data were later used to justify a regulatory waiver permitting medical/surgical use of rehabilitation and psychiatric beds during the 2000–01 influenza season. REMI provided unexpected but useful surveillance information on health-care utilization and capacity that, linked to action, helped build stronger relationships between public health professionals and health-care providers.

MHD adapted REMI in 2000 for heat-illness surveillance during heat waves and in 2002 for short-term surveillance of biologic terrorism syndromes during international-profile sporting events (*14*). This helped MHD establish multi-ED surveillance for SARS 3 days after CDC urgently requested surveillance in 2003. After successful deployment in Milwaukee, the SARS screening form was downloaded for use by hundreds of clinicians. Because the REMI application was then used in >25 cities, SARS surveillance was offered to other jurisdictions; 27 EDs reported surveillance of >146,500 visits to LPHAs in four states, and CDC staff were able to download these data for aberration analysis. REMI permitted agile deployment of a new syndromic surveillance system across widely distributed jurisdictions (*15*).

In summer 2003, SurvNet received a report of a febrile blister illness in an animal dealer associated with sick prairie dogs. Wisconsin authorities linked this report to a similar case elsewhere in the state, triggering immediate trace-forward and trace-back investigation of animal sales. The illness was subsequently diagnosed as the hemisphere's first outbreak of monkeypox. Action to protect the public began before diagnosis. However, lack of interoperable data systems impeded information-sharing among the many health and veterinary agencies involved across multiple states.

Lessons Learned — Linking Better Surveillance to Better Action and Outcomes

These experiences indicated that 1) more syndromic information (input) is available than typically used, 2) information processing can improve the timeliness and quality of decision-making, and 3) action can be accelerated by good information-management practices. Recommendations follow for better integrating surveillance information into each of these business process steps (input, processing, and action).

Improving the Input

LPHAs can easily increase the type, quality, and sustainability of surveillance by 1) mining information from daily business processes found within or near the organization; 2) integrating information flow into routine business practices before an emergency; 3) reusing information in multiple business processes, and ensuring that information-management systems enable such recycling; 4) fostering relationships with information providers by reducing burdens and meeting their needs; and 5) using agile tools to focus surveillance on pressing problems.

Within local agencies, diverse information streams on symptoms, environmental conditions (e.g., heat, water quality, and animal illness), health-care utilization (e.g., prescriptions, laboratory orders, ambulance diversion, and 911 dispatch calls), and behaviors (e.g., absenteeism or travel) are often readily available. Internal information sources might be as useful as more elaborate data gathering (e.g., MHD uses routine food-safety inspections to track the number of restaurants permitting smoking). Other local entities (e.g., the water utility or fire/EMS) also possess important, readily available information. Finally, the Milwaukee examples illustrate how environmental, health-care utilization, and other types of data can

provide earlier warning or more robust validation of problems than clinical signs and symptoms alone. For surveillance of waterborne *Cryptosporidium*, heat-related illness, and monkeypox, environmental information provided longer alert lead-times than clinical findings.

Although syndromic surveillance is often inspired by emergencies, an emergency is not the best time to begin work with unfamiliar information. Without daily practice, systems can fall into disuse and might complicate emergency response as much as facilitate it. Ideally, surveillance systems are both derived from and support daily local public health operations, thereby strengthening relationships and communications, which become even more critical during emergencies.

The health agency that uses every datum for multiple purposes can improve alertness and effectiveness at minimal cost. Ideally, information is "entered once, used often," instead of being locked inside applications and unavailable for reuse. Various Internet applications collect information but do not permit local analysis of entered data. Internet-served applications also often fail to permit uploading of information from a local agency's own information-management system. This results in duplicate entry, poorer data quality, and difficulty using or reusing information efficiently or creatively. These considerations support the argument for full and rapid implementation of CDC's Public Health Information Network (PHIN) (*16*) vision of interoperable applications that truly exchange rather than hoard health information.

The quality and quantity of surveillance information relies on the willingness of busy people to provide it. One way to improve surveillance is to make it less burdensome. Combining disease reporting for 14 jurisdictions in SurvNet made reporting easier, while also increasing the surveillance catchment area and exploiting economies of scale for more sophisticated data management. Calling one reporting site often instead of 14 infrequently helped infection-control professionals build relationships with SurvNet staff; such relationships can increase willingness to share observations of uncertain significance that enhance recognition of unusual outbreaks (e.g., monkeypox). However, such relationships are less likely within an office covering 300 jurisdictions; therefore, appropriate local scale remains important. Another way to minimize reporting burden is to use those communications tools already used by health professionals in their own day-to-day work (e.g., REMI) rather than expect busy professionals to log onto stand-alone public health utilities (e.g., certain health alert networks).

Eliminating altogether the need for conscious effort in reporting is the goal of such surveillance initiatives as electronic laboratory reporting and secondary mining of health-information-management systems. However, engaging health providers in well-designed surveillance activities has other benefits. The SARS screening form was designed to trigger infection-control protection as well as to alert public health. Its use was also reported by ED managers to improve clinicians' index of suspicion.

Providers are most likely to comply with surveillance when it aids them in activities on which they place high value, such as improving diagnosis. MHD attempts to issue timely situational alerts to cue clinicians to problems they might see in their practices (e.g., heat-related illness during a heat wave, biologic agents such as anthrax after September 11, 2001, or *E. coli* infection during an outbreak). Such alerts help focus surveillance while also helping clinicians appreciate that surveillance provides, as well as demands, useful information. Providing timely information that helps providers defend themselves from infection (e.g., SARS), send the right test, or offer special resources for affected patients also helps improve awareness of the benefits of surveillance. Finally, providers enjoy learning how surveillance contributes to healthy public policies, not just to tables and graphs.

In a rapidly changing world, surveillance should be flexible enough to focus on the most immediate threat, based on warnings as diverse as weather forecasts, law-enforcement or international intelligence, global disease trends, or nearby outbreaks. This flexibility requires agile tools for surveillance. Agility is especially important for unexpected emerging diseases (e.g., SARS or monkeypox). Milwaukee EDs have become accustomed to implementing temporary surveillance by using REMI; the threat of the day might change, but the system used remains familiar. Networks of providers or laboratories already engaged in one surveillance system (e.g., for influenza-like illness) might also be amenable to participating in emergency surveillance for other agents, providing another source of agile surveillance.

Improving Processing

Surveillance information must be processed in a timely, meaningful way for providers to be guided by knowledge instead of overwhelmed with data. Effective processing is aided by 1) combining diverse information in well-organized visual displays ("surveillance dashboards"); 2) creating alerts that warn of unusual patterns; 3) using secure Internet sites to view and share information on demand; 4) using diverse expertise to interpret complex information; 5) assembling surveillance so as to be scalable (from local to global); and 6) ensuring sufficient environmental, laboratory, and clinical capacity for confirmation and response.

Each different surveillance information stream provides only a fragmentary view of a complex world. The cryptosporidiosis

outbreak illustrated how assembling and comparing different types of existing information might be more important than collecting new information. Fragmented views occur not only between organizations but just as often within a single agency. Until recently, different MHD units produced or received statistics on pneumonia and influenza deaths, influenza-like illness, and influenza laboratory cultures, but the agency had no single coherent view of respiratory illness. Creating a single visual display of all three types of data on a common time axis and mounting it on the Internet enabled MHD to transform little-used data into rich knowledge for multiple users, accessible on demand, day or night (Figure 2).

Sharing different expertise might be as important as sharing different information. Milwaukee's Health Care Capacity Alert Committee and Water-Health Taskforce are multidisciplinary, multiagency groups that interpret and act on complex information. The Taskforce meets monthly for other tasks, which keeps it functioning smoothly, and convenes in response to unusual information or situations.

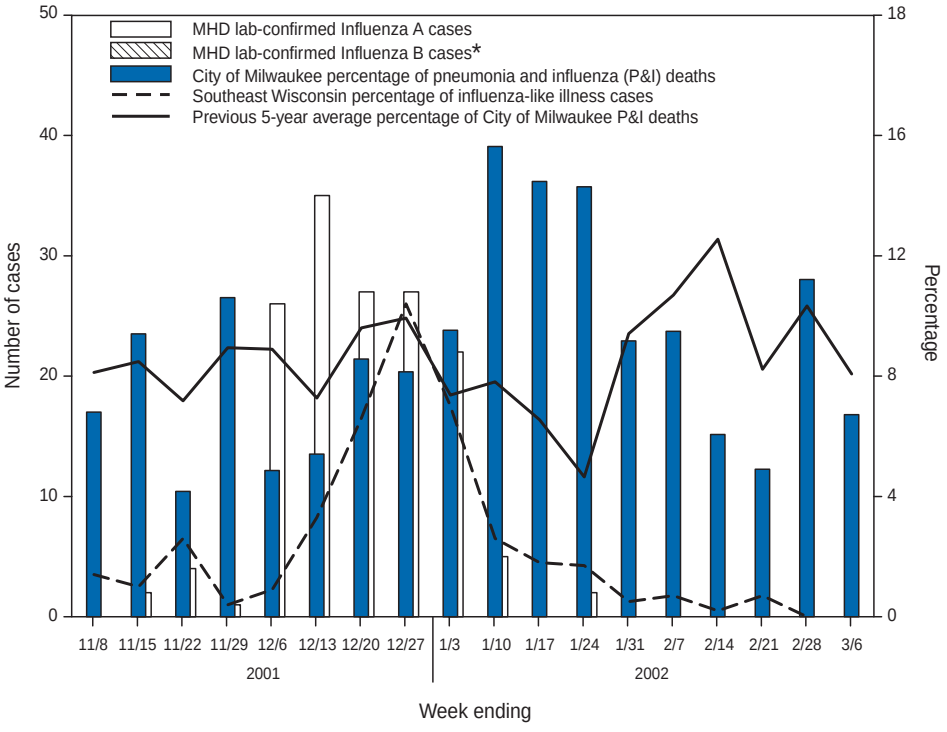
Scale is important, but optimum scale varies from one situation to another. Milwaukee's SurvNet one-stop reporting system speeds detection of and response to outbreaks that cross local jurisdictional boundaries, but the system might not recognize rare events if implemented statewide. The capacity to build scalable surveillance across regions and states is enhanced by the growth of managed care networks, multiregional REMI systems, multistate surveillance systems (e.g., FoodNet), and the interoperable information environment promised by PHIN. Combined with automated tools (e.g., SaTScan™ (*17*) analysis) to test the significance of events over variable geographic and temporal scales, potential flexibility in the scale of surveillance might approach infinity. However, more often than not, local insight is needed to interpret local surveillance information intelligently, which is why national and international surveillance systems will only be as strong as their local building blocks. Confirmation (and control) of suspected events relies heavily on well-prepared clinical, laboratory, and environmental expertise. Unless these local capabilities are in place and integrated for rapid response, even the best and earliest surveillance alert will fail to generate timely effective action.

Faster, Surer Action

Better information inputs and processing matter only when they lead to effective action. Effectiveness can be improved by 1) building surveillance directly into response plans; 2) feeding surveillance information directly into emergency response systems; and 3) employing information and communication systems used for everyday practice to the greatest extent possible.

Considerable time and effort can be saved when enhanced surveillance systems are specifically referenced in emergency plans. For example, certain types of health, environmental, or intelligence data automatically trigger higher stages of readiness in Milwaukee emergency response plans. Ideally, information from surveillance systems can directly feed information systems used for emergency response. For example, if a cluster of persons with febrile vesicular rash is detected, the next steps (investigation, laboratory diagnosis, isolation, and contact tracing) each require similar information, including names, addresses, clinical information, employers, travel, and contacts. Downloading such data from surveillance systems directly into the line lists used for outbreak investigation would reduce work and improve data quality in a rapidly evolving emergency.

FIGURE 2. Example of a surveillance "dashboard" that combines different types of influenza-related data to enhance side-by-side analysis — Milwaukee Health Department (MHD), 2001–2002



* No Influenza B cases were reported for this period.

Emergencies are not optimal times to learn how to use unfamiliar information systems. To the extent possible, surveillance and communications with community partners should employ the same systems they use everyday, as close to the point-of-service as possible. This again emphasizes the need for information exchange between the systems used routinely in clinical and public health settings, rather than forcing users to switch to new systems.

Conclusion

Public health's primary role goes beyond preparing for intermittent emergencies to reducing the leading causes of death, illness, and injury. If increased public health funding for homeland security is short-lived, resulting surveillance systems will be most sustainable if they also address long-term, common problems as well as extraordinary ones. Health departments that set quantifiable community outcome goals (e.g., to reduce enteric disease or smoking rates) place surveillance at the core of all work, not just communicable disease control. Syndromic and other surveillance systems that become an integral part of day-to-day business processes become indispensable. They don't just detect problems but also measure successes and identify what works. This doubles the value and sustainability of any surveillance system.

Acknowledgments

A partial list of those contributing to cited Milwaukee-area innovations includes Kathleen Blair, Paul Biedrzycki, Geoffrey Swain, CapriMara Fillmore, Douglas Gieryn, Susan Schultz, Sandy Coffaro, Marion Lee, Donna Rowe, Mary Ellen Bruesch, Alki Burdett, Jeffrey Davis, Edward Barthell, Kim Pemble, Nancy Healy-Haney, and the health officers of suburban Milwaukee County.

References

1. CDC. Updated guidelines for evaluating public health surveillance systems: recommendations from the guidelines working group. MMWR 2001;50(No. RR-13):2.
2. Brammer TL, Murray EL, Fukuda K, Hall HE, Klimov A, Cox NJ. Surveillance for influenza—United States, 1997–98, 1998–99, and 1999–00 seasons. MMWR 2002;51(SS-7):1–10.
3. World Health Organization Expanded Programme on Immunization. Field guide for supplementary activities aimed at achieving polio eradication. Geneva, Switzerland: World Health Organization, 1995.
4. Pettifor A, Walsh J, Wilkins V, Raghunathan P. How effective is syndromic management of STDs? A review of current studies. Sex Transm Dis 2000;27:371–85.
5. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5):1–11.
6. 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.
7. US Department of Health and Human Services. Tracking Healthy People 2010. Washington, DC: US Government Printing Office, 2000.
8. School of Public Health, University of North Carolina at Chapel Hill, CDC. Food safety: a challenge for everyone in public health. Videoconference available at http://www.publichealthgrandrounds.unc.edu/foodsafety/index.htm.
9. CDC. Heat-wave-related mortality—Milwaukee, Wisconsin, July 1995. MMWR 1996;45:505–7.
10. Milwaukee Health Department and Milwaukee Heat Task Force. Plan for excessive heat conditions 2004. Available at http://www.milwaukee.gov/citygov/health/heat.
11. Rajpal RC, Weisskopf MG, Rumm PD, et al. Wisconsin, July 1999 heat wave: an epidemiologic assessment. WMJ 2000;99:41–4.
12. Weisskopf MG, Anderson HA, Foldy SL, et al. Heat wave morbidity and mortality, Milwaukee, Wisconsin, 1999 versus 1995: an improved response. Am J Public Health 2002;92:830–3.
13. Barthell EN, Foldy SL, Pemble KR, et al. Assuring community emergency capacity with collaborative internet tools: the Milwaukee experience. J Public Health Manag Pract 2003;9:35–42.
14. Foldy SL, Biedryzcki P, Barthell E, et al. The Milwaukee biosurveillance project: real time syndromic surveillance using secure regional internet [Abstract]. J Urban Health 2003;80(2 Suppl 1):126.
15. Foldy SL, Barthell E, Silva J, et al. SARS Surveillance Project: Internet-enabled multiregional surveillance for rapidly emerging disease. MMWR 2004;53(Suppl):215–20.
16. CDC. Public Health Information Network. Atlanta, GA: US Department of Health and Human Services, CDC, 2004. Available at http://www.cdc.gov/phin.
17. Kulldorff M, Mostashari F, Hartman J, Heffernan R. Syndromic surveillance without denominator data: the space-time permutation scan statistic. Presentation, 2003 National Syndromic Surveillance Conference, New York, NY, October 2003.

Review of the 2003 National Syndromic Surveillance Conference — Lessons Learned and Questions To Be Answered

James W. Buehler
Emory University, Atlanta, Georgia

Corresponding author: James W. Buehler, Rollins School of Public Health, Rm. 416, Emory University, 1518 Clifton Rd., NE, Atlanta, Georgia 30322. Telephone: 404-712-8345; E-mail: jbuehle@sph.emory.edu.

Abstract

Syndromic surveillance is a rapidly evolving field within public health practice. Substantial experience has been gained in learning how to conduct syndromic surveillance, informed by a growing body of research and practice, including refinement of surveillance methods, development of new tools for analysis and evaluation, findings from statistical models and applied evaluations, and expansion of syndromic surveillance to uses beyond preparedness for biologic terrorism. Despite these advances, additional evaluation is needed to help health departments determine whether to conduct syndromic surveillance. This paper summarizes the lessons learned from the 2003 National Conference on Syndromic Surveillance, which provided a foundation for defining a research and evaluation agenda and for developing preliminary guidance for public health agencies planning to implement syndromic surveillance.

Introduction

Participants in the 2003 National Syndromic Surveillance Conference were junior- and senior-level professionals from multiple disciplines, including epidemiology, statistics, informatics, health care, and public health practice. Conference presentations outlined the substantial progress that has been made in understanding how to conduct syndromic surveillance. Methods are being refined, and additional health departments are gaining experience with syndromic surveillance. However, additional evaluation is needed before guidelines can be developed to help other health departments decide whether to conduct syndromic surveillance. This paper follows the outline used by the summary of the 2002 conference (*1*) to summarize the lessons learned at the 2003 conference and make recommendations for the future.

What Is Syndromic Surveillance?

The term *syndromic surveillance* describes the growing array of surveillance methods aimed at early detection of epidemics related to biologic terrorism. Although syndromic surveillance originated before 2001, the field grew substantially after the terrorist attacks of 2001 generated fears of future attacks. The word *syndromic* has been applied because the majority of such systems monitor different *syndromes* that might herald the early stages of epidemics (*2*). Other syndromic surveillance systems monitor health indicators of different actions persons might take or consequences they might suffer (e.g., miss work, use outpatient services, purchase medications, or require ambulance transport for emergency care) from the early stages of illness until death. Although certain syndromic surveillance systems depend on manual data collection, the 2003 conference emphasized systems that use automated methods to harvest data stored electronically and then transmit and analyze these data. The majority of presenters described ongoing surveillance, not systems designed to operate only during specific high-profile events.

The 2003 conference focused on describing the utility of syndromic surveillance, which remains, primarily, the early detection of an epidemic caused by deliberate release of a biologic agent. Syndromic surveillance also enables public health officials to provide reassurance that terrorism-related or other epidemics are not occurring, to detect the onset of expected seasonal upswings in viral respiratory and gastrointestinal infections, to detect common epidemics, and to conduct surveillance for a growing spectrum of health-related events.

Data Sources

Multiple data sources are being used for syndromic surveillance, limited only by the imagination of investigators. These sources can be classified into two broad categories: 1) clinical data arising from the use of health-care services (e.g., emergency department visits, clinic visits, or ambulance trip logs), and 2) all other indicators (e.g., pharmacy sales, calls to emergency numbers or information hotlines, and work or school absentee rates). Multiple health departments use a combination of data sources that complement one another.

The benefits of clinical data are twofold. First, productive relationships can arise between public health staff and clini-

cians as they establish and conduct syndromic surveillance. Second, the majority of clinical data sources enable investigators to follow up with individual patients when surveillance detects an unusual trend. Nonclinical data can complement clinical information by providing indicators of events (e.g., purchase of over-the-counter medications) that might occur before persons seek health care, by describing groups not represented at selected clinical facilities, or by validating trends observed in clinical data. One disadvantage of nonclinical data sources is that they typically do not readily allow for follow-up with affected persons.

Analytic Methods

Although various analytic methods are being used, two utilities are emerging as the statistical workhorses of syndromic surveillance: CDC's Early Aberration Reporting System (EARS), which detects unusual trends by time (*3*), and SaTScan™ (*4*), a program originally developed for detection of cancer clusters that identifies clustering by time and geographic location. As described elsewhere in these proceedings, substantial work is under way to develop new statistical methods for aberration detection and to refine syndrome categories.

Evaluation of Syndromic Surveillance Systems

After the 2002 conference, at which draft guidelines for evaluating syndromic surveillance systems were introduced, CDC engaged a panel to assist in revising these guidelines. A revised draft was distributed to participants at the 2003 conference, and the final version was published in *Morbidity and Mortality Weekly Report* (*5*). The guidelines rely on established CDC recommendations for evaluating surveillance systems but emphasize detection of epidemics rather than cases of illness. Presenters at the conference used the guidelines to describe surveillance systems and assess the balance between predictive value (i.e., the likelihood that a statistical alert represents a problem of public health importance) and sensitivity and timeliness (i.e., the likelihood that all epidemics are detected at the earliest possible stages).

Investigation of Signals

After being established, a syndromic surveillance system will inevitably generate alerts, indicating that a monitored indicator has surpassed a statistical threshold. When this happens, someone (typically an epidemiologist working in a local public health department) must decide whether, or to what extent, an investigation is warranted. Multiple conference presenters described their experiences with responding to signals, illustrating both the science and art of syndromic surveillance. Practitioners are developing graduated approaches to follow-up, ranging from closer examination of surveillance data to aggressive field investigation. They also report developing a sense of when signals merit more or less aggressive reactions. Certain practitioners wait to see whether aberrant trends persist for >1 day; others wait until more than one data source yields a signal before responding more aggressively. These varying approaches highlight the hard-to-quantify local rules that are evolving to maximize predictive value while minimizing losses in timeliness or sensitivity. What is known is that statistical alerts are common, certain alerts represent true public health emergencies, and substantial work is needed to characterize and quantify the relation between the presence or absence of an alert and the presence or absence of an outbreak.

Protecting Confidentiality

Protecting confidentiality while maximizing the usefulness of surveillance raises concerns regarding public health law, surveillance procedures, and relationships with the public. In the arena of public health law, one of the most important events of 2003 was the implementation of the Privacy Rule of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). HIPAA governs the ways that health-care providers can share patient information but provides specific exemptions that allow for reporting of confidential health information to public health agencies for surveillance and other authorized disease prevention and control purposes (*6*). Clinicians, health-care managers, public health officials, and their attorneys are struggling to achieve an understanding of HIPAA, including how the provisions for reporting to public health agencies apply to syndromic surveillance. The distinction between syndromic surveillance, which is a public health practice and thus exempt from certain HIPAA privacy provisions, and research, which is governed differently under HIPAA, has emerged as a key concern. Presenters at the 2003 conference described certain successes in conducting surveillance in the HIPAA era but also reported difficulties.

Virtually all syndromic surveillance systems shun the collection of names or other identifiable information to ensure that privacy and confidentiality are not violated in the event of a security lapse. Systems also use multiple methods to encrypt data and ensure secure transmission and storage. Certain clinical systems assign numbers to patient-level surveillance records and provide only those numbers in reports to health departments so that identifying information is retained by the individual health-care facility. For those systems, any

follow-up investigation is conducted through and with the assent of reporting site staff, who control access to identifying information. Other systems limit the detail reported to decrease the likelihood that patients will be identified inappropriately. Together, such measures reflect adherence to two principles in public health surveillance: collect information judiciously, and collect and retain identifying information as locally as possible. When describing syndromic surveillance systems based on automatic medical record systems, conference presenters referred to this practice as "the distributive data model" (*7*) because access to data is distributed in a manner commensurate with the respective roles of care providers and public health staff. The result is that epidemiologists have information needed to monitor community-level trends in selected syndromes. If surveillance indicates that further investigation is warranted, including review of individual patient records, then access can be requested from the healthcare providers.

Long-term public support for syndromic surveillance will depend on both the public's perception that public health agencies are responsible stewards of any information with which they are entrusted, and on the perception that syndromic surveillance serves a useful public good. Thus, public health agencies must be diligent in communicating with the public about the utility of syndromic surveillance and about their strategies for protecting health information.

National and Local Data

Health departments seeking to establish syndromic surveillance can either develop data sources locally or tap national systems that provide local information. The question is no longer one of selecting one source versus another but of determining the right mix of local and national sources (e.g., the systems offered by the Real-Time Outbreak Disease Surveillance System group at the University of Pittsburgh [*8*] or the resources being developed by CDC under its BioSense program [*9*]). A critical question concerning these national sources is whether they will allow for rapid local follow-up with facilities or patients when they yield an aberrant signal that merits investigation.

Who Owns Syndromic Surveillance?

The question of who "owns" syndromic surveillance was raised at the 2002 conference because the leadership roles of different governmental, academic, and private participants were unclear (*1*). As demonstrated by presenters at the 2003 conference, innovative projects are being conducted or supported by multiple entities, including local, state, and national agencies; the U.S. Department of Defense; the U.S. Department of Homeland Security; and the U.S. Department of Health and Human Services. National coordination is increasingly being provided by CDC, as evident from its role in coordinating the development of evaluation guidelines and syndrome definitions, implementing BioSense, supporting national pilot projects, and providing state funding for surveillance under its terrorism-preparedness program.

Multiple Uses for Syndromic Surveillance

Compared with the 2002 conference, the 2003 meeting included considerably less discussion of whether syndromic surveillance, traditional surveillance, or astute clinicians would most likely be the first to detect an epidemic. Instead, the emphasis was on interactions among different epidemic-detection strategies, including how syndromic surveillance can alert clinicians to community trends and improve their diagnostic assessments (*10*). Syndromic surveillance and the usefulness of the resulting information can foster better relations among health departments, clinicians, and laboratorians, thereby enhancing the reporting of notifiable diseases or suspected clusters.

Another difference during the 2003 conference was that greater attention was given to nonterrorism-related applications of syndromic surveillance, for multiple reasons. In 2002, the events of 2001 were much fresher in our minds. Since 2001, the United States has not suffered another domestic terrorist attack, and the public's fears about domestic terrorism as the nation prepared for war in Iraq have not been realized. When the Federal government directed resources toward terrorism preparedness, public health officials recognized immediately that, to justify their expense, these efforts must extend beyond surveillance of terrorism-related syndromes. Furthermore, every naturally occurring outbreak is a limited rehearsal for responding to a terrorist attack. The emergence of severe acute respiratory syndrome (SARS) in 2003 demonstrated the nation's vulnerability to new infectious diseases and their potential for epidemic spread. Presenters at the 2003 conference discussed the feasibility of adapting syndromic surveillance for SARS detection, particularly emergency-department–based systems (*11*).

Finally, those who conduct syndromic surveillance are exploring other innovative uses of this new tool. For example, New York City used its pharmacy system to assess the impact of smoking cessation interventions by tracking sales of nicotine patches (*12*), and the U.S. Department of Defense examined the mental health effects of the terrorist attack on the Pentagon (*13*).

Importance of Partnerships

Those in the vanguard of this field represent successful partnerships between public health practitioners and academics. Syndromic surveillance is more complex than traditional surveillance and benefits from expertise in informatics, statistics, and advanced epidemiologic methods — skills that health departments might not be able to maintain as a result of budget and mission constraints but that are readily available in universities. In turn, public health departments bring a familiarity with community resources, their relations with healthcare providers, and their expertise in conducting surveillance and applying it to meet public health objectives.

Relation Between Surveillance and Disease Epidemiology

One theme that was less prominent at the 2003 conference was the epidemiology of potential agents of biologic terrorism. Usually, the conduct of surveillance is shaped by the epidemiology of the condition under surveillance, including how it is diagnosed, treated, or prevented. Relevant questions regarding the detection of terrorist-related epidemics include the following:

- What is the likely shape of an epidemic curve?
- How rapidly will different stages of illness occur?
- How will the spectrum of illness become manifest with respect to different surveillance indicators?
- How will these patterns vary among the potential agents of biologic terrorism?

In the absence of terrorist attacks, the answers will likely come from epidemiologic models that simulate a range of hypothetical scenarios and that test the usefulness of data sources and aberration-detection methods. Critical groundwork for conducting such investigations was described at this meeting (*14*).

Next Steps

Evaluation

The syndromic surveillance evaluation criteria developed by CDC (*5*) should be used in multiple ways. First, the criteria should be used to describe the field's rapidly growing experience in conducting syndromic surveillance. For example, how frequently do different syndromic surveillance methods generate statistical alerts, and what is learned when alerts are investigated? Conversely, how frequently are epidemics detected through other means also identified by syndromic surveillance? How does timeliness of detection compare with timeliness of other detection methods? CDC might request grantees conducting syndromic surveillance to add this information to required periodic reports. Aggregating, summarizing, and disseminating such reports will allow for a more comprehensive assessment of the usefulness of syndromic surveillance. Second, more in-depth evaluations of syndromic surveillance should be conducted in partnership with those states or localities that have the capacity to conduct such evaluations. Third, historic data should be used to test the utility of different detection algorithms; the work presented by the Defense Advanced Research Projects Agency and its collaborators illustrates the benefits of this approach (*15*). Fourth, epidemiologic models should be constructed to test the timeliness, sensitivity, and predictive value of detection strategies under different hypothetical scenarios; progress is being made in model development (*14*).

Research and Evaluation Funding

During the 2003 conference, representatives from three federal agencies — CDC, the Agency for Healthcare Quality and Research, and the U.S. Department of Homeland Security — described the research and evaluation activities they have funded or plan to fund. These funding agencies should take guidance from this conference to define a research and evaluation agenda for syndromic surveillance and, if necessary, update their funding priorities and clarify their roles accordingly. This would help applicants by clarifying practice and evaluation objectives and increase the likelihood that investigations funded by different agencies complement one another. Federal agencies should promote government and academic partnerships by making evidence of such collaboration part of funding criteria. One strategy might be to create centers of excellence in syndromic surveillance that would focus on methods development and evaluation and provide technical assistance to health departments.

Guidelines

Despite the advances highlighted during this conference, considerable questions remain to be answered, particularly for those agencies that have not yet initiated syndromic surveillance:

- Where should syndromic surveillance be conducted? Should all states conduct a form of syndromic surveillance?
- Within a state, should syndromic surveillance be conducted in only the largest cities or in medium-sized cities and rural areas as well?
- If syndromic surveillance is conducted, what are the minimum standards for the selection or number of data sources?
- What are the recommended methods for data analysis?

These questions are difficult to answer because experience and evaluation thus far are insufficient and because quantifying the risk of a terrorist attack for a given locality is impossible. As the field gains experience with syndromic surveillance, such decisions might ultimately be based on the usefulness of syndromic surveillance in detecting outbreaks not related to terrorism, with potential detection of terrorist-related events becoming a secondary use.

In the meantime, health department officials should feel assured that a decision not to conduct syndromic surveillance is justifiable. For those who have decided to implement syndromic surveillance, expecting definitive answers to the preceding questions is premature, but preliminary guidance can be developed. Because of its increasing role in coordinating syndromic surveillance and its history of leadership in public health surveillance, CDC is the logical agency to take the lead in developing such guidance, which should include articulation of the following:

- planning steps, including whom to involve;
- advantages and disadvantages of different data sources and commonly used or readily available statistical tools;
- strategies for responding to alerts;
- what utility to expect, and what is unknown; and
- a plan to document experience and evaluate performance.

Partnerships with Community Representatives

The 2003 conference revealed a mix of partnerships involving public health professionals, clinicians, health-care administrators, emergency responders, legal experts, law enforcement, and companies that provide data and other surveillance resources. Thus far, however, the perspective of community representatives has not been prominent in deliberations about syndromic surveillance. For the majority of health problems, risk is not distributed proportionately among diverse populations. Biologic terrorism might not be an equal opportunity threat; the consequences of a terrorist attack are likely to affect most severely those populations that have long suffered the adverse consequences of health disparities. Involving community advocates is not always easy for public health professionals because advocates sometimes ask questions that are difficult to answer. However, they often have good questions, and their perspectives help ensure that surveillance meets community needs.

Conclusion

The field of syndromic surveillance has advanced considerably. An urgent need remains for continued evaluation of syndromic surveillance to define its utility and develop recommendations for its practice. Evaluation criteria developed by CDC should be used to the extent possible to guide assessments of syndromic surveillance based on both experience and hypothetical scenarios. The 2003 conference provided a basis for defining a comprehensive research and evaluation agenda. Although developing definitive guidelines on syndromic surveillance is premature, experience to date should enable the development of preliminary guidance to help those interested in stepping into this arena.

References

1. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. J Urban Health 2003;80(2 Suppl 1):i1–7.
2. Henning K. What is syndromic surveillance? MMWR 2004;53 (Suppl):7–11.
3. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response early aberration reporting system (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96.
4. Kulldorff M and Information Management Services, Inc. SaTScan™ verion 4.0: software for the spatial and space-time scan statistics, 2004. Available at http://www.satscan.org.
5. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5).
6. CDC. HIPAA Privacy Rule and public health: guidance from CDC and the US Department of Health and Human Services. MMWR 2003;52(Suppl):1–20.
7. Yih WK, Caldwell B, Harmon R, et al. National Bioterrorism Syndromic Surveillance Demonstration Program. MMWR 2004;53(Suppl):43–6.
8. Wagner MM, Espino J, Tsui F-C, et al. Syndrome and outbreak detection using chief-complaint data: experience of the Real-Time Outbreak and Disease Surveillance Project. MMWR 2004;53(Suppl):28–31.
9. Loonsk J. BioSense—a national initiative for early detection and quantification of public health emergencies. MMWR 2004;53(Suppl): 53–5.
10. Foldy SL. Linking better surveillance to better outcomes. MMWR 2004;53(Suppl):12–7.
11. Foldy SL, Barthell E, Silva J, et al. SARS Surveillance Project—Internet-enabled multiregion surveillance for rapidly emerging disease. MMWR 2004;53(Suppl):215–20.
12. Metzger K, Mostashari F, Heffernan R, Das D. Use of pharmacy data to evaluate impact of smoking regulations on sales of nicotine replacement therapies in New York City. Presentation, 2003 National Syndromic Surveillance Conference, New York, New York, October 2003. Available at http://www.syndromic.org/pdf/con2-KM-4.pdf.
13. Pavlin J. Conducting population behavioral health surveillance using automated diagnostic and pharmacy data systems. MMWR 2004; 53(Suppl):166–72.
14. Mostashari F. Analytic methods track: a summary. Presentation at the 2003 National Syndromic Surveillance Conference, New York, New York, October 2003. Available at http://www.syndromic.org/pdf/con2-FM-9.pdf.
15. Siegrist D. Bio-ALIRT biosurveillance detection algorithm evaluation. MMWR 2004;53(Suppl):152–8.

System Descriptions

New York City Syndromic Surveillance Systems

Richard Heffernan, F. Mostashari, D. Das, M. Besculides, C. Rodriguez, J. Greenko, L. Steiner-Sichel, S. Balter, A. Karpati, P. Thomas, M. Phillips, J. Ackelsberg, E. Lee, J. Leng, J. Hartman, K. Metzger, R. Rosselli, D. Weiss
New York City Department of Health and Mental Hygiene, New York, New York

Corresponding author: Richard Heffernan, Director, Data Analysis Unit, Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, 125 Worth St., Room 318, Box 22A, New York, NY 10013. Telephone: 212-788-5391; Fax: 212-788-5470; E-mail: rheffern@health.nyc.gov.

Abstract

New York City's first syndromic surveillance systems were established in 1995 to detect outbreaks of waterborne illness. In 1998, daily monitoring of ambulance dispatch calls for influenza-like illness began. After the 2001 World Trade Center attacks, concern about biologic terrorism led to the development of surveillance systems to track chief complaints of patients reporting to emergency departments, over-the-counter and prescription pharmacy sales, and worker absenteeism. These systems have proved useful for detecting substantial citywide increases in common viral illnesses (e.g., influenza, norovirus, and rotavirus). However, the systems have not detected more contained outbreaks earlier than traditional surveillance. Future plans include monitoring school health and outpatient clinic visits, augmenting laboratory testing to confirm syndromic signals, and conducting evaluation studies to identify which of these systems will be continued for the long term.

Introduction

The New York City (NYC) Department of Health and Mental Hygiene (DOHMH) has conducted prospective surveillance of nonspecific health indicators (syndromes) since 1995. This paper briefly describes syndromic surveillance systems in operation.

Syndromic Surveillance Systems

Diarrheal Disease Surveillance

NYC's first syndromic surveillance systems were implemented in 1995 to detect substantial outbreaks of diarrheal illness, particularly those caused by waterborne *Cryptosporidium* and *Giardia*. The program included three components: 1) surveillance for diarrheal illness at nursing homes, 2) surveillance of stool submissions at clinical laboratories, and 3) over-the-counter (OTC) pharmacy sales. An evaluation of these systems conducted in 2001 recommended transition to electronic reporting and use of standardized analytic methodology to detect aberrations in the data (*1*). Lessons learned from this evaluation were incorporated into the design of subsequent systems.

Emergency Medical Services Ambulance Dispatch Calls

Monitoring of ambulance dispatch calls for indicators of biologic terrorism began in 1998. Approximately 1 million calls received annually by the NYC emergency medical services (EMS) system are categorized into 52 call types and entered into a centralized database. The main outcome of interest is the percentage of calls categorized as influenza-like illness (ILI), which includes four call types: respiratory, difficulty breathing, sick, and sick pediatric. An adaptation of the excess influenza mortality cyclical (linear) regression model (*2*) detects aberrations in this daily percentage citywide. The model controls for season, day of the week, holidays, positive influenza tests, and heat waves. Daily regressions with ≤3 years of baseline data have been performed since 1999 and have identified widespread influenza epidemics 2–3 weeks before traditional influenza surveillance systems (*3*). A review of 2,294 emergency department (ED) charts determined that patients brought in by ambulance were more likely to be older, more seriously ill, and admitted to the hospital than patients arriving by other means (*4*).

Emergency Department Visits

Syndromic surveillance of ED visits was established after the 2001 World Trade Center attacks to track the acute health effects of the attacks and to detect possible biologic terrorism (*5*). The initial labor-intensive system, which relied on manual data collection, was replaced in November 2001 by an electronic system that has operated daily since then. DOHMH receives data from 48 hospitals encompassing approximately 86% of annual ED visits in NYC. Data files contain the following information for all ED visits logged during the previous day: date and time of visit, age, sex, residential zip code, and free-text chief complaint. Certain hospitals also provide a visit number or medical-record number. Other personal

identifiers are not included. Files arrive via direct file transfer protocol (FTP) or as e-mail attachments. Data are converted to a standard format, and chief complaints are coded by syndrome by using a computer algorithm that searches for key text strings (available at http://www.syndromic.org/work.html). Citywide temporal and spatial clustering in syndrome visits, by hospital location or residential zip code, is assessed by using adaptations of temporal and spatial scan statistics (*6,7*). Results are usually available before noon each day (including weekends). If an unusual cluster is detected, follow-up is conducted the same day. Follow-up involves reviewing the age, sex, and chief complaints of patients in the cluster and telephoning staff at affected EDs to alert them of the cluster and ask whether they have noted unusual presentations or higher-than-usual volume. When necessary, field investigations are conducted. A review of the methods and first year of operation of the ED surveillance system has been published previously (*8*).

Retail Pharmacy Sales

In August 2002, DOHMH established a comprehensive OTC pharmacy sales tracking system. Data from 248 NYC pharmacies (representing approximately 30% of citywide sales) are transmitted to DOHMH daily by FTP from a central pharmacy database and consist of the number of OTC units sold the previous day, grouped by drug name and store. The two syndromes monitored routinely are ILI, which includes cough and influenza medications whose sales correlate most strongly with annual influenza epidemics, and antidiarrheal medicines, including generic and brand-name loperamide. Citywide trends are evaluated by using a linear regression model similar to that used in the EMS system (*3*), controlling for season, holidays, day of the week, promotional sales, positive influenza tests, and temperature. Analysis is conducted weekdays only, with results for the preceding day available by mid-afternoon. In May 2003, DOHMH began receiving OTC pharmacy sales data from the National Retail Data Monitor (*9*).

Worker Absenteeism

Since November 2001, DOHMH has monitored worker absenteeism from a single employer (employee population: approximately 15,000) with multiple locations throughout the city for unusual patterns of illness. The workers' reasons for absence are categorized by a computer algorithm into three syndrome categories: fever/influenza, gastrointestinal, and cold (upper respiratory infection). Agencywide trends are graphed and temporal aberrations assessed by using the cumulative sum (CUSUM) method (*10*) with a 14-day baseline. Analysis is conducted weekdays only, and results for the previous day are usually available by mid-afternoon.

Staffing for Syndromic Surveillance Systems

With exceptions as noted, these systems operate 7 days/week and are staffed by a rotation of eight analysts and five medical epidemiologists. Each day an analyst with master's- or doctoral-level training in public health and statistical software programming experience dedicates 2–3 hours to collect, process, and analyze data and disseminate results. A medical epidemiologist reviews the results daily and, when indicated, directs an investigation with assistance from a public health nurse or field epidemiologist. Approximately 30 additional DOHMH public health epidemiologists and nurses have been trained to assist in signal investigations but have rarely been used. Hospital staff are occasionally enlisted to provide information on patients, perform diagnostic testing on subsequent patients, and assist with other aspects of an investigation. Annual direct costs to DOHMH to maintain the existing systems, including routine follow-up of signals, are approximately $150,000 (not including costs associated with research and development, surveillance for noninfectious disease, or data-transmission costs incurred by hospitals).

Usefulness

Syndromic surveillance has been most useful for detecting citywide increases in illness. Syndromic data have been used to augment health alerts for communitywide gastrointestinal illness caused by norovirus (*11*), annual influenza epidemics, and diarrheal illness following the August 2003 blackout (*12*). Although DOHMH has observed an average of two spatial signals per month for each syndrome, to date none has led to early detection of a localized outbreak. The occurrence of simultaneous signals for the same syndrome from multiple systems has been rare. Experience indicates that ED surveillance has the greatest value because it can track multiple illnesses and enable follow-up with individual patients at the source of care.

Future Projects

DOHMH is developing data sources and testing new analytic methods for outbreak detection. Data sources being explored include school health nurse visits, laboratory-order submissions, and outpatient encounters. Promising methodologic advances include the space-time–permutation method (*13*) and

the use of regression modeling to adjust for known sources of variation before calculating scan or CUSUM statistics. DOHMH continues to explore how syndromic data can be used for general public health surveillance (e.g., detecting carbon monoxide poisonings or examining the impact of smoking legislation on nicotine replacement sales [*14*]).

Conclusion

Syndromic surveillance is one component of overall disease-surveillance and terrorism-preparedness efforts. Formal evaluations will help DOHMH determine which of the current systems will become a permanent public health surveillance activity in NYC.

References

1. Mostashari F. Evaluation of New York City's syndromic surveillance for diarrheal illness. New York: New York Academy of Medicine, 2001.
2. Serfling RE, Sherman IL, Houseworth WJ. Excess pneumonia-influenza mortality by age and sex in three major influenza A2 epidemics, United States, 1957–58, 1960, and 1963. Am J Epidemiol 1967;86:433–41.
3. Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for community-wide influenza-like illness, New York City. J Urban Health 2003;80(2 Suppl 1):i43–9.
4. Greenko J, Mostashari F, Fine A, Layton M. Clinical evaluation of the emergency medical services (EMS) ambulance dispatch-based syndromic surveillance system, New York City. J Urban Health 2003;80(2 Suppl 1):i50–6.
5. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. J Urban Health 2003;80(2 Suppl 1):i76–88.
6. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. J R Statist Soc A 2001;164:61–72.
7. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. Emerg Infect Dis 2003;9:641–6.
8. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice: New York City. Emerg Infect Dis 2004;10. Available at http://www.cdc.gov/ncidod/EID/vol10no5/03-0646.htm.
9. Wagner MM, Robinson JM, Tsui F-C, Espino JU, Hogan WR. Design of a national retail data monitor for public health surveillance. J Am Med Inform Assoc 2003;10:409–18.
10. Hutwagner L, Thompson W, Seeman M, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96.
11. CDC. Norovirus activity—United States, 2002. MMWR 2003;52:41–4.
12. Marx MA, Rodriguez C, Greenko, J, et al. Investigation of diarrheal illness detected through syndromic surveillance after a massive blackout — New York City, August 2003 [Abstract]. MMWR2004;53 (Suppl):251.
13. Kulldorff M, Mostashari F, Heffernan R, Hartman JJ. Syndromic surveillance without denominator data: the space-time permutation statistic. Presentation, 2003 National Syndromic Surveillance Conference, New York City, October 2003.
14. Metzger K, Mostashari F, Das D, Heffernan R. Use of pharmacy data to evaluate impact of smoking regulations on sales of nicotine replacement therapies in New York City. Presentation, 2003 National Syndromic Surveillance Conference, New York City, October 2003.

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. Hutman[1]
[1]*University of Pittsburgh, Pittsburgh, Pennsylvania;* [2]*Intermountain Health Care Institute for Health Care Delivery Research, Salt Lake City, Utah;* [3]*University of Utah, Salt Lake City, Utah;* [4]*Carnegie 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 No.	Sensitivity (95% CI*)	Specificity No.	Specificity (95% CI)	Positive likelihood ratio (LR+) No.	Positive likelihood ratio (LR+) (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 (0.82 (0.75–0.90))	(8.3–18)
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.*
** $LR+ = \frac{sensitivity}{1-specificity} = \frac{0.61}{1-1}$

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

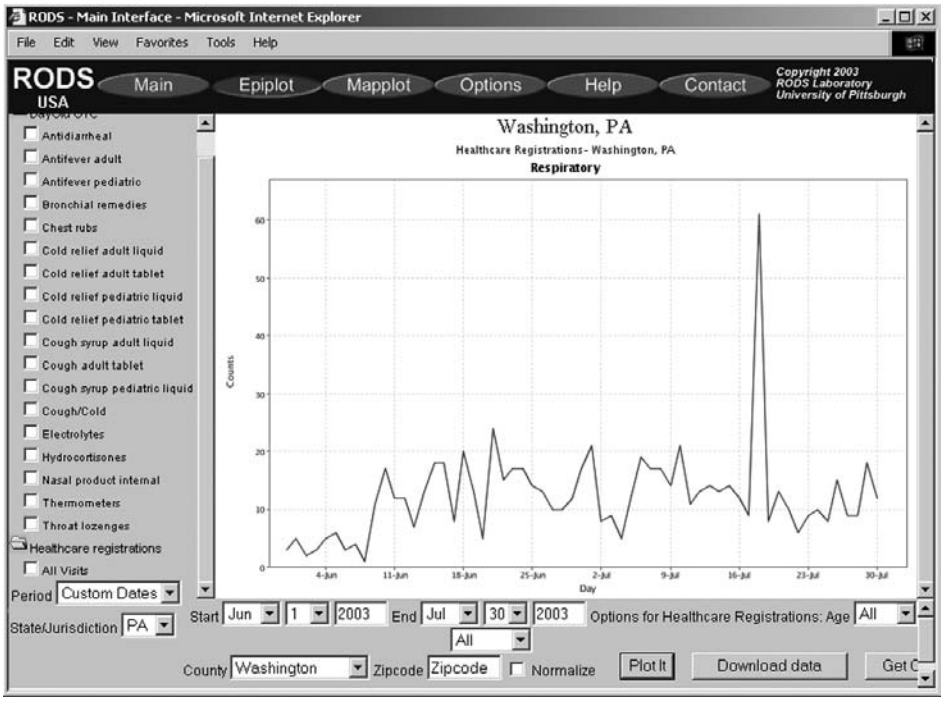
Retrospective studies cannot prove that, in field use, this type of system will lead to earlier detection than existing methods. For this reason, the project initiated a prospective evaluation.

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.

Removing a Barrier to Computer-Based Outbreak and Disease Surveillance — The RODS Open Source Project

Jeremy U. Espino, M. Wagner, C. Szczepaniak, F-C. Tsui, H. Su, R. Olszewski, Z. Liu, W. Chapman, X. Zeng, L. Ma, Z. Lu, J. Dara
University of Pittsburgh, Pittsburgh, Pennsylvania

Corresponding author: Jeremy U. Espino, Real-Time Outbreak and Disease Surveillance Laboratory, University of Pittsburgh, Suite 500, Cellomics Building, 500 Technology Dr., Pittsburgh, PA 15219. Telephone: 412-383-8130; Fax: 412-383-8135; E-mail: jue@cbmi.pitt.edu.

Abstract

Introduction: *Computer-based outbreak and disease surveillance requires high-quality software that is well-supported and affordable. Developing software in an open-source framework, which entails free distribution and use of software and continuous, community-based software development, can produce software with such characteristics, and can do so rapidly.*

Objectives: *The objective of the Real-Time Outbreak and Disease Surveillance (RODS) Open Source Project is to accelerate the deployment of computer-based outbreak and disease surveillance systems by writing software and catalyzing the formation of a community of users, developers, consultants, and scientists who support its use.*

Methods: *The University of Pittsburgh seeded the Open Source Project by releasing the RODS software under the GNU General Public License. An infrastructure was created, consisting of a website, mailing lists for developers and users, designated software developers, and shared code-development tools. These resources are intended to encourage growth of the Open Source Project community. Progress is measured by assessing website usage, number of software downloads, number of inquiries, number of system deployments, and number of new features or modules added to the code base.*

Results: *During September–November 2003, users generated 5,370 page views of the project website, 59 software downloads, 20 inquiries, one new deployment, and addition of four features.*

Conclusions: *Thus far, health departments and companies have been more interested in using the software as is than in customizing or developing new features. The RODS laboratory anticipates that after initial installation has been completed, health departments and companies will begin to customize the software and contribute their enhancements to the public code base.*

Introduction

In October 1999, researchers at the University of Pittsburgh began developing the Real-Time Outbreak and Disease Surveillance system (RODS), with the goal of improving public health agencies' capability to detect a specific threat: a large-scale, surreptitious release of *Bacillus anthracis*. The rate of this technology's adoption, although accelerating, is not commensurate with the severity of the health threats posed by biologic terrorism, emerging infections, and common disease outbreaks. Such threats warrant rapid deployment; therefore, barriers to the technology's adoption need to be identified and removed.

This paper describes the evolution of the RODS system, previous efforts to transition the technology, and the rationale behind the creation of an open-source project. It also describes how the software is licensed, the infrastructure created to enable growth of the RODS open-source community, efforts to publicize the project, metrics collected to assess its progress, the software architecture of the latest version of RODS, and plans for additional software development.

RODS System Description

The first version of RODS collected patient chief-complaint data from eight hospitals in a single health-care system via Health Level 7 (HL7) (*1*) messages in real time, categorized these data into syndrome categories by using a classifier based on *International Classification of Diseases, Ninth Revision* (ICD-9) codes, aggregated the data into daily syndrome counts, and analyzed the data for anomalies possibly indicative of disease outbreaks. The system provided an Internet-based interface enabling users to view the data in graphs and maps (Figure 1). After demonstrating the feasibility of such a system within a single health-care system in Pittsburgh and conducting research to support the hypothesis that such a system could detect disease outbreaks (*2,3*), RODS' developers expanded the system to collect additional data types and then deployed RODS in multiple states. The application service provider (ASP) version of RODS at the University of Pittsburgh collects de-identified chief complaints from 76 hospitals in Pennsylvania, Utah, and Ohio (*4,5*) and also serves as the user interface for the National Retail Data Monitor

FIGURE 1. Sample time-series graphs* of syndromic surveillance data as displayed on the Epiplot user interface of the Real-Time Outbreak and Disease Surveillance (RODS) system

* In this example, graphs of pediatric gastrointestinal emergency department visits are shown alongside graphs depicting unit sales of antidiarrheals and electrolytes.

(NRDM), which collects and analyzes daily sales data for over-the-counter (OTC) medication sales (*6,7*).

The feasibility of rapid deployment of RODS was demonstrated during the 2002 Winter Olympics in Salt Lake City, Utah (*4,8,9*). In addition, the capability to integrate other surveillance data types (e.g., electronic laboratory reports [*10*], free-text chief complaints (*11,12*), laboratory orders, dictated radiology reports, dictated hospital reports [*13–15*], and poison control center calls [*16*]) was added. Much of the code (originally in Perl and C) was rewritten in Java,™ and basic research was conducted on data and algorithms relevant to this emerging science (*17*).

Technology Transition

The initial effort to make RODS software available involved licensing it for noncommercial use. In December 2002, the University of Pittsburgh began offering the RODS system as compiled byte code, free of charge to public health departments. To date, >180 downloads of this version of the RODS system and >200 downloads of the Bayesian parser have been counted. Despite reports of successful installations in Hong Kong [David Wong, Hong Kong RODS Team, personal communication, May 15, 2003] and Missouri [Terry Tabor, Missouri Department of Health and Senior Services, personal communication, January 28, 2003], certain state health departments expressed interest in accessing the RODS source code.

Giving the software away without providing technical support soon proved insufficient. Using the RODS software requires expertise in database, network, geographic information system (GIS), HL7, and system management, capabilities not widely available at that time. Users made multiple requests for customization, support, and assistance with

installations, for which resources were not available. Therefore, in September 2003, the University of Pittsburgh released the RODS software under an open-source license, thereby creating the RODS Open Source Project to catalyze the sharing of knowledge and skills related to the software, including its design, installation, configuration, and customization.

Materials and Methods

This section describes the RODS Open Source Project, including the particular license under which RODS is distributed, the infrastructure created to enable growth of the RODS open-source community, methods for publicizing the project and recruiting developers, and the metrics collected to assess its progress.

GNU General Public License

RODS is distributed as open-source software under the GNU General Public License (GPL) (*17*), the same open-source license under which Linux® is distributed (*18*). Unlike the license under which RODS was initially released in December 2002, GPL permits anyone to use, copy, and modify RODS freely. GPL allows consultants and companies to use, install, support, and customize RODS and permits these entities to redistribute their enhanced versions of RODS, provided they make the source code available. This requirement fosters continuous software improvement, benefiting all users and preventing companies from creating proprietary, closed-source versions of RODS.

Support for Developers and Users

To coordinate community-based development of the code, the RODS Laboratory organized the Open Source Project. The RODS modules were classified into six functional areas: data collection, syndrome classification, data warehousing, database encapsulation, outbreak detection, and user interface. Specialists from the laboratory's research and development group named development leaders for each functional area. These development leaders are responsible for recommending new features based on user requests and evaluating whether a developer has the qualifications to contribute source code.

Online resources were created to support the Open Source Project, including the RODS Laboratory website (http://www.health.pitt.edu/rods) and a project website hosted on Sourceforge™ (http://openrods.sourceforge.net). The latter site provides standard software project management tools (a concurrent versions system server and patch submission area enabling developers to contribute code), e-mail lists enabling developers and users to communicate, a software-bug reporting system, contact information for the development leaders, and source code for stable versions of the system.

Recruitment of Developers and Users

E-mail announcements were sent to 181 persons who had previously downloaded the byte-compiled releases and to all 226 users in the United States who held passwords to the RODS ASP system. Users were given an opportunity for a face-to-face meeting with the core developers at two national conferences, the 2003 National Syndromic Surveillance Conference in New York City and the 2003 American Medical Informatics Fall Symposium in Washington, D.C. Project leaders of other computer-based surveillance projects were also invited.

Metrics

The following metrics are collected monthly to manage the project and assess its progress:

- cumulative number of installations;
- cumulative number of developers who have contributed code;
- number of new features;
- funding sources;
- cumulative number of mailing list subscribers (one general mailing list, one for announcements, and one for development questions);
- total website page views;
- total downloads of source code;
- number of e-mail announcements sent;
- cumulative number of inquiries from consultants and companies;
- cumulative number of inquiries from health departments;
- cumulative number of inquiries from academics; and
- cumulative number of inquiries from other groups.

The number of installations and the number of contributing developers are considered the two most important metrics.

Results

Current Software Architecture of RODS Version 2.0 and Features in Development

A complete technical description of RODS has been published (*8*). This section describes the system's software architecture and how the modules that comprise that architecture can be used to accomplish different surveillance tasks.

RODS 2.0 consists of >42,000 lines of Java code contributed by a team of eight programmers. RODS is a modular system that adheres to CDC's National Electronic Disease Surveillance System (NEDSS) (*19*) and Public Health Information Network (PHIN) (*20*) standards so that any of the components can be incorporated into a foreign surveillance system or used to create a native end-to-end RODS system.

The RODS software architecture consists of six functional areas: data collection, syndrome classification, data warehousing, database encapsulation, outbreak detection, and user interface (Figure 2). Within the following categories, additional modules are being developed under the Open Source Project (Table 1):

- **Data collection.** The data-collection modules consist of 1) an HL7 listener that accepts and maintains connections from a hospital's HL7-integration engine; 2) an HL7 parser that extracts patient-visit data from HL7 messages; and 3) a text-file parser that extracts patient-visit data from text files uploaded in batches by non-HL7–capable hospitals. In addition to modules to parse patient data from HL7 messages, modules are being developed to parse microbiology culture results from HL7 messages and to import poison center call data to RODS.

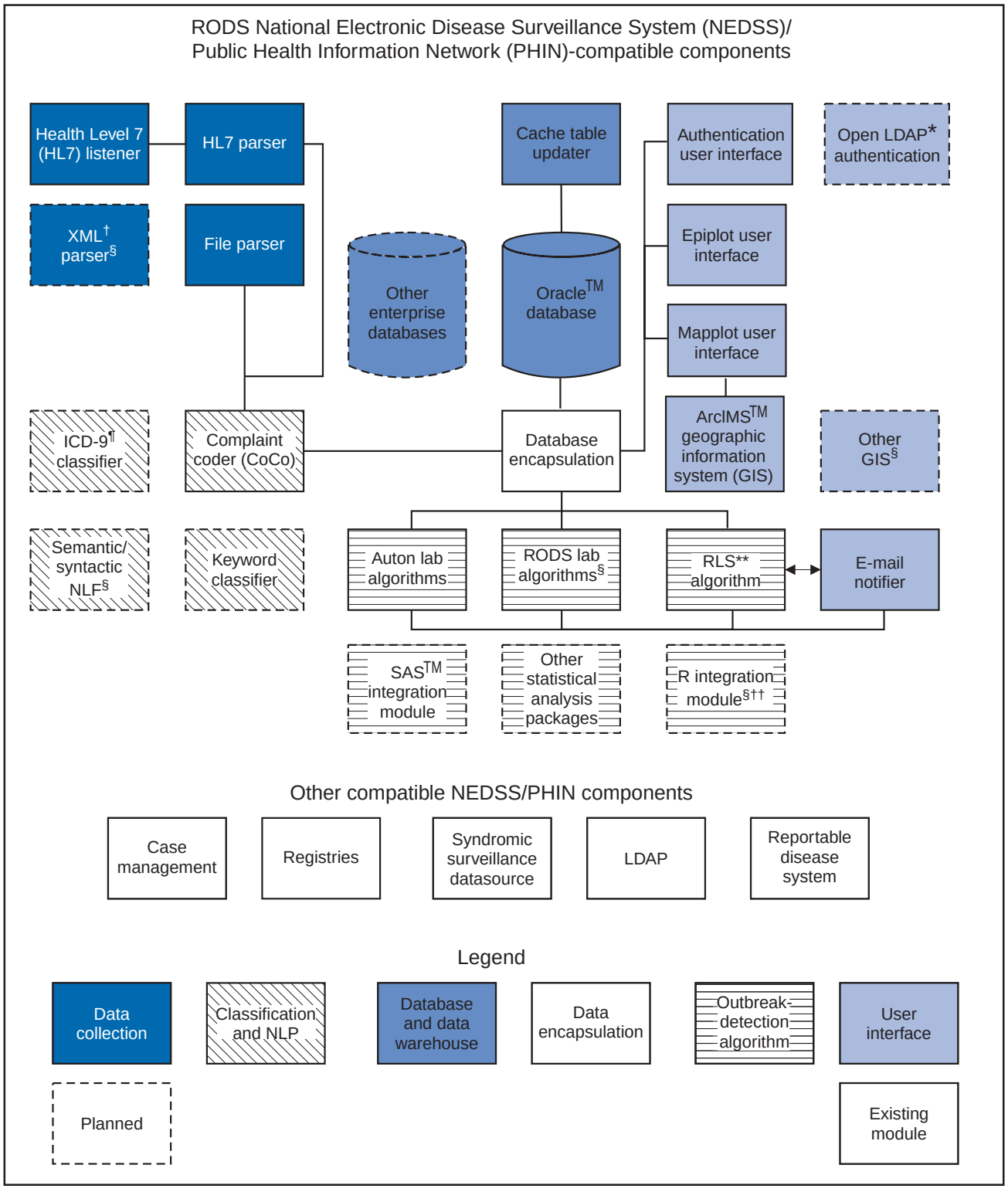
  Another module is proposed that will fully integrate detailed OTC medication sales data from the NRDM. Also planned is an extensible markup language (XML) module that works with proposed or currently used XML-document–type definitions for public health surveillance data (*21,22*).
- **Syndrome classification.** RODS Version 2.0 consists of a single module for syndrome classification, Complaint Classifier (CoCo) (*12*). CoCo uses a naïve Bayesian classifier to assign a free-text chief complaint to a syndrome category. These syndrome categories are user-specifiable, and the mappings are created automatically through machine learning from a user-provided training set.

  The RODS Laboratory has rewritten (in Java) and intends to release a module for ICD-9–based classification (*8*). Additional classification modules, including keyword-based methods and additional natural language processing modules to identify radiology reports indicative of inhalational anthrax (*15*), are in development.
- **Data warehousing.** These modules function to store and provide efficient access to surveillance data. RODS efficiently stores and retrieves time-series data from the database through a data warehouse. The data-warehousing module consists of a cache table updater that keeps running counts of the number of visits for each syndrome, stratified by age and sex.

  RODS 2.0 assumes the existence of an Oracle™ database. However, RODS does not use Oracle-specific structured query language (SQL) functions (e.g., database triggers), and a port to an alternative relational database system (e.g., PostgreSQL or Microsoft SQL Server™) should be straightforward.
- **Database encapsulation.** The database-encapsulation modules, written as Enterprise Java Beans™ (EJBs), function to retrieve preprocessed time-series data and case details (e.g., the patient's free-text chief complaint) from the database. In Java, EJBs provide a framework for creating readily accessed software objects that incorporate standard methods for security, database access, transactions, scalability, and communication. The EJBs shield developers from the database schema and standardize how the surrounding modules (e.g., the user interface modules) access the database.
- **Detection algorithm.** The detection-algorithm modules provided in the current open-source release include an implementation of the recursive least-squared (RLS) algorithm (*23*) and an initial implementation of a wavelet-detection algorithm. The RLS algorithm can detect sudden increases in daily surveillance data counts (e.g., an increase in the number of respiratory-type visits that would accompany a large-scale, covert release of *Bacillus anthracis*). The wavelet algorithm can automatically model weekly, monthly, and seasonal data fluctuations. NRDM uses wavelet modeling to indicate zip-code areas in which OTC medication sales are substantially increased; this algorithm will be applied to the analysis of health-care registration data.

  Another set of modules are planned that will enable any outbreak-detection algorithm to analyze data from the system. Currently, the architecture allows algorithms written or wrapped in Java to retrieve data directly from the database-encapsulation modules. A module will be released that outputs data as common text files so that stand-alone algorithms and statistical software packages can be used to analyze the data. This method was used by the What's Strange About Recent Events algorithm (WSARE) to analyze data from RODS during the Salt Lake 2002 Olympic Winter Games (*24*).
- **User interfaces.** These modules 1) authenticate users, 2) display surveillance data as time-series graphs, and 3) work with a GIS to depict data spatially. The graphing and GIS modules consist of Java server pages and servlets that use JFreeChart, an open-source graphing package, and the GIS functions of Environmental Systems Research Institute's ArcIMS™ software.

FIGURE 2. Software architecture of the Real-Time Outbreak and Disease Surveillance (RODS) system



* LDAP = Lightweight directory access protocol.
† XML = extensible markup language.
§ Has General Public License-compatible open-source components.
¶ ICD-9 = *International Classification of Diseases, Ninth Revision.*
** RLS = Recursive least-squared.
†† An open-source statistical computing software program.

TABLE 1. Existing features of the Real-Time Outbreak Disease Surveillance (RODS) system, version 2.0, and features awaiting development

RODS feature	Exists in RODS 2.0	Exists as GPL*-compatible source code	Needs to be developed or tested
Data-collection modules			
Health Level 7 (HL7) listener	X		
HL7 parser for microbiology reports			X
HL7 parser for admissions, discharge, and transfer (ADT) messages	X		
Text file parser	X		
XML parser		X	
Syndrome-classification modules			
Simple Bayesian syndrome classifier	X		
Syntactic/semantic natural language processing (NLP) classifier		X	
Keyword classifier			X
ICD-9† classifier			X
Multiple data-type classifier		X	
Data-warehousing modules			
Diverse database options		X	
Integrated data-warehouse engine			
Aggregation by sex and age	X		
Outbreak-detection modules			
Integrates with external statistical analysis tools	X	X	X
Recursive least-squared (RLS) detection algorithm	X		
User-interface modules			
Manual data-entry interface		X	
Diverse geographic information system software options		X	
Lightweight directory access protocol (LDAP) interface		X	
Time-series graphing	X	X	
Options and preferences	X		
Custom jurisdictions	X		
E-mail notifier	X		
Database-encapsulation modules			
Database encapsulation	X		

*GNU General Public License.
† *International Classification of Diseases, Ninth Revision.*

Certain state health departments have requested Lightweight Directory Access Protocol (LDAP) support to enable the creation of seamless links between existing state surveillance systems and the surveillance functions provided by RODS; outside development of such a module is encouraged.

State, local, or national health departments can use RODS modules to collect, analyze, and view hospital surveillance data and to view OTC medication sales data from NRDM. A health department can use a subset of these modules to accomplish a specific surveillance task (e.g., receiving and processing free-text chief complaints from hospitals), or it can use all of them (with the RODS database, analytic modules, and user interface) to create an end-to-end surveillance solution. (Examples of how health departments can mix and match RODS modules for different surveillance tasks are available at http://openrods.sourceforge.net.)

Project Metrics

A total of 480 e-mail announcements about the RODS Open Source Project were sent during the first 3 months of the project. This publicity generated 5,370 page views of the project website, 59 downloads of the source code, and 14 new members to the project mailing lists. One additional installation is using the open-source version of RODS.

To date, users are more interested in using the software "as is" and less interested in collaborative feature development. For example, users have asked when the ICD-9 classifier module will be released or whether the system yet works with Microsoft SQL Server.™ Developers at the RODS Laboratory contributed four new features (drilldown of age and sex, customized jurisdictions, a simplified GIS interface, and user preferences) (Table 2). However, at least one health department and one consulting company have expressed interest in collaborating to develop a module that will import XML data into RODS.

TABLE 2. Monthly metrics for the Real-Time Outbreak Disease Surveillance (RODS) Open Source Project — September–November 2003

Metric	September	October	November
Number of e-mail announcements sent out	406	0	74
Total page views on website	1,968	1,764	1,638
Total downloads of source code	18	19	22
Cumulative number of members on site mailing lists	7	8	14
Cumulative number of installations	3	4	4
Cumulative number of inquiries from consultants and companies	7	8	9
Cumulative number of inquiries from health departments	5	6	7
Cumulative number of inquiries from academics	1	1	1
Cumulative number of inquiries from other groups	2	3	3
Cumulative number of developers	8	8	8
Number of new features	0	0	4
Funding sources	1	1	1

Discussion

The goal of the RODS Open Source Project is to accelerate the deployment of computer-based outbreak and disease surveillance systems by writing high-quality surveillance software and catalyzing the formation of a community of users, developers, consultants, and scientists. In the initial years of computer-based outbreak and disease surveillance system development, the main barriers to deployment appeared to be doubts about its efficacy, cost of the technology, concerns about the cost and effect of false alerts on the practice of public health, and legal and administrative issues (*25,26*). Basic research about data and detectability has been conducted to address concerns about efficacy (*2,3,27–29*). To address concerns about the effects of false alerts, the RODS laboratory has deployed systems and discovered that persons working in health departments could incorporate the output of these systems into their workflows (*4,7*). The deployments also established that the cost and effort of deployment is much lower than expected. Finally, the deployments demonstrated that certain concerns about privacy could be addressed. The Health Information Portability and Accountability Act of 1996 (HIPAA), which had not yet become law, nevertheless had a substantial inhibitory effect on hospitals and other covered entities that had data needed by the project. The enactment of the final privacy rule, precedents set by system deployments (*4,30–32*), and new state laws have helped address certain concerns of data providers (*33*).

Open-source projects can create a community of like-minded persons — scientists, programmers, consultants, and users — who have the vision of creating innovative, well-supported software. The importance of catalyzing such a community cannot be overstated. It can strengthen the position of information technology (IT) managers and public health officials who wish to deploy computer-based surveillance systems during planning deliberations. They will be able to assure their supervisors that source code is available, that a pool of developers and consultants exists who can be hired to support the health department if needed, and that ongoing projects in other health departments can help them predict project costs and set appropriate timelines.

The RODS Open Source Project enables public health professionals to have a greater role in developing IT solutions to the problem of early detection. Just as public health researchers publish their results in scientific journals, so can they contribute publicly available IT solutions to the RODS Open Source Project. This role might become more apparent as public health personnel become increasingly knowledgeable about public health informatics and work more closely with IT subcontractors and consultants.

Continued goals for the RODS Open Source Project are to increase the number of deployments, developers, and supporters of the software. The proposed path for RODS software development is to increase the number of data types the system can accept and implement a range of high-performance outbreak-detection algorithms. One consulting company and one health department have separately expressed interest in collaboratively developing an XML module that can parse non-RODS data sources. The RODS Laboratory and its collaborators at the Auton Laboratory will continue to develop outbreak-detection algorithms (e.g., the wavelet-detection module and WSARE, respectively).

Conclusion

The RODS Open Source Project is making software modules available that span the spectrum of processing tasks involved in public health surveillance. Through open source, the project hopes to accelerate the deployment of real-time public health surveillance by lowering costs, increasing reliability, preventing vendor lock-in, and ensuring software customizability. By catalyzing the formation of a community of open-source public health surveillance software advocates, this approach will result in a high-quality software product that achieves mainstream acceptance.

Acknowledgments

The RODS Open Source Project is supported by the Pennsylvania Department of Health Bioinformatics Grant ME-107.

References

1. Health Level Seven, Inc. Health Level Seven. Ann Arbor, MI: Health Level Seven, Inc., 2001. Available at http://www.hl7.org.
2 Espino JU, Wagner MM. The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness. Proc AMIA Symp 2001:164–8.
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. 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.
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. Wagner MM, Robinson J, Tsui F-C, Espino JU, Hogan W. Design of a national retail data monitor for public health surveillance. J Am Med Inform Assoc 2003;10:409–18.
7. Wagner MM, Espino J, Hersh J, et al. National retail data monitor for public health surveillance. MMWR 2004;53(Suppl):40–2.
8. 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.
9. 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.
10. 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.
11. Chapman WW, Christensen LM, Wagner MM, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing. CBMI report series. Pittsburgh, PA: University of Pittsburgh, Center for Biomedical Informatics, 2002.
12. Olszewski R. Bayesian classification of triage diagnoses for the early detection of epidemics. In: Russell I, Haller S, eds. Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference. Menlo Park, CA: AAAI, 2003;412–7.
13. Chapman WW, Dowling JN, Wagner MM. Fever detection from free-text clinical records for biosurveillance. J Biomed Inform 2004;37:120–7.
14. Chapman WW, Espino JU, Dowling JN, Wagner MM. Detection of acute lower respiratory syndrome from chief complaints and ICD-9 codes. CBMI report series. Pittsburgh, PA: University of Pittsburgh, Center for Biomedical Informatics, 2003.
15. Chapman WW, Wagner M, Cooper G, Hanbury P, Chapman B, Harrison L. Creating a text classifier to detect chest radiograph reports consistent with features of inhalational anthrax. J Am Med Inform Assoc 2003;10:494–503.
16. Zeng X, Wagner MM. Accuracy, speed, and completeness of data collection by poison control center nurses for the investigation of outbreaks of acute diarrhea. CBMI report series. Pittsburgh, PA: University of Pittsburgh, Center for Biomedical Informatics, 2003.
17. Free Software Foundation, Inc. GNU general public license. Boston, MA: Free Software Foundation, Inc., 1991. Available at http://www.gnu.org/licenses/gpl.html.
18. Raymond ES. The cathedral and the bazaar: musings on Linux and Open Source by an accidental revolutionary. Rev. ed. Beijing; Cambridge, MA: O'Reilly, 2001.
19. CDC. National Electronic Disease Surveillance System: the surveillance and monitoring component of the Public Health Information Network. Atlanta, GA: US Department of Health and Human Services, CDC, 2004. Available at http://www.cdc.gov/nedss/.
20. CDC. Public Health Information Network. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/phin/.
21. Lober WB, Trigg LJ, Karras BT, et al. Syndromic surveillance using automated collection of computerized discharge diagnoses. J Urban Health 2003;80(2 Suppl 1):i97–106.
22. Barthell EN, Cordell WH, Moorhead JC, et al. The Frontlines of Medicine Project: a proposal for the standardized communication of emergency department data for public health uses including syndromic surveillance for biological and chemical terrorism. Ann Emerg Med 2002;39:422–9.
23. Hayes M. Statistical digital signal processing and modeling. New York, NY: John Wiley & Sons, Inc, 1996.
24. Wong WK, Moore A, Cooper G, Wagner M. WSARE: What's Strange About Recent Events? J Urban Health 2003;80(2 Suppl 1):i66–75.
25. Broome CV, Pinner RW, Sosin DM, Treadwell TA. On the threshold. Am J Prev Med 2002;23:229.
26. Reingold A. If syndromic surveillance is the answer, what is the question. Biosecur Bioterr 2003;1:1–5.
27. Hogan WR, Tsui F-C, Ivanov O, et al. Early detection of pediatric respiratory and diarrheal outbreaks from retail sales of electrolyte products. J Am Med Inform Assoc 2003;10:555–62.
28. 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.
29. CDC. Annotated bibliography for syndromic surveillance. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/epo/dphsi/syndromic/index.htm.
30. Moran GJ, Talan DA. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. Ann Emerg Med 2003;41:414–8.
31. Lober WB, Karras B, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc 2002;9:105–15.
32. Lewis M, Pavlin J, Mansfield J, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC, area. Am J Prev Med 2002;23:180.
33. Broome CV, Horton HH, Tress D, Lucido SJ, Koo D. Statutory basis for public health reporting beyond specific diseases. J Urban Health 2003;80(2 Suppl 1):i14–22.

National Retail Data Monitor for Public Health Surveillance

Michael M. Wagner,[1] F-C. Tsui,[1] J. Espino,[1] W. Hogan,[1] J. Hutman,[1] J. Hersh,[2] D. Neill,[3] A. Moore,[1,3] G. Parks,[1] C. Lewis,[4] R. Aller[5]
[1]University of Pittsburgh, Pittsburgh, Pennsylvania; [2]Pennsylvania Department of Health, Harrisburg, Pennsylvania; [3]Carnegie Mellon University, Pittsburgh, Pennsylvania; [4]Massachusetts Department of Public Health, Boston, Massachusetts; [5]Los Angeles Department of Health, Los Angeles, California

Corresponding author: Michael M. Wagner, Real-Time Outbreak and Disease Surveillance 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

The National Retail Data Monitor (NRDM) is a public health surveillance tool that collects and analyzes daily sales data for over-the-counter (OTC) health-care products. NRDM collects sales data for selected OTC health-care products in near real time from >15,000 retail stores and makes them available to public health officials. NRDM is one of the first examples of a national data utility for public health surveillance that collects, redistributes, and analyzes daily sales-volume data of selected health-care products, thereby reducing the effort for both data providers and health departments.

Introduction

The National Retail Data Monitor (NRDM) is a public health surveillance tool that collects and analyzes daily sales data for over-the-counter (OTC) health-care products from >15,000 retail stores nationwide. NRDM makes aggregated and analyzed data available to public health officials free of charge (*1*).

A key rationale for building NRDM is that persons with infectious diseases often purchase OTC health-care products early in the course of their illnesses (*2,3*). Furthermore, retrospective studies of certain outbreaks have indicated that monitoring OTC sales might have led to earlier detection (*4–6*). After decades of investment into developing Universal Product Codes (UPCs), optical check-out scanners, and analytic data warehouses, the retail industry has in effect constructed 95% of a surveillance-system pyramid onto which a capstone of data integration and analytic capability can be added to produce NRDM.

NRDM's objectives are to 1) enlist participation of retailers to achieve 70% coverage of OTC sales nationally; 2) influence the industry toward real-time data collection; 3) obtain supplemental information needed for spatial analysis, adjustment for promotional effects, and maintenance of UPC analytic categories (e.g., liquid cough medications); 4) promote and develop this type of surveillance practice; 5) achieve fault and load tolerance; and 6) develop detection algorithms for the data.

Methods

The methods used to acquire and analyze retail data have been described in detail elsewhere (*1*). This paper summarizes and updates that information.

Data Acquisition

Data-sharing agreements between retailers and the University of Pittsburgh enable the university to collect daily sales counts by store and by UPC. Retailers transmit data to NRDM by secure file transfer protocol daily by 3:00 pm Eastern Time for the previous day's sales. NRDM aggregates the data by zip code and product category.

Data Analysis

Health departments receive either aggregated data or access to data-analysis tools via a secure Internet interface. The tools allow users to view sales of OTC health-care products on maps (Figure 1) and timelines.

Various NRDM algorithms are under development, including 1) temporal and 2) spatio-temporal. The temporal algorithm involves univariate time-series analyses, one for each combination of category and zip code. Where u_{zct} represents the unit sales of category c in zip code z on day t, the univariate detector learns a model from the set of sales before today $\{u_{zc1}\ u_{zc2} \ldots u_{zc,t-2}\ u_{zc,t-1}\}$. NRDM uses a specially tailored wavelet model (*7*) to predict units sold today. The advantages of wavelets are their ability to account for long-term trends (e.g., seasonal effects) and short-term properties (e.g., day-of-week effects). In its simplest form, the model predicts a Gaussian distribution for today's sales, with mean and variance learned from sales before today. The actual sales for today can be compared with this Gaussian distribution to produce a z-score (i.e., the number of standard deviations by which today's sales lie above the mean). The z-score can be converted to a p-value to signal alerts.

The spatio-temporal algorithm runs a specially tailored spatial scan statistic (*8*) over all regions. Each region is evaluated according to the likelihood ratio of the data under the assump-

FIGURE 1. Sample map accessible to users of the National Retail Data Monitoring System*

* This map depicts over-the-counter (OTC) sales of cough and cold products in the continental United States on March 24, 2004, by county. Different colors are used to indicate the standard deviations between actual and expected sales.

tion of an increased product demand in the region versus no such increase. Because the data are on a national level, computational tractability is a major concern for such a use of the scan statistic. A fast multiresolution method is used (*9*).

Fault and Load Tolerance

A key requirement for NRDM is fault and load tolerance. NRDM is fault-tolerant, with the exception of the server site and Internet connection, which are single and therefore subject to loss of connection. These vulnerabilities will be addressed by creation of a second site and second Internet connection. Load tolerance refers to NRDM's ability to handle simultaneous access by a substantial number of users. Preliminary load-tolerance tests using Apache JMeter (*10*) have identified certain bottlenecks, which have since been rectified. Complete load testing is planned to determine the maximum number of simultaneous users NRDM can accommodate.

Project Administration

NRDM requires substantial administrative work, including managing contacts with retailers, executing data-sharing agreements, coordinating meetings, handling press inquiries, developing fact sheets, and raising and dispensing funds. This work is handled jointly by volunteers from state and local health departments, staff of the Real-Time Outbreak and Disease Surveillance Laboratory, and a University of Pittsburgh associate general counsel.

Initially NRDM was organized as a university-based, grant-funded project. In May 2003, representatives from four state health departments (Pennsylvania, New York, Ohio, and Georgia) founded an informal association to provide leadership and guidance that holds monthly conference calls; the association is open to any health department.

Results

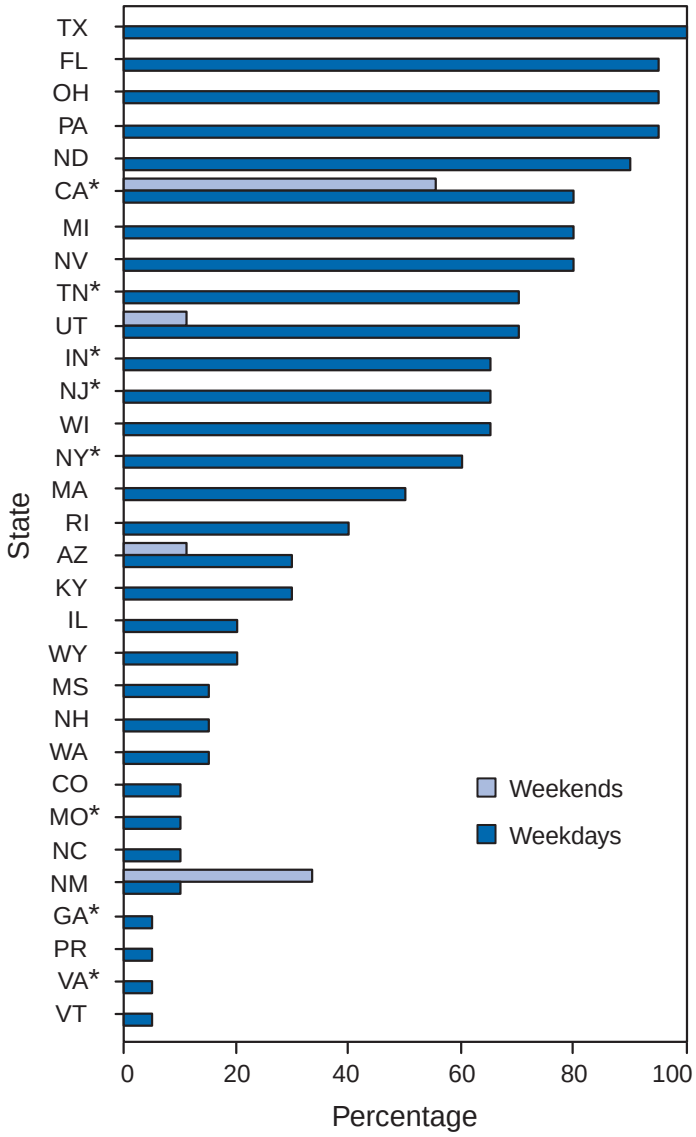
NRDM has operated continuously since December 2002. The project uses explicit measures of progress and reports them monthly to the working group, including

- number of retail stores participating;
- time latency;
- number of states with accounts for the NRDM user interface;
- proportion of weekdays and weekends that NRDM user interfaces are accessed; and
- number of states receiving raw data from NRDM.

As of March 2004, progress towards the goal of 70% data coverage (a level achievable using data from national chains) has reached approximately 40% of total national sales. The time latency is 1 day for all retailers (with one exception that provides a feed every 2 hours). The project has created >400 user accounts for health department employees in 44 states and Puerto Rico. Ten entities receive aggregate data feeds from the system. Progress towards integration of NRDM into public health practice is measured by the number of system logins. Analyses are conducted to track daily and monthly usage and to compare weekday and weekend logins (Figure 2). A level of 100% usage means that at least one user in the state logged in each day. Weekend checking remains low but might increase as public health departments recognize the need to evaluate surveillance data as it becomes available, 7 days/week.

Prospective evaluation of NRDM as a public health surveillance tool is underway. For example, NRDM has demonstrated the marked effect of influenza on sales of pediatric cough and cold remedies and pediatric antipyretics, or the effect of fires in southern California on sales of bronchial remedies. (Authorized public health users can access case studies of these and other outbreaks by using the NRDM Internet interface. To obtain access, please send e-mail to nrdmaccounts@cbmi.pitt.edu).

FIGURE 2. Percentage of weekdays and weekend days on which at least one user accessed the National Retail Data Monitoring System, by state — selected states, February 2004



* States that receive raw data feeds are more likely to conduct their own data analyses and therefore less likely to log in to the NRDM user interface.

Future Plans

From an early warning perspective, the single most important improvement to NRDM will be a reduction in reporting latency after the time of purchase. Better detection performance might also be achieved through improved algorithms, which are under development.

Because they share geographic borders, the United States and neighboring countries need interoperable public health surveillance capability. Retail data monitoring is feasible in Canada, Mexico, and other countries where retailers use the UPC system or the European Article Numbering system, with which it is interconvertible. A permanent organizational home for NRDM is also being explored, with an estimated annual operating cost of approximately $1 million.

Conclusions

NRDM is a data utility that collects, redistributes, and analyzes daily sales-volume data of selected health-care products. A national-level, data-utility approach reduces the effort required for health departments to monitor sales of OTC health-care products. Health departments can instead concentrate on analysis of data and investigation of anomalies.

Acknowledgments

Grant support for NRDM is provided by Pennsylvania Department of Health Bioinformatics Grant ME-01 737; Alfred P. Sloan Foundation; Passaic Water Commission; and the New York State, Washington, Ohio, and Utah departments of health. Participating corporations include ACNielsen, Information Resources, Inc., National Association of Chain Drug Stores, and Global Strategic Solutions.

References

1. Wagner MM, Robinson JM, Tsui F-C, Espino JU, Hogan W. Design of a national retail data monitor for public health surveillance. J Am Med Inform Assoc 2003;10:409–18.
2. Labrie J. Self-care in the new millenium: American attitudes towards maintaining personal health. Washington, DC: Consumer Healthcare Products Association, 2001.
3. McIsaac WJ, Levine N, Goel V. Visits by adults to family physicians for the common cold. J Fam Pract 1998;47:366–9.
4. Corso PS, Kramer MH, Blair KA, Addiss DG, Davis JP, Haddix AC. Cost of illness in the 1993 waterborne *Cryptosporidium* outbreak, Milwaukee, Wisconsin. Emerg Infect Dis 2003;9:426–31.
5. Stirling R, Aramini J, Ellis A, et al. Waterborne cryptosporidiosis outbreak, North Battleford, Saskatchewan, Spring 2001. Can Commun Dis Rep 2001;27:185–92.
6. Mac Kenzie WR, Hoxie NJ, Proctor ME, et al. A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. N Engl J Med 1994;331:161–7.
7. Zhang J, Tsui F-C, Wagner MM, Hogan WR. Detection of outbreaks from time series data using wavelet transform. Proc AMIA Symp 2003;748–52.
8. Kulldorff M. A spatial scan statistic. Communications in Statistics—Theory and Methods 1997;26:1481–96.
9. Neill D, Moore AW. A fast multi-resolution method for detection of significant spatial overdensities. Pittsburgh: School of Computer Science, 2003;CMU-CS-03-154.
10. Apache Software Foundation. Apache JMeter, version 1.9 [Software]. Forest Hill, MD: Apache Software Foundation, 2003. Available at http://jakarta.apache.org/jmeter.

National Bioterrorism Syndromic Surveillance Demonstration Program

W. Katherine Yih,[1] B. Caldwell,[2] R. Harmon,[3] K. Kleinman,[1] R. Lazarus,[4] A. Nelson,[5] J. Nordin,[5] B. Rehm,[6] B. Richter,[4] D. Ritzwoller,[7] E. Sherwood,[8,9] R. Platt[1,4]

[1]Harvard Medical School/Harvard Pilgrim Health Care, Boston, Massachusetts; [2]Division of Healthcare Quality Promotion, National Center for Infectious Diseases, CDC; [3]Optum/Ingenix, Eden Prairie, Minnesota; [4]Harvard Medical School, Boston, Massachusetts; [5]HealthPartners Research Foundation, Minneapolis, Minnesota; [6]America's Health Insurance Plans, Washington, D.C.; [7]Kaiser Permanente Colorado, Denver, Colorado; [8]Austin/Travis County Health and Human Services Department, Austin, Texas; [9]Williamson County and Cities Health District, Georgetown, Texas

Corresponding author: W. Katherine Yih, Department of Ambulatory Care and Prevention, Harvard Medical School/Harvard Pilgrim Health Care, 133 Brookline Ave., 6th Floor, Boston, MA 02215. E-mail: katherine_yih@harvardpilgrim.org.

Abstract

The National Bioterrorism Syndromic Surveillance Demonstration Program identifies new cases of illness from electronic ambulatory patient records. Its goals are to use data from health plans and practice groups to detect localized outbreaks and to facilitate rapid public health follow-up. Data are extracted nightly on patient encounters occurring during the previous 24 hours. Visits or calls with diagnostic codes corresponding to syndromes of interest are counted; repeat encounters are excluded. Daily counts of syndromes by zip code are sent to a central data repository, where they are statistically analyzed for unusual clustering by using a model-adjusted SaTScan™ approach. The results and raw data are displayed on a restricted website. Patient-level information stays at the originating health-care organization unless required by public health authorities. If a cluster surpasses a threshold of statistical aberration chosen by the corresponding public health department, an electronic alert can be sent to that department. The health department might then call a clinical responder, who has electronic access to records of cases contributing to clusters.

The system is flexible, allowing for changes in participating organizations, syndrome definitions, and alert thresholds. It is transparent to clinicians and has been accepted by the health-care organizations that provide the data. The system's data are usable by local and national health agencies. Its software is compatible with commonly used systems and software and is mostly open-source. Ongoing activities include evaluating the system's ability to detect naturally occurring outbreaks and simulated terrorism events, automating and testing alerts and response capability, and evaluating alternative data sources.

Introduction

The National Bioterrorism Syndromic Surveillance Demonstration Program covers a population of >20 million persons, monitoring and analyzing numbers of new cases of illness derived from electronic patient-encounter records from participating health-care organizations. It was created on the premise that early detection of acute illness in populations would be useful to public health and that primary care sites and nurse call centers might register the first evidence of such conditions.

This CDC-funded program grew out of collaborative projects between multiple health plans and their respective state health departments (*1–3*). It currently includes eight health-care organizations (Table). The coordinating center, referred to as the data center, is run by Harvard Medical School and Harvard Pilgrim Health Care. Elements of the program have been described elsewhere (*4,5*).

Objectives

The program's primary goal is to create a flexible, open-source surveillance system that uses ambulatory care data to identify unusual clusters of illness and support rapid public health follow-up. Secondary goals are to 1) reduce barriers to private health-care organizations' voluntary participation, 2) develop and test optimal signal-detection methods, and 3) develop communication and response methods that enable local public health agencies to obtain detailed clinical information about cases that are part of clusters.

System Operation

Data Sources and Processing at Data-Providing Sites

Data on patient encounters (visits or calls), including demographic information and diagnostic codes, are recorded electronically at each health-care organization as part of rou-

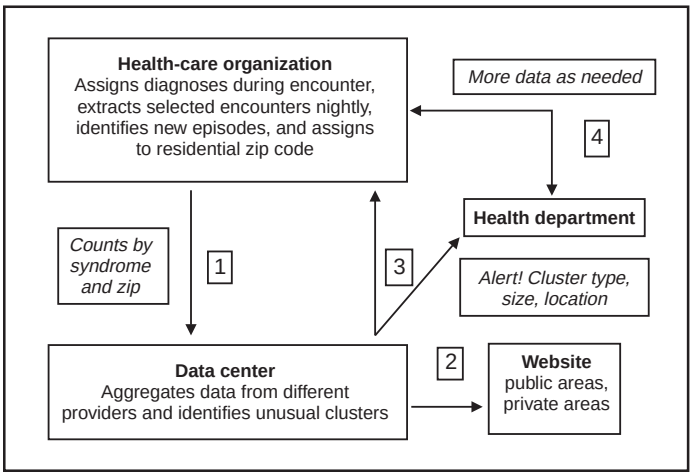
TABLE. Participating health-care organizations and populations served by the National Bioterrorism Syndromic Surveillance Demonstration Program

Health-care organization	Type of organization	Patient encounter types captured	Metropolitan area covered	Population served	Proportion of catchment area's population included
Optum	Nurse telephone triage and health information services	Calls to nurse call centers	Multiple	22,000,000	7% of U.S. population, unevenly distributed
Harvard Pilgrim Health Care and Harvard Vanguard Medical Associates	Health plan	Ambulatory visits and telephone calls	Boston, Massachusetts	140,000	6%
Health Partners Research Foundation	Health plan	Ambulatory visits	Minneapolis–St. Paul, Minnesota	240,000	8%
Kaiser Permanente Colorado	Health plan	Ambulatory visits	Denver, Colorado	380,000	15%
Scott and White Healthcare System, Austin Regional Clinic, and Austin Diagnostic Clinic	Physician organizations	Ambulatory visits	Austin, Texas	384,000	26%–30%
America's Health Insurance Plans	National trade association of companies providing health insurance to >200 million persons	Not applicable (N/A)	N/A	N/A	N/A

tine patient care, usually on the same day as the visit or call (Figure). Each night, patient encounters with codes of interest are extracted automatically from clinical data systems. The extracted encounter files are created to uniform specifications and are kept on a directory accessible to software (the console) provided by the data center.

The console maps patient encounters to syndromes (e.g., respiratory) defined by a CDC-led working group (*6*) and then identifies illness episodes by omitting patient encounters in any syndrome that occurred within 42 days of an earlier visit in the same syndrome. Episodes are mapped to patients' residential zip codes, and a single file is created containing counts of new episodes in each syndrome and zip code for each day. In addition, historic episode files are created and provide a basis for modeling. Transmission of count data to the data center in extensible markup language (XML) format is safeguarded by means of electronic security certificates and encryption. During the processing of encounter files into episode files, the console produces encounter lists containing demographic and clinical information that remain at the originating site, where they are available in the event of a query from public health authorities.

FIGURE. Information flow for the National Bioterrorism Syndromic Surveillance Demonstration Program



Statistical Analysis

For each syndrome and clinical site, daily counts are modeled over a multiyear period, and clusters are evaluated by using a model-adjusted SaTScan™ approach, which scans multiple contiguous zip codes over a specified number of consecutive days of surveillance (*7,8*). SaTScan is adjusted by using generalized linear mixed models that take into account day of the week, holidays, seasons, secular trends, and the unique characteristics of each zip code area, based upon historic data (*9*).

The recurrence interval (i.e., the number of days between predicted occurrences by chance alone within each organization's catchment area) is used to characterize the degree of statistical aberration of any cluster in the contemporary daily episode data. It is the inverse of the cluster's p-value.

Thus, the larger the value of the measure, the rarer (and possibly more worthy of investigation) the cluster is.

Data Display, Alerts, and Response

Almost immediately upon receipt, raw data and modeled results are displayed in table, graph, and map form on a restricted website designed and administered by the data center. If a signal exceeds the threshold of statistical aberration specified by the public health department in whose jurisdiction it occurs, the data center will automatically send an electronic alert to designated persons at the health department. This system is being implemented first in Massachusetts, using the state's electronic health alert network. If contacted by the health department, the clinical organization's responder can provide detailed clinical information about persons in the cluster.

System Experience

Validity for Detection of Naturally Occurring Outbreaks

In November 2003, the system detected unusual respiratory illness clusters in Colorado, Texas, and Massachusetts heralding early severe influenza outbreaks, at least in Colorado. An evaluation is being conducted of the system's ability to detect naturally occurring outbreaks of gastrointestinal illness on the basis of known outbreaks identified by the Minnesota health department.

Data Quality Potentially Affecting Validity

The proportion of the population covered by the surveillance system for different metropolitan areas is provided (Table). Persons without health insurance are not represented. Historic comparisons and simulations are being conducted to assess the minimum proportion of an area's population needed by the surveillance system to detect outbreaks of different types and sizes.

Usefulness

The system's performance in apprehending the 2003 influenza outbreak in Colorado and clusters of gastrointestinal illness in Minnesota is being evaluated. Extensive simulation is also being conducted to describe sensitivity to potential acts of biologic terrorism. Usefulness in practice will be assessed systematically in collaboration with health departments after the alerting system has operated for 1 year in at least one state.

Flexibility

The system is highly adaptable. Alert thresholds can be set at any degree of statistical aberration, can be different for different syndromes and in different locales, and can be changed. Different statistical methods can be applied to the counts by date, syndrome, and zip code. With the consent of the organizations that hold the data, new syndromes categories can easily be created, and customized queries of the originally extracted encounters (encompassing approximately 700 *International Classification of Diseases, Ninth Revision* [ICD-9] codes) are feasible.

Acceptability and Cost

The system entails no extra work for clinicians. Because patient-level data stay with the organization and are shared only when a public health need exists, the system's distributed-data model has been accepted by participating health-care organizations. Health plans consider the aggregate data to be either de-identified or limited data sets as defined by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule. Additionally, they consider this aggregated-data model to allow them greater control over their proprietary information.

Resources needed by clinical organizations include a networked Microsoft Windows® personal computer (or comparable) with Internet access, system administrator effort to create the routine data extract from host systems and to maintain connectivity, project programmer effort to install and run programs, administrative effort to review and approve new software updates before they are installed on local computers and to develop communication and response protocols with health agencies, and clinical responder training and availability. Because organizations' cost structures vary widely, predicting actual costs is difficult.

Openness, Compatibility, and Portability

The program is designed to be open, maximally compatible with elements of commonly used surveillance systems, and easy for additional health-care organizations to join. Syndrome definitions of the CDC-led working group (*6*) and open-source software and development are used wherever possible, and all protocols and computer code are available to other investigators and public health agencies.

Health-care organizations use software provided by the project (written in Python [http://www.python.org]) that can run on the majority of common operating systems, including Windows,® Macintosh,® and Linux,® to process their own

data for transmission to the data center. Uniform file specifications and console-based uploading allow the system to work at virtually any site where diagnostic codes are available electronically on the day of encounter.

Data files created by this system are also directly usable by health departments and are compatible with the emerging standards of CDC's BioSense initiative (*10*). This allows healthcare organizations to make their data directly available to local and national health agencies if they so choose.

Acknowledgments

This work has been supported by CDC cooperative agreement UR8/CCU115079; Massachusetts Department of Public Health contracts 5225 4 160002, 5223 3 160001, and 5225 3 337HAR; Minnesota Department of Health contract A57182; and a grant awarded by the Texas Association of Local Health Officials. The Colorado Department of Public Health and Environment, Austin/Travis County Health and Human Services Department, and Williamson County and Cities Health District have provided in-kind support.

References

1. Lazarus R, Kleinman K, Dashevsky I, DeMaria A, Platt R. Using automated medical records for rapid identification of illness syndromes: the example of lower respiratory infection. BMC Public Health 2001;1:1–9.
2. Lazarus R, Kleinman K, Dashevsky I, et al. Use of automated ambulatory care encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis 2002;8:753–60.
3. Martinez B. Questions of security: HealthPartners use reach, speedy data to hold watch for bioterrorism attacks. Wall Street Journal, Nov. 1, 2001:A10.
4. Platt R, Bocchino C, Caldwell B, et al. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. J Urban Health 2003;80(2 Suppl 1):i25–31.
5. Platt R. Homeland security: disease surveillance systems. Testimony submitted to the US House of Representatives Select Committee on Homeland Security. September 24, 2003. Available at https://btsurveillance.org/btpublic/publications/house_testimony.pdf.
6. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.bt.cdc.gov/surveillance/syndromedef/index.asp.
7. Kulldorff M. Prospective time periodic geographic disease surveillance using a scan statistic. J Royal Stat Soc A 2001;164:61–72.
8. Kulldorff M and Information Management Services, Inc. SaTScan™ version 4.0: software for the spatial and space-time scan statistics, 2004. Available at http://www.satscan.org.
9. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism (with invited commentary). Am J Epidemiol 2004;159:217–24.
10. CDC. BioSense: PHIN's early event detection component. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/phin/components/index.htm.

Daily Emergency Department Surveillance System — Bergen County, New Jersey

Marc Paladini
Bergen County Department of Health Services, Paramus, New Jersey

Corresponding author: Marc Paladini, New York City Department of Health and Mental Hygiene, 125 Worth St., New York, NY 10013. Telephone: 212-788-4320; Fax: 212-788-5470; E-mail: mpaladin@health.nyc.gov.

Abstract

The purpose of the Daily Emergency Department Surveillance System (DEDSS) is to provide consistent, timely, and robust data that can be used to guide public health activities in Bergen County, New Jersey. DEDSS collects data on all emergency department visits in four hospitals in Bergen County and analyzes them for aberrant patterns of disease or single instances of certain diseases or syndromes. The system monitors for clusters of patients with syndromes consistent with the prodrome of a terrorism-related illness (e.g., anthrax or smallpox) or naturally occurring disease (e.g., pandemic influenza or food and waterborne outbreaks). The health department can use these data to track and characterize the temporal and geographic spread of a known outbreak or demonstrate the absence of cases during the same period (e.g., severe acute respiratory syndrome [SARS] or anthrax). DEDSS was designed to be flexible and readily adaptable as local, state, or federal surveillance needs evolve.

Introduction

In 2001, the Bergen County Department of Health Services instituted a countywide syndromic surveillance system that uses hospital emergency department (ED) data. Located in northeast New Jersey across the Hudson River from New York City, Bergen County has a population of approximately 884,000 persons (U.S. Census 2000) living within 234 square miles.

The first step in creating the Daily Emergency Department Surveillance System (DEDSS) was to identify the appropriate stakeholders. Within the health department, the creative team consisted of an epidemiologist, an information technology (IT) professional, and the director of planning. Next, immediate external stakeholders, including the infection-control practitioner (ICP), the ED director, the hospital IT professional, and the hospital director of security, were brought into the discussion. After the system was developed, local health officers, health department nurses, and state and regional health department epidemiologists were updated on its progress.

System Operation

Four of six Bergen County hospitals provide daily data to DEDSS, representing 85% of all daily ED visits. Early each morning, the hospital's computer system generates a text file containing the following fields for each person who visited the ED the previous day: date of visit, residential zip code, age, chief complaint, and admission status. The file, abstracted from the hospital's database, uses data produced during normal, clinical ED workflow. The text file is then automatically sent to a password-protected file transfer protocol (FTP) server, where it is stored. The size of each file differs, ranging from a four-hospital total of 400 to 600 visits/day. At 8:00 a.m. each morning, the epidemiologist's computer automatically starts DEDSS. The program connects to the FTP site and downloads, formats, integrates, and analyzes the data. DEDSS then creates standardized reports and e-mails them to the epidemiologist along with an alert to his cellular telephone indicating the system ran successfully. The epidemiologist can then access the reports remotely and determine any needed follow-up.

Data are analyzed daily by using a modified version of the cumulative sum statistic (*1*) programmed in SAS® (*2*). For each syndrome in each hospital, a ratio is calculated by dividing the number of visits caused by the syndrome by the total number of ED visits. This ratio is then compared with the mean of an 11-day moving baseline that precedes the day of interest. The first 3 days before the current observation are ignored to act as a buffer for an outbreak that might grow slowly over 1–2 days, and the mean is tabulated for days 4–14 before the day of interest. Because the data are not transformed and any signals that might arise remain in the data set, the health department uses both a buffer and an 11-day moving average to offset the effects that days of increased activity would have on the analysis.

If an observation is higher than expected, on the basis of the moving average plus 3 standard deviations, a signal is created and two reports are generated. The first report includes the syndrome signaled, hospital (if the signal has occurred at a single hospital) or county (if the signal has occurred at ≥2

hospitals), date, total number of visits, total number in the syndrome, ratio for that day, and baseline ratio with which it was compared. For each signal, a corresponding report is generated that features a line listing of all persons who were part of the signal.

The first step, as in any outbreak investigation, is to verify the diagnosis. Because using text strings to identify affected patients can result in inclusion of patients who do not have the chief complaints of interest (e.g., *no fever* instead of *fever*), the chief-complaint field for each member of the line listing is examined. This field contains a mixture of triage information, clinical diagnoses, and patient statements. For example, a case of viral respiratory disease (e.g., influenza) might be coded as *fever and cough*, *viral syndrome*, or *I don't feel well*, depending on the hospital. After an investigation determines the system properly identified appropriate chief complaints and all of the observations appear to be valid, a level of concern is assigned.

Three levels of concern can be assigned to signals, *low*, *moderate,* or *elevated*, each with corresponding steps. The epidemiologist assigns the level after reviewing each day's report, which usually takes <10 minutes. If a signal is attributable to low numbers (<10), is just above the baseline, is attributable to seasonality (e.g., pneumonia in winter), and exhibits no obvious epidemiologic links (e.g., age or zip code), then the signal level assigned is *low,* and no action is taken.

A level of *moderate* is assigned if multiple signals occur on the same day in different hospitals; if two, consecutive, low-level signals occur in the same hospital; if a low-level signal arises with possible epidemiologic links (e.g., geographic clustering); or if the signal is substantially but not exceptionally higher than the baseline (on the basis of experience rather than statistics, until an algorithm is developed to quantify this). Response to a moderate signal includes e-mail notification of possible activity to hospital ICPs and epidemiologists in surrounding counties. Those epidemiologists and ICPs then decide whether to investigate their jurisdiction's conditions.

If a signal is exceptionally higher than the baseline (on the basis of experience rather than statistics) or if moderate signals occur at more than one hospital on a given day, a signal level of *elevated* is assigned. An elevated signal entails immediate notification of hospital ICPs, internal chain of command, regional epidemiologists, and state health department officials that further investigation is warranted. Status of hospitals involved in an elevated-level signal is determined through phone consultation, and if disease activity remains high, an epidemiologic investigation is initiated. Depending on the number of persons and hospitals involved, either the epidemiologist or the epidemiologic response team are sent to the hospital to review charts, interview patients, and confer with hospital personnel regarding next steps.

System Experience

Although the burden to Bergen County has been minimal, the system's cost and maintenance requirements need to be better quantified, both in terms of resources spent and person-hours used to respond to system alerts. Furthermore, the better the system operators (e.g., epidemiologists and IT personnel) understand hospitals' coding and triage practices, the better they will understand the system's output and be able to alter it as needed. To date, no elevated signals have occurred. Moderate signals have occurred but none that required more than a telephone consultation with hospital ICPs. In all cases, the numbers decreased substantially after 1 day, and no specimens were collected by hospital physicians.

DEDSS monitors two primary syndromes: influenza-like illness (ILI) and gastrointestinal illness (GI). Each syndrome has a corresponding case definition, complaint group (i.e., a list of chief complaints being monitored), and diagnostic group (i.e., a list of *International Classification of Diseases, Ninth Revision* [ICD-9] codes for validation studies). Preliminary comparisons of chief complaint to ICD-9-coded diagnoses indicate sensitivity of 76%, specificity of 96%, and positive predictive value of 53% for ILI and sensitivity of 61%, specificity of 97%, and positive predictive value of 32% for GI. Specific results need to be analyzed further to identify and quantify the source of noise and discrepancies within the syndrome definitions, especially when examining positive predictive value.

As the system is fine-tuned and case definitions and complaint groups revised, the epidemiologist can easily change the coding as needed. The system's malleability enables the health department to monitor seasonal or short-term disease-activity trends. During a crisis, the epidemiologist can request that hospitals place a keyword in the complaint field for all visits relating to a certain event (e.g., alleged anthrax exposures) to monitor visits more precisely.

DEDSS is designed to accommodate inclusion of new fields when necessary. If the system were also able to link the clinical aspects of a patient's visit (e.g., X-ray results, medications prescribed, laboratory results, or blood work) to each observation, the epidemiologist reviewing the day's data would have more information to examine when assigning the level of concern. Because the infrastructure is already in place, establishing future projects that capture different data will be even easier.

Obstacles and Benefits

The primary obstacles encountered during development and maintenance of DEDSS involve IT and resources. The ability to troubleshoot technical and programmatic computer prob-

lems has been limited by departmental resources. Although the system is intended to be automated and electronic, certain hospitals had difficulty scheduling tasks and transferring the files. Fortunately, the fundamental act of creating the daily data file was not a problem for any hospitals. However, because hospital IT personnel are instrumental to the mechanics of file creation, automation, and transfer, including them in early planning is essential.

After establishing standard analytic methods and reporting protocols within a jurisdiction, the next step is to coordinate surveillance systems within the region; as multiple systems come online, maintaining communication and methodologic developments in real time is crucial. Conducting surveillance and validation regionally would enable joining of resources to accomplish similar goals.

Beyond DEDSS' stated goals, the system has had additional benefits. The process of meeting with the hospital personnel and setting up the data transfer generated excellent working relations between the health department and the hospitals. It increased the timeliness of reporting routine incidents and fostered communication around unusual occurrences. Furthermore, an infrastructure supporting the electronic transfer of data between hospitals and the health department is now in place. Unfortunately, redundant capabilities are not yet built into the system; currently, when one aspect of the system fails, the entire system goes offline. The system also lacks a single, dedicated manager. These limitations can result in periods of system inactivity.

The health department hopes the system will be useful for more than terrorism-preparedness purposes. Its goal is to have a multifaceted system that uses multiple analytic processes and creates reports for multiple users on different aspects of public health and health-care delivery.

References

1. Hutwagner, LC, Maloney EK, Bean NH et al. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. Emerg Infect Dis 1997;3:395–400.
2. SAS Institute. SAS,® Version 8.2 [Software]. Cary, NC: SAS Institute.

Hospital Admissions Syndromic Surveillance — Connecticut, September 2001–November 2003

Zygmunt F. Dembek, K. Carley, A. Siniscalchi, J. Hadler
Connecticut Department of Public Health, Hartford, Connecticut

Corresponding author: Zygmunt F. Dembek, Epidemiology Program, Connecticut Department of Public Health, 410 Capitol Ave., MS # 11EPI, P.O. Box 340308, Hartford, CT 06134-0308. Telephone: 860-509-7994; Fax: 860-509-7910; E-mail: zygmunt.dembek@po.state.ct.us.

Abstract

On September 11, 2001, the Connecticut Department of Public Health (CDPH) initiated daily, statewide syndromic surveillance based on unscheduled hospital admissions (HASS). The system's objectives were to monitor for outbreaks caused by Category A biologic agents and evaluate limits in space and time of identified outbreaks. Thirty-two acute-care hospitals were required to report their previous day's unscheduled admissions for 11 syndromes (pneumonia, hemoptysis, respiratory distress, acute neurologic illness, nontraumatic paralysis, sepsis and nontraumatic shock, fever with rash, fever of unknown cause, acute gastrointestinal illness, and possible cutaneous anthrax, and suspected illness clusters). Admissions for pneumonia, gastrointestinal illness, and sepsis were reported most frequently; admissions for fever with rash, possible cutaneous anthrax, and hemoptysis were rare. A method for determining the difference between random and systemic variation was used to identify differences of ≥3 standard deviations for each syndrome from a 6-month moving average. HASS was adapted to meet changing surveillance needs (e.g., surveillance for anthrax, smallpox, and severe acute respiratory syndrome). HASS was sensitive enough to reflect annual increases in hospital-admission rates for pneumonia during the influenza season and to confirm an outbreak of gastrointestinal illness. Follow-up of HASS neurologic-admissions reports has led to diagnosis of West Nile virus encephalitis cases. Report validation, syndrome-criteria standardization among hospitals, and expanded use of outbreak-detection algorithms will enhance the system's usefulness.

Introduction

On September 11, 2001, the Connecticut Department of Public Health (CDPH) developed and initiated a syndromic surveillance system based on unscheduled hospital admissions called HASS. The system's initial objective was to monitor for a concurrent terrorist event caused by Category A biologic agents (*1,2*). All hospitals were required to submit standardized reports to CDPH regarding the number of patients admitted the previous day with acute respiratory or neurologic problems and perceived illness clusters among newly admitted patients. Another objective was to evaluate the spatio-temporal limits of identified outbreaks and other public health threats.

Methods

HASS Description

All 32 acute-care hospitals within Connecticut are required to report to CDPH on a standardized form the number of unscheduled admissions from the previous day. Reporting is required for 11 syndromic categories, including pneumonia, with a subcategory for health-care workers with clinical responsibilities; hemoptysis; respiratory distress syndrome, with a subcategory for health-care workers with clinical responsibilities; acute neurologic illness, including meningitis, encephalitis, or unexplained acute encephalopathy; nontraumatic paralysis; nontraumatic shock, including sepsis; fever with rash; fever of unknown cause; acute gastrointestinal illness, including vomiting, diarrhea, or dehydration; skin infection indicating possible anthrax; and apparent illness clusters.

HASS has been modified to meet changing disease surveillance needs as follows:

- On the basis of feedback from hospitals in October 2001, better-differentiated syndrome categories were created (e.g., pneumonia, hemoptysis, and respiratory-distress categories replaced a total-respiratory category).
- Reporting categories for gram-positive rod isolates and radiographic findings consistent with inhalational anthrax were added for 1 month in late November 2001 after a case of inhalational anthrax was identified.
- Case follow-up was instituted for all reports of fever with rash illness beginning November 2002 to enhance smallpox surveillance.
- Subcategories for health-care workers with clinical responsibilities were added in May 2003 in response to severe acute respiratory syndrome (SARS).

Initially, hospitals reported to CDPH by fax or e-mail. In May 2003, a secure website with the report form was inaugurated. Since October 2003, all 32 acute-care hospitals have reported their data by using the secure website. Each hospital has access to its data on the website.

CDPH investigates all detected or reported disease clusters and all cases of selected syndromes. Case follow-up is routine for the following syndromic categories:

- pneumonia in clinical health-care workers (potential SARS cases);
- acute respiratory distress or respiratory failure in clinical health-care workers (potential SARS cases); and
- fever and rash illness (potential smallpox cases).

Analysis

The HASS data set is transformed into an Excel™ spreadsheet and analyzed with SAS® for Windows™ Version 8e (*3*) by using the Shewhart method (*4*) of analysis to determine the difference between random and systemic variation. A 3-standard–deviation difference (i.e., statistically significant) is calculated by using a 6-month moving average of all data collected. This analysis is performed for each syndrome for all hospitals combined and for hospitals in each of the three largest of Connecticut's eight counties (Figure).

A CDPH epidemiologist inspects data daily. Analysis is conducted weekly, whenever a peak in rates is noted, whenever disease-surveillance questions occur (e.g., when an outbreak is detected through routine reporting mechanisms), after unusual events (e.g., the August 2003 electrical blackout), or when determining whether influenza activity has increased.

System Experience

During August 2002–July 2003, unscheduled admissions were reported most frequently for pneumonia (an average of 82.7 admissions/million population/week), followed by acute gastrointestinal illness (26.0), sepsis and nontraumatic shock (16.8), fever of unknown origin (13.1), and respiratory distress (11.3) (Table). Syndromes reported least commonly were disease clusters (0.006), possible cutaneous anthrax (0.1), fever and rash illness (0.4), and hemoptysis (1.2). No significant difference was found by day of the week for admission rates for the most frequent syndromic categories.

During August 2002–July 2003, a total of 59 spikes in activity >3σ were noted. All spikes were detected from county-specific (not statewide) analysis. By syndrome category, all were either gastrointestinal (35) or pneumonia (24). All gastrointestinal spikes were limited, single-day events. With one exception (a January 2002 1-day spike correlated with a norovirus outbreak affecting 116 persons), no spikes were associated with known gastrointestinal outbreaks. Spikes in pneumonia clustered during the winter months and were likely caused by seasonal influenza.

During November 2002–November 2003, a total of 58 cases of fever and rash illness were reported and subsequently investigated to rule out smallpox. These cases had other diagnoses, including viral syndrome, serum sickness, meningococcemia, pustular psoriasis, uticaria, toxic shock syndrome,

FIGURE. Information flow for Connecticut's hospital admissions syndromic surveillance (HASS) system

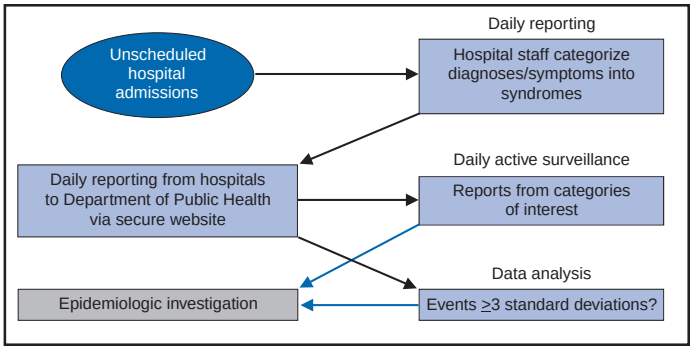


TABLE. Average number and range of unscheduled hospital admissions per week, by syndrome — Connecticut acute-care hospitals, August 2002–July 2003

Syndrome	Average no. of admissions*	Range of admissions*
Pneumonia	82.7	24–122
Hemoptysis	1.2	0–5
Acute respiratory distress syndrome or respiratory failure of unknown origin	11.3	4–19
Meningitis, encephalitis, or unexplained acute encephalopathy	1.9	0.3–5
Nontraumatic paralysis, Guillian-Barré syndrome, or descending paralysis	2.0	1–4
Sepsis and nontraumatic shock	16.8	6–24
Fever and rash illness	0.4	0–2
Fever of unknown origin	13.1	8–18
Gastrointestinal illness, vomiting, diarrhea, dehydration	26.0	7–39
Skin infection; possible cutaneous anthrax	0.1	0–1
Clusters of illness	0.006	0–1

* In millions.

scarlet fever, tickborne disease, staphylococcal infection, parvovirus, human immunodeficiency virus infection, and chickenpox.

During May–November 2003, two cases of acute respiratory distress and nine cases of pneumonia among health-care workers with clinical responsibilities were reported to HASS and investigated. None met then-current CDC or World Health Organization criteria for suspected SARS (*5*).

Individual hospitals have reported four illness clusters since HASS's inception, all gastrointestinal illness of unknown etiology. Laboratory-based surveillance detected 15 different gastrointestinal illness clusters. No increase in gastrointestinal illness was observed in Connecticut hospitals serving those areas affected by the Northeast power blackout during or after August 14–15, 2003.

A health director who regularly monitored HASS data in his municipality discovered the first two of Connecticut's 17 confirmed human West Nile virus cases during 2002. He requested investigation of two late-summer neurologic syndrome reports. Both patients had encephalitis and subsequently tested positive for West Nile virus infection.

Discussion

CDPH chose to design and implement a system based on hospital admissions for multiple reasons. Hospital admissions measure severe illness; the biologic agents of greatest concern (Category A) all cause illness severe enough to require hospitalization. Obtaining additional clinical, follow-up, and laboratory information on these patients is possible because they are hospitalized in a known place and are monitored. HASS is easy and inexpensive to implement and modify, requiring no special computer equipment or programming and 10–15 minutes/day for most hospitals to review the previous day's admissions and prepare and submit data. It can be readily implemented statewide, a desirable feature in a state with discrete population centers (compared with a densely populated area such as New York City). HASS requires someone at each hospital to be aware of admission patterns, increasing the potential to recognize and report unusual events. Finally, unlike systems based on outpatient visits, HASS enables detection and investigation of outbreaks as limited as a single case (e.g., smallpox or SARS).

Baseline information is now available on the frequency of admissions for a range of syndromes. The system is sensitive enough to reflect important community events (e.g., concurrent increases in a monitored syndrome in city or county hospitals). A sizable laboratory-reported gastrointestinal outbreak was also evident with HASS. Admission rates for pneumonia have been observed to vary by season and increase markedly during an active influenza season. CDPH has increased confidence that HASS can be used for statewide surveillance and to monitor an outbreak that results in hospitalizations.

HASS has been used successfully to identify and rapidly investigate individual cases of relatively unusual syndromes (e.g., the detection of two cases of West Nile virus encephalitis by following up on reports of encephalitis in one hospital). Investigation of cases of fever with rash has identified chickenpox. Continued investigation of admissions for fever and rash illness is a reasonable way to conduct enhanced smallpox surveillance.

HASS has important limitations. First, it is insensitive to slight changes in the syndromes most frequently reported (i.e., pneumonia, gastrointestinal illness, and sepsis). Second, HASS has yet to detect an outbreak not also detected by other means. Third, it is insensitive to outbreaks that primarily produce outpatient illness. Fourth, because it depends on patient admissions, identifying an outbreak with a time lag between symptom onset and admission (e.g., anthrax) can be delayed by 1–2 days when compared with an outpatient syndromic surveillance system. Finally, because HASS obtains only case counts rather than individual demographic data, increases in illness among a demographic subset of the population (e.g., children or women) cannot be detected without obtaining additional information.

References

1. CDC. Bioterrorism agents/diseases. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.bt.cdc.gov/agent/agentlist-category.asp.
2. Franz DR, Jahrling PB, McClain DJ, et al. Clinical recognition and management of patients exposed to biological warfare agents. Clin Lab Med 2001;21:435–73.
3. SAS Institute, Inc. SAS Online Doc,® version 8. Cary, NC: SAS Institute, Inc., 1999.
4. Montgomery, DC. Introduction to statistical quality control. 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1996.
5. CDC. Updated interim US case definition for severe acute respiratory syndrome. Atlanta, GA: US Department of Health and Human Services, CDC, 2004. Available at http://www.cdc.gov/ncidod/sars/casedefinition.htm.

BioSense — A National Initiative for Early Detection and Quantification of Public Health Emergencies

John W. Loonsk
Information Resources Management Office, CDC

Corresponding author: John W. Loonsk, Associate Director for Informatics, Information Resources Management Office, CDC, 1600 Clifton Rd, Mailstop D-45, Atlanta, GA 30333. Telephone: 404 639-7600; Fax: 404 639-7711; E-mail: jloonsk@cdc.gov.

Abstract

BioSense is a national initiative to enhance the nation's capability to rapidly detect, quantify, and localize public health emergencies, particularly biologic terrorism, by accessing and analyzing diagnostic and prediagnostic health data. BioSense will establish near real-time electronic transmission of data to local, state, and federal public health agencies from national, regional, and local health data sources (e.g., clinical laboratories, hospital systems, ambulatory care sites, health plans, U.S. Department of Defense and Veterans Administration medical treatment facilities, and pharmacy chains).

Introduction

BioSense is a national initiative to support the advancement of early detection capabilities by promoting greater and timelier acquisition of relevant data and by advancing technologies associated with near real-time reporting, automated outbreak identification, and analytics. It is one of three initiatives recently advanced by the President of the United States to improve national preparedness; others include BioShield, which focuses on rapid development of vaccines and therapeutics, and BioWatch, which places environmental air samplers in key locations.

To enhance consistency of public health surveillance nationally, BioSense will facilitate the sharing of automated detection and visualization algorithms and approaches by promoting national standards and specifications developed by such initiatives as the Public Health Information Network (PHIN) (*1*) and the eGov activities of Consolidated Health Informatics (*2*). Finally, the initiative will encourage integration of early detection systems with outbreak management and response systems. Because the benefits of early detection emanate from early response, standards for early detection systems will help them share data and integrate with information systems that support the management of possible and confirmed cases, laboratory results, isolation, prophylaxis, and vaccination.

BioSense is a component of PHIN, which seeks to use industry data and technical standards to develop specifications and software elements, allowing for a national electronic network to support public health needs. In addition to inclusion of functional and technical specifications for early event detection, PHIN also provides routine public health surveillance (e.g., the National Electronic Disease Surveillance System [NEDSS]), secure communications, analysis and visualization, information dissemination and knowledge management, health alerting, outbreak management, laboratory information systems, and vaccine and prophylaxis administration (Figure 1).

BioSense will include an Internet-based software-system implementation to enable public health officials in major cities to view data for their communities (Figure 2). The software system will implement identified industry standards and provide a platform for integrating and evaluating different outbreak-detection approaches. The BioSense software system includes both spatio-temporal and temporal analysis algorithms and approaches to visualizing unusual events in data (Figure 2). Phase I of the BioSense system is operating in >20 cities nationally.

Supporting Early Event Detection

Discussion around early event detection over recent years has focused on the relative value of data sources that are prediagnostic or syndromic in nature. BioSense seeks to advance public health capabilities for both prediagnostic and diagnostic data sources in near real time. Given the ongoing controversy about prediagnostic surveillance, BioSense will support rigorous evaluation of these data sources. Where available, BioSense will prioritize early detection data on the basis of diagnostic skills of clinical personnel. Frequently, tension exists between getting data early and having them be inclusive of clinical judgment, but progress can be made in advancing real-time reporting of diagnostic and prediagnostic data that emanate from settings in which an experienced medical professional originates the data.

At the same time, BioSense will seek to minimize reporting burden by extracting early detection information from data sources that exist for purposes other than public health

FIGURE 1. Public Health Information Network (PHIN) component functions and initiatives

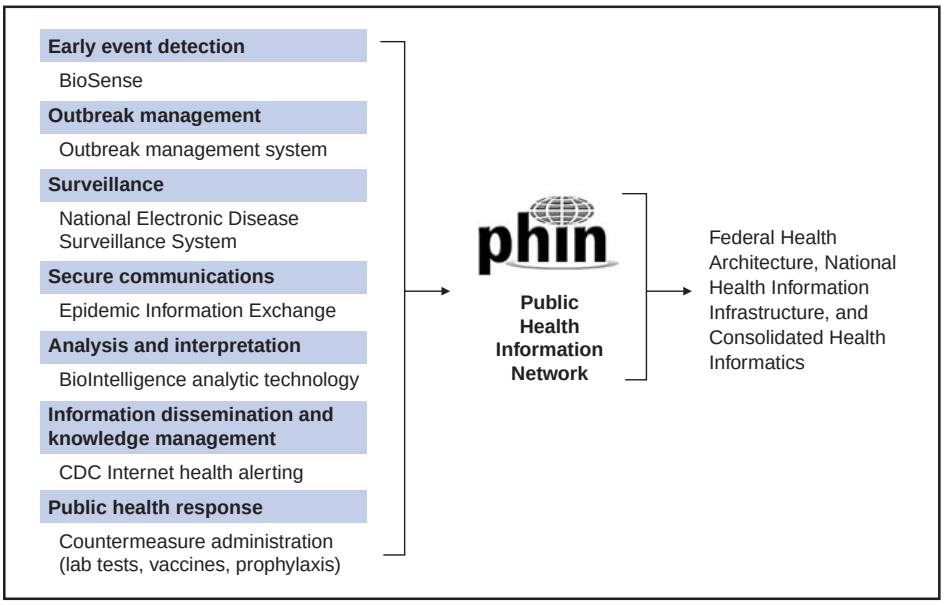
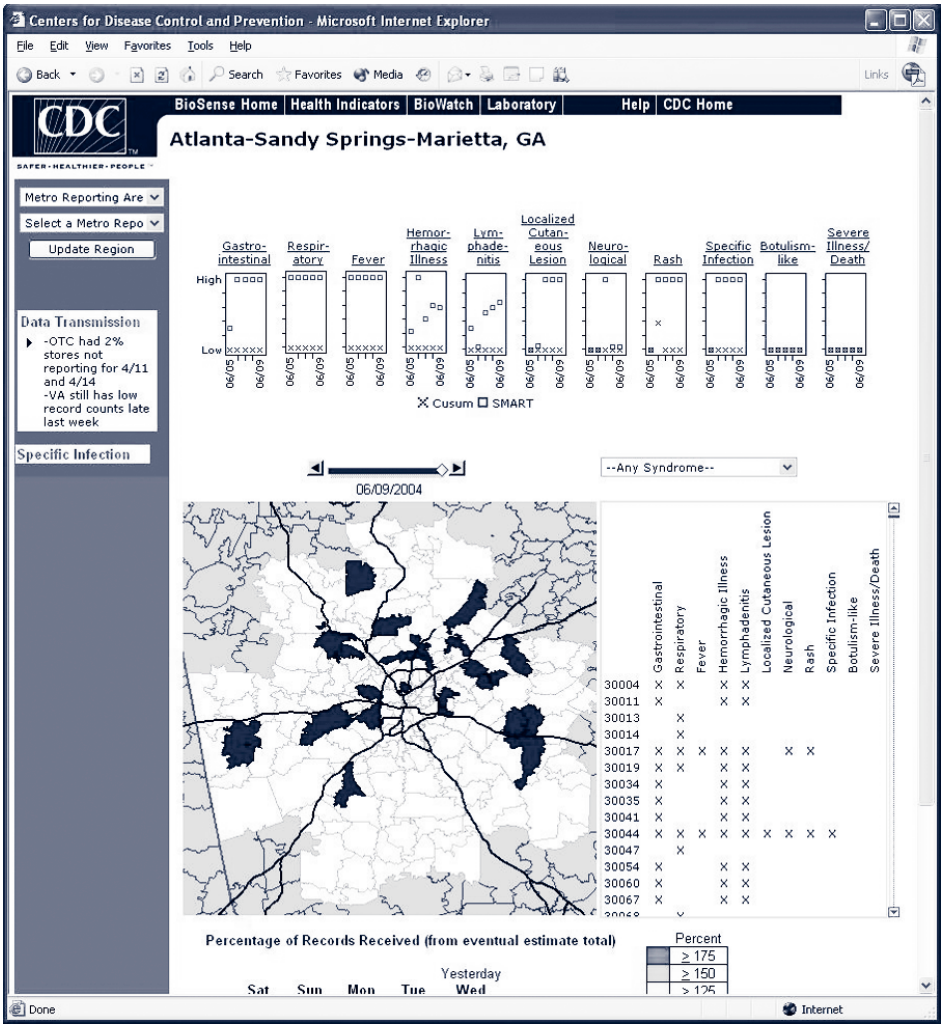


FIGURE 2. Demonstration data displayed on the BioSense software system



reporting. For example, it will use clinical-care information-system data rather than asking medical personnel to enter data manually for the sole purpose of detecting an outbreak.

BioSense will support early event-detection capabilities at the local, state, and national levels. Because routine public health reporting systems are inconsistent across the United States, early detection is usually implemented, if at all, at only one of these levels for any given area. To maximize national ability to detect and manage events early, to leverage expertise at local, state, and national levels, and to take advantage of data sources that are aggregated locally, regionally, and nationally, capabilities need to be advanced at all three levels and data need to flow rapidly and easily among them.

Guiding Investigation Decisions

Consequence management is a key concern for public health, and although electronic detection systems might be useful in assisting public health professionals, they can also create a tremendous burden. BioSense seeks to address these concerns in the short term by avoiding the forced consequence management of predetermined alerts. Instead of necessitating a series of responses to an alert that is identifying only a possible occurrence, BioSense seeks to capitalize on the analytic capabilities of public health professionals, including their abilities to compare and interpret multiple data sources and determine the likelihood of an event. It should also enable them to create and manage thresholds and circumstances for alerting to avoid forced consequence management. To support these capabilities of public health professionals, BioSense should coordinate viewing of multiple data sources and leverage these sources into greater sensitivity and greater specificity.

Addressing Privacy Concerns

Although prediagnostic data sources remain to be rigorously evaluated, researchers using such data sources should anticipate concerns about privacy from the public. To address such concerns, BioSense promotes the use of data that do not contain direct patient identifiers, even though public health authorities are eligible under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) to receive identified data under certain circumstances. All BioSense data will be securely managed for access by authorized public health professionals with appropriate jurisdictional access controls, and data providers will retain any directly identifiable information. An anonymous data linker will enable an authorized public health investigation in the event of a potential outbreak.

Supporting Public Health Needs

Finally, BioSense seeks to pursue early detection in the context of the multiple needs of public health. Initial detection of an event by identifying patterns of health-seeking behavior should be followed by case identification and quantification of the number, locations, and density of cases. Identifying a possible outbreak requires investigating symptoms across multiple cases, travel history, and possible environmental exposures, and then tracing contacts relative to people and disease vectors. These capabilities should be integrated with early detection systems and with systems for isolation, prophylaxis, accelerated vaccination, and adverse-event follow-up and management.

Conclusion

The initial focus of BioSense has been to advance early detection and management technologies and capabilities in a way that considers public health needs and ongoing efforts to use and evaluate early detection technology and data sources. It intends to support this work at national, regional, and local levels and provide a test bed for further evaluation and implementation.

References

1. CDC. Public Health Information Network (PHIN). Atlanta, GA: US Department of Health and Human Services, CDC, 2004. Available at http://www.cdc.gov/phin/.
2. eGov. Government to business portfolio: consolidated health informatics. Washington, DC: eGov, 2003. Available at http://www.whitehouse.gov/omb/egov/gtob/health_informatics.htm.

Syndromic Surveillance at Hospital Emergency Departments — Southeastern Virginia

Christine M. Yuan,[1,3] S. Love,[2] M. Wilson[1]
[1]*University of Michigan, Ann Arbor, Michigan;* [2]*Virginia Department of Health, Norfolk, Virginia;*
[3]*District of Columbia Department of Health, Washington, D.C.*

Corresponding author: Christine M. Yuan, Bureau of Epidemiology and Health Risk Assessment, District of Columbia Department of Health, 825 North Capitol St. NE, Third Floor, Washington, D.C., 20002. Telephone: 202-442-5859; Fax: 202-442-4796; E-mail: christine.yuan@dc.gov.

Abstract

Hospital emergency department (ED) syndromic surveillance has been proposed for early detection of a large-scale biologic terrorist attack. However, questions remain regarding its usefulness. The authors examined the use of active syndromic surveillance at hospital EDs in Virginia for early detection of disease events and analyzed the effectiveness of the cumulative sum (CUSUM) algorithm in identifying disease events from syndromic data. Daily chief-complaint data were collected for 10 months at seven hospital EDs in southeastern Virginia. Data were categorized into seven syndromes (fever, respiratory distress, vomiting, diarrhea, rash, disorientation, and sepsis), and the CUSUM algorithm was used to detect anomalies in each of the seven syndromes at each hospital. Fever and respiratory distress syndromes exhibited monthly and ambient-temperature–specific trends consistent with southeastern Virginia's influenza season. Furthermore, preliminary frequencies of hospital ED patient chief complaints in southeastern Virginia during a 10-month period were produced by using syndromic data. This system represents an example of a local syndromic surveillance program serving multiple cities in a limited geographic region.

Introduction

Syndromic surveillance in hospital emergency departments (EDs) involves monitoring incoming patients with nonspecific syndromes to determine whether an unusual excess of any group of symptoms exists. Although syndromic surveillance might prove useful for detecting a deliberate release of a biologic agent, baseline ED chief-complaint data first need to be better characterized to create a surveillance instrument that can detect unusual disease incidence of any cause (*1,2*). Lives might be lost if an untested surveillance system misses a disease event (*3*). Therefore, syndromic surveillance systems should be investigated critically to determine whether ED data can serve these purposes. Accordingly, syndromic surveillance was performed at seven hospital EDs in southeastern Virginia, and the value of ED-based syndromic surveillance was explored by analyzing the effectiveness of the cumulative sum (CUSUM) algorithm for detecting unusual disease events.

Syndromic Surveillance System

Population

The Tidewater or Hampton Roads region of southeastern Virginia has a substantial military presence, consisting of a major U.S. Air Force base and a naval amphibious base. Approximately 13% of the population of the four Virginia cities from which data were collected (Norfolk, Chesapeake, Newport News, and Virginia Beach) were either in reserve or on active military duty in the year 2000 (*4*). In addition, the military is responsible for approximately 25% of the region's economy (*5*). The syndromic surveillance system established in this region involved seven civilian hospitals serving approximately 1 million residents (*6*).

Data Collection and Aberration Detection

ED data were collected from seven hospitals during September 2001–June 2002. Chief-complaint data (i.e., the patient's stated reason for visiting the ED) were faxed daily from hospitals to the health department. These data were then categorized manually into one of seven syndromes (fever, respiratory distress, vomiting, diarrhea, rash, disorientation, and sepsis). A CUSUM algorithm (*7*) was used to analyze unusual increases in each of the seven syndromes at each hospital. The CUSUM algorithm used three different moving average calculations (*mild*, *medium*, and *ultra*) to identify unusually high occurrences of each syndrome. The *mild* calculation used a moving average of syndrome counts for the 7 days preceding the ED visit. The moving average for the *medium* calculation was for 3–9 days previous, and the moving average for the *ultra* calculation was for 3 preceding days. Upper limits for all three calculations were set to the moving average plus 3 standard deviations, and observed daily syndrome counts were compared with each upper limit.

A working database was created for the Tidewater region that combined daily entry of syndrome counts with the CUSUM anomaly-detection algorithm. Daily syndrome counts were dichotomized as *high occurrence* or *low occurrence* on the basis of daily CUSUM calculations. On *high occurrence* days, the health department performed patient chart reviews and reported information on patient ED visits (e.g, discharge diagnosis, laboratory testing, and patient disposition) to the regional epidemiologist. Monthly reports were also generated on syndrome counts and distributed to participating hospitals' infection-control practitioners, personnel involved in emergency response to biologic terrorism, and ED personnel.

Detection of Influenza

The CUSUM algorithm detected trends in fever and respiratory distress occurrences indicative of influenza at hospital C (Figure 1) and by month and temperature (Figure 2). According to the sentinel influenza surveillance system, which consists of a designated group of reporting physicians in the region, influenza occurrence in eastern Virginia increased during the week of January 23, 2002. However, syndromic data on fever and respiratory distress revealed an increase in these two syndromes during the week of January 14, 2002, indicating an earlier start to the influenza season.

Experience

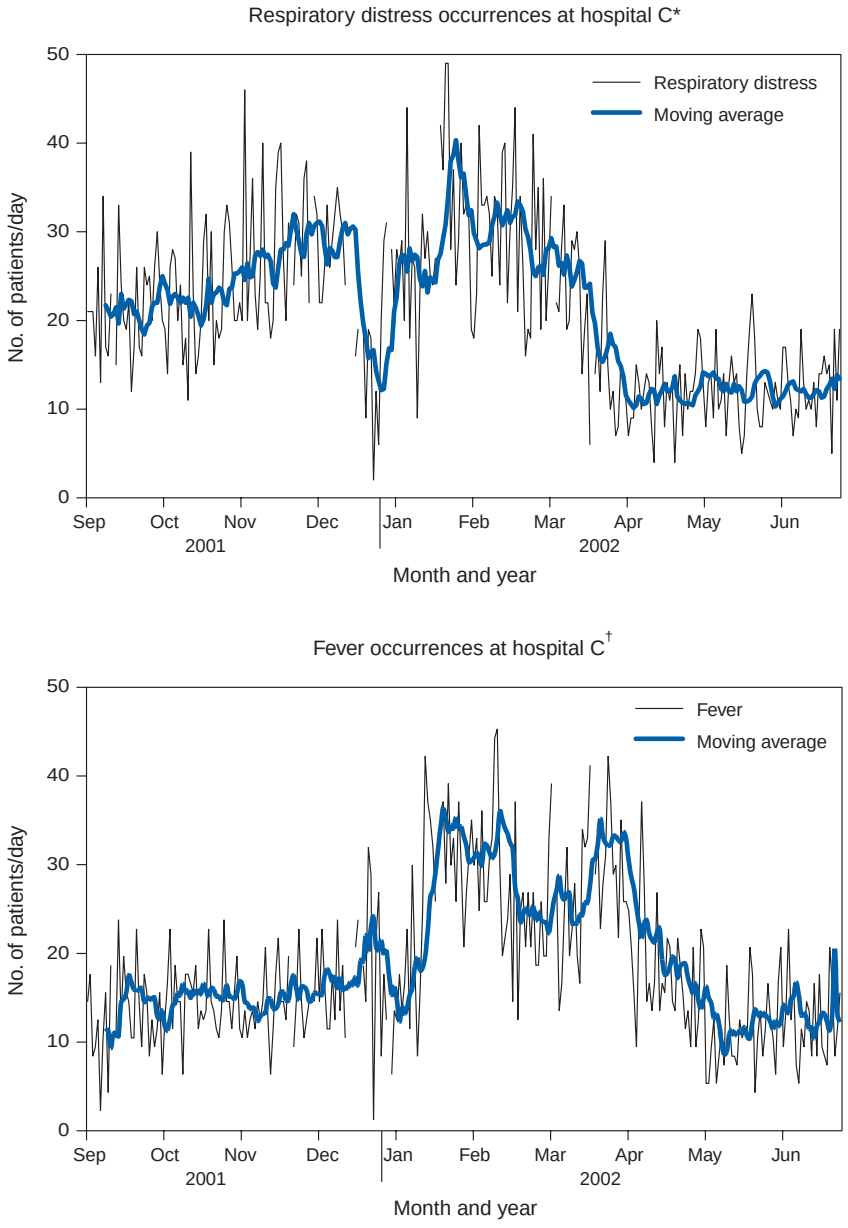
Challenges

Syndromic surveillance presented certain challenges. The collected data spanned only a 10-month period that included both a biologic terrorist event involving anthrax in a nearby region as well as an influenza season. Thus, syndromic data might have reflected both seasonally expected trends and unexpected syndrome occurrences. Moreover, the lack of an electronic method for rapid and accurate data transfer often delayed the collection process. Syndromic surveillance retrospectively detected disease occurrences (e.g., the influenza season); however, without timely data reporting, acute disease events might not be detected quickly enough to permit rapid response.

Extrinsic Value

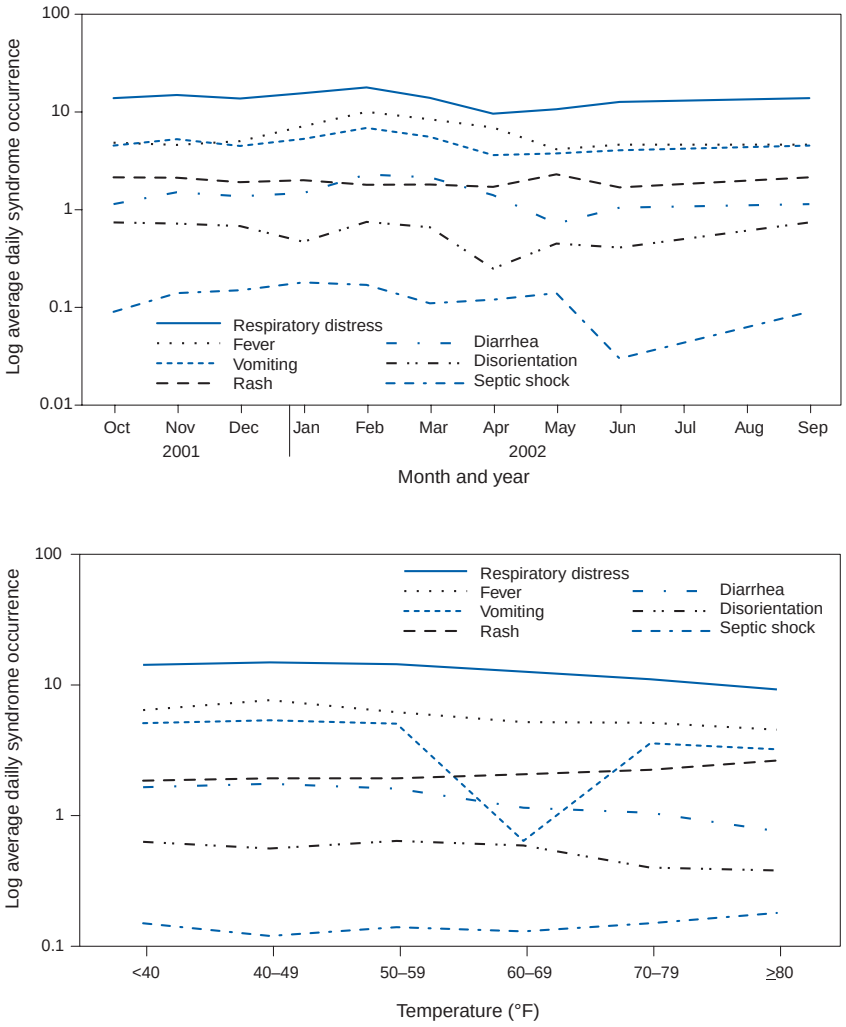
Despite certain difficulties, a preliminary characterization of hospital ED populations and syndrome occurrences in the Tidewater region was produced, the first such effort in southeastern Virginia. Because syndromic surveillance has only recently been introduced into public health, patterns from different surveillance systems have rarely been compared. This surveillance system compared seven different hospitals and

FIGURE 1. Detection of seasonal influenza by syndromic surveillance at one hospital in southeastern Virginia, September 2001–June 2002



*N = 5,969.
†N = 4,991.

FIGURE 2. Average daily occurrence* of seven syndromes, by month and by temperature — Virginia, September 2001–June 2002



* N = 50,144.

syndromes and identified substantial pattern differences in two syndromes at only one hospital. This indicates that recognizing anomalies in any one place and for any one syndrome might require analysis of local circumstances (e.g., the populations served by particular hospitals) to enhance syndromic surveillance and improve detection of the unusual. The CUSUM algorithm identified increased influenza activity (i.e., respiratory distress and fever). With refinement and longer time series, CUSUM should become more sensitive and eventually be able to provide earlier recognition of natural outbreaks or terrorist events.

Intrinsic Value

Syndromic surveillance can increase communication among professionals in public health and clinical medicine. Through greater interaction between the public health and medical fields, ED physicians and other healthcare personnel realize the value of a public health specialist *(8)*. Furthermore, partnering of public health professionals with physicians, law enforcement and other disaster-management workers can improve a jurisdiction's preparedness for any disease event (*9*). The effectiveness of a surveillance system requires the cooperation and collaboration of multiple persons. As part of syndromic surveillance, EDs might capture sudden, subtle changes in the magnitude and distribution of diseases in a population *(8)*. Meanwhile, public heath departments are responsible for continuously monitoring surveillance reports and findings (*1*). For syndromic surveillance to enhance rapid detection of anomalous events, clear communication among hospitals and public health agencies, as well as preparedness and response capacities, must be in place.

Acknowledgements

J. Davies-Cole, C. Glymph, and A. Diallo of the District of Columbia Department of Health provided guidance and support. Reviewers also provided invaluable comments and suggestions.

References

1. Green M, Kaufman Z. Surveillance for early detection and monitoring of infectious disease outbreaks associated with bioterrorism. Israel Med Assoc J 2002;4:503–6.
2. Connolly C. In New York, on alert for bioterrorism: city's tracking system viewed as model. Washington Post, November 24, 2001:A1.
3. Kohane I. The contributions of biomedical informatics to the fight against bioterrorism. J Am Med Inform Assoc 2002;9:116–9.
4. Rhein M, Lafronza V, Bhandari E, Hawes J, Hofrichter R. Advancing community public health systems in the 21st century. Washington, DC: National Association of County and City Health Officials, 2001.
5. Foshey D, ed. Hampton Roads statistical digest. Richmond, VA: Virginia Business, 2000. Available at http://www.virginiabusiness.com/market_research/digests/hampton/digest_hamp03.shtml.
6. US Census Bureau. Race, Hispanic or Latino, and age. Available at: http://factfinder.census.gov/servlet/SAFFPeople?_sse=on.
7. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. Emerg Infect Dis 1997;3:395–400.
8. Pollock D, Lowery D, O'Brien P. Emergency medicine and public health: new steps in old directions. Ann Emergency Med 2001;38:675–83.
9. Khan A, Morse S, Lillibridge S. Public-health preparedness for biological terrorism in the USA. Lancet 2000;356:1179–82.

Research Methods

Bivariate Method for Spatio-Temporal Syndromic Surveillance

Al Ozonoff, L. Forsberg, M. Bonetti, M. Pagano
Harvard School of Public Health, Boston, Massachusetts

Corresponding author: Marcello Pagano, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115. Telephone: 617-432-4911; Fax: 617-739-1781; E-mail: pagano@hsph.harvard.edu.

Abstract

Introduction: *Statistical analysis of syndromic data has typically focused on univariate test statistics for spatial, temporal, or spatio-temporal surveillance. However, this approach does not take full advantage of the information available in the data.*

Objectives: *A bivariate method is proposed that uses both temporal and spatial data information.*

Methods: *Using upper respiratory syndromic data from an eastern Massachusetts health-care provider, this paper illustrates a bivariate method and examines the power of this method to detect simulated clusters.*

Results: *Use of the bivariate method increases detection power.*

Conclusions: *Syndromic surveillance systems should use all available information, including both spatial and temporal information.*

Introduction

In 2002, CDC advised health departments to seek routinely collected electronic data as part of early warning systems for biologic terrorism (*1*). The potential cost-effectiveness of such systems might explain why certain major metropolitan areas (e.g., Boston and New York) are beginning to implement CDC's recommendation (*2,3*). The primary concern of a biosurveillance system is to analyze and interpret data as they are collected and then decide whether further investigation is required. This report proposes a statistical methodology needed to make such a system efficient and effective and focuses on how to use information about the number of patients affected and where they live to detect outbreaks or other deviations from the normal pattern of disease.

Two statistical concerns are fundamental to surveillance: 1) determining a reasonable definition of "normal" behavior, and 2) being vigilant for deviations from this normalcy. CDC's weekly surveillance for pneumonia and influenza mortality in 122 U.S. cities is one example of an attempt to put this into practice (see *MMWR Weekly* at http://www.cdc.gov/mmwr). In that model, historic data allow for time-series modeling of seasonal fluctuations in deaths; the model represents an attempt to define normalcy. Building on a sinusoidal model for the seasonal baseline, standard statistical methods (*4*) provide a confidence band outside of which mortality can be considered a deviation from the norm. Such a definition of normalcy is too stringent because deviations from normalcy occur almost every year; therefore, its usefulness for a surveillance system might be questionable. However, a too-lenient definition of normalcy might then never detect a deviation from normal.

Combining Univariate Statistics

Combining more than one test statistic from a single data source poses problems. In certain situations, multiple testing without an appropriate statistical adjustment leads to an inflation of the false-positive rate. However, such adjustments can be conservative and adversely affect the power of the tests.

One approach that avoids the multiple-testing problem involves investigating the joint distribution of the test statistics. As a result, the information encoded in each statistic is used, but the false-positive rate can still be carefully controlled. The bivariate methodology described in this paper is one example of combining univariate statistics. Although the concept generalizes easily to other settings, implementation of this methodology will necessarily differ, depending on the situation. The requirements and assumptions (as well as the strengths and weaknesses) of the particular univariate models and statistics used will affect the power and robustness of any implementation of this bivariate approach.

Data

Data for this study were obtained from a major health-care provider in eastern Massachusetts. As patients arrive for emergency care, their cases are geocoded (typically by using the patient's residential or billing address); this information is centralized electronically on a daily basis. For this study, a subset of the data was selected, consisting of upper respiratory infections (URIs) during January 1, 1996–October 30, 2000, for a period of 1,399 days. (For protection of confidentiality, the spatial data provided in this report were aggregated by census

tract and white noise was added to the centroids of the tracts.) Thus, the data stream provides the temporal patterns of disease (i.e., the number of cases arriving each day), as well as the spatial patterns of disease (i.e., the locations of patients over time).

Using all available information should provide better detection power than using just the number of patients or only their locations. Thus, the proposal is to analyze the temporal series first, then the spatial series, and, finally, to conduct a joint analysis of the two.

Methods

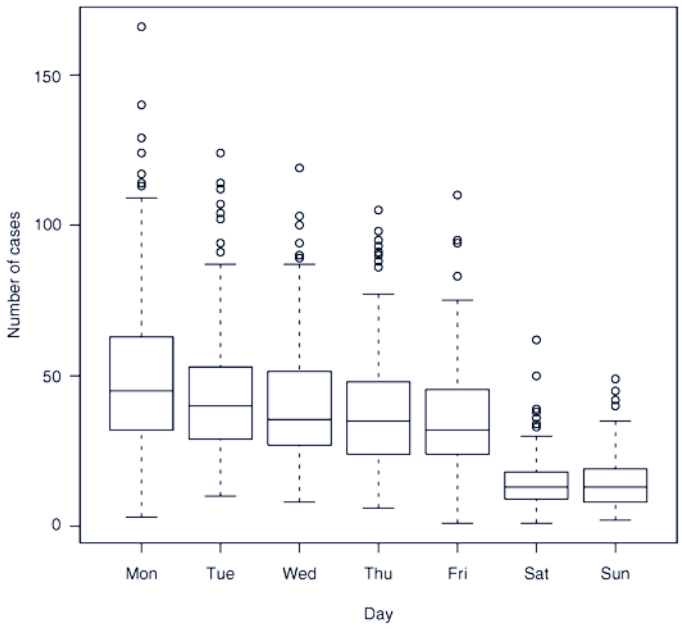
Time-Series Modeling

Time-series modeling is one approach for analyzing temporal data. Certain trends in the number of patients reporting daily with URIs make modeling challenging. One such trend is a seasonal effect, which can be modeled efficiently. Superimposed on the seasonal effect is a substantial daily effect, including a slight downward trend in the number of URIs from Monday through Friday, as well as a substantially higher variance from the start of the week to the end (Figure 1). Weekends and holidays must be analyzed separately because certain clinics and other locations are closed on those days, resulting in lower case volume and a different spatial distribution of patients. Health-care demand for weekends and holidays is often satisfied on Mondays or weekdays immediately after holidays, resulting in a higher case volume on those days.

For the time series N(t) of number of URIs to be accurately modeled, a sinusoidal baseline curve must first be fitted to account for seasonal variations. Each data point can then be considered as a residual departure from the baseline prediction. The residuals are then modeled to find a best predicted value of N(t). Because patient behavior varies by day of week, days are categorized as follows: 1) weekend days or holidays; 2) Mondays or days after holidays; and 3) all other weekdays. Seasonal and daily effects are incorporated into a linear model. The residuals from this mean function are autocorrelated; therefore, a third-order autoregressive component and a first-order moving average component (Autoregressive Moving Average [ARMA] [3,1] are used to model this autocorrelation. Thus, the final model is formulated as

$$\log[N(t)] = (\text{seasonal sinusoid} + \text{daily indicators} + \text{interactions}) + e(t) + \beta_1 e(t-1) + \beta_2 e(t-2) + \beta_3 e(t-3) + \gamma \log[N(t-1)]$$

FIGURE 1. Sample box plots of daily case volume of upper respiratory infections, by day



Note: Caseload on weekends is lower, when certain clinics are closed. Monday counts are, on average, slightly higher but are also more variable because Mondays are often holidays (which, in turn, results in an elevated average Tuesday caseload).

where e(t) is the residual (observed or predicted value) at time t, and the β, γ are ARMA coefficients estimated from a standard statistical package. The standard deviation of the residuals is used as a measure of the model's goodness-of-fit. After inclusion of the ARMA terms, the standard deviation of the residuals was reduced from 0.732 to 0.321 (on the log scale), indicating that the ARMA series has a better fit than the simple sinusoid. Standard deviations for holidays and weekends, Mondays and days after holidays, and other weekdays are all comparable; however, these are measured on the log scale, and thus, the higher case volume on Mondays and days after holidays, together with greater variation on those days (Figure 1), reduces the model's predictive power for those days as compared with weekends and holidays, which have lower mean case counts.

The time series N(t) is an attempt to describe normal behavior. The residuals are distributed approximately normally with mean 0, and a nominal alpha level can be chosen on the basis of historic data, and any observation falling outside of a particular critical region can be considered worthy of investigation.

Spatial Statistic

Temporal analysis provides only one perspective, albeit a classic one, of the information in the surveillance data (i.e., the number of patients) The geocoded portion of the data set (i.e., the location of the patients) provides a second perspective. Other researchers have used spatial analytic approaches (*2,3,5*) on the assumption that terrorist attacks might produce a pattern of disease with a distinctive spatial signature (*6*).

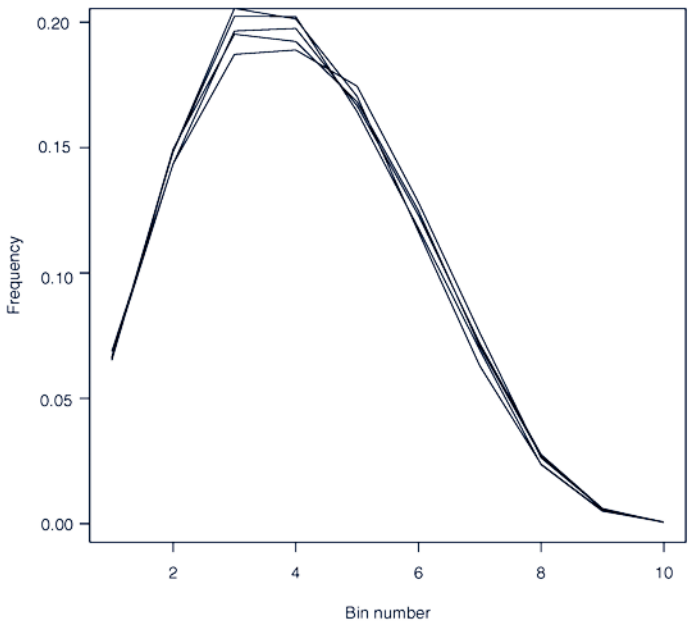
Multiple spatial statistics have been designed to detect distinctive spatial patterns (*7,8*). Because the particular disease pattern that a terrorist attack might produce remains unknown, a statistic should be sufficiently flexible to detect multiple distortions from normalcy without requiring *a priori* knowledge of how such a distortion might appear. For this analysis, simple application of the M-statistic (*9*), which is based on the distribution of distances between patients, was chosen. To compute the M statistic for detection of outbreaks, all pairwise distances between locations of patients arriving for care each day are calculated. An empirical cumulative distribution function (ECDF) of these distances can then be compared with the historically determined distribution of distances to yield a test statistic, M. Asymptotic properties of the M statistic (*9*) or empirical simulation allow for a nominal alpha level to determine substantial deviations from the norm.

Fundamental to use of the M statistic is the remarkable stationarity of the distribution of distances over time. The frequency polygon of distances, derived from the ECDF, for five randomly chosen, nonoverlapping 30-day periods distributed across seasons and throughout the approximate 4-year study period, is illustrated (Figure 2). The ECDF is sufficiently stable from season to season and year to year to establish a definition of normalcy.

Daily geocoded data enables 1) calculation of the ECDF $\hat{F}(D)$ (where $F(D)$ denotes the cumulative distribution function of interpoint distances determined from historic data) for each day's disease cases, and 2) calculation of a test statistic measuring the departure from $F(D)$. To avoid complexities, the daily case load is used to calculate distances between patients; typically, memory can be incorporated into the system by extending a temporal window within which to calculate distances. This extension would be especially important when dealing with a contagious ailment that has an incubation distribution. To facilitate calculation of the statistic, all of the interpoint distances are placed into 10 bins that are equiprobable under the distribution $F(D)$, and a Mahalanobis-like distance is calculated as

$$M = (o - e)^t S^- (o - e)$$

FIGURE 2. Frequency polygons of distances for five nonoverlapping periods, illustrating seasonal stability of the empirical cumulative distribution function of interpoint distances



Note: Although equiprobable bins are used when calculating the M statistic, they are displayed here as a standard (equal width) format for ease of viewing.

where o is the 10-dimensional vector of observed proportions of distances in each bin; e is the vector of expected proportions (equal to [0.1, … , 0.1]) under the null distribution; and S is an estimator of the variance-covariance matrix $\sum$ of the bin proportions calculated under the null. S is calculated from the historic data and a generalized inverse S^- is used because S is not of full rank.

Because the distribution of distances between patients is stationary, an alert based on M can be instituted so that large values of M generate the alert; exactly how large these values must be is determined by the desired false-positive rate. The null distribution of M is determined by the null distribution of the distances; however, asymptotically, NM has a χ^2 distribution with degrees of freedom equal to the rank of the covariance matrix $\sum^- \sum$ (where NM refers to the product of the test statistic M(t) and number of cases N at a time t). Thus, the distribution of NM is asymptotically independent of the number of cases used to calculate the statistic. As the degrees of freedom increase, the log of a χ^2 random variable approximates a normal distribution, and experience has confirmed that the values log(NM) give a close normal approximation.

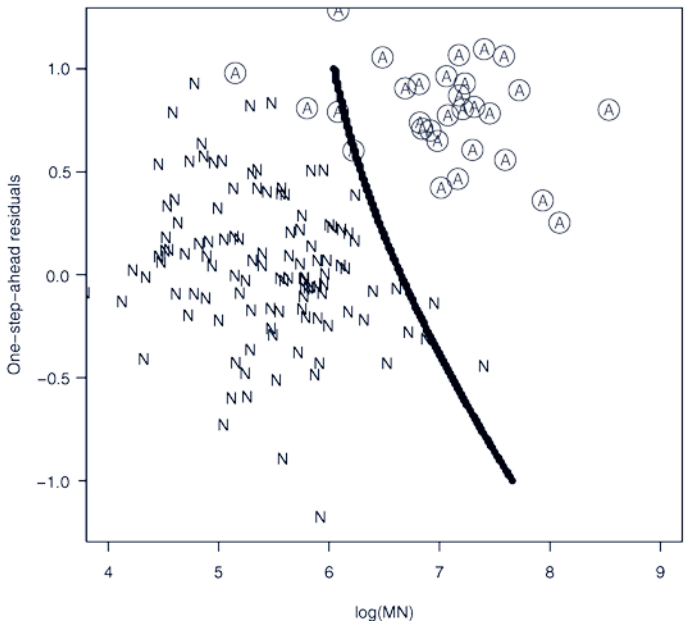
More importantly, this demonstrates that the random variables NM and N are approximately independent for large N (i.e., N >40). Thus, the temporal information and spatial information are orthogonal (for large N).This substantiates combining the two to produce an even more powerful statistic, as discussed in the following section.

Bivariate Test Statistic

Use of a bivariate test statistic, composed of the two statistics described previously, is proposed to increase the power of outbreak detection. N(t) permits calculation of a residual value for the number of cases arriving, on the basis of the time-series prediction for that day, with residuals that are approximately normal. Log(NM) expresses the deviation of the spatial distribution of cases from normalcy, and this statistic is approximately normal as well. Standard techniques from multivariate analysis can be used to construct an elliptical rejection region for a bivariate normal population at prespecified alpha level (false-positive rate) that can be used to detect deviations from normalcy. However, this might not offer particular protection against the alternative of interest (i.e., an outbreak resulting from release of a biologic agent).

As another approach, potential biologic attacks can be modeled to simulate bivariate values in the event of an attack; in this case, an optimal discriminator (the quadratic classification rule) exists between two bivariate normal populations: 1) the bivariate distribution under the null, and 2) the modeled bivariate distribution under the alternative of a biologic attack (*10*). The classification rule is a quadratic form that, given log(NM) and the one-step-ahead time-series residuals, assigns one day's observations to either the null or alternative population. This rule minimizes the expected error of misclassification. The false-positive rate can be controlled by shifting the quadratic boundary appropriately, as determined through simulation or resampling of the historic record. A typical case of the null and alternative populations, together with the boundary of the discriminator, is illustrated (Figure 3).

FIGURE 3. Subset of the null (N) and alternative (A) populations used to train the quadratic discriminator for using the bivariate test statistic to perform power calculations for spatio-temporal disease surveillance



Note: The horizontal axis measures the spatial component of the data, the vertical axis measures the temporal component, and the solid black line (a portion of the classification boundary) is used to decide whether a particular day's observation falls into the null (normal) or alternative (unusual/outbreak) population.

Results

Because no biologic terrorism events occurred in eastern Massachusetts during the period of study, an outbreak simulation was necessary. To this end, for each of four locations, either six, nine, or 12 additional URIs were added to the existing data set. The range of 6–12 cases represents approximately 0.25–1.25 standard deviations of the original caseload, depending on the day of the week (mean daily case count is approximately 15 cases/day on weekends, 55 cases/day on Mondays, and 40 cases/day on other weekdays). The signal was dispersed across adjacent census tracts (i.e., adding six cases at a particular location amounted to choosing six nearby tracts and adding one case to each tract). (For brevity, such a simulated signal is called a cluster.) By using the statistics discussed previously, power was calculated on the basis of this simulated disease signal. Although other methods might have higher power to detect a concentrated cluster (e.g., six additional cases in one tract), they are less likely to perform as well when the signal is dispersed.

A simulated cluster was added to each of the 1,399 days of data, 1 day at a time, to assess how frequently different statistics might detect such a signal. Power calculations were performed separately for each of the three daily categories (weekend days or holidays, Mondays or days after holidays, and all other weekdays) because prediction and behavior differ within each of these categories. A detection threshold was set for each statistic on the basis of an alpha level of 0.05. For daily observations (as are illustrated here), this is equivalent to one false alert every 20 days. Power equals the ratio of detections to the total number of observations.

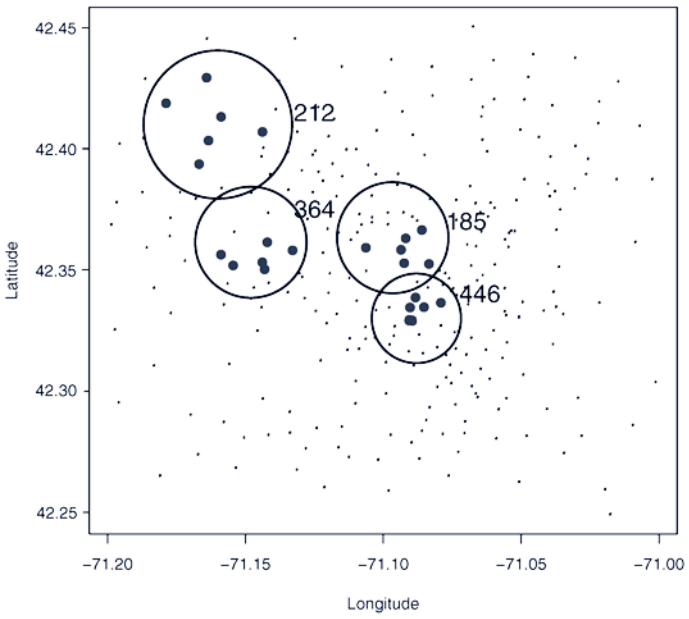
The four locations chosen for the simulation are in different geographic areas covered by the data. Previous simulations have demonstrated that power to detect a cluster might

depend on the local geography and location of the signal source (*11*). This effect is confounded by the population distribution in the data available. Locations on the outskirts of the region covered tend to be more sparsely populated; hence, the signal is more widely dispersed. The census-tract locations in the study area, together with the four locations at which clusters were simulated, are illustrated (Figure 4). The cluster at location 446 corresponds to an area approximately circular with radius 0.5 miles; at locations 185 and 364 with radius 1 mile; and at location 212 with radius 1.5 miles. These radii reflect population densities.

Power calculations for the three test statistics are provided (Table). Results for the univariate test statistic N based on time-series modeling are not stratified by location because the statistic depends only on the number of cases and not on locations. Power to detect an additional six, nine, or 12 cases added to the case counts of the final 399 days of data was then calculated by using the first 1,000 days to train the model (Table).

Next, a training sample was generated based on a modeled signal consisting of 12 cases near location 446, superimposed on each of the first 1,000 days of data. This permitted generation of two distinct bivariate normal populations of values, consisting of N(t) residuals together with log(NM) calculations, as a training sample. Next, for a simulated cluster in the final 399 days of data, the corresponding bivariate test statistic was calculated, and the quadratic classification rule was used to place each day's simulated cluster into the null (no signal) population or the alternative (signal present) population (Table). Power in this case equals the number of clusters classified in the alternative divided by the total number of observations.

FIGURE 4. Simulated clusters for use in outbreak-detection power calculations involving spatial and bivariate test statistics



Note: Four different sets of simulations were performed, using different cluster locations; these are indicated by the circles. Within each circle, large dots indicate census tracts for which cases were added to simulate a disease cluster. The small dots represent census tract locations across the Greater Boston area.

TABLE. Powers for three statistical tests in detecting disease outbreaks when simulated clusters of size six, nine, and 12 are superimposed on original data from four locations (census tracts 446, 185, 364, and 212)

Location, cases and cluster size	Overall	Holidays/ weekends	Weekdays	Days after holidays
Temporal test*				
N + 6	0.128	0.168	0.112	0.100
N + 9	0.213	0.304	0.187	0.117
N + 12	0.286	0.408	0.234	0.217
Spatial test using the M-statistic				
446, N + 6	0.141	0.162	0.138	0.108
185, N + 6	0.141	0.148	0.151	0.090
364, N + 6	0.093	0.103	0.092	0.075
212, N + 6	0.054	0.078	0.044	0.042
446, N + 9	0.258	0.299	0.264	0.156
185, N + 9	0.254	0.256	0.276	0.175
364, N + 9	0.187	0.237	0.171	0.142
212, N + 9	0.064	0.087	0.051	0.061
446, N + 12	0.383	0.422	0.395	0.258
185, N + 12	0.382	0.397	0.410	0.250
364, N + 12	0.292	0.349	0.283	0.203
212, N + 12	0.072	0.075	0.071	0.071
Bivariate statistic				
446, N + 6	0.441	0.536	0.453	0.200
185, N + 6	0.456	0.520	0.514	0.117
364, N + 6	0.373	0.424	0.416	0.117
212, N + 6	0.308	0.360	0.327	0.133
446, N + 9	0.659	0.776	0.682	0.333
185, N + 9	0.652	0.776	0.682	0.283
364, N + 9	0.564	0.728	0.575	0.183
212, N + 9	0.391	0.464	0.416	0.150
446, N + 12	0.777	0.904	0.790	0.467
185, N + 12	0.807	0.896	0.850	0.467
364, N + 12	0.747	0.864	0.780	0.383
212, N + 12	0.509	0.608	0.537	0.200

* Results for this test are not stratified by location because the statistic depends only on the number of cases and not on location.

Conclusions

The power of the univariate statistic N, which detects deviations from the predicted number of cases daily, illustrates the difficulties of time-series modeling for public health surveillance. The behavior of the time series N(t) is nonstationary, with differing variation according to season and day of the week. Rather than relying on a simple autoregression, detection results could be improved by considering a multivariate

periodic autoregression (*12*). Meanwhile, the spatial statistic M has exhibited promise in other contexts to detect spatial deviations from the norm (*3*,*9*). Further research into the characteristics of this and other spatial statistics is needed, as different complementary spatial methods exist that can be used in conjunction with differing detection power.

Development of additional statistical methods and research into those methods are critical to the terrorism surveillance effort. Because routinely collected electronic data are often available to public health departments and researchers, efficient analysis of these data provides a low-cost method for surveillance. Although one cannot make any claims as to the robustness or generalizability of the bivariate method to other data sets or other univariate statistics, the power calculations provided here demonstrate that information on the number of cases as well as the spatial distribution of those cases can be used effectively in combination to improve the efficiency of surveillance systems.

Acknowledgments

This project was funded in part by grants from the National Institutes of Health (Grant no. RO1AI28076) and the National Library of Medicine (Grant no. RO1LM007677).

References

1. CDC. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. MMWR 2002; 51:13–5.
2. Kleinman K. Lazarus R. Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. Am J Epidemiol 2004;159:217–24.
3. Mandl KD, Overhage JM, Wagner MM, et al. Syndromic surveillance: a guide informed by the early experience. J Am Med Inform Assoc 2004;11:141–50.
4. Serfling R. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963;78:494–506.
5. Olson, KL, Mandl, KD. Use of geographical information from hospital databases for real time surveillance. Pediatric Research 2002;51:94a.
6. Meselson M, Guillemin J, Hugh-Jones M, et al. The Sverdlovsk anthrax outbreak of 1979. Science 1994;266:1202–8.
7. Kulldorff M. A spatial scan statistic. Communication Statistics—Theory and Methods 1997;26:1481–96.
8. Thompson HR. Distribution of distance to the n^{th} nearest neighbor in a population of randomly distributed individuals. Ecology 1956; 37:391–4.
9. Bonetti M, Pagano M. The interpoint distance distribution as a descriptor of point patterns: an application to cluster detection. Stat Med 2004 (in press).
10. Johnson RA, Wichern DW. Applied multivariate statistical analysis. 5^{th} ed. Upper Saddle River, NJ: Prentice Hall, 2002:590–8.
11. Bonetti M, Forsberg L, Ozonoff A, Pagano M. The distribution of interpoint distances. In: Banks HT, Castillo-Chavez C, eds. Bioterrorism: mathematical modeling applications in homeland security. Philadelphia, PA: SIAM, 2003:87–106.
12. Pagano M. On periodic and multiple autoregressions. Annals of Statistics 1978;6:1310–7.

Role of Data Aggregation in Biosurveillance Detection Strategies with Applications from ESSENCE

Howard S. Burkom,[1] Y. Elbert,[2] A. Feldman,[1] J. Lin[1]
[1]*Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland;* [2]*Walter Reed Army Institute for Research, Silver Spring, Maryland*

Corresponding author: Howard S. Burkom, 11100 Johns Hopkins Road, Mailstop 8-224, Laurel, MD 20723. Telephone: 443-778-4361; Fax: 443-778-5950; E-mail: Howard.Burkom@jhuapl.edu.

Abstract

Introduction: *Syndromic surveillance systems are used to monitor daily electronic data streams for anomalous counts of features of varying specificity. The monitored quantities might be counts of clinical diagnoses, sales of over-the-counter influenza remedies, school absenteeism among a given age group, and so forth. Basic data-aggregation decisions for these systems include determining which records to count and how to group them in space and time.*

Objectives: *This paper discusses the application of spatial and temporal data-aggregation strategies for multiple data streams to alerting algorithms appropriate to the surveillance region and public health threat of interest. Such a strategy was applied and evaluated for a complex, authentic, multisource, multiregion environment, including >2 years of data records from a system-evaluation exercise for the Defense Advanced Research Project Agency (DARPA).*

Methods: *Multivariate and multiple univariate statistical process control methods were adapted and applied to the DARPA data collection. Comparative parametric analyses based on temporal aggregation were used to optimize the performance of these algorithms for timely detection of a set of outbreaks identified in the data by a team of epidemiologists.*

Results: *The sensitivity and timeliness of the most promising detection methods were tested at empirically calculated thresholds corresponding to multiple practical false-alert rates. Even at the strictest false-alert rate, all but one of the outbreaks were detected by the best method, and the best methods achieved a 1-day median time before alert over the set of test outbreaks.*

Conclusions: *These results indicate that a biosurveillance system can provide a substantial alerting-timeliness advantage over traditional public health monitoring for certain outbreaks. Comparative analyses of individual algorithm results indicate further achievable improvement in sensitivity and specificity.*

Introduction

A working definition of syndromic surveillance is the monitoring of available data sources for outbreaks of unspecified disease or of specified disease before identifying symptoms are confirmed. Its goal is to complement existing sentinel surveillance by identifying outbreaks with false-alert rates acceptable to the public health infrastructure. After data sources are chosen, multiple data-aggregation decisions follow. Foremost among these decisions are which data records to monitor, how data will be aggregated in space and time, and how other covariates (e.g., age and sex) will be managed. In data aggregation, a thematic tradeoff exists between expanding the space or time window to increase structure for background modeling and masking a potential outbreak signal with the additional counts.

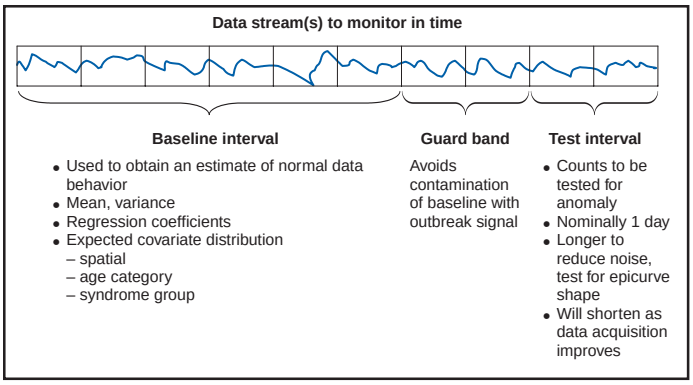
This paper explores data aggregation by space, time, and data category; discusses the relevance of data aggregation to the effectiveness of alerting algorithms; describes approaches selected for use by the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) (*1*); and discusses these approaches' performance in a detection evaluation exercise conducted in 2003 by the Bio-Event Advanced Leading Indicator Recognition Technology (Bio-ALIRT) program of the Defense Advanced Research Project Agency (DARPA) (*2*).

Background

Sliding Buffer Concept

A temporal-aggregation concept underlying certain surveillance algorithms (*3,4*), including those used by ESSENCE, is the separation of recent data into three segments that slide forward in time (Figure 1). These segments include 1) a baseline period to estimate expected data behavior; 2) the recent test period, typically 1–7 days, of potentially anomalous data; and 3) a guard band between them to avoid contamination of the baseline by an outbreak signal. Whether

FIGURE 1. Conceptual sliding buffers for temporal data aggregation



the quantities of interest are simple means and standard deviations, regression coefficients, spatial distributions, or distributions of covariate strata (e.g., age groups), these temporal subdivisions are used to determine whether the test-period data violate the null hypothesis of expected behavior inferred from the baseline.

Data Aggregation and Purely Temporal Surveillance

Purely temporal surveillance monitors data time series for outbreak-induced anomalies without using spatial information. Categorical- and spatial-aggregation decisions determine both the time series to be monitored and the regression-based or process-control–based approaches to be implemented for monitoring. Historic data analysis is used to choose the baseline lengths, and the expected data effects of outbreaks are used to determine the length of the test period and guard band. These aggregation decisions (e.g., to stratify among neighboring regions or data subtypes) might result in the monitoring of multiple time series. Multivariate algorithms using the data-covariance matrix can exploit the correlation among these time series but might be sensitive to changes in data relationships (e.g., changes caused by informatics or organizational changes) that are irrelevant to monitoring for disease.

Data Aggregation and Scan Statistics

Spatial-aggregation decisions for purely temporal methods can be driven by jurisdictional or logistical considerations, but such decisions can decrease the early warning advantage of syndromic surveillance (e.g., when early cases are scattered among the chosen regions). Use of scan statistics (*5*,*6*), notably in SaTScan™ software (*7*), has become popular because it avoids preselection bias and can choose the most important among possible outbreak locations and extents without oversensitivity caused by multiple testing. Use of scan statistics guides spatial aggregation and can direct limited public health resources to localities of anomalous case distributions. Temporal aggregation becomes a concern in ESSENCE adaptations of scan statistics when the underlying assumption of uniform spatial incidence fails. In such cases, historic data are used to obtain expected spatial distributions; temporal baseline and test-period decisions are then necessary. For example, the New York City Department of Health and Mental Hygiene successfully used a 28-day baseline and 7-day guard band and test periods in West Nile virus surveillance (*3*). An enhanced scan-statistics implementation in ESSENCE enables treatment of other aggregation problems (e.g., the distance measure for generating candidate clusters). The distance matrix is usually formed by using the Euclidean distance between centroids of component subregions. Although this distance measure might be appropriate for monitoring threats caused by atmospheric risk factors (e.g., an aerosolized release of a biologic agent), driving distance might be a more suitable measure for monitoring an increase in communicable endemic disease. Test-bed implementations have demonstrated that direct, heuristic modifications to the distance matrix can avoid undesirable clustering. An ESSENCE enhancement also permits use of multiple data sources to search for anomalous clusters (*8*). The different data sources need not have the same spatial partitioning, and their baseline and test intervals might differ. A stratified scan-statistics approach is used to avoid the signal masking caused by mismatched scales or variances in the respective data sources. A performance measure, described and tested with various signal distributions (*8*), demonstrates that the stratified approach retains power to detect signals in both single and multiple data sources.

Objectives

ESSENCE's biosurveillance systems attempt to fuse information from multiple data sources that vary in their medical specificity, spatial organization, scale, and time-series behavior. True denominator data specifying the number of persons at risk are rarely available. These systems are increasingly used at multiple jurisdictional levels; therefore, the system alerts should be appropriate to the purview of the user. Specific objectives are to 1) present aggregation and detection strategies that were applied to the city-level DARPA evaluation exercise (see Methods), 2) present the ESSENCE results from this exercise, and 3) draw conclusions about potential system capability and identify areas for enhancement.

Methods

For temporal-detection algorithms, statistical process control (SPC) and multiple statistical process control (MPSC) algorithms are applied to raw or normalized time-series data.

Data Normalization Strategies

Normalization is required if the raw time-series data exhibit systematic features (e.g., day-of-week effects). These features are most often seen in counts of large syndrome-group diagnoses collected from well-represented regions; an approximate quantitative rule for these features is a median of ≥5 counts per day. When such data features occur, SPC algorithms are applied to the residuals of linear or Poisson regression. Current ESSENCE systems apply goodness-of-fit statistics to automate the choice of whether to use regression residuals; regression-predictor variables include time, day-of-week indicators, and other data-dependent quantities.

Aggregation and Fusion Concerns

Monitoring multiple series might be necessary for three reasons: 1) multiple, disparate data sources might be available; 2) time series for a data source might be divided among political regions or treatment facilities; 3) the need to monitor for multiple outbreak types might require stratification of available data by syndrome or product group. These circumstances are increasingly intertwined in ESSENCE systems as the surveillance areas and number of available data sources increase. Two combined monitoring approaches are taken. In the multiple univariate approach, detection algorithms are applied separately to each time series, and alerting depends on how the separate results are combined. The combination method must retain sensitivity while avoiding excessive alerts caused by multiple testing. In the multivariate approach, MSPC algorithms are applied to the set of time series to produce a single statistic. These algorithms usually depend on a recent estimate of the covariance matrix of the input streams, and the challenge is to avoid alerts caused by changes to data interrelationships that are irrelevant to potential outbreaks.

Multiple Univariate Strategies

Univariate SPC methods used by recent ESSENCE systems include 1) an exponential weighted moving average (EWMA) algorithm (*9*), with baseline and guard band optimized for timely alerting of an epicurve-like signal, and 2) the nonhistoric cumulative sum (CUSUM) algorithms from the Early Aberration Reporting System (EARS) (*10*) used by many local health departments. Alerting based on the maximum value of the chosen univariate method over input data streams leads to excessive alerting as the number of these streams increases. Using Edgington's consensus method (*11*) for multiple experiments reduces this problem. Bayes Belief Networks (BBNs) (*12*), a more versatile means for combining algorithm outputs, were used in the DARPA evaluation exercise to calculate a composite p-value for alerting. BBNs provide a compact encoding of the joint probability distribution of algorithm outputs along with other synoptic evidence. This approach uses a directed graphical structure to represent knowledge of conditional independences among variables to simplify the representation of the overall joint probability distribution. Because variables (nodes in the graph) usually depend on a limited number of other variables, estimates of probabilities are needed only for the local (connected) relationships. The overall probability distribution is then determined from all local distributions. Thus, the BBN approach permits environmental evidence and heuristic rules to be included in alerting decisions.

Multivariate Methods

The use of MSPC methods for surveillance against cyber attacks by adopting Hotelling's T^2 is described elsewhere (*13*). Certain published discussions (*14,15*) state that multivariate EWMA (*16*) and CUSUM (*17*) methods are preferable to Hotelling's T^2 for detecting changes in the multivariate mean because they have shorter average run lengths before the process is declared out of control. For the application of finding outbreak signals in outpatient-visit data, all of these methods were determined to be oversensitive because they generated alerts from irrelevant changes in the covariance matrix estimate. To illustrate, the T^2 statistic can be written

$$(X - \mu)^T S^{-1} (X - \mu)$$

where X = multivariate data from the test interval; μ = vector mean estimated from the baseline interval; and S = estimate of covariance matrix calculated from the baseline interval.

Certain nuisance alerts caused by relative data dropoffs were eliminated by implementation of a one-sided test in which the test statistic was replaced with 0 whenever the sum of current z-scores over the data streams was negative. These z-scores were calculated by using the current baseline mean and standard deviation in each stream. This procedure naturally reduced the number of alerts in all MSPC methods, and the resulting T^2 statistic performed well in the Bio-ALIRT evaluation. Additional work is needed to improve the specificity of certain methods (*16,17*) for biosurveillance applications.

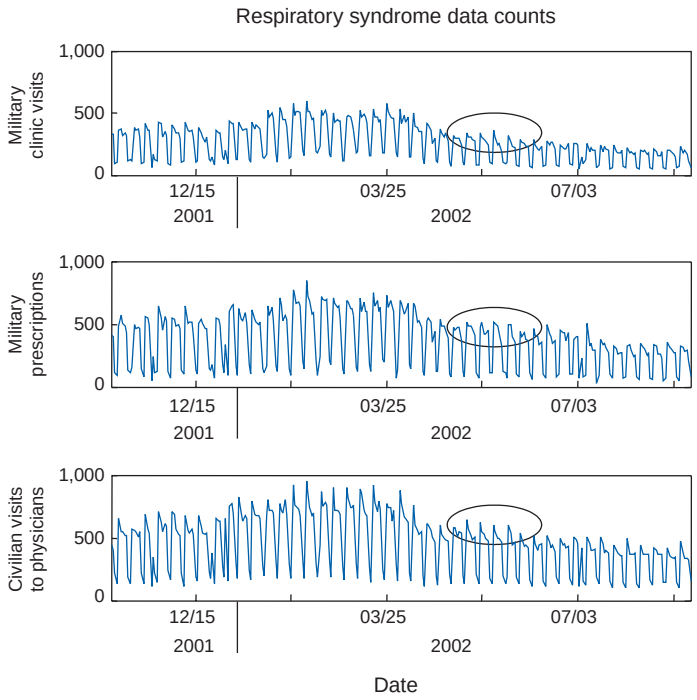
Bio-ALIRT 2003 Detection System Evaluation

The DARPA evaluation exercise was a comprehensive comparison of the effectiveness of detection methodologies used by participating contractor teams in a large, complex, authentic data environment. The exercise is discussed elsewhere in detail (*2*), and its main features are summarized here. The task for the contractor teams was to find authentic outbreaks when given daily records from three data sources: military clinic visits, physician office visits by civilians, and military prescriptions. Only records of visits or prescriptions that could be classified with a respiratory or gastrointestinal (GI) diagnosis were included in the sample; for simplicity, respiratory and GI data were analyzed separately for outbreaks. The principal covariates included in the records were patient age, sex, residential zip code, and specific *International Classification of Diseases, Ninth Revision* (ICD-9) codes (or, for prescriptions, Specific Therapeutic Class [GC3] codes and National Drug Codes [NDC]), along with the respiratory/GI classification. Data sets from five cities were processed separately. The outbreak detection group (ODG), a committee of epidemiologists and physicians, chose these data sets and identified sample outbreaks for training purposes.

Fourteen months of training data from all five cities were supplied to Bio-ALIRT detection teams for learning the data features and for choosing and calibrating optimal detection methods. The resulting methods were to be applied without further modification to the next 9 months of data. ODG then examined the 9-month test period of these data sets independently and, for each outbreak identified, specified a start date, nominal date when traditional public health monitoring would have recognized the outbreak, peak date, and end date. The ODG findings of eight respiratory outbreaks and seven GI outbreaks in the test period were treated as the standard for the exercise, against which the algorithm outputs of each detection team were scored. The positive and negative aspects of applying human medical professional judgment to authentic, noisy data for performance-evaluation purposes have been discussed elsewhere (*2*).

Sample plots of the training data for each data source are presented (Figure 2). These time series of patient encounters indicated substantial respiratory syndrome data counts, distinct day-of-week effects, and seasonal trends. ODG directed the detection teams to look for city-scale outbreaks of any duration. Faint outbreaks in the training set were detected, not completely synchronized among the data streams, which could be found only with multivariate methods.

FIGURE 2. Training data sample from the Defense Advanced Research Project Agency detection evaluation exercise



Note: The circles in the figure indicate a faint outbreak in the training set, not completely synchronized among the data streams, which could be found only with multivariate methods.

Performance Assessment Tools

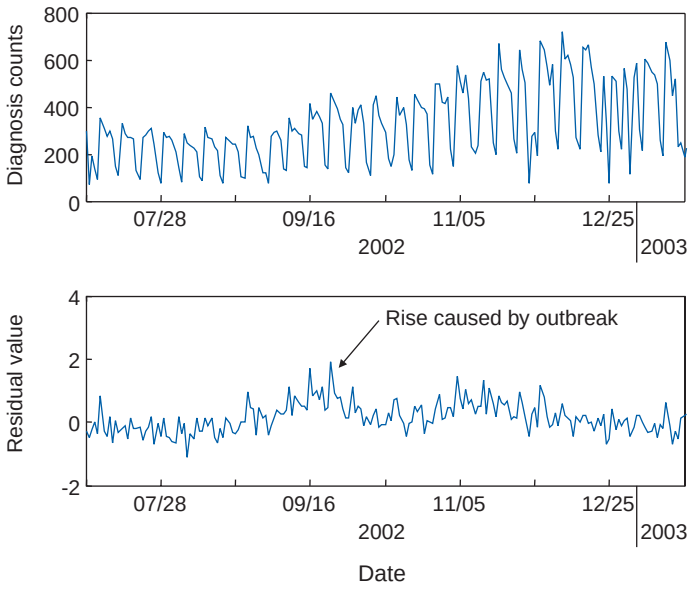
The methodology used to measure the performance of the detection algorithms in this exercise is described elsewhere in computational detail (*2*). The two measures used were algorithm sensitivity (i.e., the number of outbreaks detected) and timeliness (i.e., the number of days between the outbreak start and subsequent alert). However, instead of being assessed at fixed algorithm thresholds at uncontrolled specificity, both measures were calculated for fixed false-alert rates seen as practical for public health surveillance. False-alert rates of 1 per 2 weeks, 1 per 4 weeks, and 1 per 6 weeks were chosen for this purpose. Series of trials were conducted on the training data sets to choose algorithms that were effective at these false-alert rates with parameters that were approximately optimal for the surveillance context of this exercise.

Data Conditioning Using Provider-Count Regression

In terms of the performance measures adopted, a particularly effective data-conditioning procedure was a linear regression of the daily syndrome counts in which the count of providers reporting each day was used as a predictor. The daily

reporting provider counts were calculated according to the data type (i.e., the count of clinics for the military outpatient data, of pharmacies for the military prescription data, and of individual physicians for the civilian office-visit data). Residuals from this regression were used as input to the alerting algorithms. Substitution of the count data with these residuals probably improved algorithm performance because the daily provider counts can reflect both known data features (e.g., holiday and weekend dropoffs) and unknown ones (e.g., special military events and severe weather effects). Thus, the regression can remove such features, which are irrelevant for public health purposes, from the algorithm inputs (Figure 3). In effect, the algorithms operate on the difference of the observed counts from the expected counts given the number of reporting providers. In comparison plots of actual count data and regression residuals, the day-of-week effect is strongly attenuated in the residual plot (Figure 4). Baseline lengths of 1–10 weeks were tested on the training data, and a 5-week baseline gave the best detection performance on a chosen set of outbreak signals.

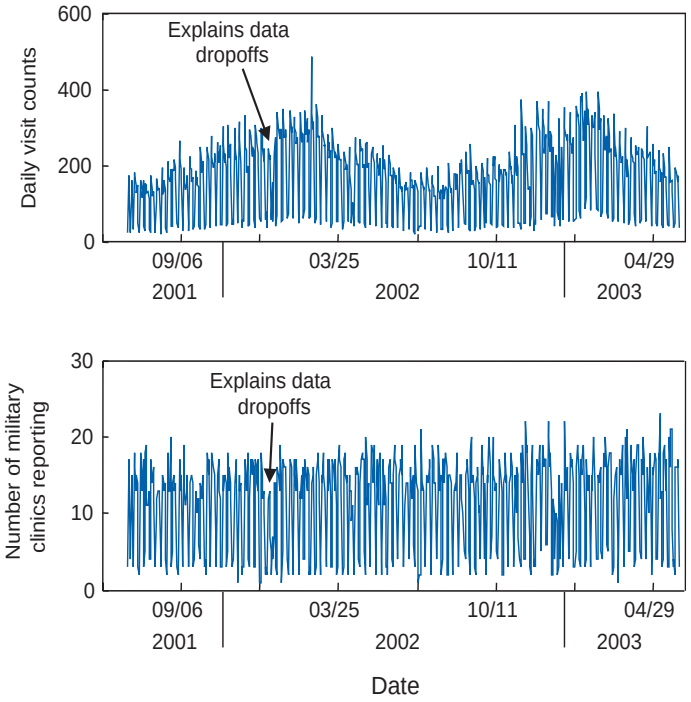
FIGURE 4. Day-of-week-effect attenuation in provider-count regression residuals



Results

Two algorithmic methods gave robust performance in detection testing on the evaluation training data sets, using a candidate set of outbreak events and the false-alert criteria described previously. The first method was to precondition all three data streams by using provider count regression and then to apply Hotelling's T^2 algorithm. The second method was a multiple univariate EWMA algorithm similar to the EARS C2 method (*10*), with the baseline length chosen by empirical testing (Figure 3). Because these methods differed in the limited number of outbreaks not detected at the chosen false-alert rates, their outputs were combined by applying a BBN based on the joint probability distribution of the outputs calculated from the training period data.

FIGURE 3. Daily counts of total patient encounters and number of military clinics reporting



These two methods and the BBN composite were applied to the exercise-test data sets for comparison with ODG outbreak findings. Performance results are summarized separately for the respiratory and GI outbreaks (Table). For GI outbreaks at the specificity level of one false alert per 4 weeks, the median detection time was 1 day after the start date chosen by ODG epidemiologists, whereas their median unaided recognition date was 2 weeks after the start date. For the two individual algorithms, the median detection time increased to 5 days for the most constrained false-alert rate, whereas the BBN improved timeliness by 2 days. The BBN also detected an additional outbreak at the lowest specificity. Corresponding results for the respiratory outbreaks indicated that the multiple univariate method was superior in both sensitivity and timeliness at the higher specificity levels.

TABLE. Performance of three methods for detecting two outbreak types — Defense Advanced Research Project Agency detection evaluation exercise

		Sensitivity			Median timeliness		
	Gastrointestinal outbreaks	**Alerts/7 events**			**Days before alert**		
	False-alert rate (expected days between alerts)	14	28	42	14	28	42
Methods	Provider-count–adjusted MSPC*	6	6	6	1	1	5
	Multiple univariate SPC†	6	6	6	1	1	5
	Bayes Belief Network combination	7	6	6	1	1	3
		Sensitivity			**Median timeliness**		
	Respiratory outbreaks	**Alerts/8 events**			**Days before alert**		
	False-alert rate (expected days between alerts)	14	28	42	14	28	42
Methods	Provider-count–adjusted MSPC	8	7	6	1	4.5	4.5
	Multiple univariate SPC	8	8	8	1	1	1
	Bayes Belief Network combination	8	7	7	1	1	4.5

* Multiple statistical process control.
† Statistical process control.

Conclusions

Judicious data-aggregation strategies have important functions in improving detection performance of biosurveillance systems. Choosing the appropriate scope for monitored time series, stratifying and filtering patient-encounter data, and tuning algorithms effectively can improve these systems' sensitivity for early outbreak detection. The DARPA evaluation exercise provided a useful test bed for quantifying these improvements by using authentic data streams from five geographic regions.

The focus on city-level outbreaks in this exercise led to an emphasis on temporal alerting methods. Both multiple univariate and multivariate approaches yielded good detection sensitivity and timeliness, and both presented challenges that indicate a need for further improvement. As ESSENCE surveillance systems become more complex, enhancement of these approaches will be important for managing the multiple-testing problem while preserving sensitivity. For the multiple univariate problem, the BBN approach appears versatile for combining separate algorithm-output streams. BBNs are also robust in that they can handle missing data in a mathematically consistent way, an important feature in syndromic surveillance, where data dropouts are common. Another advantage of BBNs is the capability to combine other evidence (e.g., sensor or environmental data) with the algorithm outputs for a fused assessment of the probability of an outbreak. Multivariate methods might have the best potential for finding faint signals distributed over multiple data sources, but adaptations are needed for specificity in the biosurveillance context.

The DARPA exercise results should be understood in perspective. Using authentic clinical data from five cities, the epidemiologist team specified start dates and unaided public health recognition dates for 15 disease outbreaks. The best algorithms generated alerts within days of the start date, whereas the median gap between the start dates and recognition dates was 2 weeks. The focus on city-level outbreaks and the restriction of outbreaks to respiratory or gastrointestinal symptoms probably boosted the algorithm performance. For the more difficult challenge of a multisource, multilevel system to detect outbreaks of unconstrained symptomatology, a comprehensive evaluation with authentic data would be extremely complex. Finally, if detection algorithms can truly give advance warning of ≥1 week for certain outbreaks, the matter of how to respond to these early warnings is critical for public health decision-makers.

Acknowledgments

This research is sponsored by DARPA and managed under Naval Sea Systems Command (NAVSEA) contract N00024-03-D-6606. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of DARPA, NAVSEA, or the U.S. Government. The authors wish to acknowledge the technical contributions of Anshu Saksena, Johns Hopkins University Applied Physics Laboratory.

References

1. Lombardo JS, Burkom HS, Elbert YA, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health 2003;80(2 Suppl 1):i32–42.
2. Siegrist D, Pavlin J. Bio-ALIRT biosurveillance detection algorithm evaluation. MMWR 2004;53(Suppl):152–8.
3. Mostashari F, Kulldorff M, Hartman J, Miller J, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. Emerg Infect Dis 2003;9:641–6.
4. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. Emerg Infect Dis 1997;3:395–400.

5. Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic. J R Stat Soc A 2001;164:61–72.
6. Wallenstein S. A test for detection of clustering over time. Am J Epidemiol 1980;104:576–84.
7. Kulldorff M, Information Management Services, Inc. SaTScan™ version 4: software for the spatial and space-time scan statistics, 2003. Available at http://www.satscan.org.
8. Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. J Urban Health 2003;80(2 Suppl 1):i57–65.
9. Ryan TP. Statistical methods for quality improvement. New York, NY: John Wiley & Sons, 1989, Section 5.5.
10. Hutwagner LC, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96.
11. Edgington ES. A normal curve method for combining probability values from independent experiments. J Psychol 1972;82:85–9.
12. Hayek CS, Pineda FJ, Doss OW, Lin JS. Computer-assisted interpretation of mass spectra. Johns Hopkins APL Tech Dig 1999;20:363–71.
13. Ye N, Cheng Q, Emran S, Vilbert S. Hotelling's T2 multivariate profiling for anomaly detection. IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, West Point, NY, June 6–7, 2000. Piscataway, NJ: IEEE Press, 2000:180–6.
14. Thompson JR, Koronacki J. Statistical process control for quality improvement. New York, NY: Chapman & Hall, 1993, Ch 7.
15. Wierda SJ. Multivariate statistical process control. Groningen, Netherlands: Wolters-Noordhoff, 1994, Ch 7.
16. Lowry CA, Woodall WH. A multivariate exponentially weighted moving average control chart. Technometrics 1992;4:46–53.
17. Crosier RB. Multivariate generalizations of cumulative sum quality-control schemes. Technometrics 1988;30:291–303.

Scan Statistics for Temporal Surveillance for Biologic Terrorism

Sylvan Wallenstein,[1] J. Naus[2]
[1]*Mount Sinai School of Medicine, New York, New York;* [2]*Rutgers, the State University, New Brunswick, New Jersey*

Corresponding Author: Sylvan Wallenstein, Box 1023, 1 Gustave Levy Place, Mount Sinai School of Medicine, New York, NY 10029. Telephone: 212-241-1526; Fax: 212-860-4630; E-mail: sylvan.wallenstein@mssm.edu.

Abstract

Introduction: *Intentional releases of biologic agents are often designed to maximize casualties before diagnostic detection. To provide earlier warning, syndromic surveillance requires statistical methods that are sensitive to an abrupt increase in syndromes or symptoms associated with such an attack.*

Objectives: *This study compared two different statistical methods for detecting a relatively abrupt increase in incidence. The methods were based on the number of observations in a moving time window.*

Methods: *One class of surveillance techniques generates a signal based on values of the generalized likelihood ratio test (GLRT). This surveillance method is relatively well-known and requires simulation, but it is flexible and, by construction, has the appropriate type I error. An alternative surveillance method generates a signal based on the p-values for the conventional scan statistic. This test does not require simulation, complicated formulas, or use of specialized software, but it is based on approximations and thus can overstate or understate the probability of interest.*

Results: *This study compared statistical methods by using brucellosis data collected by CDC. The methods provided qualitatively similar results.*

Conclusions: *Relatively simple modification of existing software should be considered so that when GLRTs are performed, the appropriate function will be maximized. When a health department has data that indicate an unexpected increase in rates but its staff lack experience with existing software for surveillance based on GLRTs, alternative methods that only require computing Poisson probabilities can be used.*

Introduction

Traditional surveillance systems tend to focus on compulsory reporting of specific diseases. However, in recent years, syndromic surveillance based on emergency department admissions, hospital bed occupancy, pharmaceutical sales, and other correlates of disease has increased to detect possible biologic terrorism attacks (*1*). This study analyzed methods useful in detecting surges in illness (*1*), particularly when these increases are abrupt, as might occur during a biologic attack.

This study was based on the assumption that, according to historic data, events occur on the basis of a known pattern of events (e.g., seasonal, specific day of the week, or weather). Methods used to estimate this pattern based on historic data have been addressed by others (*1–3*) and are not the focus of this paper, although one simple fitting method is illustrated. Although multiple statistical approaches to surveillance have been proposed and compared before 2001 (*4,5*), interest in these methods has recently increased (*6–8*).

This study's overall approach scans time, seeking unusual incidence within a short period. The symbol *t* represents current time, and *w* represents a window of time used for surveillance, usually a limited number of days. $Y_t(w)$ is the number of events in the last *w* days before and including *t*, and $E_t(w)$ is the expected number of such events, usually based on historic data. The proposed methods result in an alert being generated at time *t*, if $Y_t(w)$ is substantially greater then $E_t(w)$. The procedures are designed so that, if the event rates are the same as the historic rates, the probability of generating one or more false signals in a period T is α. The total time frame T is under the investigator's control.

The procedures described in this paper can be contrasted with what are termed quadrat-based tests (*9*) or cell procedures. In such procedures, time is subdivided into nonoverlapping periods of days, weeks, or months, and the data analyst searches for substantial increases in these periods. The Communicable Disease Surveillance Centre (CDSC) in London uses such a system (*10*) to automatically scan weekly reports to provide early warning of disease outbreaks. CDSC staff compare observed counts of a disease in a given week with historically fitted expected counts. However, equally concerning is a cluster of cases that occurs during a 7-day period that overlaps 2 calendar weeks. In a monitoring system that continuously updates reports, advantages exist, both with power and speed of detection, in using scan-like statistics and

examining the number of cases in a moving time interval instead of just looking at nonoverlapping intervals. This is particularly true for monitoring disease organisms that can be used for a biologic terrorism event, during which an early warning might be critical. If the reported effect of a release of a biologic agent is expected to spread over a 7-day period, then health department staff use a 7-day scanning window rather than a calendar week for monitoring.

This study focused on how staff decide that an observed count in a limited window of width w (measured in days or weeks) is more substantial than expected, taking into account multiple testing during a longer surveillance period T. Two functions of the observed and expected values were used to judge what constitutes more substantial counts. One function was based on generalized likelihood ratio tests (GLRTs), and the other was based on p-values calculated from the classical, constant-risk scan statistics.

Both of these approaches can be viewed as extensions of the classical scan statistic, the maximum number of observations in an interval of width w. One of the defects of the classical scan statistic is that it assumes a constant baseline rate (*4*). This difficulty can be overcome by scanning on the basis of GLRT (*11–14*). The first procedure discussed in this paper shares a common theoretical background with this surveillance method but differs in that the type I error refers to a period of time (e.g., the time of a limited objective surveillance, or a month, or a year) rather than the instant at which an alert might be generated. The second procedure, based on p-values, does not require simulation and thus can be more easily applied.

For this study, both of these procedures were applied to brucellosis data collected by CDC during 1997–2002. The point of using these example data is not to evaluate brucellosis but to illustrate how such an analysis can be performed.

Methods

For this study, the authors assumed that the incidence of events follows a Poisson process. In this description of the methods, the notation concerning the process was suppressed, and focus was placed on $E_t(w)$, the expected number of events in a window of w days ending at time t. The first test requires that the window width w be fixed before the surveillance; this condition is then removed.

In a biologic terrorism event, the difference between an early signal and an obvious outbreak might be days. A critical period exists, d days, within which the data analyst should detect the increase. Multiple authors (*7,8*) have reported that special techniques are needed when only a limited time delay can be tolerated. Therefore, the signal decision should be based on observations within the past d days. In this context, the window size is in the range $w \leq d$. Alternatively, for increased power, a fixed window of $w = d$, or $w = d - 1$, can be used.

G-Surveillance Methods

If the window width, w, is fixed in advance, G-surveillance used to detect an abrupt increase, on the basis of a fixed type I error for a given period, generates an alert for substantial values of the statistic $G_t(w)$,

$$G_t(w) = Y_t(w)\, ln\, [Y_t(w)/E_t(w)] - [Y_t(w) - E_t(w)]$$

where *ln* is the natural logarithm. (Details of the proof are available from the corresponding author upon request.) An alert will be sounded at time t, if $G_t(w)$ is larger than a threshold (i.e., the critical value) obtained through simulation.

The extension to the case where w is not fixed but is within a certain range (e.g., 1–3 days) follows the same pattern as previously described (*9,15,16*). G-surveillance with variable window widths will signal an alert at time t if

$$G_t(u\ to\ v) = \max_{u \leq w \leq v} G_t(w)$$

is larger than a new critical value.

When data are recorded daily, u in the previous equation corresponds to the smallest number of days of interest (presumably, $u = 1$), and v to the largest number of days. When surveillance is continuous, u should not be set so small that it picks up artifacts of data collection and, in certain contexts, might be ≥24 hours. If the expected values depend only on past history, the threshold can be obtained before surveillance begins by generating realizations of the complete process. For numerous local health departments to avoid having to develop expertise in simulating the process, this critical value can be computed once a year at a central location and then transmitted to local health departments. In other cases, the expected values depend on current data (e.g., weather conditions), and the user might have to re-do simulations at each time point t.

P-Surveillance Methods

An alternative method is a fixed-window scan surveillance method, P surveillance, that does not require simulation but instead is based on p-values from the classical scan statistic (*17*). The traditional fixed window scan statistic, S_w, is the largest number of cases to be found in any subinterval of length w (for w, a known constant) of the surveillance interval (0,T). Two recent books (*18,19*) summarize results on finding the exact probability (*20*), finding bounds (*21*), and finding approximations (*21,22*) for the distribution of S_w. For the atypical surveillance application, in which the expected

number of events in any interval of width w is a constant, λ, the approximation (*22*) is given by

$$\Pr(S_w > k) = (T/w)(k-\lambda)\, p(k,\lambda) + s(k,\lambda)$$

where $p(k,\lambda)$ is the Poisson probability of observing exactly k events, $p(k,\lambda) = \exp(-\lambda)\,\lambda^k / k!$, and $s(k,\lambda)$ the probability of observing k + 1 or more events.

The limited usefulness of the classic scan statistic in surveillance, because of its assumption of constant baseline risk, has been noted (*4*). One early method to overcome this limitation involved stretching or contracting time (*23*), which has the disadvantage that it would not allow surveillance in 24-hour units. G-surveillance is another way to overcome the limitation.

P-surveillance is based on computing a p-value at time t, focusing on what is happening at that time and ignoring all other information. The same p-value should be used if the baseline risk over the whole period is constant at the local rate at time t.

Under continuous surveillance, an alert is signaled at time t, if

$$(T/w)\,[Y_t(w) - E_t(w)]\, p[Y_t(w), E_t(w)] + s[Y_t(w), E_t(w)] < \alpha$$

Under this procedure, α is the probability of generating a false alert in time frame T (e.g., T = 1 year) and will usually be set to 0.05 or 0.10. In surveillance applications, loss of precision will be limited if the second term in the last equation is ignored so that an alert will be signaled if

$$(T/w)\,[Y_t(w) - E_t(w)]\,[exp[-E_t(w)]\, E_t(w)^{Y(w)} / Y_t(w)!] < \alpha$$

Thus, P-surveillance in continuous time requires calculating the left side of the previous equation each time an event occurs and deciding if it is less than a prespecified α.

Conceptually, a different test based on the ratchet scan statistic (*24*) should be performed when the data are collected daily or weekly instead of continuously. The principle underlying the test would be the same.

Justification for the use of P-surveillance requires 1) demonstrating formally that theoretical (mathematical) reasons exist to assume that P-surveillance has the claimed false-alert rate, and then substantiating it by simulation, and 2) using theory or simulations to demonstrate that P-surveillance had power somewhat comparable to G-surveillance. Work on the first assertion has already been performed (*18*), and limited numerical work by the authors supports the second assertion.

Results (Example)

A study of disease characteristics of microbiologic agents with particular potential for biologic terrorism lists brucellosis among critical biologic agents reported to the National Notifiable Disease Surveillance System (*25*). For this paper, weekly national reports of brucellosis are used (for illustration purposes only) as a proxy for the type of daily totals that might arise for certain more common conditions in limited geographic areas.

Provisional (and for years 2001 and 2002, revised) cumulative data can be obtained from *Morbidity and Mortality Weekly Report* (available at http://www.cdc.gov/mmwr). The data are revised to adjust for delayed reporting because certain states submit reports in batches and include suspected cases in addition to confirmed ones. In using the provisional cumulative data, distinguishing between negative adjustments caused by removing previous suspected cases and new suspected or confirmed cases is impossible. This study used revised data for 1997–2001 provided by CDC (Table) as a proxy for the analysis possible if the provisional data provided the number of new cases/week.

Of these 260 weekly baseline counts, all but three are in the range of 0–7. These three cases are all in different years and occur at the end of the year. Careful scrutiny of the counts reveals certain yearly and seasonal patterns; however, to obtain an overall impression of the magnitude, the mean (1.60; standard deviation: 1.45) of the remaining 257 counts was computed (Table).

The following procedure was used to calculate the estimated value per week (Table). The average (or for weeks 49 and 52, the median) number of cases of brucellosis per week during 1997–2001 was calculated. The averages were then smoothed by fitting a spline to the means (or for weeks 49 and 52, the medians) for the first 51 weeks of data. No adjustment was made for a possible secular trend.

G-surveillance (i.e., GLRT-based, scan-type methods) was based on 1) a fixed 3-week window size and 2) on a window that can be either 1, 2, or 3 weeks. Because the model postulated does not involve any factors unknown at the start of the year, percentiles of interest can be computed once before the surveillance period begins. To obtain the percentiles for both statistics, 100,000 realizations of the process were simulated for the period of T = 52 weeks, in which weekly counts were generated on the basis of a Poisson distribution with the expected value (see last column of Table).

	Percentiles of statistics				
	0.50	**0.75**	**0.90**	**0.95**	**0.99**
G(3 weeks)	1.98	2.82	3.77	4.36	5.94
G(1–3 weeks)	2.67	3.49	4.48	5.19	6.77

TABLE. Revised brucellosis counts per week and predicted values

Week	1997	1998	1999	2000	2001	2002	Average 1997–2001	Spline
1	1	0	1	1	1	1	**0.8**	**0.81**
2	0	0	1	3	0	1	**0.8**	**0.81**
3	2	0	1	0	2	2	**1.0**	**0.8**
4	0	1	0	0	2	2	**0.6**	**0.8**
5	0	1	1	0	2	1	**0.8**	**0.81**
6	1	0	2	2	1	1	**1.2**	**0.83**
7	0	0	1	1	0	0	**0.4**	**0.88**
8	0	0	2	0	1	4	**0.6**	**0.96**
9	0	0	0	2	4	3	**1.2**	**1.08**
10	3	0	0	1	3	1	**1.4**	**1.18**
11	3	0	1	0	5	1	**1.8**	**1.23**
12	0	0	1	2	2	1	**1.0**	**1.23**
13	1	1	2	1	1	1	**1.2**	**1.2**
14	1	1	2	1	2	1	**1.4**	**1.17**
15	1	1	0	0	2	2	**0.8**	**1.16**
16	1	1	0	3	1	5	**1.2**	**1.2**
17	0	1	0	2	0	0	**0.6**	**1.3**
18	3	0	0	3	2	0	**1.6**	**1.46**
19	0	3	1	3	4	5	**2.2**	**1.62**
20	0	1	3	3	1	5	**1.6**	**1.74**
21	4	2	0	3	5	9	**2.8**	**1.82**
22	1	1	1	1	3	2	**1.4**	**1.87**
23	0	0	1	3	2	0	**1.2**	**1.96**
24	0	2	1	2	2	—	**1.4**	**2.13**
25	5	3	1	0	6	—	**3.0**	**2.34**
26	3	0	1	3	6	—	**2.6**	**2.51**
27	4	4	5	5	2	—	**4.0**	**2.6**
28	1	2	2	3	0	—	**1.6**	**2.59**
29	2	2	1	1	4	—	**2.0**	**2.56**
30	5	3	1	3	1	—	**2.6**	**2.53**
31	4	1	1	4	5	—	**3.0**	**2.46**
32	1	2	2	2	6	—	**2.6**	**2.32**
33	1	2	3	2	5	—	**2.6**	**2.13**
34	2	0	0	1	1	—	**0.8**	**1.93**
35	1	1	1	3	3	—	**1.8**	**1.79**
36	0	1	1	2	4	—	**1.6**	**1.71**
37	3	2	0	1	3	—	**1.8**	**1.65**
38	4	2	2	1	1	—	**2.0**	**1.59**
39	2	0	2	2	3	—	**1.8**	**1.52**
40	1	1	0	1	2	—	**1.0**	**1.47**
41	0	0	0	2	2	—	**0.8**	**1.46**
42	0	1	0	1	4	—	**1.2**	**1.5**
43	4	5	2	2	3	—	**3.2**	**1.54**
44	2	0	2	0	2	—	**1.2**	**1.52**
45	0	1	2	3	1	—	**1.4**	**1.48**
46	3	0	1	0	3	—	**1.4**	**1.46**
47	0	1	0	0	1	—	**0.4**	**1.5**
48	2	3	2	1	3	—	**2.2**	**1.62**
49	2	1	(14)	1	2	—	**2.0***	**1.74**
50	2	1	2	2	3	—	**2.0**	**1.86**
51	2	0	0	2	5	—	**1.8**	**1.96**
52	(24)	(12)	0	2	7	—	**7.0***	**7**†
Mean	**1.5**	**1.1**	**1.1**	**1.7**	**2.6**	—	**1.6**	**1.6**
Standard deviation	**1.5**	**1.2**	**1**	**1.2**	**1.7**	—	**0.8**	**0.5**

Note: The three values in parentheses are presumptive outliers and were not included directly in descriptive statistics for the year or in the average.
* The median was used instead of the mean because of presumptive outliers.
† The value was used in generating counts but was not used to fit the spline.

G-surveillance was applied to the 2002 data. The most noteworthy feature (up to week 23) is the observed counts of 5, 5, and 9 for weeks 19, 20, and 21, respectively, contrasted with expected counts of 1.62, 1.74, and 1.82, respectively. For weeks 19–21, the observed 3-week count is 19, and the expected is 5.18. Assuming use of surveillance with a fixed 3-week window, G-surveillance at week 21 is based on the value 19 *ln* (19/5.18) – (19 – 5.18) = 10.88, where *ln* is the natural logarithm. Because this statistic exceeds 5.94, the probability of observing such a substantial 3-week excess during a period of 52 weeks was <0.01.

P-surveillance (i.e., corresponding to the p-value) for a 3-week window starting at an arbitrary day cannot be determined exactly from this data. Using the weekly tabulations, the p-value is less than or equal to that associated with the 19 events in weeks 19–21. The p-value associated with the 19 cases in weeks 19–21 is

$$[(52/3)\ (19 - 5.18)]\ [\exp(-5.18)\ (5.18)^{19}/19!] = 239.2\ (0.00001182) = 0.0004$$

Thus, if surveillance were performed for a year, the chance of finding such a substantial excess, relative to the assumed expected values, is approximately 0.0004. This example is extreme, and no formal analysis might be required. Statistical significance at p<0.05 would be noted if 14 or 15 cases existed in the 3-week period.

Discussion and Conclusion

Two surveillance procedures associated with a set error rate over a period T are described. G-surveillance as described is a modification of a statistic used by others and implemented in SaTScan™ software (*11,26*). The procedure in this report differs from that previously implemented in terms of the function maximized, the events to which type I errors refer, and the logistics of implementation. G-surveillance, as described here, can have different properties by setting T to values of 21– 30 days somewhat akin to average run lengths proposed for implementation of CUSUM (*7*) or setting it to 1 year, which would result in substantially fewer false alarms but decreased sensitivity. A comparison with a statistic (e.g., CUSUM) using both data from real outbreaks and simulated data would identify the properties of the proposed statistics under both abrupt increases and gradual increases. For the latter scenario, CUSUM-like statistics might have superior properties over the methods proposed here.

Apparently, the P-surveillance method is new, but it potentially frees the investigator from performing any simulation. However, three caveats exist, as follows:

1. No reason exists to assume that P-surveillance is better than G-surveillance, although reasons might exist to prefer the reverse on the basis of presumed optimal properties of GLRT.
2. The p-values for P-surveillance are approximate, whereas those for the G-surveillance are exact. (The exactness to a given number of decimal places is attributable to performing enough simulations.)
3. G-surveillance is more flexible, its variants have been described extensively, and it has withstood multiple tests over time.

Nevertheless, the p-value computed by P-surveillance, using either the method described here for continuous surveillance or the ratchet scan for daily or weekly surveillance, should give an overall indication of the likelihood of observing a given excess over expected values in a certain time window, taking into account that the surveillance is performed for a specified period.

Acknowledgments

The two authors are listed in order of who presented the poster session on which this paper is based. The authors thank the reviewers; Martin Kulldorff, Ph.D., for his helpful comments; Sam Groseclose, M.D., David Williamson Ph.D., and Willie J. Anderson of CDC for providing the brucellosis data; and Elana Silver, M.S., for her comments on an early draft.

References

1. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11:141–50.
2. Lazarus R, Kleinman K, Dashevsky I, et al.Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis 2002;8:753–60.
3. Williamson GD, Hudson GW. A monitoring system for detecting aberrations in public health surveillance reports. Stat Med 1999;18:3283–98.
4. Stroup DF, Williamson GD, Herndon JL. Detection of aberrations in the occurrence of notifiable diseases surveillance data. Stat Med 1989;8:323–9.
5. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. Emerg Infect Dis 1997;3:395–400.
6. Rogerson P. Monitoring point patterns for the development of space-time clusters. J R Stat Soc A 2001;164:87–96.
7. Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. J R Stat Soc A 2003;166:5–12.
8. Frisén M. Statistical surveillance, optimality and methods. Int Stat Rev 2003;71:403–34.
9. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. Stat Med 1995;14:799–810.
10. Farrington CP, Andrews NJ, Beale AA, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. J R Stat Soc A 1996;154:61–74.
11. Kulldorff M. A spatial scan statistic. Communications in Statistics—Theory and Methods 1997;26:1481–96.
12. Kulldorff M, Feuer E, Miller B, Freedman L. Breast cancer in Northeastern United States: a geographical analysis. Am J Epidemiol 1997;146:161–70.
13. Kulldorff M, Athas WF, Feuer E, Miller B, Key C. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos. Am J Public Health 1998;88:1377–80.
14. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. J R Stat Soc A 2001;164:61–72.
15. Loader C. Large deviation approximations to the distribution of scan statistics. Advances in Applied Probability 1991;23:751–71.
16. Nagarwalla N. A scan statistic with a variable window. Stat Med 1966;15:845–50.
17. Naus J. The distribution of the size of the maximum cluster of points on a line. J Am Stat Assoc 1965;60:532–38.
18. Glaz J, Naus J, Wallenstein S. Scan statistics. New York, NY: Springer Verlag, 2001.
19. Glaz J, Balakrishnan N, eds. Scan statistics and applications. Boston, MA: Birkhäuser, 1999.
20. Naus J. Some probabilities, expectations, and variances for the size of the largest clusters and smallest intervals. J Am Stat Assoc 1966;61:1191–9.
21. Glaz J. Approximations and bounds for the distribution of the scan statistic. J Am Stat Assoc 1989;84:560–9.
22. Wallenstein S, Neff N. An approximation for the distribution of the scan statistic. Stat Med 1987;6:197–207.
23. Weinstock M. A generalized scan statistic for the detection of clusters. Int J Epidemiol 1981;10:289–93.
24. Wallenstein S, Weinberg C, Gould M. Testing for a pulse in seasonal event data. Biometrics 1989;45:817–30.
25. Chang M, Glynn MK, Groseclose SL. Endemic, notifiable bioterrorism-related diseases, United States, 1992–1999. Emerg Infect Dis 2003;9:556–64.
26. National Cancer Institute. SaTScan version 3.0: software for the space and space-time scan statistics. Bethesda, MD: National Cancer Institute, 2003.

Approaches to Syndromic Surveillance When Data Consist of Small Regional Counts

Peter A. Rogerson, I. Yamada
University at Buffalo, Buffalo, New York

Corresponding author: Peter A. Rogerson, University at Buffalo, Department of Geography, Wilkeson Hall, Buffalo, NY 14261. Telephone: 716-645-2722; Fax: 716-645-2329; E-mail: rogerson@buffalo.edu.

Abstract

Introduction: *Statistical systems designed for syndromic surveillance often must be able to monitor data received simultaneously from multiple regions. Such data might be of limited size, which would eliminate the possibility of using more common surveillance methods that assume data from a normal distribution.*

Objectives: *The objectives of this study were to design and illustrate a multiregional surveillance system based on data inputs consisting of small regional counts, where frequencies are typically on the order of ≤5.*

Methods: *Cumulative sum (CUSUM) methods designed for cumulating the sum of the deviations between observed and expected Poisson-distributed data were modified to account for changing expectations over time, including weekly and monthly effects. Data on lower respiratory tract infections during 1996–1999 at multiple Boston clinics among residents from 287 census tracts were used to illustrate the approach.*

Results: *When each region was monitored, 19% of the census tracts signaled a departure during 1999 from the base period (1996–1998) rates. When local statistics were used to monitor tracts and neighborhoods consisting of surrounding tracts, 60% of tracts experienced departures during 1999 from the base period. These results imply that the increases in lower respiratory tract infection that occurred during 1999 were geographically pervasive.*

Conclusions: *Poisson CUSUM methods are useful for monitoring small regional counts over time. The methods can be generalized to account for time-varying expectations in the counts.*

Introduction

Detecting the locations of statistically significant increases in the rates of health syndromes among multiple geographic areas as rapidly as possible is a critical public health need (*1*). Multiple systems are being designed to achieve this goal; comprehensive discussion of the desirable features of a statistical health surveillance system has been published previously (*2*). This paper focuses on two characteristics of such systems: 1) systems should be capable of detecting increases in regional rates quickly while keeping the number of false alerts at an acceptable level, and 2) observations might consist of limited frequencies that would necessitate the use of binomial or Poisson variables instead of normally distributed variables.

Multiple approaches to spatial surveillance in a public health context have been taken previously. One approach is to use cumulative sum (CUSUM) methods to monitor disease counts in geographic areas of interest (*3*). Another is to perform surveillance by detecting outliers in a temporal sequence of observed binomial variables for multiple geographic regions (*4*). Other investigators take existing spatial statistical methods used for retrospective detection of geographic clusters of disease and modify them for use in surveillance, which requires repeated tests for emergent clusters (*5–7*).

This paper uses and develops further a CUSUM approach for small counts (i.e., where frequencies are typically on the order of ≤5) assumed to follow a Poisson distribution. CUSUM methods cumulate deviations between observed and expected counts during a given period and generate an alert or signal when cumulated observed counts exceed expected counts by a predetermined threshold (*8*).

This paper reviews CUSUM methods for normal and Poisson-distributed variables. It then describes how to modify the Poisson CUSUM approach to allow the expected counts to vary from one period to the next. It also indicates how the approach can be used to monitor neighborhoods consisting of a set of contiguous regional units. These approaches are applied to data on lower respiratory infection episodes reported by Boston-area clinicians during January 1996–October 1999. The paper concludes with a discussion of findings.

CUSUM Methods

CUSUM methods are designed to detect sudden changes in the mean value of a quantity of interest; they are widely used in industrial process control to monitor production quality. The basic methods rely on two assumptions: 1) the quan-

tity being monitored is distributed normally, and 2) the variable exhibits no serial autocorrelation.

If the variable of interest is converted to a *z*-score with mean 0 and variance 1, the CUSUM, following observation *t*, is defined as follows:

$$S_t = \max(0, S_{t-1} + z - k)$$

where k is a parameter. A change in mean is signaled if $S_t > h$, where h is a threshold parameter.

Values of z in excess of k are cumulated. The parameter k in this instance, in which a standardized variable is being monitored, is often chosen to be equal to one half; in the more general case, k is often chosen to be equal to one half the standard deviation associated with the variable being monitored.

The parameter h is chosen in conjunction with a predetermined acceptable rate of false alerts; high values of h lead to a low probability of a false alert but also a lower probability of detecting a real change. The time between false alerts is the in-control average run length and is designated by the notation ARL_0. When $k = ½$, an approximation for ARL_0 is

$$ARL_0 = 2(e^a - a - 1)$$

where $a = h + 1.166$ (*9*). One can choose the parameter h by first deciding upon a value of ARL_0, and then solving the approximation for the corresponding value of h. This expression for the average run length can be solved, approximately, for h (P. Rogerson, University at Buffalo, unpublished data):

$$h \approx \left(\frac{ARL_0 + 4}{ARL_0 + 2}\right)\ln\left(\frac{ARL_0}{2} + 1\right) - 1.166$$

The choice of $k = ½$ minimizes the time required to detect a 1 standard-deviation increase in the mean. More generally, k is chosen to be equal to one half the size of the change (in units of standard deviations) sought for rapid detection. For this case (i.e., when k might take on a value other than one half)

$$h \approx \left(\frac{2k^2 ARL_0 + 2}{2k^2 ARL_0 + 1}\right)\frac{\ln(1 + 2k^2 ARL_0)}{2k} - 1.166$$

CUSUMs for Poisson Variables

When the assumption of normality is not a good one, transformations to normality are sometimes possible. One such normalizing transformation for data consisting of small counts is (*10*):

$$y = \frac{x - 3\lambda + 2\sqrt{\lambda x}}{2\sqrt{\lambda}}$$

where x is the observed count and λ is the expected count.

This transformation can be misleading for small values of λ. In particular, the actual ARL_0 values might differ substantially from the desired nominal values. For example, when desired values of $ARL_0 = 500$ and $ARL_1 = 3$ (where ARL_1 is the average time taken to detect an increase) are used in situations where $\lambda < 2$, simulations demonstrate that using this transformation will almost always yield actual values of ARL_0 substantially lower than the desired value of 500. In certain cases (e.g., $\lambda \approx 0.15$), the actual ARL will be <100, indicating a much higher rate of false alerts than desired. The performance is better when $ARL_0 = 500$ and $ARL_1 = 7$, but use of the transformation will again lead to substantially more false alerts than desired when λ is less than approximately 0.25. Also troubling is the instability with respect to similar values of λ; $\lambda = 0.56$ will lead to an ARL_0 of approximately 400, whereas $\lambda = 0.62$ is associated with an ARL_0 of >700. This is also true when $ARL_1 = 3$; $\lambda = 0.96$ has an ARL of approximately 212, whereas $\lambda = 0.98$ has an ARL of 635.

When the variable being monitored has a Poisson distribution, the CUSUM is

$$S_t = \max(0, S_{t-1} + X_t - k)$$

New considerations are necessary to determine the parameters k and h (*12*). If λ_0 is the mean value of the in-control Poisson parameter, the k-value that minimizes the time to detect a change from λ_0 to a prespecified out-of-control parameter λ_1 is

$$k = \frac{\lambda_1 - \lambda_0}{\ln\lambda_1 - \ln\lambda_0} \quad (1)$$

Then, h can be determined from the values of the parameter k and the desired ARL_0 by using either a table (*11*), Monte Carlo simulation, or an algorithm that makes use of a Markov chain approximation (*12*).

Poisson CUSUM methods have been applied previously in a public health context, primarily in surveillance of congenital malformations (*13,14*); the approach has also been recommended in surveillance for *Salmonella* outbreaks (*15*).

Poisson CUSUM Methods with Time-Varying Expectations

The expected in-control value associated with the Poisson variable might vary with time ($\lambda_{0,t}$; t = 1, 2, ...) (e.g., as a result of seasonal effects). Simply implementing a CUSUM scheme with constant parameters would have misleading results if the actual values of λ_0 fluctuated from period to period about the constant assumed parameter. Instead, time-specific values of the parameters k and h were used. The observed values, X_t, were then used in the CUSUM as follows:

$$S_t = \max[(0, S_{t-1} + c_t (X_t - k_t)] \quad (2)$$

where the parameters c_t and k_t change from one period to the next, and their values are now discussed.

First h is chosen on the basis of the mean of the time-varying Poisson parameter, an associated value of k, and the desired ARL_0. Once h is chosen, next choose k_t on the basis of $\lambda_{0,t}$ and $\lambda_{1,t}$,

$$k_t = \frac{\lambda_{1,t} - \lambda_{0,t}}{\ln \lambda_{1,t} - \ln \lambda_{0,t}} \quad (3)$$

Then, c_t is chosen as the ratio h to h_t, where the latter is the value of the threshold associated with the desired ARL_0, k_t, and constant values of $\lambda_{0,t}$ and $\lambda_{1,t}$. Thus, $c_t = h/h_t$. The quantity c_t is chosen so that observed counts X_t will make the proper relative contribution toward the signaling parameter h that is used in the actual CUSUM. If, for example, $h > h_t$, then the contribution $X_t - k_t$ is scaled up by the factor h/h_t. An alternative approach is to apply a multiplicative factor to the baseline, or average value of λ (*16*).

Poisson CUSUM Methods for Neighborhoods Consisting of Contiguous Regional Units

An extension is to construct local statistics in association with each geographic unit. These are defined as a weighted sum of the region's observation and surrounding observations, where the weights could decline with increasing distance from the region. CUSUMs associated with these local statistics would be monitored. Local statistics are spatially autocorrelated, and Monte Carlo simulation of the null hypothesis can be performed to determine appropriate thresholds for the CUSUMs if no deviation from expected values of the Poisson parameters exists.

Application to Boston Data on Lower Respiratory Infection

Data

Harvard Vanguard Medical Associates (Boston, Massachusetts) uses an automated record system for its 14 clinics. After each patient office visit, the clinician records diagnoses and *International Classification of Disease, Ninth Revision* (ICD-9) codes. Patient addresses are recorded; these have been geocoded and assigned to census tracts.

Data on lower respiratory infection episodes were available for January 1996–October 1999. During this period, 47,731 episodes occurred that could be assigned to one of the 287 census tracts in the study region.

Model for Expected Counts

The first 3 years of data (January 1996–December 1998) were used to calibrate logistic regression models for each census tract. The logistic transform of the probability of a visit is taken to be a linear function of the explanatory variables:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_m x_{mi} + \varepsilon_i$$

where p_i is the probability of a visit in region i; x_{li} is the value of explanatory variable l in region i; and the βs are the regression coefficients; m explanatory variables and m + 1 coefficients are estimated in each tract.

Compared with the random-effects model described previously (*4*), this modeling approach has coefficients that are specific to individual regions. However, constructing a model for each region might result in region-specific coefficients that might not be reliable over time, especially when they are estimated from a limited number of observations. An alternative might be to have region-specific dummy variables in a single equation, but this could use a substantial number of degrees of freedom relative to the number of observations.

In each census tract, the unit of observation was the day. During the 3-year base period (i.e., 1,096 days), expected counts on each day were modeled as a function of time trend (i.e., the logistic transform of the probability of a visit was taken to be a linear function of the day number). Eleven dummy variables were created for the months of the year; December was taken as the arbitrary, omitted category. Finally, a dummy variable was also included for visits that occurred on weekends, with weekday observations as the reference category. Another potential variable capturing temporal autocorrelation in the counts was also considered, but in the majority of cases it was not significant. Inclusion of such a variable would be a way to address violations of the assumption of independence in the CUSUM method.

The average coefficient for each of the explanatory variables (in which the average is taken over the 287 census tracts) is provided (Table 1). Visits are most likely in December; the probability of visits declines steadily thereafter until July. In August, the probability of a visit begins to increase, until reaching its maximum in December. The likelihood of weekend visits is substantially lower than weekday visits, as expected. Finally, the average time trend is positive.

TABLE 1. Average coefficients in logistic regression model, by month* and day†

Variable	Coefficient
January	–0.211
February	–0.472
March	–0.692
April	–0.976
May	–1.113
June	–1.527
July	–2.039
August	–1.419
September	–1.369
October	–0.657
November	–0.270
Weekend	–1.262
Day	0.00345
Intercept	–7.640

* December is the omitted reference month.
† Refers to the time trend; the coefficient indicates the daily increase in the log-odds of a visit.

Poisson CUSUM Method

For an illustration of how the modified CUSUM approach might be applied, the estimated parameters for each tract were used, together with the relevant explanatory variables, to derive the expected probability of a visit for each day, for each census tract, for the 303-day period beginning January 1, 1999. These expected probabilities were multiplied by the number of patients in each tract on each day to derive the expected number of visits on each day. The latter quantity is the time-varying, in-control Poisson parameter, $\lambda_{0,t}$. To minimize the time to detect a one half standard-deviation change in this parameter, the out-of-control Poisson parameter is chosen to be

$$\lambda_{1,t} = \lambda_{0,t} + \frac{1}{2}\sqrt{\lambda_{0,t}}$$

Although minimizing the time to detecting a one standard-deviation change is probably more common, one half of a standard deviation is used here because the standard deviation is so large relative to the mean. For example, when

$$\lambda_{0,t} = 0.1,\ \sqrt{\lambda_{0,t}} = 0.32$$

for detecting a 1 standard-deviation change,

$$\lambda_{1,t} = 0.1 + 0.32 = 0.42$$

and for detecting a one half standard-deviation change,

$$\lambda_{1,t} = 0.1 + 0.16 = 0.26$$

An overall probability of 0.05 was desired for an alert, under the null hypothesis of no change in the visit probabilities. In addition, because 287 CUSUMs are being tested simultaneously, adjustment is needed for multiple testing (because 287×303 values of the CUSUM are examined). A Bonferroni adjustment can be made by using 287×303 instead of 303 in the run-length calculations. In particular, because run lengths have an exponential distribution (*17*), $p(\text{run length} < 287 \times 303 = 1 - \exp(-287 \times 303 \times \mu) = 0.05$, which implies an average run length of $1/\mu$ =1,695,366.

Next, the value of the tract-specific threshold (h) that is consistent with this average run length and with the tract-specific values of λ_0 and k was determined by using an algorithm described elsewhere (*13*). Then, time-varying tract-specific values of h_t were determined by either of the following methods:

1. If $\lambda_{0,t}$ was close to any of the average values of λ_0, the associated value of h was adopted; if not,
2. A regression equation relating h and λ was estimated by using the 287 average values of λ_0 and the 287 associated values of h. The regression equation was

$$h = 8.18 + 32.04\lambda$$

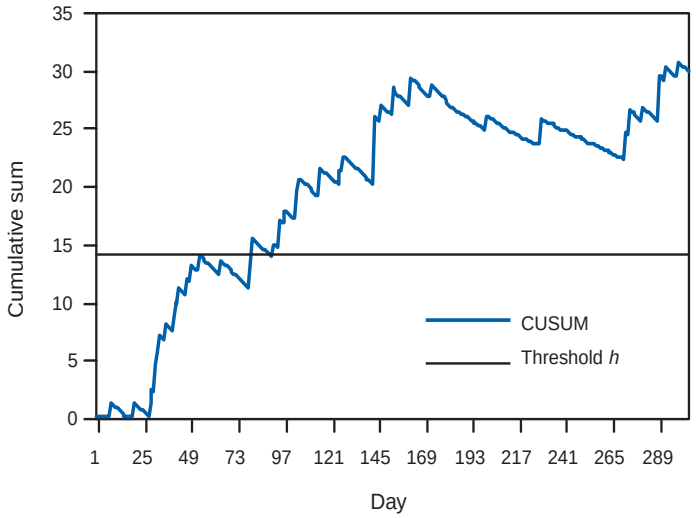
The Poisson CUSUM (equation 2) was then started for each tract on January 1, 1999, by using the observed number of daily visits, the expected number of daily visits ($\lambda_{0,t}$), and values of h, h_t, k, and k_t, as described previously.

Results

Of 287 census tracts, 58 (19%) had ≥1 signal during the 303-day monitoring period. In 19 (37%) tracts, the signals were short-term and continued no longer than 30 days. Of the remaining 39 tracts with signals, the majority were either sustained for approximately the latter half of the monitoring period (12 tracts) or characterized by a rapid increase in the CUSUM near the end of the monitoring period (14 tracts).

Tract 26 had an average 0.111 cases/day during the 3-year base period, which increased to an average of 0.145 cases/day during the monitoring period (Figure 1). The initial increase in the CUSUM began in late January. Cases were observed on January 28, 29, and 31; additional cases were observed on February 1, 2, and 3. Thirteen cases were observed during a 27-day period that began on January 28, for an average of 0.481 cases/day, substantially higher than the baseline of 0.111 cases/day. The CUSUM continued to increase until June, indicating a sustained period of higher-than-average visitation rates, and then declined slightly until September.

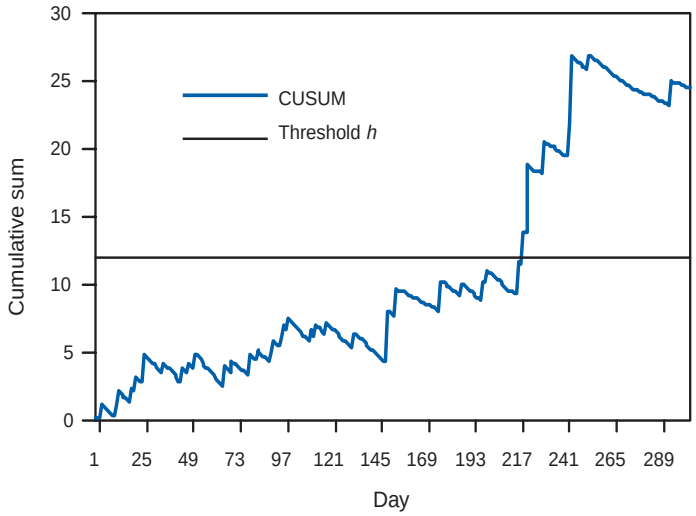
FIGURE 1. Cumulative sum (CUSUM) chart for lower respiratory infection episodes — Census Tract 26, Boston, Massachusetts, January–October 1999



During the base period, tract 83 had an average of 0.120 cases/day; this rose to 0.135 cases/day during 1999 (Figure 2). Cases leading to the alert occurred on August 4, 6, and 9 (two cases were observed on August 9). These four cases in 6 days (0.67 cases/day) were sufficient to generate an alert, particularly because the CUSUM had been increasing slowly during the preceding months.

During the calibration period (1996–1998), 33.4 cases/day occurred in the study region; during the first 303 days of 1999, an average of 36.8 cases/day occurred. The daily increase was >10%, and this is easily picked up by the CUSUMs in multiple subregions.

FIGURE 2. Cumulative sum chart for lower respiratory infection episodes — Census Tract 83, Boston, Massachusetts, January–October 1999



Results for Monitoring Regional Neighborhoods

Neighborhoods consisting of each individual region and its immediately adjacent neighboring regions were monitored to illustrate the surveillance of local regional statistics. Of 287 census tracts, 173 (60%) had at least one signal during the monitoring period, and 43 also signaled under the original Poisson CUSUM. Among the 173 signaling tracts, 90 (52%) sustained signals for the latter half of the monitoring period, and 25 (14%) witnessed rapid increases in their CUSUMs near the end of the monitoring period. The distribution of regions that had CUSUMs above the threshold on the last day of the monitoring period (i.e., day 303), under both the original Poisson CUSUM and the local statistics CUSUM, is illustrated (Figure 3). More regions signal when the local statistic is used; here the search for spatial patterns occurs on a broader geographic scale. The northern, southwestern, and southeastern portions of the study area emerge as subareas that deviate substantially from baseline expectations established during 1996–1998.

The statistical significance of the local statistics was derived by using a Bonferroni correction for the number of regions. This is conservative because the local statistics are correlated. Monte Carlo simulations were conducted by using 30- and 100-region subsets of the original study area to determine more appropriate thresholds for the local statistics CUSUM. To achieve a 0.05 probability of a false alert during the 303-day monitoring period, the target ARL_0 under the null hypothesis can be calculated by using

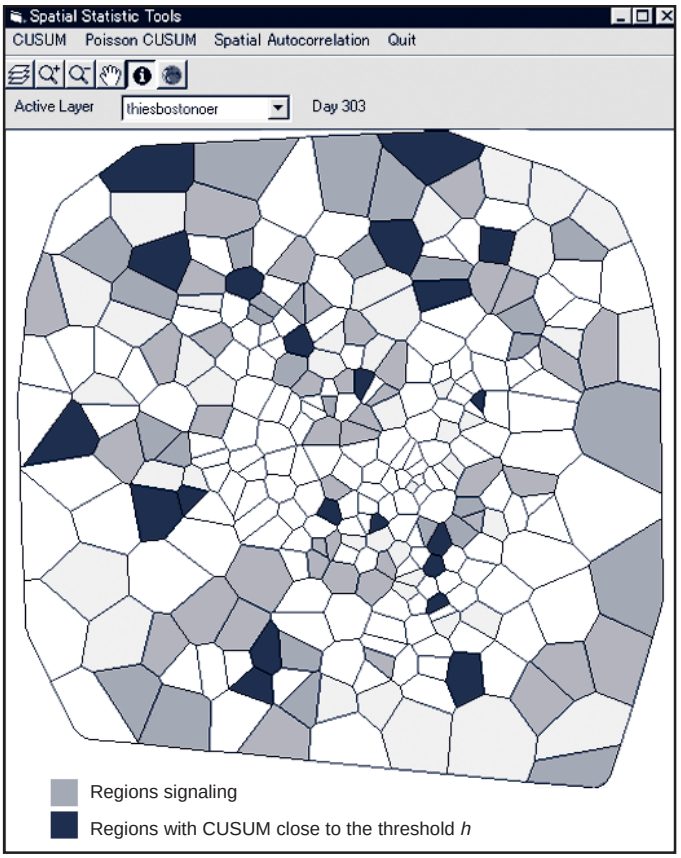
$$p(\text{run length} < 303\times n) = 1 - e^{(-303\times s\times \mu)} = 0.05$$

where n is the number of regions and $\mu = 1/ARL_0$. For multiple values of s, the target ARL_0 was calculated and the corresponding CUSUM parameters were obtained. The false-alert rates obtained by the simulations under the null hypothesis are provided (Table 2). Apparently, the appropriate value of s is 50%–60% of the number of regions n when the neighborhood is defined by the binary adjacency described previously. Using different definitions of the neighborhood would change the appropriate value of s.

On the basis of this result, local statistics CUSUM analysis was conducted on the Boston data by using $s = 160$, which is approximately 55% of the total number of tracts. This time, 183 census tracts, 10 tracts more than before, had at least one signal during the monitoring period, but no change was noted in terms of the day and the tract of the first signal.

FIGURE 3. Distributions of regions that signaled on day 303 of the monitoring period, indicating lower respiratory infection episodes — Boston, Massachusetts, January 1–October 30, 1999

Poisson cumulative sum (CUSUM) for individual regions



Poisson CUSUM for local statistics

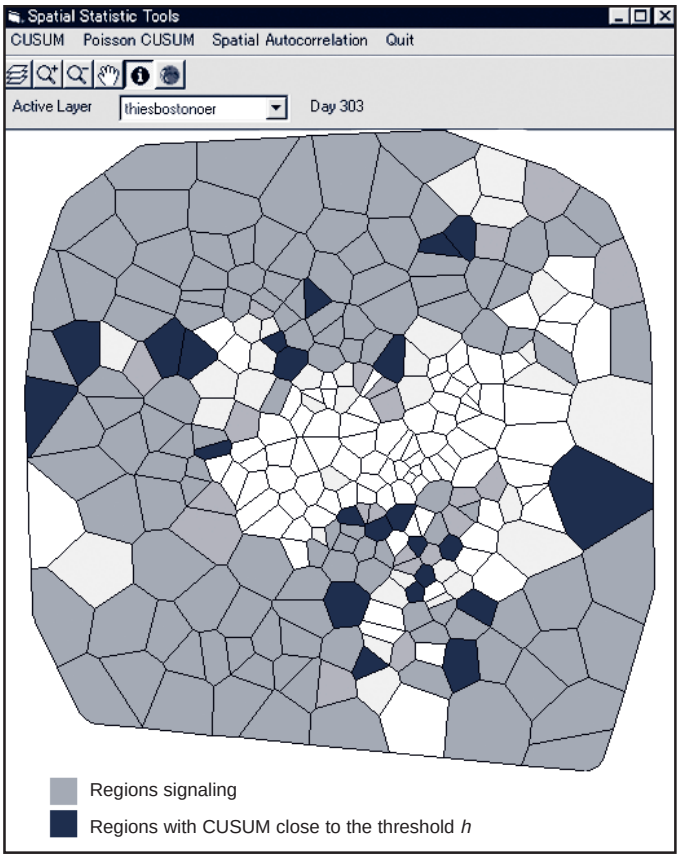


TABLE 2. False-alert rates* simulated under the null hypothesis

30 regions			100 regions		
s†	ARL_0§	Signaling probability	*s*	ARL_0	Signaling probability
30	177,216	0.024	100	590,720	0.035
20	118,144	0.038	60	354,432	0.048
15	88,608	0.055	50	295,360	0.056
10	59,072	0.075			

* Average: >4,000 trials.
† *s* = number of effectively independent regions.
§ ARL_0 = average run length, or time between false-alerts under the null hypothesis.

Discussion

This paper demonstrates how the Poisson CUSUM can be used in the context of spatial surveillance. In particular, it focuses on two developments: 1) an extension to allow the use of Poisson CUSUM methods when expectations vary over time, and 2) an extension along lines originally discussed previously (*3*) that permits monitoring of CUSUMs in subregions and their surrounding neighborhoods. Software for the Poisson CUSUM method is available at http://wings.buffalo.edu/~rogerson.

An important question raised by the implementation of these methods in the context of public health surveillance is whether accurate expectations of disease counts can be formed. To the extent that expected counts are not well-modeled, the CUSUM tends to increase, and alerts caused by deviations from expectations will be attributable more to inability to model expectations and less to any real public health problem.

The methods are ultimately better suited for certain public health problems than for others. For example, for certain biologic agents, a single case is sufficient to generate an alert, and a sophisticated statistical system is not needed. In other situations, monitoring symptoms might reveal patterns that would otherwise remain hidden in the data. In the 1993 gastroenteritis outbreak in Milwaukee, a substantial number of cases went unnoticed for an extended period (*18*); quick detection of spatial patterns in symptoms might have allowed a quicker public health response.

Acknowledgments

This work was funded by National Institutes of Health Grant 1R01 ES09816-01 and National Cancer Institute Grant R01 CA92693-01. Ken Kleinman (Department of Ambulatory Care and Prevention, Harvard Medical School, Harvard Pilgrim Health Care, and Harvard Vanguard Medical Group) provided the data on lower respiratory infection episodes. Two reviewers also provided helpful comments.

References

1. Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. J R Stat Soc A 2003;166:5–21.
2. Farrington P, Beale AD. The detection of outbreaks of infectious disease. In: Lierl L, Cliff AD, Valleron A, Farrington P, Bull M, eds. Geomed '97. Stuttgart: BG Teubner, 1998.
3. Raubertas RF. An analysis of disease surveillance data that uses the geographic locations of the reporting units. Stat Med 1989;8:267–71.
4. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease: biological terrorism and other surveillance. Am J Epidemiol 2004;156:217–24.
5. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. J R Stat Soc A 2001;164:61–72.
6. Rogerson P. Surveillance methods for monitoring the development of spatial patterns. Stat Med 1997;16:2081–93.
7. Rogerson P. Monitoring point patterns for the development of space-time clusters. J R Stat Soc A 2001;164:87–96.
8. Montgomery D. Introduction to statistical quality control. New York, NY: John Wiley, 1996.
9. Siegmund D. Sequential analysis: tests and confidence intervals. New York, NY: Springer-Verlag, 1985.
10. Rossi G, Lampugnani L, Marchi M. An approximate CUSUM procedure for surveillance of health events. Stat Med 1999;18:2111–22.
11. Lucas JM. Counted data CUSUMs. Technometrics 1985;27:129–44.
12. White CH, Keats JB. ARLs and higher order run length moments for Poisson CUSUM. Journal of Quality Technology 1996;28:363–9.
13. Hill GB, Spicer CC, Weatherall JAC. The computer surveillance of congenital malformations. BMJ 1968;24:215–8.
14. Weatherall JAC, Haskey JC. Surveillance of malformations. BMJ 1976;32:39–44.
15. Hutwagner LC, Maloney EK, Bean, NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. Emerg Infect Dis 1997;3:395–400.
16. Hawkins DM, Olwell DH. Cumulative sum charts and charting for quality improvement. New York, NY: Springer, 1998.
17. Brook D, Evans DA. An approach to the probability distribution of CUSUM run length. Biometrika 1972; 59:539–49.
18. Eisenberg JN, Seto EY, Colford JM Jr, Olivieri A, Spear RC. An analysis of the Milwaukee cryptosporidiosis outbreak based on a dynamic model of the infection process. Epidemiol 1998;3:228–31.

Algorithm for Statistical Detection of Peaks — Syndromic Surveillance System for the Athens 2004 Olympic Games

Urania G. Dafni,[1,2] S. Tsiodras,[1,2] D. Panagiotakos,[1] K. Gkolfinopoulou,[1,2] G. Kouvatseas,[1] Z. Tsourti,[1] G. Saroglou[2]
[1]*University of Athens, Athens, Greece;* [2]*Hellenic Center for Infectious Diseases Control, Ministry of Health, Athens, Greece*

Corresponding author: Urania G. Dafni, Division of Public Health, Department of Nursing, University of Athens, 123 Papadiamantopoulou St., Athens 11527, Greece. Telephone: 30-210-746-1456; Fax: 30-210-746-1489; E-mail: dafni@cc.uoa.gr.

Abstract

Introduction: *No generally accepted procedure exists for detecting outbreaks in syndromic time series used in the surveillance of natural epidemics or biologic attacks.*

Objectives: *This report evaluates the usefulness for syndromic surveillance of the Pulsar approach, which is based on removing long-term trends from an observed series and identifying peaks in the residual series of surveillance data with cutoffs determined by using a combination of peak height and width.*

Methods: *Simulations were performed to evaluate the Pulsar method and compare it with other approaches. The daily syndromic counts in emergency departments of four major hospitals in the Athens area during August 2002–August 2003 were analyzed for two common syndromes. A standardized residual series was generated by omitting trends and noise in the original data series; this series was examined for the presence of peaks (i.e., points having magnitude higher than at least one of three probabilistically determined cutoffs). The whole process was iterated, and the baseline was recalculated by assigning reduced weight to the identified peaks.*

Results: *For the specific simulation schema used, the Pulsar method fared well when compared with other approaches in meeting the performance criteria of sensitivity, specificity, and timeliness.*

Conclusions: *Although the suggested algorithm needs further validation regarding the correspondence between detected peaks and true biologic alerts, the Pulsar technique appears effective for observing peaks in time series of syndromic events. The simplicity of the algorithm, its ability to detect peaks based not only on height but also on width, and its performance in the simulated data sets make it a promising candidate for further use in syndromic surveillance.*

Introduction

Syndromic time series are used in surveillance of natural epidemics or biologic attacks. CDC and the New York City Department of Health and Mental Hygiene used syndromic surveillance systems for detection of biologic terrorism after the September 2001 terrorist attacks (*1–3*). Almost simultaneously, other systems emerged (*4*), including those developed by the Boston Department of Health (*5*) and the University of Pittsburgh (*4,6*), CDC's drop-in surveillance systems (*7*), the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) (*8,9*), and others (*10–12*).

The Athens 2004 Olympic Games (August 13–29, 2004) have made critical the need for a real-time surveillance system that can alert public health officials to unexpected communicable-disease outbreaks and likely clinical presentations of a biologic terrorist attack, as has been used for other major athletic events (*13–17*). Therefore, in July 2002, a drop-in syndromic surveillance system was established in Greece similar to that used during the Salt Lake 2002 Olympic Winter Games (*13,18*).

Different outbreak-detection algorithms are used in operating syndromic surveillance systems (*2,19–23*). Ideally, all alert mechanisms generate an alert whenever the number of observed events exceeds the expected number of events while minimizing the frequency of false alerts. However, no generally accepted procedure exists for outbreak detection in syndromic surveillance (*24*). This paper proposes an algorithm for statistical detection of peaks. The method is based on removing long-term trends from the series of observations and identifying peaks in the residual series of data. This approach was developed for studying episodic hormonal secretion and has been used for other applications (*25–27*). An important feature of the proposed algorithm is that it generates alerts, taking into consideration both height and breadth of signals. The proposed method was applied in the Athens Olympic syndromic surveillance system database (*18*) and was compared through simulations with other methods currently applied in syndromic data series (*19–23,28*).

Methods

Data Acquisition

Drop-in syndromic surveillance in emergency departments (EDs) of major hospitals was first established in Greece by the Hellenic Center for Infectious Diseases Control in July 2002. The project's primary aims were to assess system feasibility and data-collection timeliness, establish a 2-year background database, and enhance collaboration with and sensitization of ED personnel of major hospitals (*18*). During August 2002–August 2003, the syndromic surveillance system operated in eight hospitals and one major health-care center in the greater Athens area. Surveillance was conducted for the following 10 syndromes: 1) *respiratory infection with fever*; 2) *bloody diarrhea*; 3) *gastroenteritis (diarrhea, vomit) without blood*; 4) *febrile illness with rash*; 5) *meningitis, encephalitis, or unexplained acute encephalopathy/delirium*; 6) *suspected acute viral hepatitis*; 7) *botulism-like syndrome*; 8) *lymphadenitis with fever*; 9) *sepsis or unexplained shock*; and 10) *unexplained death with history of fever*. These syndrome categories were used by the Salt Lake City Department of Health for syndromic surveillance during the 2002 winter Olympics (*13*). Trained personnel visited EDs and identified syndromic cases from chief complaints as recorded in ED visit books. All syndromes identified daily in the ED were recorded, as were the total number of visits. Data were entered into a database, and data management and analysis were performed centrally. In the work presented here, the time series for the two most commonly encountered syndromes (*respiratory infection with fever* and *gastroenteritis* [*diarrhea, vomit*] *without blood*) are used.

Algorithm Description

The Pulsar method is based on identifying peaks in the syndromic time series that exceed a specified threshold. Long-term changes are first screened out, and then peaks are identified in the screened series. This approach has been previously suggested for studying episodic hormonal secretion (*25*).

First, a baseline is defined for the original syndromic series by using the locally weighted smoothing scatterplots method (LOWESS) (*29*), in which a fixed proportion of observations (the smoothing parameter) is used, and a baseline value is calculated from the observations closest in time to the point. Weights are assigned to the observations, depending on their distance from the point. The fraction of observations in the window is selected so that the window's average width minimizes the bias-corrected Akaike's information criterion (AIC), which incorporates both the tightness of the fit and the model complexity. This criterion often selects better models than AIC in small samples (*30*). Then, a weighted nonparametric regression of syndromic counts versus time within the window provides the initial baseline value estimate for that time point. After the initial estimation of baseline values, new weights giving less influence to observations far from the corresponding baseline values are assigned, and the weighted regression is repeated. This procedure produces baseline estimates that are not influenced by extreme outlier observations.

A residual series, containing short-term variations but not trends, is obtained by subtracting the smoothed data from the original counts and is standardized by dividing the residuals by an estimate of the noise level, to yield a scaled residual series, expressed in signal-to-noise units. The peaks in the standardized residual series are identified on the basis of a combination of height and width, with no assumption for the shape of the peak. To be classified as a peak, an elevation should either be substantially high, even if it is narrow, or span multiple points in width, even if it is moderately high. For a point in the signal-to-noise series to be considered part of a peak, it should exceed a certain cutoff value G(1); or it should exceed a lower cut-off value G(2) along with one adjacent point; or it should exceed an even lower cut-off value G(3) along with two adjacent points; and so forth. The specific choices of n and G(n)s depend on the time series used for calibration purposes, the relative choice between higher but narrow peaks as opposed to lower but broad ones, and the desired false-alert rate. After the initial identification of peaks, the baseline is recalculated. Reduced weight is assigned to observations previously identified as part of a peak. Iterations of the whole process are performed until the same assignment of points to peaks is achieved.

Algorithm Customization

In the 13-month syndromic series, LOWESS smoothing was applied with optimal smoothing parameter equal to 15% for *respiratory infection with fever* and 52% for *gastroenteritis (diarrhea, vomit) without blood*. Alternative estimates were used for the standardization, including the standard deviation and the mean absolute deviation in the original series, as well as the 7-day moving standard deviation and the 7-day mean absolute deviation in the simulated series. The latter were based either on the seven most recent observations to the current time point or on the tenth to fourth most recent observations (i.e., not taking into account the three most recent ones). The procedure is performed iteratively to weigh down extreme values and detect outliers appearing in clusters. In this data set, extreme clustered observations do not appear to exist, and two iterations were sufficient to obtain a smoothed series (the resulting detected peaks of the two iterations differ by <2.5%). G(1), G(2), and G(3) cutoffs were chosen under the assump-

tion of normality for the standardized residual series to derive 97% specificity in the whole series and take into account the effect of multiple testing on the significance level. The threshold is given by G(n) = probit(1-α^*[n/6]/d) where d = number of days that a false alert occurs with probability α = 0.10 and n = 1 or 2 or 3, whereas the factor n/6 provides the necessary adjustment for multiple testing (*25*).

Alternative Methods

The Pulsar approach was evaluated by comparison through simulations with other commonly used syndromic surveillance methods (*19–23,28*). All parameters for each model used in the comparisons were set so that the specificity (true nonalerts/nonoutbreaks) in the original time series was fixed at 97%, assuming no outbreak condition (*20,21*). For each method, the day of an outbreak on which an alert was first generated was recorded. Sensitivity (true alerts/outbreaks) across all simulated series for each syndrome and the timeliness for each method (i.e., the percentage of the first alert per day of outbreak) were compared among the alternative approaches. The three performance criteria (sensitivity, specificity, and timeliness) were reported and compared through the Wilcoxon signed rank or Friedman nonparametric tests. Bonferroni-adjusted α^* are reported. The methods mentioned here have been used in syndromic surveillance and were evaluated in this syndromic data series.

The temporal aberration detection (TAD) approach used by the Early Aberration Reporting System (EARS), a program provided by CDC to all interested health departments, uses cumulative sum (CUSUM) methods from the quality-control literature. CUSUM compares the proportion of syndrome counts to total visits on each of the most recent 3 days to the mean proportion plus 1 standard deviation, during a 7-day moving baseline. CUSUM of positive differences is calculated based on a 3-day interval, and an alert is considered to occur if it exceeds 2 standard deviations (*2,22,23*). Time-series methods (e.g., autoregressive integrated moving average [ARIMA] time-series models) were proposed for describing 10-year syndromic data from a major Boston-area hospital (*20,21*). Different filters were evaluated in data sets with simulated outbreaks, using a fixed specificity rate of 97%. The linear 7-day filter proved superior in simulations (*20*). Standard one-sided CUSUM methods have also been proposed for detecting outbreaks in surveillance data (*19,28*).

Simulation Schema

For evaluation of the performance of the proposed methodology, 100 simulated series were created. The original time series of counts is considered to include no outbreaks. A scenario involving a terrorist attack depends on the biologic agent, quality, and quantity released; the method of dispersion; and population characteristics. A 4-day outbreak was chosen to represent a probable period between symptom presentation and diagnosis (i.e., the window of opportunity for possible earlier detection because of syndromic surveillance) (*31*). However, different durations of that window are also possible.

Each simulated time series was produced by randomly injecting 4-day–long outbreaks to the original time series of daily counts for each syndrome of interest with probability of 15% per day (leading to 18.5 4-day outbreaks on average among simulated series). An outbreak led to duplication of the observed counts of the syndrome for that day (respiratory infection: median size = 27; 5th and 95th percentiles: 24, 29 and gastroenteritis: median size = 15; 5th and 95th percentiles: 14, 16). Two adjacent outbreaks were forced to be ≥15 days apart to ensure that a previous outbreak did not adversely affect the alert-detection mechanism of the next (*20*). The detection algorithms should detect an outbreak as if it is the first one that occurs in the original time series.

An outbreak was considered successfully detected if an alert was generated on ≥1 day of the outbreak. Alternative patterns of outbreaks were also examined, including 1) constant increase for all 4 days, equal to the median counts of the syndrome (23.5 for *respiratory infection with fever* and 15.5 for *gastroenteritis*) or 2) constant increase for all 4 days, equal to the 75th percentile of the counts of the syndrome (35 for *respiratory infection with fever* and 22 for *gastroenteritis*); 3) linear increase for the 4 days: (increase of one median/day); 4) exponential increase for the 4 days: increase of 1, 1.5, 2.5, and 4 medians for day 1–4, respectively; or 5) exponential increase for the first 3 days (1, 1.5, 2.5 medians) and subsequent decrease on day 4. All statistical computations were performed by using SAS® software, version 8.2 (*32*).

Results

The original 13-month time series for four major hospitals in metropolitan Athens sharing the same catchment area for the *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromes were used to illustrate and evaluate the proposed method. A total of 305,039 ED visits (mean: 770/day) were recorded during August 2002–August 2003 in these hospitals. The corresponding mean total syndrome counts were 26 and 15 per day for each syndrome, respectively.

The six different standardization estimates already described for the Pulsar algorithm, leading to different threshold speci-

fications, were compared. The best approach for both syndromes with respect to the achieved sensitivity was the one that used the standard deviation in the original series (Table 1; see Model 1). The corresponding parameter *d* to the G(1), G(2), and G(3) thresholds was 0.5 and 1 for *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromic series, respectively, whereas α was set equal to 0.10 for both syndromes. The standardized residuals from the original time series and from a sample simulated series (number 10) for each syndrome along with the thresholds are illustrated (Figure 1).

The TAD approach was used both on the count series and on the proportion of counts of syndromes to total ED visits (*2,22,23*). The results for the count series were superior to those for the proportion series. The fixed specificity of 97% in the original time series of counts, for both syndromes, was reached by using 3 standard deviations for the alert mechanism (Model 1) instead of the 2 used in EARS (Model 2) (*2,22,23*) (Table 2).

ARIMA models were used for describing the 13-month original series of syndromic data (*20,21*). For *respiratory infection with fever*, the autoregressive (AR) order, the moving average (MA) order, and the integration (I) order were all equal to 4 days. Weekend was also statistically significant and used as an explanatory variable in the model. For *gastroenteritis (diarrhea, vomit) without blood*, AR = 4 days, MA = 2 days, and I = 4 days. The filters evaluated were seventh-order MA (Model 1), seventh-order linear average (Model 2), and seventh-order exponential average (Model 3). The threshold was again set so that specificity of 97% was achieved in the original time series, and the best filter regarding sensitivity was the seventh-order MA filter (Model 1). The corresponding thresholds are equal to probit(1-α/7) and probit(1-α/8) for each syndrome, respectively, with α equal to 0.10.

For the one-sided CUSUM method used here, a 7-day moving average and standard deviation used for standardization proved superior to the standard approach (*19,28*). The cumulative sum was calculated by $S_t = \max\{0,(S_{(t-1)}+z_t-k)\}$, where $k = 0.5$. The specified threshold h was set so that specificity 97% was achieved in the original time series, and the corresponding values for the two syndromes were set to 3.5 and 2.75, respectively (Model 1). A second approach employing values from the literature that actually minimize the average run length (ARL) of the process was also used ($k = 0.5$ and $h = 2.5$) (*19*) (Table 2).

The sensitivity and specificity of the alternative methods (TAD, the time-series approach, and CUSUM) were compared (Table 2). Performance criteria for the best models with respect to sensitivity for each approach, among the ones using a set specificity of 97% in the original time series, are directly compared (Figures 2 and 3). Box-plots of the model's sensitivity and specificity (Figure 2) and timeliness (Figure 3) are presented.

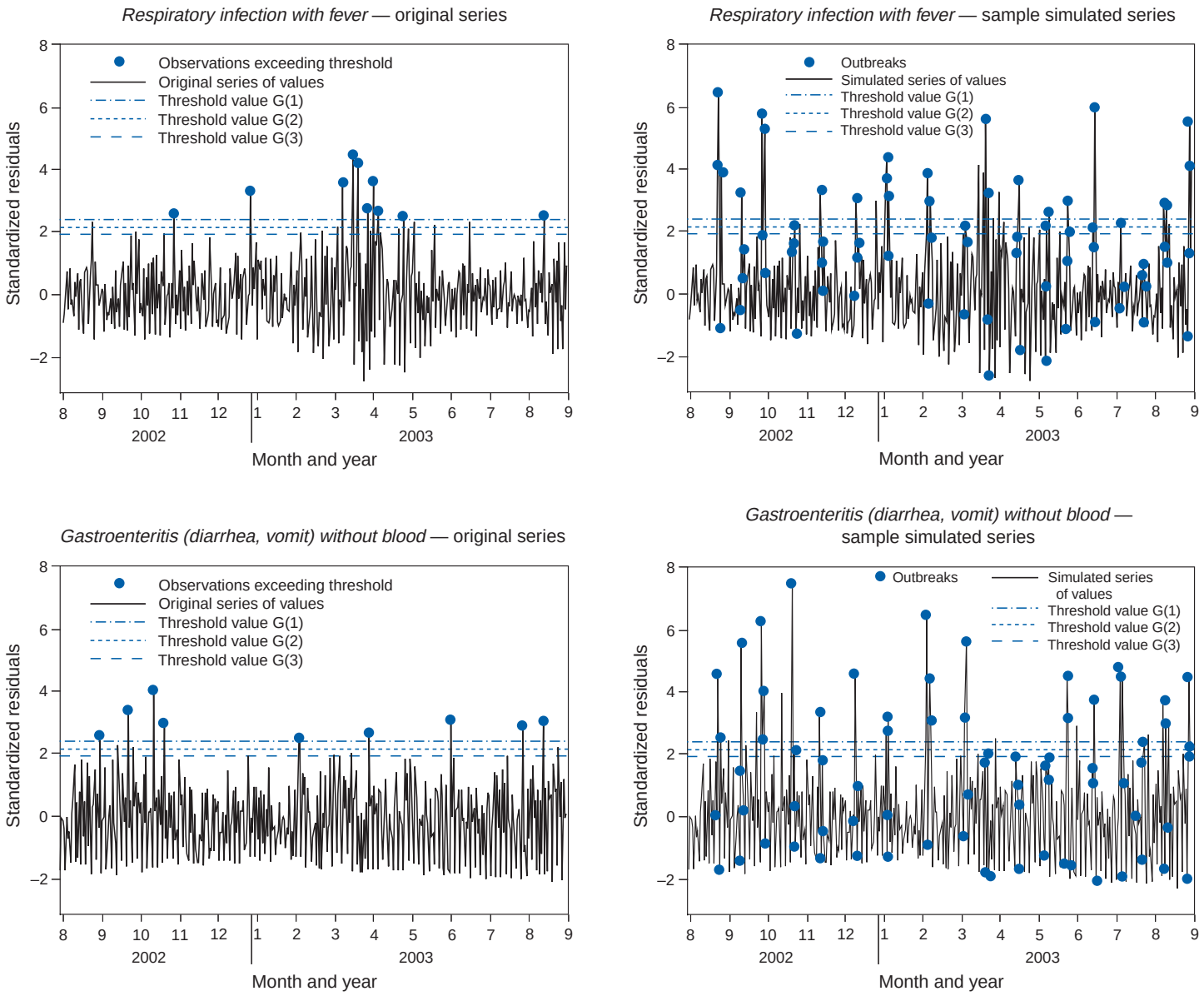
The Pulsar approach fared well in comparison with the other methods for each evaluation criterion. In particular, mean sensitivity was statistically significantly higher (Bonferroni $\alpha^* = 0.0056$) for the Pulsar approach when compared with the other approaches for both syndromes (Wilcoxon signed rank, $p<0.001$ for all comparisons). Furthermore, mean specificity for the Pulsar method was significantly higher (Bonferroni $\alpha^* = 0.0056$) than the specificity of the one-sided CUSUM method (Wilcoxon signed rank, $p<0.001$). This finding holds for both syndromes examined. In addition, in the case of *respiratory infection with fever*, the specificity of Pulsar was significantly higher than the specificity of TAD (Wilcoxon signed rank, $p<0.001$), whereas in the case of *gastroenteritis*, the specificity of Pulsar was higher than the specificity of the ARIMA approach (Wilcoxon signed rank, $p<0.001$). No other signifi-

TABLE 1. Sensitivity and specificity of the Pulsar approach for *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromes (simulated series)

	Syndrome											
	Respiratory infection with fever						*Gastroenteritis (diarrhea, vomit) without blood*					
	Sensitivity			Specificity			Sensitivity			Specificity		
Method	Mean	(SE*)	Median (Min–Max†)	Mean	(SE)	Median (Min–Max)	Mean	(SE)	Median (Min–Max)	Mean	(SE)	Median (Min–Max)
Pulsar analysis												
Model 1§	0.854	(0.0071)	0.85 (0.667–1)	0.981	(0.0004)	0.981 (0.972–0.991)	0.812	(0.0074)	0.818 (0.625–1)	0.982	(0.0003)	0.981 (0.975–0.991)
Model 2¶	0.744	(0.0081)	0.737 (0.5–0.9)	0.982	(0.0004)	0.981 (0.972–0.991)	0.768	(0.0088)	0.771 (0.55–1)	0.978	(0.0004)	0.978 (0.972–0.988)
Model 3**	0.656	(0.009)	0.649 (0.45–0.895)	0.985	(0.0004)	0.985 (0.975–0.997)	0.536	(0.0112)	0.538 (0.211–0.789)	0.991	(0.0004)	0.991 (0.981–1)
Model 4††	0.523	(0.0104)	0.5 (0.263–0.842)	0.987	(0.0004)	0.988 (0.976–0.997)	0.496	(0.0109)	0.5 (0.2–0.789)	0.99	(0.0004)	0.991 (0.981–1)
Model 5§§	0.718	(0.0089)	0.722 (0.5–0.9)	0.993	(0.0004)	0.994 (0.979–1)	0.701	(0.0101)	0.706 (0.375–0.944)	0.991	(0.0004)	0.991 (0.981–1)
Model 6¶¶	0.719	(0.0093)	0.737 (0.5–0.9)	0.992	(0.0004)	0.994 (0.982–1)	0.7	(0.0104)	0.706 (0.375–0.944)	0.99	(0.0005)	0.991 (0.978–1)

* Standard error.
† Minimum–Maximum.
§ Standardization by the standard deviation of the original series.
¶ Standardization by the mean absolute deviation of the original series.
** Standardization by the 7-day moving standard deviation of the simulated series.
†† Standardization by the 7-day mean absolute deviation of the simulated series.
§§ Standardization by the 7-day moving standard deviation of the simulated series with a 3-day lag.
¶¶ Standardization by the 7-day mean absolute deviation of the simulated series with a 3-day lag.

FIGURE 1. Time-series plots of standardized* residuals for the Pulsar method — G(1), G(2), G(3) thresholds and outbreaks



* Standardization by the standard deviation of the original series.

cant differences regarding specificity between the Pulsar method and the others were identified for either syndrome. Timeliness for the first day (proportion of alerts at the first day of an outbreak) differed significantly among the four approaches (Friedman test, $p<0.001$) for both syndromes. Timeliness of the Pulsar method was lower than the timeliness of the ARIMA model (Wilcoxon signed rank, $p<0.001$; Bonferroni $\alpha^* = 0.0056$). However, for *respiratory infection with fever*, Pulsar's timeliness was higher than TAD and CUSUM ($p<0.001$), and for *gastroenteritis*, Pulsar's timeliness was higher than TAD's ($p<0.001$). Results were similar when the mentioned alternative patterns of outbreaks were used.

Discussion

This paper proposes an algorithm for outbreak detection in the context of syndromic surveillance time-series data, based on alert criteria for both height and breadth of signals (*25*). The performance of the Pulsar approach and other suggested methods for outbreak detection (*19–23,28*) were assessed

TABLE 2. Sensitivity and specificity for alternative outbreak-detection approaches for *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromes (simulated series)

	Syndrome											
	Respiratory infection with fever						*Gastroenteritis (diarrhea, vomit) without blood*					
	Sensitivity			Specificity			Sensitivity			Specificity		
Method	Mean	(SE*)	Median (Min–Max†)	Mean	(SE)	Median (Min–Max)	Mean	(SE)	Median (Min–Max)	Mean	(SE)	Median (Min–Max)
Temporal aberration detection												
Model 1§	0.774	(0.0098)	0.778 (0.526–0.947)	0.977	(0.0008)	0.978 (0.956–0.99)	0.701	(0.0088)	0.696 (0.471–0.889)	0.981	(0.0007)	0.981 (0.956–0.997)
Model 2¶**	0.912	(0.0067)	0.914 (0.737–1)	0.955	(0.001)	0.955 (0.931–0.975)	0.831	(0.0082)	0.833 (0.611–1)	0.957	(0.001)	0.956 (0.924–0.978)
Auto-regressive integrated moving average												
Model 1††	0.738	(0.0083)	0.737 (0.5–0.9)	0.981	(0.0005)	0.981 (0.969–0.997)	0.679	(0.01)	0.684 (0.35–0.867)	0.979	(0.0005)	0.979 (0.966–0.991)
Model 2§§	0.667	(0.0088)	0.667 (0.5–0.895)	0.978	(0.0005)	0.978 (0.969–0.994)	0.629	(0.0105)	0.637 (0.35–0.833)	0.976	(0.0005)	0.975 (0.963–0.988)
Model 3¶¶	0.55	(0.0096)	0.55 (0.3–0.737)	0.979	(0.0005)	0.979 (0.969–0.997)	0.57	(0.0106)	0.579 (0.294–0.8)	0.975	(0.0003)	0.975 (0.966–0.985)
Cumulative sum												
Model 1***	0.711	(0.011)	0.706 (0.444–0.895)	0.926	(0.0023)	0.923 (0.877–0.978)	0.728	(0.0095)	0.737 (0.474–0.941)	0.94	(0.0015)	0.941 (0.906–0.975)
Model 2**†††	0.84	(0.0088)	0.842 (0.6–1)	0.87	(0.0029)	0.869 (0.803–0.946)	0.781	(0.0091)	0.789 (0.556–0.947)	0.927	(0.0016)	0.927 (0.887–0.96)

* Standard error.
† Minimum–Maximum.
§ Specificity set at 97%.
¶ Threshold of 2 standard deviations.
** Models presented only for purposes of illustration.
†† 7-day moving average filter.
§§ 7-day linear filter.
¶¶ 7-day exponential filter.
*** Specificity set at 97%.
††† Threshold set so that k = 0.5 and h = 2.5.

through simulations on the basis of direct comparison of sensitivity, specificity, and timeliness. For these performance criteria, Pulsar appears to be at least as effective as the other methods.

The Pulsar approach, first suggested for studying of episodic hormonal secretion, was successfully used in the context of syndromic surveillance data. Syndromic data are expressed initially in signal-to-noise units; then, through an iterative process, peaks are identified. Point elevations that are substantially high or elevations only moderately high but spanning multiple points in width are identified as peaks. The thresholds for peak detection are determined probabilistically on the assumption of normally distributed residuals. The idea of stochastically determining the thresholds is extended to the other methods under comparison. The thresholds are chosen so that a specificity of 97% is achieved in the original syndromic time series (*20,21*).

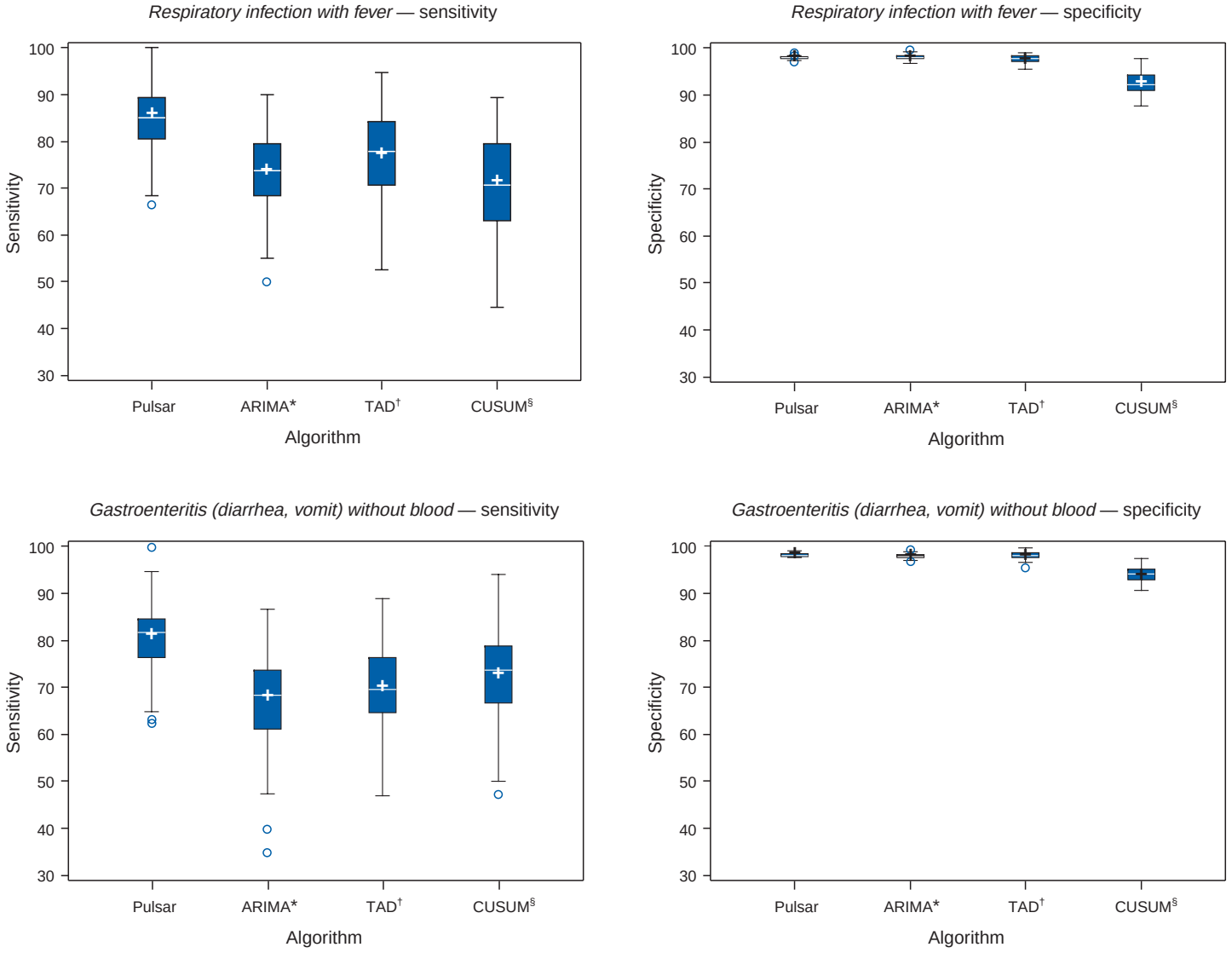
In the simulated data sets, the 97% specificity was most closely reproduced when using the Pulsar method as compared with the other methods (Tables 1 and 2). Sensitivity for the chosen Pulsar model (Model 1) for *respiratory infection with fever* ranged from 67% to 100%, with a mean of 85%, whereas sensitivity for *gastroenteritis (diarrhea, vomit) without blood* ranged from 62.5% to 100%, with a mean of 81%. The mean sensitivity for the Pulsar approach was higher than the sensitivity for the other methods. This method compared well with the others as far as specificity. All methods held specificity close to the 97% benchmark, with the exception of the one-sided CUSUM. In all methods evaluated, the higher percentage of alerts was generated on the first day of the outbreak with the exception of the TAD model, for which alerts occurred with similar frequency on the first 3 days of the outbreak. The ARIMA model exhibited the best timeliness results, followed by the Pulsar approach.

Of note, the performance evaluation criteria led to uniformly worse results for all methods when applied to the daily proportion of syndrome counts to total visits as opposed to syndrome counts. Methods adapted to proportion are under investigation and could be evaluated simultaneously. In addition, a specific simulation schema was used to compare methods, with varying outbreak sizes of fixed duration affecting the generalization of the comparison under other simulation settings. However, the critical comparison is always the one based on the detection performance of real outbreaks (*33*). Finally, this analysis did not consider other methods that have been proposed for analysis of syndromic data (*34,35*), including spatial statistical methods (e.g., spatial scan statistic, Bayesian approaches, and multivariate methods) (*36–40*).

Conclusion

The performance results of the Pulsar method are overall comparable with the other methods examined for the specific simulation schema used. The simplicity of the algorithm, its ability to be modified regarding choice of standardization and distributional assumptions for the signal-to-noise ratio, and its ability to detect peaks based not only on height but also on width (which more closely addresses the epidemic shapes that one would expect to last for >1 day) make it a promising candidate for further use in syndromic surveillance. The abrupt

FIGURE 2. Box plots of sensitivity and specificity of 100 simulated series for *respiratory infection with fever and gastroenteritis (diarrhea, vomit) without blood* syndrome counts



* Autoregressive integrated moving average.
† Temporal aberration detection.
§ Cumulative sum.

increase in population anticipated for the Athens 2004 Olympic Games will provide an ideal prospective surveillance setting for comparing the behavior of all proposed methods regarding alert mechanisms.

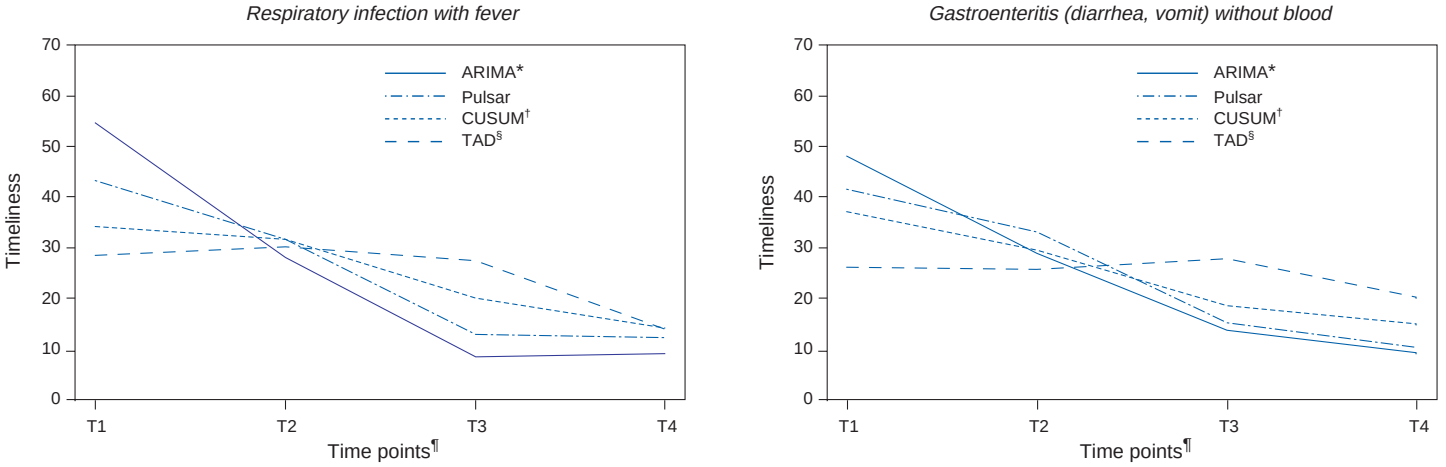
Acknowledgments

The authors thank the members of the syndromic surveillance team, Dimitris Papamihail, Aggeliki Lambrou, and Ioannis Karagiannis, as well as the >50 health professionals who made every effort to gather quality data from the EDs.

References

1. Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for communitywide influenza-like illness, New York City. J Urban Health 2003;80 (2 Suppl 1):i43–9.
2. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. J Urban Health 2003;80(2 Suppl 1):i76–88.
3. CDC. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. MMWR 2002;51 (Special Issue)13–5.

FIGURE 3. Timeliness plots of outbreak-detection methods for *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromes



* Autoregressive integrated moving average.
† Cumulative sum.
§ Temporal aberration detection.
¶ T1 = first day of outbreak; T2 = second day of outbreak; T3 = third day of outbreak; T4 = fourth day of outbreak.

4. 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.
5. 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.
6. Liu Z, Tsui F-C, Zeng X. Cache table design for disease surveillance system [Abstract]. In: Kohane IS (ed). Proceedings of the Annual Fall Symposium of the American Medical Informatics Association, November 9–13, 2002, San Antonio, Texas. Philadelphia, Pennsylvania: Hanley & Belfus, Inc, 2002;1086.
7. Ackelsberg J, Layton M. Update #5: terrorist attack at the World Trade Center in New York City: medical and public health issues 2001. Available at http://www.nyc.gov/html/doh/html/cd/wtcf.html.
8. Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC, area. Am J Prev Med 2002;23:180–6.
9. Department of Defense Global Emerging Infections System. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE). Available at http://www.geis.fhp.osd.mil/geis/surveillanceactivities/essence/essence.asp.
10. Schafer K. LEADERS (Lightweight Epidemiology Advanced Detection & Emergency Response System), presentation W203. Available at http://www.tricare.osd.mil/conferences/2001/agenda.cfm.
11. Zelicoff A, Brillman J, Forslund DW, et al. The rapid syndrome validation project (RSVP). Proc AMIA Symp 2001:771–5.
12. Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc 2002;9:105–15.
13. Mundorff MB, Gesteland PH, Rolfs RT. Syndromic surveillance using chief complaints from urgent care facilities during the Salt Lake 2002 Olympic Winter Games [Abstract]. MMWR 2004;53 (Suppl):254.
14. 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.
15. 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.
16. Jorm LR, Thackway S, Churches TR, Hills MW. Watching the Games: public health surveillance for the Sydney 2000 Olympic Games. J Epidemiol Community Health 2003;57:102–8.
17. Meehan P, Toomey KE, Drinnon J, Cunnigham S, Anderson N, Baker E. Public health response for the 1996 Olympic Games. JAMA 1998; 280:1229–30.
18. Dafni U, Gkolfinopoulou K, Tsiodras S, Saroglou G. Planning syndromic surveillance for the Athens 2004 Olympic Games: a pilot study [Abstract]. J Urban Health 2003;80(2 Suppl 1):i125.
19. Rogerson PA. A cumulative sum approach to syndromic surveillance in geographic regions [Abstract]. J Urban Health 2003;80(2 Suppl 1):i130.
20. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci U S A 2003;100:1961–5.
21. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak 2003;3:2.
22. Hutwagner L. Centers for Disease Control, Early Aberration Reporting System (EARS). Presentation, National Syndromic Surveillance Conference, New York, NY, September 23–24, 2002.
23. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96.
24. Lawson A. Temporal-spatial methodologies and data sources. Presentation, National Syndromic Surveillance Conference, New York, NY, September 23–24, 2002.
25. Merriam GR, Wachter KW. Algorithms for the study of episodic hormone secretion. Am J Physiol 1982;243:E310–18.

26. Nindl BC, Hymer WC, Deaver DR, Kraemer WJ. Growth hormone pulsatility profile characteristics following acute heavy resistance exercise. J Appl Physiol 2001; 91:163–72.
27. Lee WM, Meij BP, Bhatti SF, Mol JA, Rijnberk A, Kooistra HS. Pulsatile secretion pattern of growth hormone in dogs with pituitary-dependent hyperadrenocorticism. Domest Anim Endocrinol 2003;24:59–68.
28. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. Emerg Infect Dis 1997;3:395–400.
29. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74:829–36.
30. Hurvich CM, and Simonoff JS. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J R Stat Soc B 1998;60:271–93.
31. Heffernan R, Mostashari F, Das D, et al. New York City syndromic surveillance systems. MMWR 2004;53(Suppl):25–7.
32. SAS Institute, Inc. SAS,® version 8.2 [software]. Cary, NC: SAS Institute, Inc.
33. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5):1–11.
34. Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. J R Stat Soc A 2003;165:5–21.
35. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. J Urban Health 2003;80(2 Suppl 1):i1–7.
36. Kulldorff M. Spatial scan statistics: models, calculations and applications. In: Glaz J and Balakrishnan N, eds. Scan statistics and applications. Boston, MA: Birkhäuser, 1999;303–22.
37. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. Emerg Infect Dis 2003;9:641–6.
38. Lawson AB, Cressie N. Spatial statistical methods for environmental epidemiology. In: Sen PK, Rao CR, eds. Handbook of statistics. Amsterdam, Holland: Elsevier, 2000;18:357–96.
39. Clark AB, Lawson A. Change-point detection using directional derivatives [Abstract]. MMWR 2004;53(Suppl):233.
40. Wessman P. Some principles for surveillance adopted for multivariate processes with a common change point. Communications in Statistics 1998;27:1143–61.

Taming Variability in Free Text: Application to Health Surveillance

Alan R. Shapiro
New York University School of Medicine, New York, New York

Corresponding author: Alan R. Shapiro, Department of Medicine, New York University School of Medicine, 5 Pheasant Run, Pleasantville, NY 10570. Telephone: 914-747-1804; E-mail: alan.shapiro@med.nyu.edu.

Abstract

Introduction: *Use of free text in syndromic surveillance requires managing the substantial word variation that results from use of synonyms, abbreviations, acronyms, truncations, concatenations, misspellings, and typographic errors. Failure to detect these variations results in missed cases, and traditional methods for capturing these variations require ongoing, labor-intensive maintenance.*

Objectives: *This paper examines the problem of word variation in chief-complaint data and explores three semi-automated approaches for addressing it.*

Methods: *Approximately 6 million chief complaints from patients reporting to emergency departments at 54 hospitals were analyzed. A method of text normalization that models the similarities between words was developed to manage the linguistic variability in chief complaints. Three approaches based on this method were investigated: 1) automated correction of spelling and typographical errors; 2) use of* International Classification of Diseases, Ninth Revision, Clinical Modification *(ICD-9-CM) codes to select chief complaints to mine for overlooked vocabulary; and 3) identification of overlooked vocabulary by matching words that appeared in similar contexts.*

Results: *The prevalence of word errors was high. For example, such words as diarrhea, nausea, and vomiting were misspelled 11.0%–18.8% of the time. Approximately 20% of all words were abbreviations or acronyms whose use varied substantially by site. Two methods, use of ICD-9-CM codes to focus searches and the automated pairing of words by context, both retrieved relevant but previously unexpected words. Text normalization simultaneously reduced the number of false positives and false negatives in syndrome classification, compared with commonly used methods based on word stems. In approximately 25% of instances, using text normalization to detect lower respiratory syndrome would have improved the sensitivity of current word-stem approaches by approximately 10%–20%.*

Conclusions: *Incomplete vocabulary and word errors can have a substantial impact on the retrieval performance of free-text syndromic surveillance systems. The text normalization methods described in this paper can reduce the effects of these problems.*

Introduction

Syndromic surveillance using existing free-text sources (e.g., electronic medical records or emergency department [ED] chief complaints) offers potential advantages in the timeliness and richness of the information that can be provided (*1*). In particular, capturing surveillance information as free text does not incur the human effort, delay, or drastic reduction in information incurred by coding. However, using free text to track symptom occurrence incurs four particular challenges caused by linguistic variation: 1) a single symptom can be described in multiple ways by using synonyms and paraphrases; 2) medical concepts are often recorded using abbreviations and acronyms that are idiosyncratic to individual hospitals; 3) the same concept can be indicated with different parts of speech; and 4) words are frequently misspelled or mistyped in busy medical settings, causing the continual appearance of new, previously unseen errors. This paper discusses new approaches to address these four challenges.

Failure to detect linguistic variations results in missed cases. This problem is potentially severe enough to motivate efforts to develop surveillance systems based on apparently unambiguous numerical codes or standardized vocabularies. One goal of the current study is to analyze ED chief complaints empirically to explore the extent of variation present.

Certain efforts to manage linguistic variations and to increase system sensitivity can produce their own false positives, thereby lowering specificity, increasing false alarms, and ultimately wasting limited public health resources. Most importantly, monitoring symptoms adequately in the presence of such variability requires ongoing, costly, labor-intensive maintenance.

The problem of linguistic variation in surveillance systems designed for early detection of covert attacks deserves attention. Increasingly sensitive statistical methods are available with the potential to detect an outbreak affecting a local geographic area (e.g., a single hospital). These methods are most effective when clean data are provided. In addition, syndromic surveillance systems have been used not only for outbreak detection but for case finding and outbreak monitoring; these functions can also be compromised when substantial numbers of cases are missed. Even if a surveillance system contains minimal errors when used in the site where it was developed, word usage can vary substantially among sites, making algorithms developed for one site inadequate for others. Efforts to combine systems for extensive regional surveillance need to be able to detect and address performance differences caused by word variation from one site to another.

This paper examines the extent of word variation in the text of ED chief complaints. It then reviews different approaches for managing word variation, discusses their limitations, and outlines a new approach to text normalization on which three approaches to handling linguistic variation are based. The performance of these approaches when combined is then compared to a common approach in free-text surveillance systems based on word-stem matching.

Extent of Word Variation in Chief-Complaint Databases

Chief-complaint databases from the New York City (NYC) Department of Mental Health and Hygiene (DOHMH), Emergency Medical Associates of New Jersey (EMA), and Boston Beth Israel Deaconess Medical Center (AEGIS) were used in these studies. Collectively, the data consist of the chief complaints from approximately 6 million patient encounters at 54 hospitals over a period of 1–7 years, depending on the hospital.

Types of Word Variation

The word variation in these approximately 6 million chief complaints can be grouped into two types. The first, orthographic variation, includes variations in spelling attributable either to different grammatical forms of the same word (e.g., coughs, coughed, or coughing) or to spelling errors, transcription errors, or typographic errors. In principle, orthographic variation might be addressed, at least in part, through the use of string-matching algorithms that group similarly spelled words.

The second type of word variation, nonorthographic (or semantic) variation, unfortunately cannot be managed merely by looking at the arrangement of letters in a word. The same chief complaint can usually be described in multiple ways by using acronyms, word truncations, idiosyncratic abbreviations, or legitimate synonyms, all of which can differ from one hospital to another. For example, spelling-correction or string-matching algorithms cannot be expected to discover that the 869 chief complaints of *N V* in the DOHMH database should be regarded as instances of *nausea and vomiting*. Such cases in which only a limited number of letters are retained from the original word are better treated as synonyms rather than orthographic variations and are referred to here as examples of nonorthographic or semantic variation.

Orthographic Variation

Substantial orthographic variation was found among words commonly included in chief complaints (e.g., *diarrhea*, *nausea*, or *abscess*) (Table 1). These numbers were derived from the DOHMH database, but results for the EMA and AEGIS databases were similar. A word as simple as *vomiting* was misspelled at least 379 ways (Table 2).

TABLE 1. Variability in strings used to denote selected words in free-text emergency department chief-complaint data — New York City, November 2001–November 2002

Word	No. of variations	No. of instances	Incorrect (%)
Abscess	92	3,419	45.4
Diarrhea	349	4,006	11.1
Vomiting	379	16,288	16.7
Nausea	137	4,143	18.8
Headache	196	1,771	3.4

Source: New York City Department of Health and Mental Hygiene chief-complaint database.

TABLE 2. Examples of different strings* used to denote *vomiting* in free-text emergency department chief-complaint data

1. Andvomiting	100. Vomitedx5today	300. Vommioting
2. Bomiting	101. Vomiteing	301. Vommited
3. Cvomiting	102. Vomites	302. Vommitiing
—	103. Vomiteted	303. Vommiting
15. V0mitting	104. Vomitfever	304. Vommitintig
16. Vamiting	105. Vomitg	305. Vommitit
17. Vbomiting	—	—
18. Vfomiting	200. Vomitint	325. Vomti
19. Vimit	201. Vomitintg	326. Vomtied
20. Vimited	202. Vomitiny	327. Vomtig
—	—	—
50. Vomiging	250. Vomitting3xdays	377. Vvomitting
51. Vomihing	251. Vomittinga	378. Womiting
52. Vomiig	252. Vomittingab	379. Womitting

Source: New York City Department of Health and Mental Hygiene chief-complaint database.
* N = 379

Spelling-correction programs are often based on the observation that 80% of spelling errors are usually caused by a single insertion, substitution, or deletion of a letter in the word (*2*). That study, based on the performance of computer transcriptionists in 1964, did not reflect the conditions of the typical modern-day ED. By contrast, in the present study, the modal number of errors per misspelled word was two, and in 31% of instances the misspelled words contained ≥3 errors.

Nonorthographic Variation

Nonorthographic variation in the study data was common. In each of the three databases, >20% of all nonstop words (i.e., words other than common articles, conjunctions, and prepositions [e.g., *the*, *and*, *a*, or *or*]) in the chief complaints were nonstandard acronyms, abbreviations, or truncations. This number was obtained *after* first excluding such standard medical abbreviations as CHF, ECG, HCT, HBV, HIV, SOB, WBC, and 43 others. This observation necessitated this study's efforts to address the nonorthographic or semantic variation found in medical free text.

Substantial differences in usage among sites were present. Approximately 55% of the word strings in the EMA database were not contained in the DOHMH data, and 35% of the strings in the AEGIS database were not present in the DOHMH data even though the AEGIS database is only 8% the size of the DOHMH database. The words *rigors* and *myalgias* were used in the AEGIS database 211 and 76 times more frequently than in the DOHMH and EMA databases, respectively. Those words occurred so rarely in the NYC chief complaints that they were not included in, and would not have been detected by, the DOHMH algorithms. Similarly, 3,392 instances of skin rashes described in the EMA hospitals using the string *erupt* would not have been retrieved because the truncation *erupt* was used only rarely in NYC and therefore not included in their algorithms. The acronym *DIB* for *difficulty in breathing* appeared in 2,679 chief complaints from New York City but only twice in the >3.5 million chief complaints recorded elsewhere. Such differences highlight the need for more systematic, preferably automated, methods for managing site customization.

Methods for Managing Linguistic Variation

The need to clean textual data has been recognized in every discipline in which textual data is processed, and corresponding methods to deal with the problem have been developed (*3*). The majority of these methods have addressed only orthographic word variation. This paper describes the limitations of the three most commonly used methods (phonetic spelling correction, word-stem algorithms, and edit distances) and proposes the need for a fourth, more powerful approach for managing medical text.

Phonetic Spelling Correction Methods

Phonetic spelling-correction methods include algorithms such as Soundex, Editex, or Phonix (*4*). Soundex has been used for more than a century and is often used in medical applications. These methods recognize that words can be misspelled when certain letters that sound alike, such as *d* and *t* (as in *jauntice* or *pregnand*) or *g* and *j* (as in *conjested*) are substituted for one another.

Unfortunately, multiple exceptions to these pairings exist (e.g., *g* does not sound like *j* in *cough* and thus misspellings of *cough* would not be detected). More importantly, among the chief complaints examined in this study, typing and transcription errors were more common than phonetic errors. The letters *r* and *y* were substituted for *t* 5 times more frequently than the letter *d* because they are located on either side of *t* on the keyboard.

Keyword or Word-Stem Methods

The idea behind this current method in free-text syndromic surveillance is that most words contain a unique string, usually the first few letters, that is specific enough to identify the word and that is unlikely to be misspelled. For example, this method assumes that although *breathing* might be spelled 147 ways in chief-complaint data, searching for all words beginning with *breat* would capture the majority of them. Unfortunately, this strategy did not find the 56 (38%) spellings of *breathing* in the DOHMH database that did not begin with *breat*.

Relying on a word-stem approach not only misses cases but also requires an untenable level of labor-intensive maintenance. For example, for a system to recognize cases not beginning with *breat*, other word stems (e.g., *brath*, *bereath*, and *DIB*) need to be added. However, this strategy results in multiple false positives (e.g., *mandibular* fractures, *dibetes*, or the use of a *dibfulator*). Further logic is required to avoid retrieving mentions of any therapeutic *breathrough*. Eliminating such new false positives requires making further ad-hoc modifications and a continuing spiral of time-consuming maintenance and increasingly unreadable, error-prone code.

Even if a temporary state is reached in which false positives and negatives are minimal, new strings will keep arriving, making the previous logic inadequate. In the present study, even after 2 million chief complaints had been processed in the DOHMH system, approximately 750 new strings

appeared each week. Furthermore, when separate systems are joined, the complexity of the algorithmic logic must be increased again (e.g., although *diarrhea* was spelled 349 different ways in the DOHMH database, the EMA database contained an additional 154 spellings).

A system is far more maintainable if the medical logic regarding which concepts best represent a syndrome can be kept at a conceptual level, separate from the underlying technical intricacies of text processing.

Edit-Distance Methods

A third common method for matching strings is the edit-distance approach, which measures similarity as the minimum number of operations (e.g., insertions, deletions, substitutions, or transpositions) required to transform one string into another. Multiple modifications of this approach have focused primarily on computational efficiency in matching long strings (*5*). Edit distances, however, often give results inconsistent with human intuition. For example, the method would score both *azma* and *stomac* as equally close to *asthma*. Health professionals would not find this useful.

Generalized Edit-Distance Method

The text-normalization method developed for and used in this project is a generalization of the edit-distance approach — it models the similarity between two words as the minimum number of typographic errors, phonetic spelling errors, transcription errors, medical affixes (suffixes and prefixes), and concatenations that could transform one word into another. Because the method attempts to create the most plausible model of how a misspelled string could be generated, it is designed to represent the psychological distance between two strings rather than the computational distance.

As an example of the capabilities of this approach, the string *coughvomintingdiarre*, which actually appeared in a chief complaint, would be recognized by the text normalization software as an instance of the string *vomiting* (as well as of *cough* and *diarrhea*). Programs based on phonetic matching, edit distance, keywords, or the majority of other algorithms would not recognize the first string as a plausible instance of the second string.

Because the distances between words produced by the algorithm make intuitive sense (i.e., they correspond closely to the judgments about word similarity that would be made by humans), users can more easily work interactively with the computer or rely on the algorithm to make good decisions when run fully automatically. In one configuration, text-normalization software can be used as a pre-processor that passes normalized chief complaints or medical records as input into a separate program dedicated to the higher-level task of recognizing syndromes and analyzing their frequency.

Applications of Text Normalization

To improve system performance, the text-normalization method was applied to the chief-complaint databases in three ways. The first use was a straightforward application of text normalization to automatically remove typographical errors, misspellings, word concatenations, and other forms of orthographic variation in chief complaints. The other two methods used text normalization as an essential tool for vocabulary expansion, in particular to search for overlooked abbreviations, acronyms, and other relevant vocabulary. Each application is described briefly.

Normalization of Chief Complaints

Chief complaints were presented to a text-normalization program, which compared each word in each chief complaint to a list of 68 key concepts that had been identified as useful for syndrome identification in the DOHMH syndromic surveillance algorithms. For example, the list included the words *pulmonary*, *pleuritic*, *cough*, *gasping*, and *dyspnea* for respiratory syndrome identification. Words sufficiently close to a key concept were matched with that concept.

To compare the performance of the text-normalization approach to orthographic variation with the often-used word-stem approach, the DOHMH word-stem algorithm for diarrheal syndrome was applied to the EMA chief-complaint data, both with and without prior text normalization. Each instance retrieved by one algorithm but not by the other was reviewed to determine which approach was correct. Of the 38,956 cases of diarrhea in the EMA database identified by either approach, 5,217 (13%) were recorded in a nonstandard way. When previously trained on the DOHMH chief complaints, the text-normalization program was able to identify all but five of these cases, an improvement of 896 when compared with cases recognized without normalization, while incurring only 17 false positives. Orthographic normalization alone improved performance by 2.3% when compared with the word-stem approach.

Using ICD-9-CM Codes To Uncover Overlooked Vocabulary

Although orthographic normalization generated substantial improvement, the possibility remained that additional words were being used to indicate symptoms and were being missed. Two additional approaches based on text normaliza-

tion were used to uncover overlooked vocabulary (e.g., unanticipated abbreviations, acronyms, and truncations).

The first approach was to use *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) codes to select chief complaints most likely to contain vocabulary relevant to a particular syndrome. For example, the chief complaints of patient encounters assigned one of the ICD-9-CM codes in CDC's ESSENCE grouping for gastrointestinal (GI) syndrome (*6*) could be examined as a likely source for overlooked words that indicate GI syndrome. Specifically, these chief complaints were analyzed to see which words occurred more frequently in this ICD-9-CM GI syndrome group than in the cases not in that group. The procedure, in effect, searched for words with the highest relative risk of occurrence in the selected group as a means to detect words useful for designating GI syndrome.

In practice, this strategy is compromised because the seemingly innumerable misspellings and corruptions of words in the chief complaints result in an unmanageable list of word strings (e.g., 349 variations of *diarrhea*) whose relevance cannot be distinguished from the numerous irrelevant words that also occur with low frequency in the group. Used in this case, text normalization removes much of the noise and allows the relevant concepts to emerge.

The EMA database was used for this experiment because it contained both ED chief complaints and discharge ICD-9-CM codes for each case. Using the ICD-9-CM codes for GI syndrome with text normalization uncovered a number of relevant words not previously included in the DOHMH word-stem algorithms, including *cramps* (4,415 instances), *runs*, *NVD*, *LBM*, *Shigella*, *noninf* (1,689 instances, as in *noninf gastroenteritis*) and others.

Choosing a different subset of ICD-9-CM codes (e.g., only those codes that reflect intestinal rather than upper GI disease) might have uncovered yet additional words. The best strategies and criteria for choosing productive codes for synonym generation remain to be investigated. The potential benefits of using more precise and comprehensive coding schemes (e.g., SNOMED CT®) might also be explored (*7*).

Using Context To Uncover Overlooked Vocabulary

A second approach to retrieving overlooked vocabulary, as well as site-specific idiosyncratic vocabulary requiring customization, is adapted from the dictum in computational linguistics that "a word is known by the company it keeps" (*8*). This approach seeks to retrieve words with similar meanings by finding words that occur in similar contexts. Words that co-occur with the same other words tend either to have similar meanings or at least to be closely related.

In this approach, for each word in the chief-complaint database, the words that most specifically occurred with that word were tabulated, resulting in a co-occurrence profile of closely associated words for each word. Each word was then compared with every other word to identify those with the most similar co-occurrence profiles. Similarity was assessed by using rank-order correlation between profiles. Examples of word strings of ≤5 letters uncovered by this method that would have been overlooked when using current word-stem algorithms are provided (e.g., 4,970 hive-like rashes would have been overlooked because *hives* was not previously a search term) (Table 3).

Detection Performance With and Without Text Normalization

Fortified with normalized text and additional vocabulary, a syndrome classifier operating on text that has been normalized can demonstrate greater sensitivity and specificity than a word-stem algorithm operating without normalization. Even though the two approaches will agree in the majority of cases, the cases where they differ are revealing.

Word-stem algorithms with and without text normalization were applied to detect instances of lower respiratory illness syndrome. On this particular task, in 3.3 million chief complaints, 201,327 instances were retrieved, and the sensitivity of the keyword and text-normalization approaches differed by 5.6% (11,252 instances). When the word-stem algorithm without normalization indicated presence of a lower respiratory illness but the algorithm using text normalization did not, the text-normalization approach was correct in 96.4% of cases. In the instances in which the text-normalization

TABLE 3. Expanding keyword vocabulary by locating words that appear in similar contexts in free-text chief-complaint data

Key concept*	Word strings of ≤5 letters with similar contexts†
Black	Dark,§ brown, drk,§
Cough	Plegm,§ cgh,§
Enteritis	Age§
Fever	Fevr, feve, fev, cough
Nausea	N,§ NVD,§ NV§
Pneumonia	RLL, pneu,§ exac§
Rash	Rashes, hives§
SOB	DIB
Stool	Urine, dark,§ brown, black, drk,§ tarry,§ BRBPR§

Source: New York City Department of Health and Mental Hygiene chief-complaint database.
* Key concepts used in free-text syndromic surveillance.
† Word strings with similar contexts, shown in order of computed similarity.
§ Word strings that would have been overlooked when using current word-stem algorithms.

approach declared a syndrome to be present and the word-stem algorithm alone did not, human review of the chief complaints determined that text normalization was correct in 99.8% of cases. Use of text normalization thus substantially reduced the number of both false positives and false negatives (Table 4).

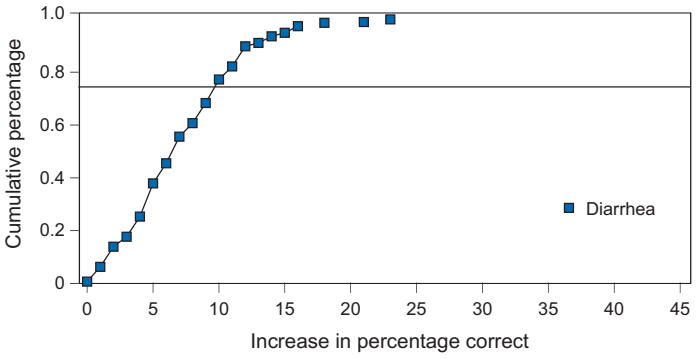
If text normalization were applied daily or in hospitals or surveillance systems with far fewer visits than 3.3 million, the differences between the two approaches might be greater still. The two approaches generated substantial differences when tracking diarrhea or bloody diarrhea syndrome in New York City hospitals with ≥100 visits for diarrhea per week (Figure). In approximately 25% of instances, the sensitivity of the word-stem approach was improved by 10%–20% when used with text normalization. In no case was the specificity decreased.

TABLE 4. Comparison of accuracy of word-stem algorithm with and without text normalization as applied to chief-complaint data in 12,270 instances in which the two approaches disagreed

	Reviewer determination	
Algorithm decision*	**Present**	**Absent**
Text normalization: present Word stem (without text normalization): absent	11,238	14
Text normalization: absent Word stem (without text normalization): present	37	981

* Decision of the algorithm regarding presence or absence of a given syndrome.

FIGURE. Effect of text normalization on free-text chief-complaint data in emergency departments with ≥100 diarrhea cases/week



A similar analysis was performed for fever/influenza syndrome (excluding upper respiratory illness), which comprises approximately 16.5% of New York City ED encounters. In 12% of instances, text normalization resulted in a 10%–20% improvement in sensitivity over the word-stem approach in tracking the number of fever/influenza chief complaints.

Conclusions

Incomplete vocabulary and word errors can have a substantial impact on the retrieval performance of free-text syndromic surveillance systems. Certain methods based on text normalization can greatly reduce the impact of these problems. New, increasingly sensitive methods of analysis will be most effective with careful attention to the quality of the data on which they rely.

Acknowledgments

Rick Heffernan and Farzad Mostashari of the New York City Department of Health and Mental Hygiene, Dennis Cochrane and John Allegra of Emergency Medical Associates of New Jersey, and Karen Olsen and Kenneth Mandl of Boston Children's Hospital provided portions of their chief-complaint data for this study.

References

1. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. JAMA 2004;11:141–50.
2. Damerau F. A technique for computer detection and correction of spelling errors. Communications of the ACM 1964;7:171–6.
3. Kukich K. Techniques for automatically correcting words in text. ACM Computing Surveys 1992;24:377–439.
4. Zobel J, Dart P. Phonetic string matching: lessons from information retrieval. In: Proceedings of the 19th International Conference on Research and Development in Information Retrieval, Zurich, Switzerland, 1996.
5. Navarro G. A guided tour to approximate string matching. ACM Computing Surveys 2001;33:31–88.
6. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.bt.cdc.gov/surveillance/syndromedef/.
7. McClay JC, Campbell J. Improved coding of the primary reason for visit to the emergency department using SNOMED. Proc AMIA Symp 2002;499–503.
8. Firth JR. Modes of meaning. In: Papers in linguistics. London, UK: Oxford University Press, 1957.

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. Cardenas[2]
[1]*Rush University Medical Center, Chicago, Illinois;* [2]*Chicago 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

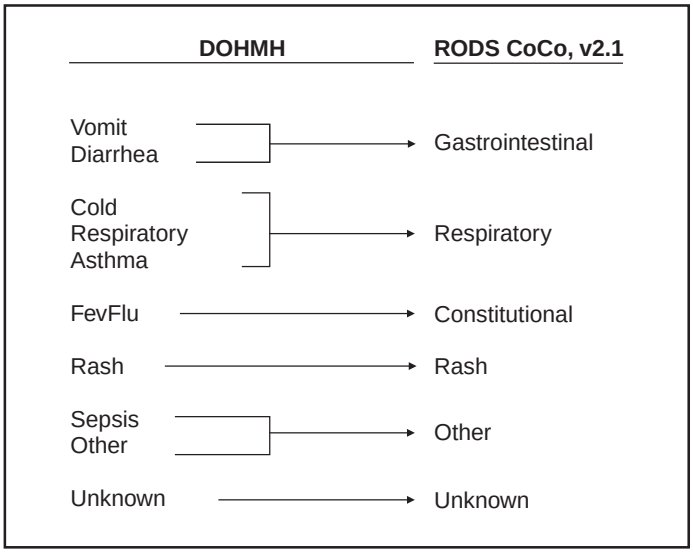
version 2.1 of the Complaint Coder (CoCo) developed by the Real-Time Outbreak Detection System (RODS) laboratory at the Center for Biomedical Informatics, University of Pittsburgh (*5*). This Bayesian classifier codes symptoms into syndromes on the basis of probability (i.e., the chances that a given 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:

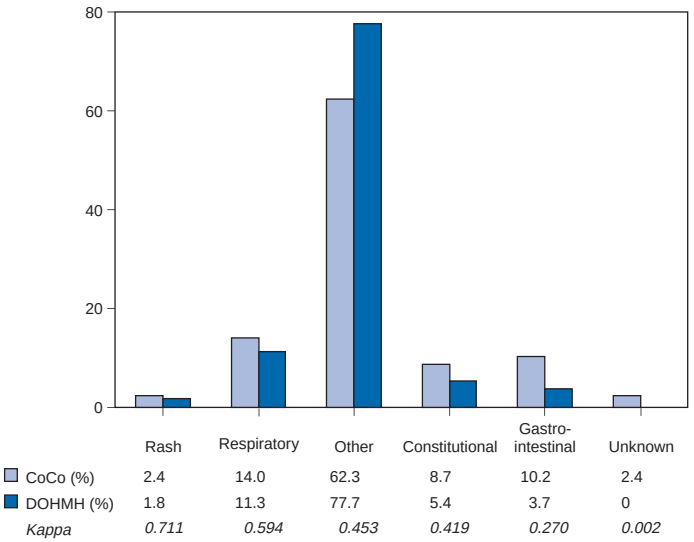
- Any symptom coded as *Vomit* or *Diarrhea* by the DOHMH algorithm was renamed *Gastrointestinal* to match CoCo.
- Any symptom classified as *Cold*, *Respiratory*, or *Asthma* by the DOHMH coder was renamed *Respiratory.*
- *FevFlu* in the DOHMH system was renamed *Constitutional*.
- Any symptom coded 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.

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



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

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	*Diarrhea*: diar, dair, diah, enteri, gastroent, 558; IF stool, bowel, bm AND loose, watery, or liquid	pain or cramps anywhere in the abdomen, nausea, vomiting, diarrhea, and abdominal distension or swelling
	Vomit: throwing up, threw up, food pois, vom, vmt, n/v, 787	
Respiratory	*Cold*: stuffi, stuffy, sneez, nasal, nasel AND cong, conj, drip, disch, runn, cong, runn, nose, congested, congestion, cold	problems of the nose (coryza) and throat (pharyngitis), as well as the lungs; examples of *Respiratory* 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 *Respiratory* and not *Constitutional*
	Respiratory: 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.	
	Asthma: asth, asmtha, ashtma, astma, asyhma, whez, azth, az, airway, whee, wheel, 490, 491, 492, 493, COPD, c.o.p.d.	

* 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.

How Many Illnesses Does One Emergency Department Visit Represent? Using a Population-Based Telephone Survey To Estimate the Syndromic Multiplier

Kristina B. Metzger, A. Hajat, M. Crawford, F. Mostashari
New York City Department of Health and Mental Hygiene, New York, New York

Corresponding author: Kristina B. Metzger, Bureau of Epidemiology Services, New York City Department of Health and Mental Hygiene, 125 Worth Street, Room 315, CN-6, New York, NY 10013. Telephone: 212-788-9309; Fax: 212-788-4473; E-mail: kmetzger@health.nyc.gov.

Abstract

Introduction: *Syndromic surveillance monitors trends in nonspecific health indicator data to detect disease outbreaks in a timely manner; however, only a limited percentage of persons with mild illness might exhibit behaviors that could be detected by syndromic surveillance.*

Objectives: *The objectives of this study were to 1) examine the demographic characteristics of New Yorkers with recent flu-like or diarrheal illness, 2) describe behaviors associated with having flu-like illness, and 3) estimate the citywide burden for selected illnesses by calculating the syndromic multiplier (i.e., the number of citywide illnesses represented by each visit to an emergency department [ED]).*

Methods: *A cross-sectional telephone survey of 2,433 adult residents of New York City (NYC) was conducted during March 19–March 31, 2003, and October 27–November 23, 2003. Respondents were asked about flu-like illness, behaviors related to flu-like illness, and diarrheal illness during the 30 days before the interview. Estimated numbers of citywide illnesses were compared with ED visits for flu-like and diarrheal illnesses that were recorded by the NYC syndromic surveillance system for the same periods.*

Results: *Every ED visit for flu-like illness represented approximately 60 illnesses among city residents; every visit for diarrheal illness represented approximately 251 illnesses. Among adults who reported a recent flu-like illness, 53.2% purchased over-the-counter (OTC) medications; 32.6% reported missing school or work; 29.1% visited a physician; 21.4% called a physician for advice; 8.8% visited an ED; and 3.8% called a nurse or health hotline for advice. Of those who reported multiple behaviors, respondents most commonly reported purchasing OTC medications as their first response to a flu-like illness.*

Conclusions: *Population-based survey data can be used in conjunction with syndromic surveillance data to better understand the relation between nonspecific health indicators and the burden of certain illnesses in the community, and to assess the representativeness of different syndromic data sources.*

Introduction

Syndromic surveillance systems are typically designed to detect increases in nonspecific health indicators that potentially signal the beginning of a disease outbreak, including an outbreak attributable to biologic terrorism. The data used by syndromic surveillance systems often represent behavioral indicators of early illness (e.g., pharmaceutical sales or employee absenteeism) and clinical indicators associated with more severe illness (e.g., emergency department (ED) visits or ambulance dispatches). Syndromic surveillance is particularly notable for its timeliness because it collects and analyzes these data daily.

Because diseases caused by potentially threatening biologic agents often have prodromes that include fever, cough, shortness of breath, muscle aches, and general malaise (*1*), syndromic surveillance systems frequently examine nonspecific respiratory and constitutional symptoms, collectively referred to here as symptoms for *flu-like illness*.* However, only a limited percentage of persons with such symptoms might exhibit behaviors that syndromic surveillance systems could detect. For example, ED visits for flu-like illness are likely to represent a fraction of the total number of flu-like illnesses in the community because persons with milder forms of the illness might not seek treatment at an ED.

Unlike in traditional disease surveillance systems, syndromic surveillance data are highly dependent on health-seeking or consumer behaviors. Better understanding of the actions people take when they become ill could highlight potential gaps in surveillance, identify promising data sources, and improve quantification of the magnitude of community-

* Use of *flu-like illness* in this paper is not synonymous with the CDC case definition for influenza-like illness, which is defined as fever >100°F and cough or sore throat (*2*).

level illness corresponding to syndromic alerts. Such information is also crucial to the development of simulated disease-outbreak models.

To better understand the relation between illness in the community and syndromic surveillance data regarding nonspecific health indicators, the New York City Department of Health and Mental Hygiene (DOHMH) applied information from a citywide survey on self-reported illnesses and behaviors to its syndromic surveillance system. During the spring and fall of 2003, DOHMH conducted a population-based survey to estimate citywide prevalence of chronic diseases and behavioral risk factors. The survey asked adult NYC residents about recent flu-like and diarrheal illnesses as well as about their health-seeking and consumer behaviors during flu-like illness. Prevalence estimates of behaviors during flu-like illness could provide an indication of the most frequent and timely sources of health-indicator data for use in syndromic surveillance in NYC. By combining information from the survey on the prevalence of illness with ED syndromic surveillance data, DOHMH was able estimate the *syndromic multiplier* — the number of citywide illnesses represented by each ED visit.

Objectives

The objectives of this study were to 1) examine the demographic characteristics of New Yorkers with recent flu-like or diarrheal illness, 2) describe health-seeking and consumer behaviors associated with flu-like illness, and 3) estimate the citywide burden of illness corresponding to syndromic surveillance ED visits by calculating the syndromic multiplier.

Methods

Community Health Survey

To assess annual trends in the health and health behaviors of New Yorkers, DOHMH conducts the New York City Community Health Survey (*3*), a citywide, cross-sectional telephone survey of 10,000 persons, modeled after the Behavioral Risk Factor Surveillance System (BRFSS) (*4*). The target population for the survey is noninstitutionalized NYC adults aged ≥18 years with telephones. A smaller, supplemental citywide survey, used to ask timely and seasonally related questions and to pilot test other questions for the larger survey, was administered twice in 2003, once in the spring and once in the fall. A total of 1,211 interviews were conducted during March 19–March 31, 2003, and 1,222 interviews were conducted during October 27–November 23, 2003. Interviews were conducted in English, Spanish, and Chinese. The minimum cooperation rate, using the definition provided by the American Association of Public Opinion Research (*5*), was 48% in the spring survey and 64% in the fall survey. The survey was a simple random sample; weights were applied to each observation such that the sum of the weights equaled the total adult population of NYC (N = 6,068,009, on the basis of the 2000 U.S. Census).

During both the spring and fall surveys, respondents were asked the following question about recent flu-like illness: "In the last 30 days, did you have a flu-like illness with high fever, muscle aches, and cough or sore throat?" If respondents answered yes, they were then asked about different behaviors: "During this illness, did you a) purchase an over-the-counter (OTC) medication; b) miss work or school; c) call a doctor's office for advice; d) call a nurse or other health hotline; e) visit with your regular doctor; f) visit a hospital emergency room or urgent care center; g) visit a health-care facility other than your doctor or an emergency room?" Questions regarding behavior were asked in random order to minimize bias associated with respondent fatigue. In the fall survey, respondents who replied yes to ≥2 behavior-related questions were asked to specify which action they took first. In the spring survey, respondents were also asked whether they experienced recent diarrheal illness: "In the last 30 days, did you have diarrhea with at least three loose bowel movements within 24 hours?"

Syndromic Surveillance Data

As part of the NYC syndromic surveillance system, data on ED visits, which include chief complaints, are transmitted daily to DOHMH from participating NYC hospitals (*6*). Each ED visit is categorized into one of several syndromes (i.e., respiratory, fever/influenza, diarrhea, vomiting, and asthma) on the basis of the free-text information contained within each chief complaint. Daily counts of ED visits for these syndromes are analyzed each day to detect citywide temporal increases or localized spatial clustering that might be indicative of a disease outbreak.

For this analysis, ED visits by adults (aged ≥18 years) for the respiratory or fever/influenza syndromes were considered to be visits for flu-like illness. Visits included in the diarrhea syndrome category were considered to be visits for diarrheal illness.

Estimation Methods

Using the survey data, DOHMH estimated the prevalence of self-reported flu-like illness, behaviors associated with flu-like illness, and diarrheal illness during the 30 days before the survey interview. The relative standard error (RSE) was used as a criterion of precision, calculated by dividing each estimate by its standard error; estimates with RSE of ≥30% have low preci-

sion and stability. The estimated numbers of New Yorkers with flu-like and diarrheal illness were calculated by using the weighted prevalence estimates from the survey.

Using the syndromic surveillance data, DOHMH summed the total counts of ED visits for flu-like and diarrheal illness over 30-day periods. Because the survey was administered over multiple weeks, 30-day weighted averages of ED visits were calculated by applying the percentage of respondents on each interview date to the corresponding count of ED visits for the previous 30 days.

The approximate number of citywide illnesses represented by each ED visit in the syndromic surveillance system was obtained by dividing the estimated number of New Yorkers with illness during the previous 30 days by the 30-day weighted average of ED visits. The syndromic multiplier was then calculated by multiplying the above estimate by the citywide coverage of the syndromic surveillance system (e.g., the percentage of ED visits reported out of all ED visits in NYC). By using the standard errors of the survey prevalence estimates, DOHMH also calculated 95% confidence limits on the syndromic multiplier. SAS® version 8.2 (*7*) and SUDAAN® version 8 (*8*) were used to conduct the analyses.

TABLE 1. Prevalence of flu-like illness and diarrheal illness, by age, sex, race/ethnicity, education level, and health-care access — New York City Community Health Survey, 2003

	Flu-like illness in last 30 days (n = 2,433)*		Diarrheal illness in last 30 days (n = 1,211)†	
Characteristic	**%§**	**(95% CI¶)**	**%**	**(95% CI)**
All respondents	19.6	(17.8–21.6)	8.7	(7.2–10.6)
Age				
18–64 years	22.0	(19.9–24.2)	8.9	(7.2–10.9)
≥65 years	6.3	(4.1–9.5)	7.9	(4.7–13.0)
Sex				
Male	17.6	(15.1–20.4)	7.2	(5.1–10.0)
Female	21.4	(18.8–24.2)	10.1	(7.9–12.7)
Race/ethnicity				
White, non-Hispanic	16.0	(13.6–18.9)	8.9	(6.1–11.6)
Black, non-Hispanic	19.7	(16.2–23.8)	7.9	(5.4–11.5)
Hispanic	25.8	(22.0–30.1)	8.7	(5.7–13.0)
Other	18.6	(13.0–25.7)	9.7**	(4.9–18.2)
Education level				
<High school	20.7	(16.4–25.7)	12.0	(7.5–18.6)
High school graduate	23.4	(19.4–27.9)	8.1	(5.6–11.6)
>High school	17.5	(15.3–19.9)	7.8	(6.0–10.1)
Health-care access				
Insured	18.0	(16.0–20.0)	8.8	(7.1–10.9)
Uninsured	26.7	(21.8–32.2)	7.5	(4.6–11.9)

* Asked during the spring (March 19–March 31, 2003) and fall (October 27–November 23, 2003) surveys.

† Asked during the spring (March 19–March 31, 2003) survey only.

§ Weighted prevalence estimates.

¶ Confidence interval.

** Estimate has a relative standard error of ≥30%, indicating low precision and stability.

Results

Survey Results

The overall prevalence of adult New Yorkers who reported a flu-like illness during the previous 30 days was 19.6% (Table 1), which corresponds to approximately 2.4 flu-like illnesses/person/year. The prevalence of flu-like illness was slightly higher during the fall survey (20.8%) than during the spring survey (18.5%). The prevalence of adult New Yorkers who reported a diarrheal illness during the previous 30 days was 8.7%, which corresponds to approximately one diarrheal illness/person/year.

Of all reported behaviors during a flu-like illness, respondents most frequently reported purchasing OTC medications (53.2%) (Table 2). Additionally, 32.6% reported missing work or school, 29.1% reported visiting a physician, and 21.4% reported calling a physician for advice. Respondents less frequently reported visiting an emergency department (8.8%) or calling a nurse or health hotline (3.8%). Only 18.5% of those with flu-like illness exhibited none of the health-seeking behaviors asked about in the survey.

Adults aged 18–64 years were significantly more likely to report a recent flu-like illness (22.0%) than adults aged ≥65 years (6.3%; p<0.001). Older adults reported calling a physician or visiting an ED during a flu-like illness more often than did younger adults, although those differences were not significant. No difference in prevalence of reported diarrheal illness by age group was observed. Females were slightly more likely than males to report both recent flu-like illness (21.4% and 17.6%, respectively; p = 0.05) and diarrheal illness (10.1% and 7.2%, respectively; p = 0.09). Although Hispanics were significantly more likely to report flu-like illness (25.8%) than whites (16.0%; p<0.001), limited differences were observed among the racial/ethic groups regarding behaviors during a flu-like illness.

The prevalence of flu-like illness was similar across education level. However, persons with less than a high school education were significantly less likely to miss work or school because of this illness (16.3%) than were those with more than a high school education (35.2%; p<0.001). A higher prevalence of respondents without any health insurance reported flu-like illness (26.7%) than those with health insurance (18.0%; p = 0.002), but fewer reported calling a physician, visiting a physician, or visiting an ED because of this illness.

Of those who reported flu-like illness during the previous 30 days, 50.7% reported carrying out ≥2 of the health-seeking or consumer behaviors examined by the survey. Of

TABLE 2. Prevalence of behaviors during flu-like illness (n = 460), by age, sex, race/ethnicity, education level, and health-care access — New York City Community Health Survey, 2003*

Characteristic	Purchased over-the-counter medication		Missed work or school		Visited physician		Called physician		Visited emergency department		Called nurse or health-line	
	%†	(95% CI§)	%	(95% CI)	%	(95% CI)	%	(95% CI)	%	(95% CI)	%	(95% CI)
All respondents	53.2	(47.7–58.5)	32.6	(27.8–37.9)	29.1	(24.7–33.9)	21.4	(17.5–25.9)	8.8	(6.5–11.9)	3.8	(2.4–6.1)
Age												
18–64 years	54.9	(49.3–60.3)	33.8	(28.7–39.2)	29.0	(24.5–34.0)	20.5	(16.6–25.1)	8.6	(6.2–11.8)	3.9	(2.4–6.3)
≥65 years	17.9¶	(8.0–35.2)	8.9¶	(2.2–30.0)	29.7¶	(14.9–50.5)	39.3	(20.7–61.6)	13.0¶	(3.7–36.7)	3.0¶	(0.4–18.6)
Sex												
Male	52.2	(43.9–60.4)	36.0	(28.4–44.3)	30.4	(23.5–38.3)	19.1	(13.8–25.9)	6.6	(3.7–11.7)	5.5¶	(2.9–10.3)
Female	53.8	(46.8–60.8)	30.2	(24.1–37.1)	28.1	(22.7–34.3)	23.0	(17.8–29.2)	10.4	(7.2–14.7)	2.6¶	(1.3–5.2)
Race/ethnicity												
White, non-Hispanic	50.2	(41.2–59.2)	37.5	(29.3–46.6)	30.4	(22.8–39.2)	25.7	(18.8–34.0)	6.9¶	(3.7–12.5)	6.6¶	(3.5–12.0)
Black, non-Hispanic	49.5	(38.8–60.3)	26.1	(18.1–36.0)	33.0	(24.1–43.3)	19.5	(12.9–28.3)	12.4	(7.2–20.5)	1.9¶	(0.6–6.3)
Hispanic	56.8	(47.8–65.4)	29.5	(21.8–38.7)	29.9	(22.7–38.2)	21.6	(15.0–30.2)	9.1	(5.2–15.2)	4.0¶	(1.6–9.60
Other	57.7	(39.3–74.2)	40.0	(23.5–59.2)	16.9¶	(7.8–32.9)	14.0¶	(6.0–29.2)	6.4¶	(2.3–17.0)	0¶	
Education level												
<High school	50.1	(37.9–62.3)	16.3	(9.0–27.8)	37.4	(26.7–49.6)	18.7	(11.0–30.0)	10.6¶	(5.4–19.8)	0.9¶	(0.1–6.0)
High school graduate	58.2	(47.6–68.2)	35.6	(26.1–46.4)	29.7	(21.4–39.5)	16.3	(10.6–24.4)	9.9	(5.9–16.2)	3.8¶	(1.5–9.1)
>High school	52.2	(45.0–59.4)	35.2	(28.8–42.2)	27.0	(21.4–33.4)	25.1	(19.5–31.6)	7.8	(4.8–12.5)	4.9	(2.7–8.6)
Health-care access												
Insured	51.1	(45.0–57.2)	32.3	(26.8–38.2)	33.7	(28.4–39.5)	23.5	(19.0–28.7)	9.7	(6.9–13.5)	3.8	(2.2–6.3)
Uninsured	61.7	(49.9–72.3)	34.3	(24.3–46.0)	14.6	(8.8–23.1)	16.0	(9.2–26.3)	6.6¶	(3.2–13.1)	4.2¶	(1.4–11.6)

* Asked during the spring (March 19–March 31, 2003) and fall (October 27–November 23, 2003) surveys.
† Weighted prevalence estimates.
§ Confidence interval.
¶ Estimate has a relative standard error ≥30%, indicating low precision and stability.

these, 36.6% of respondents reported purchasing OTC medications first, before carrying out any other behavior (Table 3). An additional 30.3% of respondents reported first missing work or school. The next most common initial behaviors were visiting a physician (16.2%) and calling a physician for advice (11.8%). Only 3.3% of respondents reported first visiting the ED before any other behavior. The least common first behavior was calling a nurse or other health hotline (0.7%).

Calculating the Syndromic Multiplier

Approximately three-quarters of all ED visits in NYC are captured by the city's syndromic surveillance system. In February and March 2003, a total of 39 hospitals provided daily ED data to the system, representing 58% of hospitals and 74% of ED visits in NYC. In October and November 2003, a total of 40 hospitals (60% of hospitals, 76% of visits) provided daily ED data. The mean (standard deviation) daily counts of adult visits were 476 (70) for flu-like illness and 46 (12) for diarrheal illness. Flu-like illness accounted for 9% of all ED visits; diarrheal illness accounted for <1% of all visits. The daily counts of ED visits for flu-like and diarrheal illness during September 1, 2002–November 30, 2003 are provided (Figure).

By calculating the syndromic multiplier, DOHMH estimated that each ED visit represented 60.0 flu-like illnesses among adult New Yorkers, including 76.5 illnesses among those aged 18–64 years and 11.1 illnesses among those aged ≥65 years (Table 4). For diarrheal illness, each ED visited was estimated to represent approximately 250.6 illnesses among adults citywide.

TABLE 3. Frequency of initial behavior during flu-like illness among persons who took ≥2 health-seeking actions (n = 108) — New York City Community Health Survey, 2003*

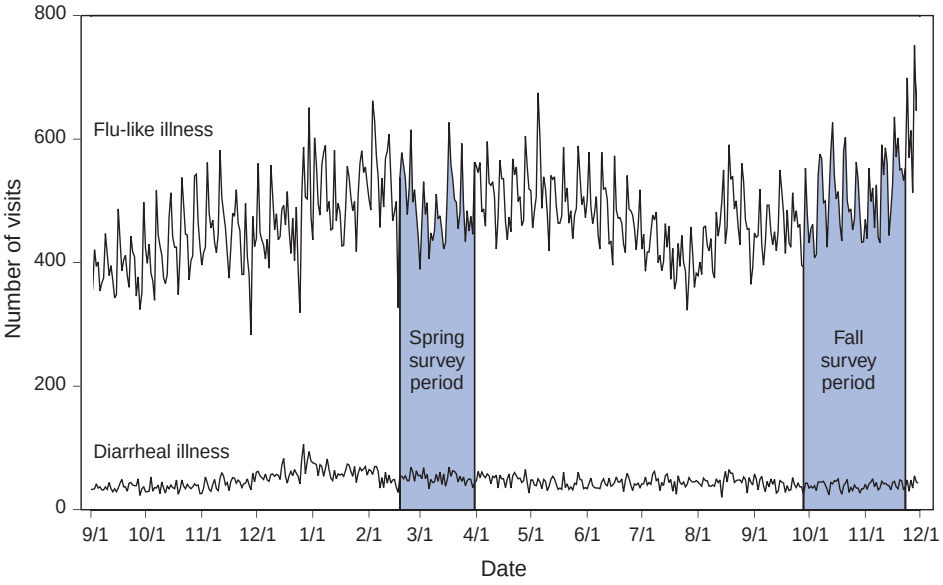
Behavior	%†	(95% CI§)
Purchased over-the-counter medication	36.6	(26.6–48.0)
Missed work or school	30.3	(20.6–42.1)
Visited physician	16.2	(10.1–25.1)
Called physician	11.8	(6.7–19.9)
Visited emergency department	3.3¶	(1.1–9.8)
Called nurse or health hotline	0.7¶	(0.2–3.0)

* Asked during the fall (October 27–November 23, 2003) survey only.
† Weighted prevalence estimates.
§ Confidence interval.
¶ Estimate has a relative standard error of ≥30%, indicating low precision and stability.

Discussion

Understanding the frequency and timing of health behaviors during illness provides valuable context for syndromic surveillance and can help guide development of simulated disease-outbreak models. The prevalence estimates of flu-like and diarrheal illnesses determined by this population-based survey of adult NYC residents are similar to those from other population-based surveys of communitywide flu-like

FIGURE. Daily emergency department visits for flu-like and diarrheal illness by adults — New York City, September 1, 2002–November 30, 2003



Source: New York City syndromic surveillance system.

illness (*9*) and diarrheal illness (*10,11*). By using these prevalence estimates, DOHMH was able to estimate the syndromic multiplier — the number of citywide illnesses that each ED visit represents.

Although OTC medication purchases and absenteeism appear to be two of the more frequent and timely health behaviors during flu-like illness, the lack of specificity and the variability caused nonhealth-related events (e.g., promotions influencing OTC medication purchases, or reasons for absenteeism other than illness) might reduce the suitability of these data sources for timely outbreak detection. The survey results also indicate that outpatient physician encounters were considerably more frequent and timely than ED visits. Where available, data on outpatient physician encounters might offer a degree of disease specificity and an ability to investigate signals equal to or greater than the more commonly monitored ED chief-complaint data.

Population-based surveys can help identify gaps in current syndromic surveillance systems. For example, this survey determined that persons without health insurance were more likely to report recent illness but less likely to seek care. Including data from outpatient sites that provide health care to medically indigent and uninsured persons might improve the representativeness of syndromic surveillance data.

Because the survey relied on self-reports, the data might suffer from bias caused by inexact recall of the timing of recent flu-like illnesses and resulting behaviors. Respondents might have reported illnesses and behaviors that occurred ≥30 days before the survey; this type of recall bias is often encountered in surveys eliciting temporal-based information (*12*).

TABLE 4. Calculation of the syndromic multiplier by using prevalence estimates of flu-like illness and diarrheal illness from the New York City Community Health Survey and 30-day counts of emergency department (ED) visits from the New York City Syndromic Surveillance System, 2003

	Community Health Survey			Syndromic Surveillance System			
Characteristic	%*	(95% CI†)	Weighted population estimate§	30-day count of ED visits	% citywide coverage of all ED visits	Syndromic multiplier¶	(95% CI)
Flu-like illness during previous 30 days							
(n = 2,433)**	19.6	(17.8–21.6)	1,187,956	14,849	75	60.0	(54.4–66.0)
Age							
18–64 years	22.0	(19.9–24.2)	1,132,361	11,105	75	76.5	(69.2–84.2)
≥65 years	6.3	(4.1–9.5)	55,595	3,743	75	11.1	(7.2–16.9)
Diarrheal illness during previous 30 days							
(n = 1,211)††	8.7	(7.2–10.6)	537,363	1,578	75	250.6	(205.8–304.0)
Age							
18–64 years	8.9	(7.2–10.9)	457,360	1,307	75	262.4	(212.0–323.3)
≥65 years	7.9	(4.7–13.0)	70,004	271	75	193.6	(114.7–319.6)

* Weighted prevalence estimates.
† Confidence interval.
§ Of adult New York City residents (N = 6,068,009).
¶ Using the following calculation: (weighted population estimate / 30-day count of ED visits × percentage of citywide coverage of all ED visits.
** Asked during the spring (March 19–March 31, 2003) and fall (October 27–November 23, 2003) surveys.
†† Asked during the spring (March 19–March 31, 2003) survey only.

As a result, the self-reported 19.6% with a recent flu-like illness and subsequent 8.8% who visited an ED might represent overestimates of the illness' true prevalence.

In addition, the syndromic surveillance case definition for flu-like illness is unlikely to identify all ED visits for flu-like illness. Previous studies have determined that the sensitivity of ED chief-complaint data ranges from 44% when medical chart review is used as the standard (*13*) to 81% when discharge diagnosis is used as the standard (*14*). However, visits for other reasons are unlikely to be misclassified as visits for flu-like illness; the specificity of chief-complaint data in the two studies were 97% and 95%, respectively. Consequently, this study's calculation of 60 citywide illnesses/ED visit for flu-like illness might overestimate the true ratio.

These surveys were conducted during periods without any known outbreaks of influenza or gastrointestinal illness. A primary objective of syndromic surveillance is to detect abnormal increases in behaviors associated with flu-like illness not necessarily attributable to influenza (e.g., to detect events of biologic terrorism). However, these estimates of citywide illness might change during an outbreak if severity of illness alters the pattern of health-seeking behaviors.

Conclusions

By combining data from a citywide survey with syndromic surveillance data, DOHMH was able to use the syndromic multiplier to estimate the number of illnesses in the community represented by each ED visit. Survey responses regarding the actions persons take during a flu-like illness provided important information about health-seeking behaviors and about the representativeness of different data sources used in syndromic surveillance.

References

1. Rotz LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM. Public health assessment of potential biological terrorism agents. Emerg Infect Dis 2002;8:225–230.
2. CDC. Flu activity: reports and surveillance methods in the United States. Atlanta, GA: US Department of Health and Human Services, CDC, 2004. Available at http://www.cdc.gov/flu/weekly/fluactivity.htm.
3. New York City Department of Health and Mental Hygiene (DOHMH). NYC community health data. New York: DOHMH, 2004. Available at http://www.nyc.gov/html/doh/html/survey/survey.html.
4. CDC. Behavioral risk factor surveillance system user's guide. Atlanta: US Department of Health and Human Services, CDC, 1998.
5. American Association of Public Opinion Research. Standard definitions: final disposition of case codes and outcome rates for surveys. Ann Arbor, Michigan: AAPOR, 2000.
6. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 2004;10:858–64.
7. SAS Institute, Inc. SAS,® version 8.2 [Software]. Cary, NC: SAS Institute, Inc., 1999.
8. RTI International. SUDAAN,® version 8. [Software]. Research Triangle Park, NC: RTI International.
9. Monto AS, Sullivan KM. Acute respiratory illness in the community. Frequency of illness and the agents involved. Epidemiol Infect 1993;110:145–60.
10. Kuusi M, Aavitsland P, Gondrosen B, Kapperud G. Incidence of gastroenteritis in Norway—a population-based survey. Epidemiol Infect 2003;131:591–7.
11. Herikstad H, Yang S, Van Gilder TJ, et al. A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996–7. Epidemiol Infect 2002;129:9–17.
12. Huttenlocher J, Hedges LV, Bradburn NM. Reports of elapsed time: bounding and rounding processes in estimation. J Exp Psychol Learn Mem Cogn 1990;16:196–213.
13. Espino JU, Wagner MM. Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. Proc AMIA Symp 2001;164–8.
14. Begier E, Sockwell D, Branch LM, et al. The National Capitol Region's ED syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? Emerg Infect Dis 2003;9:393–6.

Comparison of Office Visit and Nurse Advice Hotline Data for Syndromic Surveillance — Baltimore-Washington, D.C., Metropolitan Area, 2002

Jade Vu Henry,[1] S. Magruder,[2] M. Snyder[1]
[1]*Kaiser Permanente of the Mid-Atlantic States, Rockville, Maryland;* [2]*Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland*

Corresponding author: Jade Vu Henry, 2102 East Jefferson Street, Rockville, MD 20849. Telephone: 301-816-6891; Fax: 301-816-7465; E-mail: jade.v.henry@kp.org.

Abstract

Introduction: *Kaiser Permanente of the Mid-Atlantic States (KPMAS) is collaborating with the Electronic Surveillance System for Early Notification of Community-Based Epidemics II (ESSENCE II) program to understand how managed-care data can be effectively used for syndromic surveillance.*

Objectives: *This study examined whether KPMAS nurse advice hotline data would be able to predict the syndrome diagnoses made during subsequent KPMAS office visits.*

Methods: *All nurse advice hotline calls during 2002 that were linked to an outpatient office visit were identified. By using* International Classification of Diseases, Ninth Revision *(ICD-9) codes, outpatient visits were categorized into seven ESSENCE II syndrome groups (*coma, gastrointestinal, respiratory, neurologic, hemorrhagic, infectious dermatologic, *and* fever*). Nurse advice hotline calls were categorized into ESSENCE II syndrome groups on the basis of the advice guidelines assigned. For each syndrome group, the sensitivity, specificity, and positive predictive value of hotline calls were calculated by using office visits as a diagnostic standard. For matching syndrome call-visit pairs, the lag (i.e., the number of hours that elapsed between the date and time the patient spoke to an advice nurse and the date and time the patient made an office visit) was calculated.*

Results: *Of all syndrome groups, the sensitivity of hotline calls for* respiratory *syndrome was highest (74.7%), followed by hotline calls for* gastrointestinal *syndrome (72.0%). The specificity of all nurse advice syndrome groups ranged from 88.9% to 99.9%. The mean lag between hotline calls and office visits ranged from 8.3 to 50 hours, depending on the syndrome group.*

Conclusions: *The timeliness of hotline data capture compared with office visit data capture, as well as the sensitivity and specificity of hotline calls for detecting* respiratory *and* gastrointestinal *syndromes, indicate that KPMAS nurse advice hotline data can be used to predict KPMAS syndromic outpatient office visits.*

Introduction

Across the United States, managed care organizations operate clinical and administrative information systems to support the routine delivery of health-care services to their members. Data from these information systems offer promise for enhancing public health surveillance activities in communities where managed care organizations operate (*1*–*4*). However, despite the widely recognized potential of using managed-care data for tracking community health indicators, managed care organizations and public health departments have previously had limited incentive to form alliances to improve public health surveillance (*5*).

In 2001, when residents of the Baltimore-Washington, D.C., metropolitan area were diagnosed with inhalational anthrax, Kaiser Permanente of the Mid-Atlantic States (KPMAS) experienced a public health emergency requiring daily interaction with local health authorities. In the course of delivering health care to victims of this biologic terrorism, KPMAS employees became acutely aware of the need for stronger links with public health agencies at all government levels (*6*). While responding to the anthrax crisis, KPMAS was also able to demonstrate the public health potential of its administrative and clinical information systems, having searched its databases to identify and contact hundreds of enrollees at-risk to recommend testing and treatment options (*7*). Drawing from the lessons of this front-line experience, KPMAS began looking for substantive ways to support the local public health infrastructure.

Since 2002, the KPMAS Research Department has collaborated with the Electronic Surveillance System for Early Notifi-

cation of Community-Based Epidemics II (ESSENCE II) program. KPMAS researchers view ESSENCE II as a community health partnership — a voluntary collaboration of diverse community organizations that are pursuing a shared interest in community health (*8*). The ESSENCE II program seeks to strengthen the local public health infrastructure by developing a regional syndromic surveillance system. To achieve this objective, ESSENCE II has drawn together a multisector, multidisciplinary group of researchers, health-care providers, and public health authorities with expertise in medicine, mathematics, and public health, as well as access to health data.

KPMAS operates a secure information system that routinely captures data from an array of health-care operations, including laboratory tests, radiology procedures, pharmacy prescriptions, inpatient and outpatient visits, as well as membership demographics, appointment history, clinician notes, and nurse advice hotline calls. All of these data sources have potential value for syndromic surveillance. However, limited staff and funding are available to make KPMAS health-care information accessible for public health surveillance. Understanding the relative strengths and weaknesses of each KPMAS data stream can help prioritize information-technology investments for syndromic surveillance. To assess the performance of potential new data streams for the ESSENCE II surveillance system, this paper compares the epidemiologic properties of KPMAS nurse advice hotline data with outpatient office visit data obtained during January 1–December 31, 2002. The goal of this study was to determine whether nurse advice hotline data would be able to predict the syndromic diagnoses made during a subsequent office visit.

Methods

Population and Delivery System

KPMAS contracts with 36 local hospitals and operates 30 outpatient medical centers to deliver health services to a population of >500,000 members. To facilitate the continuity of care, each member is assigned a unique permanent identification number that can be used to retrieve and link all administrative and clinical data. Appointment scheduling and the nurse advice hotline function together within the KPMAS call center, which serves as a major entry point into the delivery system. Appointment clerks in the call center schedule routine appointments, while nurses operate an advice hotline to administer protocol-driven, medically appropriate advice over the telephone and to schedule acute-care office visits when necessary. In 2002, 369,646 members made 1,497,686 calls to the nurse advice hotline.

Data Link Between Individual Nurse Advice Hotline Calls and Outpatient Office Visits

The KPMAS information system captures not only the date and time a patient is seen for an outpatient office visit (Encounter_Date), but also the date and time that appointment is scheduled (Enc_File_Date). A third data field (Advice_Date) captures the date and time the patient contacted the nurse advice hotline. For this study, a hotline call and an office visit were defined as linked if all of the following criteria were met: 1) the patient identification number assigned to the hotline call matched the patient identification number assigned to the office visit; 2) the date of the hotline call matched the date the patient called for the appointment; and 3) the time of the hotline call either matched or preceded the time the patient called for an appointment.

Categorizing KPMAS Data into Syndromic Groups

All outpatient visits are assigned one or more diagnostic codes from the *International Classification of Diseases, Ninth Revision* (ICD-9). These diagnostic codes reflect the patients' presenting conditions, symptoms, presumed diagnoses, and definitive diagnoses. Before this study, ESSENCE II developed sets of ICD-9 codes representing clinical manifestations of potential infectious disease outbreaks (*9*). Those sets of ICD-9 codes were used to categorize all KPMAS outpatient visits during 2002 into seven syndrome groups: *coma, gastrointestinal, respiratory, neurologic, hemorrhagic, infectious dermatologic,* and *fever*.

The information system that supports the KPMAS nurse advice hotline does not use the ICD-9 coding system. Instead, as nurses speak with patients, they select one or more KPMAS advice guidelines from a drop-down menu on the computer screen. These advice guidelines are based on 586 current KPMAS nurse advice clinical-practice protocols that correspond to patient-reported symptoms and presumed diagnoses. The KPMAS advice guidelines were classified into syndrome groups corresponding to the seven ICD-9–based ESSENCE syndrome categories (Box). Of the 586 KPMAS advice guidelines, 68 were used to define syndrome categories for nurse advice hotline calls (the remaining 518 were for conditions not of interest for syndromic surveillance). Because none of the advice guidelines corresponded to the ESSENCE II coma syndrome group, this category was dropped from analysis and only six of the seven categories were used.

BOX. Classification for coding of nurse advice guidelines — Kaiser Permanente of the Mid-Atlantic States (KPMAS)

GI (gastrointestinal)
Abdominal pain, adult
Abdominal pain, pediatric
Acute GI, gastroenteritis, adult
Diarrhea, 0–24 months, pediatric
Diarrhea, >2 years, pediatric
Diarrhea, adult
Diarrhea, long-term care
Diarrhea, pediatric
Diarrhea, prenatal, OBGYN
HIV diarrhea, adult
Nausea, adult
Nausea, pediatric
Vomiting and hyperemesis, adult
Vomiting, long-term care
Vomiting, pediatric

DERMHEM (hemorrhagic manifestations)
Bruise/hematoma, adult

NEURO (neurologic)
Headache, adult
Headache, long-term care
Headache, pediatric
HIV headache, adult
HIV mental status changes, adult
Meningitis, adult
Meningitis, neonates
Meningitis, pediatric, 3 months–2 years
Meningitis, pediatric, children/young adults
Meningitis, pediatric, infants
Meningitis, pediatric, >2 years

FEVER
Fever, adult
Fever, long-term care
Fever, neonatal
Fever, pediatric
HIV fever, adult

RESP (respiratory infection)
Asthma, adult
Asthma, pediatric
Bronchiolitis, pediatric
Bronchitis, acute, adult
Croup, pediatric
Earache, pediatric
HIV dyspnea, adult
HIV pneumonia, adult
Influenza, adult
Influenza, pediatric
Laryngitis, adult
Respiratory distress, adult
Shortness of breath, long-term care
Sore throat, adult
Sore throat, pediatric
Throat culture, positive
Upper respiratory infection, adult
Upper respiratory infection, long-term care
Upper respiratory infection, pediatric

DERMINF (dermatologic, infectious)
Chicken pox, adult
Chicken pox, pediatric
Chicken pox, prenatal, OBGYN
Fifth disease, pediatric
Hand, foot, mouth disease, pediatric
Herpes zoster, adult
Herpes zoster/shingles, adult
Measles, pediatric
Rash, adult
Rash/fungal infection, adult
Rashes, pediatric
Rash, prenatal, OBGYN
Roseola, pediatric
Shingles
Smallpox

All nurse advice hotline calls received during 2002 were categorized according to the new syndrome groups. Approximately 20% of all hotline calls were assigned at least one syndromic advice guideline. The remaining 80% of hotline calls were associated with advice guidelines not relevant to syndromic surveillance and therefore not used in the hotline syndrome classification scheme.

Because more than one ICD-9 diagnosis can be assigned to a patient during a given office visit, single office visits can fall into more than one ESSENCE syndrome group and be counted multiple times. Similarly, a nurse can use more than one advice guideline during a given call, in which case the call would fall into more than one advice syndrome group and be counted more than once. Multiple calls or multiple visits by a given patient in a single day for the same syndrome are counted only once.

Analysis

The total numbers of hotline calls and outpatient visits by syndrome group received during January 1–December 31, 2002, were determined and compared. All hotline calls linked to a subsequent outpatient visit were then identified. For each syndrome group, the sensitivity, specificity, and positive predictive value of hotline calls was calculated, using office visits as a diagnostic standard.

All hotline calls whose syndrome assignments matched those of the linked office visit were identified. For each of these matching syndrome call-visit pairs, the time lag (i.e., the number of hours that elapsed between the date and time the patient spoke to an advice nurse [Advice_Date] and the date and time the patient was seen for the office visit [Encounter_Date]) was calculated. Univariate statistics on this time lag were then generated.

The study was reviewed and approved by the legal department of the Kaiser Foundation Research Institute, as well the KPMAS Institutional Review Board for the protection of hu-

man subjects. All analysis of patient-level data was performed by researchers at KPMAS. No patient-level data were released from KPMAS for this analysis.

Results

Total nurse advice hotline calls and total outpatient office visits, by syndrome group, were compared (Table 1). Of all syndrome groups, *respiratory* and *gastrointestinal* syndromes generated the highest volume of hotline calls and office visits. Within every syndrome group, patients made at least twice as many hotline calls as office visits, with the exception of the *respiratory* syndrome category, which resulted in 242,785 hotline calls and 201,402 office visits.

Approximately 570,500 hotline calls were linked to an office visit. The exact count varies slightly among syndrome groups because multiple calls or multiple visits by a given patient in a single day for the same syndrome were counted only once. The sensitivity, specificity, and positive predictive value of hotline calls for detecting syndromes diagnosed during office visits were calculated (Table 2). Of all syndrome groups, the sensitivity of hotline calls for *respiratory* syndrome was highest (74.7%), followed by hotline calls for *gastrointestinal* syndrome (72.0%). Hotline calls for *respiratory* and *gastrointestinal* syndromes also had the highest positive predictive value. Sensitivity was lowest for hotline calls in the *hemorrhagic* group. The specificity of all nurse advice syndrome groups was high, ranging from 88.9% to 99.9%.

Univariate statistics for the lag between syndromic hotline calls and their matching syndromic office visits were generated (Table 3). The mean lag between hotline calls and office visits ranged from 8.3 to 50 hours, depending on the syndrome group. Hotline calls in the *hemorrhagic* syndrome category provided the greatest mean lead time (50 hours) over corresponding office visits; however, hotline calls and office visits were both categorized as *hemorrhagic* in only 14 instances. The median lead time ranged from 4 hours for *gastrointestinal* and *dermatologic/infectious* syndromes to 25 hours for hemorrhagic syndrome. The mode for the lag ranged from 1 to 3 hours, depending on the syndrome.

TABLE 1. Number of nurse advice hotline calls and outpatient office visits, by syndrome group — Kaiser Permanente of the Mid-Atlantic States, 2002

Syndrome group	No. of hotline calls*	No. of outpatient office visits*
Gastrointestinal	72,107	26,521
Respiratory infection	242,785	201,402
Neurologic	22,957	144
Dermatologic, infectious	20,117	440
Fever	17,866	5,230
Hemorrhagic manifestation	1,580	456

* Multiple calls or multiple visits by a given patient in a single day for the same syndrome are counted only once.

TABLE 2. Validity of syndromic nurse advice hotline calls to detect syndromic outpatient office visits — Kaiser Permanente of the Mid-Atlantic States, 2002

Syndrome group	Office visit in syndrome group	Office visit not in syndrome group
Gastrointestinal		
Call in syndrome group	13,178	22,376
Call not in syndrome group	5,124	530,016
Sensitivity = 72.0%	**Specificity = 95.9%**	**PPV* = 37.1%**
Respiratory infection		
Call in syndrome group	88,410	50,078
Call not in syndrome group	29,977	402,818
Sensitivity = 74.7%	**Specificity = 88.9%**	**PPV = 63.8%**
Neurologic		
Call in syndrome group	10	12,478
Call not in syndrome group	27	557,906
Sensitivity = 27.0%	**Specificity = 97.8%**	**PPV = 0.1%**
Dermatologic, infectious		
Call in syndrome group	166	10,491
Call not in syndrome group	118	559,639
Sensitivity = 58.5%	**Specificity = 98.2%**	**PPV = 1.6%**
Fever		
Call in syndrome group	1,153	8,600
Call not in syndrome group	2,376	558,414
Sensitivity = 32.7%	**Specificity = 98.5%**	**PPV = 11.8%**
Hemorrhagic manifestation		
Call in syndrome group	14	750
Call not in syndrome group	91	569,546
Sensitivity = 13.3%	**Specificity = 99.9%**	**PPV = 1.8%**

* PPV = positive predictive value.

Discussion

Analysis indicates that nurse advice hotline data is 4–50 hours timelier for syndrome detection than outpatient office visit data, depending on the syndrome group. This lead time

TABLE 3. Syndromic nurse advice hotline calls with matched* syndromic outpatient office visit — Kaiser Permanente of the Mid-Atlantic States, 2002

Syndrome group	No. of calls with matching visit	Hours between hotline call and office visit		
		Mean	Median	Mode
Gastrointestinal	13,178	12.0	4.0	2.0
Respiratory infection	88,410	12.3	5.0	2.0
Neurologic	10	23.6	11.0	3.0
Dermatologic, infectious	166	8.17	4.5	1.0
Fever	1,153	8.3	5.0	2.0
Hemorrhagic manifestation	14	50.0	25.0	—

* A syndromic call that generates an office visit in the same syndrome group is defined as *matched*.

might even be substantially greater because ICD-9 codes for KPMAS office visits are not immediately entered into the information system; as much as a 1-month lag can exist between the time a patient is seen and the time information from that visit is available for data analysis.

Of >1 million hotline calls made during 2002, approximately 570,500 were linked to an outpatient visit. KPMAS patients made at least twice as many hotline calls as office visits within each syndrome category, with the exception of *respiratory* syndrome. A critical consideration in assessing the ability of syndromic advice calls to detect syndromic office visits is that many advice calls do not generate an office visit.

Although all of the hotline call syndrome groups demonstrated high specificity relative to office visits, sensitivity and positive predictive value varied according to syndromic group. Further research is needed to explain the differences observed between hotline data and office-visit data. The performance of hotline data might be compromised by the data stream's emphasis on symptoms rather than clinical presentations and definitive diagnoses. Alternatively, these observed discrepancies might identify opportunities to add or remove advice guidelines from syndrome classifications of nurse advice hotline calls. Patient health-seeking behavior might also account for part of the differences observed between the number of calls and similarly grouped visits (e.g., work requirements, child-care needs, or transportation barriers might lead certain patients to use the advice hotline exclusively in place of a clinical examination). Finally, coding practices and other provider behaviors, as well as delivery-system factors (e.g., appointment access), might generate differences between counts of calls and visits. Health-services research aimed at understanding how patient, provider, and delivery-system factors relate to syndrome classifications might be helpful in establishing the theoretical underpinnings for effective outbreak-detection algorithms.

Conclusions

This study examined the relative value of two alternative health-care data streams, nurse advice hotline calls and outpatient office visits, collected from a single, integrated delivery system. The timeliness of hotline data capture compared with office visit data capture, as well as the sensitivity and specificity of hotline calls for detecting *respiratory* and *gastrointestinal* syndromes, indicate that KPMAS nurse advice hotline data can be used to predict KPMAS syndromic outpatient office visits.

This analysis did not attempt to address whether KPMAS data could be used to detect epidemics in the broader Washington, D.C.-area community. Additional studies assessing the external generalizability of KPMAS data should be performed to determine whether KPMAS can serve in a sentinel surveillance capacity.

Acknowledgments

Staff from the U.S. Department of Defense, Global Emerging Infections Surveillance and Response System, provided guidance in creating the syndrome categories for KPMAS advice guidelines, and Eileen McLaughlin shared her expertise on the KPMAS Call Center operations and information systems. This work was sponsored by the Defense Advanced Research Projects Agency (DARPA) as part of the Bio-ALIRT program. This paper also benefited from the comments of an anonymous reviewer.

References

1. Stoto MA, Abel C, Dievler A, eds. Healthy communities: new partnerships for the future of public health. Washington, DC: National Academies Press, 1996.
2. Showstack J, Lurie N, Leatherman S, Fisher E, Inui T. Health of the public: the private sector challenge. JAMA 1996;276:1071–4.
3. Halverson PK, Mays GP, Kaluzny AD, Richards TB. Not-so-strange bedfellows: models of interaction between managed care plans and public health agencies. Milbank Q 1997;75:337.
4. Davis JR, ed. Managed care systems and emerging infections: challenges and opportunities for strengthening surveillance, research, and prevention, workshop summary. Washington, DC: National Academies Press, 2000.
5. Mays GP, Halverson AD, Kaluzny AD, Norton EC. How managed care plans contribute to public health practice. Inquiry 2000–2001;37:389–410.
6. Rosenbuam S, Skivington S, Praeger S. Public health emergencies and the public health/managed care challenge. Journal of Law, Medicine, & Ethics 2002;30(3 Suppl):63–9.
7. Mcgee MK. The bioterrorism threat is forcing health care to lose its aversion to IT. InformationWeek [serial on the Internet], Nov. 19, 2001. Available at http://www.informationweek.com/story/showArticle.jhtml?articleID=6507871.
8. Mitchell SM. The governance and management of effective community health partnerships: a typology for research, policy, and practice. Milbank Q 2000;78:241–89.
9. Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC, area. Am J Prev Med 2002;23:180–6.

Progress in Understanding and Using Over-the-Counter Pharmaceuticals for Syndromic Surveillance

Steven F. Magruder, S. Happel Lewis, A. Najmi, E. Florio
Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland

Corresponding author: Steven F. Magruder, Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland, 20723-6099. Telephone: 443-778-6537; Fax: 443-778-5950; E-mail: steve.magruder@jhuapl.edu.

Abstract

Introduction: *Public health researchers are increasingly interested in the potential use of monitoring data on over-the-counter (OTC) pharmaceutical sales as a source of timely information about community health. However, fundamental uncertainties persist, including how timely such information is and how best to aggregate information about hundreds of products.*

Objectives: *This analysis provides new information about OTC timeliness and illustrates a method of OTC product aggregation for surveillance purposes.*

Methods: *Timeliness measurements were made by correlating pharmaceutical sales counts with counts of physician encounters, after adjustment to remove seasonal effects from both counts. OTC product aggregations were formed by a two-stage process. In the first stage, individual products were placed into small groups based on qualitative observations. In the second stage, a clustering algorithm was used to form supergroups (i.e., product group clusters) sharing similar sales histories.*

Results: *Even after seasonal correction, OTC counts correlated with clinical measures of community illness. However, the lead time of nonseasonal fluctuations was substantially shorter than that for uncorrected data. The clustering approach produced 16 meaningful supergroups containing products that behaved approximately alike.*

Conclusions: *Measurements of OTC lead time sensitive to the timing of annual cyclic trends in the behavior of persons seeking health care do not reliably indicate the lead time observed for short-term (e. g. weekly or monthly) fluctuations in community health-care utilization.*

Introduction

Data on the sale of over-the-counter (OTC) pharmaceutical products might provide a convenient, meaningful, and timely indicator of public health conditions (*1*–*6*). Monitoring sales of OTC products offers at least three advantages as possible early indicators of public health. First, these products are widely used. Second, a reliable and detailed electronic record is made instantly at the time of each sale, and such records are aggregated regionally for commercial purposes; these electronic records can be readily transmitted to aid in health surveillance. Finally, OTC data also capture the location of sale and type of product (and, by implication, the symptom[s] that the product is intended to relieve).

Despite growing interest in OTC data, certain questions persist, including 1) how to interpret OTC sales data, 2) how much lead time these data should be expected to provide, 3) how to aggregate OTC products into informative product groupings, 4) how to control confounding factors, and 5) which product sales correlate with which types of illnesses. This report outlines progress in answering two of those questions. With respect to OTC lead time in tracking trends in health-care utilization, the analysis indicates that lead-time measurements based on the timing of annual cyclic trends can be longer than those based on short-term fluctuations, which are more relevant to public health surveillance. With respect to appropriate aggregations of OTC products, the report describes a method used by the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II) (*7*). The actual product aggregations identified might also provide insights for future study.

OTC Lead Time: Short-Term and Seasonal Observations

Multiple studies have attempted to quantify the timeliness of OTC sales compared with other indicators of public health (*1,5,6*). A 1964 study based on two outbreaks in a single city identified a substantial peak in cold remedy sales at the beginning of an increase in encounters with clinical patients known

to be infected with influenza B virus and 1 week before the peak in those encounters; an earlier increase in cold remedy sales was approximately coincident with the early winter increase in noninfluenza respiratory virus activity (*1*).

A second study compared the time series of hospital-discharge diagnoses to OTC electrolyte sales for six cities and over three annual cycles (*5*). Because these discharge diagnoses were tagged with the time of hospital admission, they could be viewed as a proxy for a chief-complaint data source. Lead times were measured by two methods: cross-correlation of the raw time series and comparison of the times for the first detectable increase each year. The two methods produced consistent estimates indicating approximately 2-week lead times for pediatric electrolyte sales relative to pediatric hospital admissions for respiratory and diarrheal disease. Lead times measured by both methods are sensitive to the seasonal variation of the two data sources; the timing of events that occur on shorter time scales might be obscured.

A third study compared a time series of outpatient insurance-claim diagnoses for acute respiratory conditions to OTC sales of influenza remedies in six different subregions of the Maryland–Washington, D.C.–Virginia area (*6*). Lead times were estimated by cross-correlation of data that were corrected for day-of-week effects and for the effect of the late-December holiday period. Measured peak correlations ranged from 0.86 to 0.93, and the average measured lead time of OTC sales relative to outpatient physician encounters was 2.8 days (range: 2–7 days). Although these results also were dominated by seasonal trends, this report presents corresponding results with seasonal effects removed.

Although certain natural and societal processes that occur annually could influence these results, such processes might not be important for short-term surveillance time scales, and the applicability of seasonal results might be questionable. For public health surveillance applications, the timing of seasonal trends is not the quantity of primary interest. More often, disease surveillance seeks timely recognition of short-term (e.g., weeks or days) health trends.

OTC Product Aggregations

Because the >1,000 OTC pharmaceutical products that are of potential interest for public health surveillance compete for customers with the same ailments, aggregation of related products is necessary to obtain statistically useful inferences about the number of people feeling ill. The goal of an aggregation method is to combine products that are used by the same demographic groups to treat the same illnesses (defined as a given combination of symptoms and by the relative severity of those symptoms). Differences in sales between products in an aggregated product group would then be irrelevant for public health surveillance. By contrast, when products are used by different demographic groups or to treat different symptoms, then aggregation of these products could compromise specificity and be less useful.

Data Sources

This analysis relied on two data sources identical to those used in a previous study (*6*). The first source was pharmacy-sales data from approximately 300 drugstores in the Maryland–Washington, D.C.–Virginia area. The pharmacy data included store location, product sold, number of units sold, and date sold; no information was provided that would identify the purchaser. For the timing study, only remedies for treating influenza were used. For the OTC aggregation study, a larger set of product categories was used, including cough, cold, allergy, sore throat, fever, "flu," antidiarrheal, bronchial, sinus, and pain remedies. The second data source was insurance-billing data from approximately 13,000 outpatient clinics and doctors offices in the Maryland–Washington, D.C.–Virginia area. These data included the patient's geographic region, the date of the patient-physician encounter, and the primary diagnostic code used for billing purposes. Not all patients from these 13,000 clinics were included. A weekly average of approximately 4,000 encounters was reported for acute respiratory conditions in all geographic regions. Only diagnostic codes of interest for syndromic surveillance were collected, and only acute respiratory diagnoses were used in the analysis.

Methods

OTC Lead Time

Both the physician acute respiratory encounter data and the OTC influenza remedy data were modeled by a Poisson regression. The covariates were a linear time ramp, a sinusoidal annual cycle (*8*), day-of-week factors, and a day-of-week/annual cycle interaction term. Holidays and heavy snow days were ignored when the regression parameters were estimated.

After the data were fitted to a model of seasonality, separate considerations were made of seasonal and nonseasonal trends in the data. Weekly cycles were removed from the OTC- and physician-encounter–model fits by smoothing with a 7-day moving average window. The resulting smoothed model fits contained only linear and seasonal variations. The seasonal contribution to lead time was measured by cross-correlating the smoothed model fits. Nonseasonal contributions were measured by correlating the residuals of the model fits (smoothed actual counts divided by smoothed model fit) for

each of the subregions in the study. A comparison of these residuals for the most populous region included in the study, the Urban National Capital Area (consisting of the urban and suburban areas near Washington, D.C., Baltimore, Maryland, and the corridor in between) is provided (Figure 1). A strong correspondence was observed between fluctuations in OTC sales and fluctuations in physician encounters, even after a sinusoidal annual cycle was removed from both. These residuals exhibited smaller correlations than did the original data, but because they were not driven by cyclic annual trends, the relevance to time-critical public health surveillance was clearer.

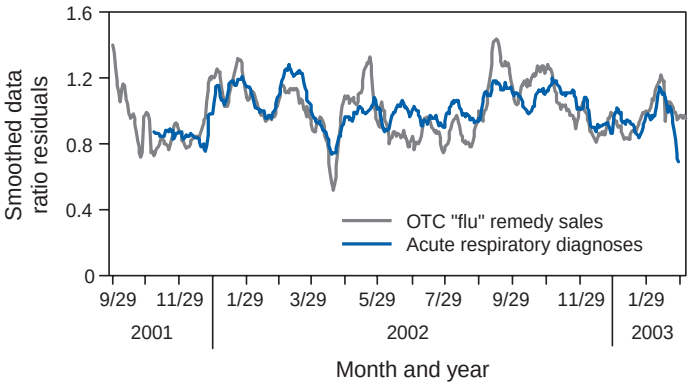
Method To Cluster Similar Sales Histories

A two-step OTC aggregation method was developed for preliminary use in the ESSENCE II surveillance system. The first step was to group individual products qualitatively into 41 adult groups, 16 pediatric groups, and four infant groups, each of which was formed by combining an indication (e.g., *allergy*, *cough*, or *fever*) with a physical type (e.g., *chest rub*, *inhaler*, or *lozenge*). Indications for the product were judged first by the product name. If the names alone left the indications ambiguous, then product descriptions were consulted.

This first step was required to obtain a high enough count of sales in each group so more quantitative methods could be applied. Although the first step was essentially qualitative, a conservative approach was taken by finely dividing the set of all OTC products into a substantial number of first-level product groups. This process was not expected to result in products with distinct uses being placed in the same group.

For the second stage of aggregation, observed sales histories (i.e., the number sold on each day during a certain period) of the different first-level groups were compared across a test period of approximately 17 months. If the ratio of sales of one product group to another was approximately constant over time, then the two product groups were assumed to be used to treat the same illnesses. Therefore, groups with approximately proportional sales histories were aggregated into supergroups for use in public health surveillance.

FIGURE 1. Comparison between residuals for physician billing claims for respiratory ailments and over-the-counter (OTC) sales of "flu" remedies, after correction for seasonal effects — urban Baltimore, Maryland–Washington, D.C., region



Note: Data were smoothed by a 7-day moving average to eliminate day-of-week effects.

The likelihood of observing the data under two different models was compared to measure the similarity of different groups' sales histories. Under model 1, the aggregated sales of product group N and M were assumed to be Poisson distributed with means that could vary from day to day. The natural log of the ratio of their means was assumed to be normally distributed with a standard deviation of 0.1. (This standard deviation was chosen to be small so the ratio between expected sales of products N and M could vary only slightly in the model.) The overall average log ratio and the daily (geometric) average of the means of product groups M and N were chosen by a maximum likelihood fit to the data. Under model 2, the sales of product groups N and M on each day were assumed to be independently Poisson-distributed, with means equal to the observed daily sales counts.

Because it was less constrained, the second model would always fit better. However, if the product groups were closely related, and if sales of product group N tended to rise and fall in proportion to sales of product group M, then model 1 would fit almost as well. The difference in data likelihood between the two models indicated the degree to which the two sales histories are not proportional. Therefore, a distance, D, was defined between product groups M and N by applying the following formula:

$$D = \log(\text{probability of observing the data under model 2}) - \log(\text{probability of observing the data under model 1})$$

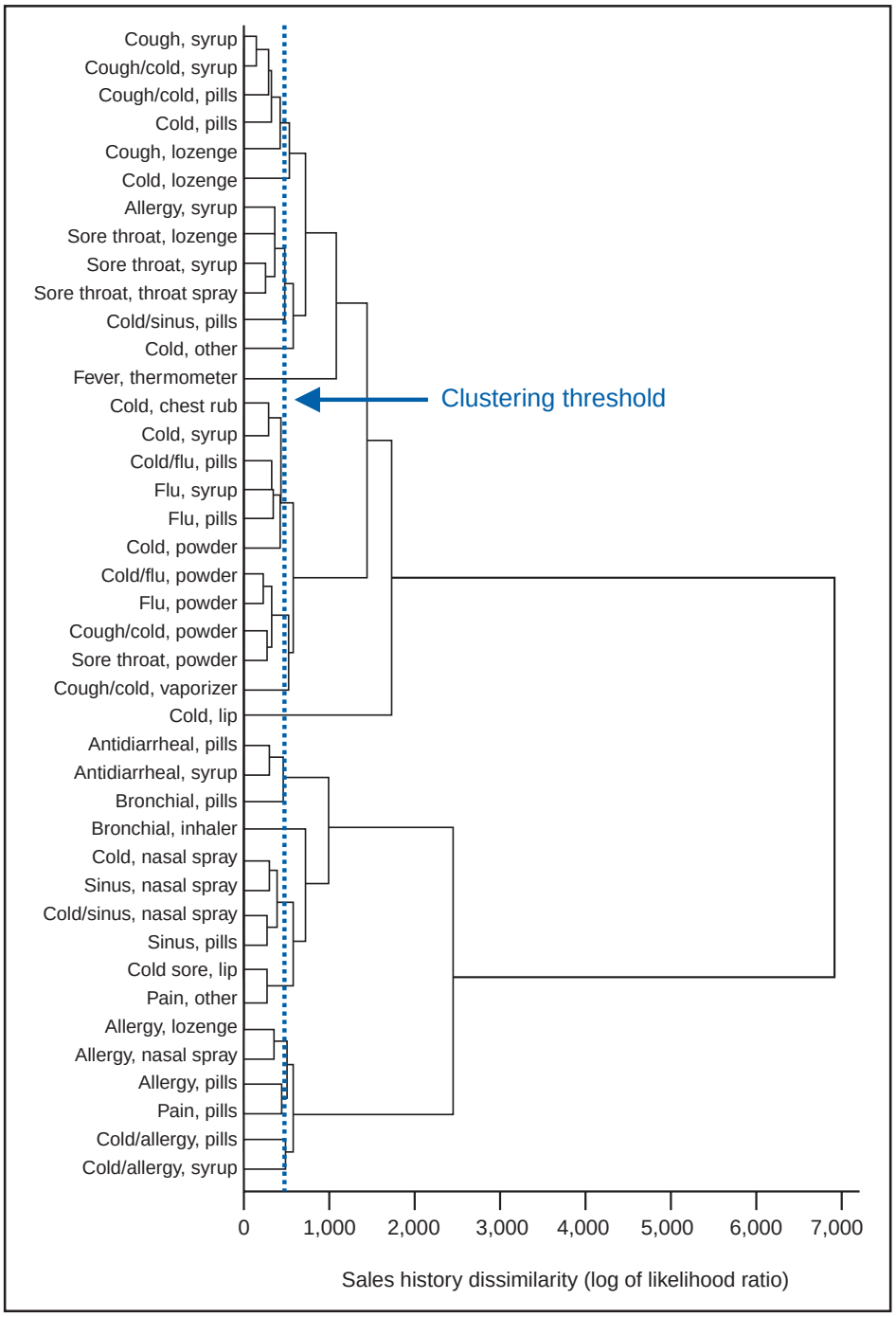
After this distance measure was obtained, standard hierarchical clustering techniques (*9*) were used to find clusters of product groups that were close together relative to the other product groups, as measured by the distance, D.

As this technique was refined, a complication was encountered that was apparently attributable to the effects of product promotions. When daily sales of cold remedies in powder form were compared with sales of cold/influenza remedies in pill form, products were found to have closely related sales histories. However, on three occasions (November 2001, September 2002, and October 2002), sales of cold powders substantially exceeded their normal level for periods of 6–7 days, whereas sales of cold/influenza pills did not. These events were assumed to be attributable to promotions and were excluded from the analysis.

An automated way to identify these 1-week aberrations was developed. First, a local background estimate was subtracted from raw OTC data, aggregated for each first-stage product group by using a trimmed-mean algorithm with a 20-day window centered on each day to create a normalized time series. Second, the normalized data were compared with a threshold, relative to a local estimate of the standard deviation. Finally, runs of threshold exceedences lasting 6–8 days were identified and excluded from the calculation of the distance, D.

The output of the clustering algorithm for the adult product groups is summarized in a dendrogram (Figure 2). By setting a threshold on this dendrogram at a specific distance value, distinct clusters of product groups (supergroups) were formed to be aggregated for health surveillance purposes. If the threshold were set too high, specificity would be lost because unrelated groups would be aggregated together. If it were set too low, statistical power would be lost because the resulting larger number of aggregated groups would have lower counts, and results would also be more susceptible to product-specific influences (e.g., promotions or introductions of new products). For ESSENCE applications, the threshold was set initially at a level to form 16 supergroups, some of which might not be selected for monitoring.

FIGURE 2. Results of clustering algorithm for group adult over-the-counter (OTC) medications for purposes of syndromic surveillance



Note: First-stage OTC product groups are listed along the y-axis. Vertical lines joining each group to a cluster at the x-axis represent the dissimilarity between that group and the most dissimilar element already included in the cluster. Clusters that are similarly joined at the x-axis represent the greatest dissimilarity between members of the two clusters joined. Product groups that are joined by vertical lines to the left of the clustering threshold are aggregated together for surveillance purposes. The indicated value of the clustering threshold is merely one option; the optimal setting for the threshold has not been determined by this analysis.

Results

OTC Lead Time

An analysis of the correlation-based measurements of OTC lead time identified high cross-correlations between the smoothed model fits for physician visits and OTC sales (Table 1). This finding reflects the fact that both model fits were 1-year–period sine waves that

TABLE 1. Peak correlations* and corresponding lead times of over-the-counter "flu" medications compared with outpatient visits for respiratory ailments for six regions in or near the National Capital Area (NCA)†

Region	Seasonal variation		Residuals	
	Correlation§	Lead time¶ (days)	Correlation**	Lead time†† (days)
Richmond	0.99	2	0.25	3
Eastern Shore	0.99	8	0.43	0
Western NCA	0.995	21	0.26	–3
Urban NCA	0.98	15	0.75	2
Southern NCA	0.95	12	0.47	–8
Northern NCA	0.97	16	0.66	–3

* Although the correlations provided here were computed from curves obtained for the period September 6, 2001–April 29, 2003, this table only includes correlations for November 2, 2001–July 1, 2002, to enable full comparison with those published earlier (*6*).
† Seasonal variations and residual, nonseasonal variations were considered separately, and snow days and holidays were ignored in both data sets.
§ Maximum cross-correlations of the fitted seasonal trend models.
¶ Time shifts that were observed to maximize the seasonal trend model correlations.
** Maximum cross-correlations of the residuals (data divided by the fitted seasonal trend model).
†† Time shifts that were observed to maximize the residual correlations.

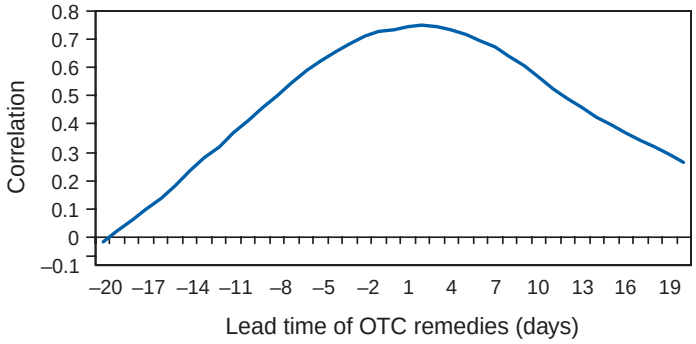
were shifted in time to maximize cross-correlation. In every case except Richmond, the sine-wave fit to the OTC data was shifted approximately 1–3 weeks earlier than the sine wave that was fit to the physician-encounter data. This indicates a repeatable 1–3 week lead in the seasonal cycle of OTC purchases, relative to the corresponding cycle in physician encounters.

Strong correlations between physician-visit and OTC residuals were observed, even though the seasonal trends were removed. The observed time-shifts in these residuals (as defined by maximum cross-correlation) were much shorter (in every case except that of Richmond) than those observed for the seasonal fits. The correlation in the best case (Urban National Capital Area) was also evident from a plot of the data (Figure 1), and the lead for this case was measurable, as indicated by the rapid decrease in correlation at other lags (Figure 3).

Clustering

A total of 16 supergroups were identified (Table 2). The sales histories represented by these groups ranged from strong winter seasonal peaks to approximately constant daily sales throughout the year to peaks in the spring and fall pollen seasons. Product groups with similar indications or similar physical forms tended to be placed in the same supergroups. This result was not guaranteed by the method but rather indicates that similar sales histories correlate with similar product use.

FIGURE 3. Cross-correlation versus time offset between physician respiratory billing claim residuals and over-the-counter (OTC) "flu"-remedy sales residuals, after correction for seasonal and day-of-week effects — urban Baltimore, Maryland–Washington, D.C., region, November 2, 2001–July 1, 2002



Note: A positive time offset indicates that OTC-sale fluctuations anticipate physician encounters.

Although this analysis took an empirical approach, certain supergroups (e.g., *cough*, *allergy*, *sore throat*, and *sinus remedies*) would have been formed anyway on the basis of intuition. However, the strength of this empirical approach is evident in the more surprising results. For example, pain pills were used heavily during the pollen season and therefore are grouped in the allergy cluster. Also, sales histories of powders sold to treat various maladies are more similar to each other than they are to other products advertised for the same mala-

TABLE 2. Empirical aggregated supergroupings of over-the-counter pharmaceutical products

Group	Supergroup members	Group	Supergroup members
1	Allergy, syrup	9	Cold, nasal spray
	Sore throat, lozenge		Cold/sinus, nasal spray
	Sore throat, syrup		Sinus, nasal spray
	Sore throat, throat spray		Sinus, pills
2	Cold/sinus, pills	10	Cold sore, lip
3	Cold/influenza, powder		Pain, other
	Cough/cold, powder	11	Cold, chest rub
	Influenza, powder		Cold, powder
	Sore throat, powder		Cold, syrup
4	Cough/cold, vaporizer		Cold/influenza, pills
5	Cold, lozenge		Influenza, pills
6	Cold, pills		Influenza, syrup
	Cough, lozenge	12	Cold, other
	Cough, syrup	13	Bronchial, inhaler
	Cough/cold, pills	14	Antidiarrheal pills
	Cough/cold, syrup		Antidiarrheal syrup
7	Allergy, lozenge		Bronchial, pills
	Allergy, nasal spray	15	Fever, thermometer
	Allergy, pills	16	Cold, lip
	Pain, pills		
8	Cold/allergy, pills		
	Cold/allergy, syrup		

dies. Monitored allergy syrups do not appear to belong with other allergy medications because sales peak during the winter cold season rather than during the pollen season. (A probable explanation, obtained after the analysis was completed, was that most allergy syrups included in the data were targeted for diabetics.) Finally, products advertised to treat chest congestion had little indication of a seasonal trend and therefore did not cluster with products advertised to treat other respiratory conditions.

Conclusions

OTC Lead Time

Persistent correlations between OTC influenza remedy sales and physician acute-respiratory encounters were determined, even after removal of the annual sinusoidal variation from both. This makes a more convincing case for the use of OTC products to monitor sudden changes in public health than do results strongly influenced by annual variations. However, these data do not indicate a repeatable positive lead time of OTC products relative to physician encounters on shorter, subannual time scales. Earlier results about OTC timeliness based on annual cycles could be misleading.

The findings outlined in this report are subject to at least two limitations. First, the lower correlations observed in certain regions might be biased by inexact spatial correspondence between physician encounter and OTC data sets; a more comprehensive data set might provide a basis for more precise measurements of correlations and lead times. Second, only the relation between influenza remedies and acute respiratory diagnoses was considered, and other OTC–physician connections might yield different results.

If other researchers are able to verify the result of no significant lead time of OTC data relative to physician encounters at subannual time scales, this would not necessarily imply that OTC data are not useful for public health surveillance. None of this analysis includes the lag in reporting the data. OTC sales data might be electronically available with a shorter reporting lag after the sales event compared with the lag to receive physician outpatient data. The number of patients seeking OTC medications early during a given outbreak might also be larger than the number seeking care from a physician. All else being equal, OTC sales data are potentially a more sensitive measure of community illness.

OTC Product Aggregation

A quantitative method was presented that can be used to enhance and validate a more qualitative approach by automatically sorting through a heterogeneous set of OTC product groups to find relatively homogenous supergroups of products. Both the method and the specific supergroups identified might be helpful to others attempting to use OTC data for surveillance of community health. This method demonstrated its value for the ESSENCE surveillance system by finding certain unexpected relationships between product groups. Appropriate aggregation of product supergroups might vary regionally or demographically. The method discussed in this report might be a good approach for identifying custom OTC aggregations for specific applications.

Acknowledgments

This work was sponsored by the Defense Advanced Research Projects Agency (DARPA) as part of the Bio-ALIRT program. The material contained in this report was cleared by DARPA and approved for public release, distribution unlimited. Ron Bloom of General Dynamics provided useful insights concerning the interpretation of lead time measurements.

References

1. Welliver RC, Cherry JD, Boyer KM, et al. Sales of nonprescription cold remedies: a unique method of influenza surveillance. Pediatr Res 1979;13:1015–7.
2. Angulo FJ, Tippen S, Sharp DJ, et al. A community waterborne outbreak of salmonellosis and the effectiveness of a boil water order. Am J Public Health 1997;87:580–4
3. 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.
4. Stirling R, Aramini J, Ellis A, et al. Waterborne cryptosporidiosis outbreak, North Battleford, Saskatchewan, Spring 2001. Can Commun Dis Rep 2001;27:185–92.
5. Hogan WR, Tsui F-C, Ivanov O, et al. Early detection of pediatric respiratory and diarrheal outbreaks from retail sales of electrolyte products. J Am Med Inform Assoc 2003;10:555–62.
6. Magruder S. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of public health. Johns Hopkins University Applied Physics Laboratory Technical Digest 2003;24:349–53.
7. Lombardo J, Burkom H, Elbert E, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health 2003;80(2 Suppl 1):i32–42.
8. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963;78:494–506.
9. The MathWorks, Inc. Statistics toolbox. Natick, MA: The MathWorks, Inc. Available at http://www.mathworks.com/access/helpdesk/help/toolbox/stats/stats.shtml.

Evaluation

Evaluation Challenges for Syndromic Surveillance — Making Incremental Progress

Daniel M. Sosin, J. DeThomasis
Epidemiology Program Office, CDC

Corresponding author: Daniel M. Sosin, Associate Director for Science, Office of Terrorism Preparedness and Emergency Response, CDC, 1600 Clifton Road, Mailstop D-44, Atlanta, GA 30333. Telephone: 404-639-1528; Fax: 404-639-7977; E-mail: dms8@cdc.gov.

Abstract

Introduction: *The 2003 National Syndromic Surveillance Conference provided an opportunity to examine challenges and progress in evaluating syndromic surveillance systems.*

Objectives: *Using the conference abstracts as a focus, this paper describes the status of performance measurement of syndromic surveillance systems and ongoing challenges in system evaluation.*

Methods: *Ninety-nine original abstracts were reviewed and classified descriptively and according to their presentation of evaluation attributes.*

Results: *System evaluation was the primary focus of 35% of the abstracts submitted. Of those abstracts, 63% referenced prospective evaluation methods and 57% reported on outbreak detection. However, no data were provided in 34% of the evaluation abstracts, and only 37% referred to system signals, 20% to investigation of system signals, and 20% to timeliness.*

Conclusions: *Although this abstract review is not representative of all current syndromic surveillance efforts, it highlights recent attention to evaluation and the need for a basic set of system performance measures. It also proposes questions to be answered of all public health systems used for outbreak detection.*

Introduction

Interest in syndromic surveillance remains high in the United States, with approximately 100 state and local health jurisdictions conducting a form of syndromic surveillance in 2003 (*1*). However, skepticism about the efficacy of syndromic surveillance for early detection of terrorism-related illness has increased (*1–4*).

At the 2002 National Syndromic Surveillance Conference, an evaluation framework (*5*) was presented that closely followed CDC's Updated Guidelines for Evaluation of Public Health Surveillance Systems (*6*). That evaluation framework described the system attributes that should be measured but provided limited guidance on how to measure those attributes consistently.

In 2003, CDC convened a national working group on outbreak-detection surveillance.* The working group clarified terminology and revised earlier frameworks to emphasize early outbreak detection, putting syndromic surveillance into context as a specialized surveillance tool. The resulting Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks (*7*) provides a structure for evaluating syndromic surveillance systems and reporting the results. The revised framework offers a task list for describing a surveillance system (Box 1) and provides visual aids to improve standard collection and reporting of evaluation information. The framework also provides a timeline with milestones in outbreak development and detection, from exposure to a pathogen to the initiation of a public health intervention. Although this timeline does not specify a single, reproducible measure to reflect the timeliness of detection, it does provide more consistent specification of intervals for comparing performance among different systems and different settings. The framework also describes two approaches, encompassing sensitivity, predictive value negative, and predictive value positive, to evaluate system validity for outbreak detection: 1) the systematic description and accumulation of experiences with outbreak detection, and 2) simulation-based methods.

The importance of evaluating syndromic surveillance systems is widely recognized (*1,3–5,8–11*), but a common set of measures have not yet been defined that will establish the added value of syndromic surveillance compared with current surveillance tools. Nonetheless, progress has been made toward uniform guidance on evaluating syndromic surveillance systems (*7*). This paper summarizes progress during 2003 and describes steps for the future.

* **Working group members:** Daniel M. Sosin, M.D., Claire Broome, M.D., Richard Hopkins, M.D., Henry Rolka, M.S., Van Tong, M.P.H., James W. Buehler, M.D., Louise Gresham, Ph.D., Ken Kleinman, Sc.D., Farzad Mostashari, M.D., J. Marc Overhage, M.D., Julie Pavlin, M.D., Robert Rolfs, M.D., David Siegrist, M.S.

BOX 1. Tasks for evaluating public health surveillance systems for early detection of outbreaks*

Task A. Describe the system

1. Purpose: What is the system designed to accomplish?
2. Stakeholders: Whom does the system serve?
3. Operation: How does the system work?
   a. Systemwide processes
   b. Data sources
   c. Data preprocessing
   d. Statistical analysis
   e. Epidemiologic analysis, interpretation, and investigation

Task B. Provide data demonstrating outbreak detection attributes

1. Timeliness: How early in the outbreak is the event detected?
2. Validity: How well does the system perform in distinguishing outbreak detection of public health significance from less important events or random variations in disease trends?
   a. Sensitivity and predictive value: What percentage of true outbreaks are detected by the system? What percentage of signals by the system are relevant (true positives)? What percentage of negative results are truly negative?
   b. Data quality: How does data quality affect validity of outbreak detection?
      i. Representativeness: How well does the system reflect the population of interest?
      ii. Completeness: What percentage of data are present for each record?

Task C. Describe the system experience

1. System usefulness: In what ways has the system demonstrated value relevant to public health?
2. Flexibility: How adaptable is the system to changing needs and risk thresholds?
3. System acceptability: Have stakeholders been willing to contribute to and use the system?
4. Portability: How readily can the system be duplicated at another location?
5. System stability: How consistent has the system been in providing access to reproducible results?
6. System costs: What are the resource requirements to deploy and maintain the system?

Task D. Summarize conclusions and make recommendations for use and improvement of systems for early outbreak detection

* **Source:** CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5).

Methods

The authors reviewed the original 99 abstracts submitted to the 2003 National Syndromic Surveillance Conference and divided them into two categories: 1) surveillance systems and 2) analytic methods. Abstracts about surveillance systems were subcategorized into 1) system descriptions, 2) implementations, and 3) evaluations. Analytic methods abstracts included those addressing detection algorithms, data modeling, and case definitions. For each abstract, the reviewers identified the geographic location of the surveillance system or primary author and the responsible entity for the system or study being described (e.g., local health department or university). Information was also gathered about the data-collection method used, the purpose of the system, and the type of data used. An abstract was classified as pertaining to system evaluation if the author indicated intent to present a system evaluation or if the abstract provided results of the system's experience in detecting outbreaks. Evaluation variables abstracted were frequency of system signals, investigations, outbreaks detected and missed, estimation of timeliness, and the effect of early detection.

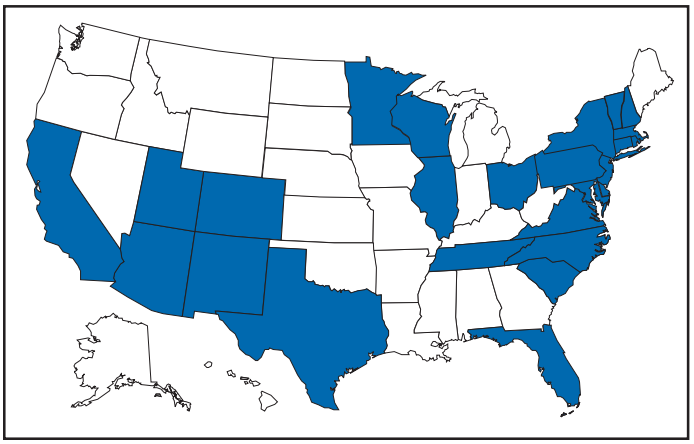
Each abstract was reviewed by both authors of this paper and results were reconciled in a meeting. Abstract forms were entered into Epi Info 2002 (http://www.cdc.gov/epiinfo/) for analysis.

Results

The 99 abstracts were submitted by authors from 23 states, the District of Columbia, and seven countries outside the United States (Figure). The bulk of the syndromic surveillance work, as reflected in these abstracts, is occurring in state and local health departments and within U.S. academic institutions. Abstract authors were based in state and local health departments (40%), universities (32%), federal government agencies (13%), health-care organizations (11%), and businesses (4%). Abstracts focused on system evaluation (35%), description of systems or their implementation (26%), data management, modeling, and detection algorithms (28%), and case definition (11%).

Of the 60 abstracts that described a full syndromic surveillance system, 30% indicated use of manual data collection

FIGURE. Location of U.S.-based syndromic surveillance systems described in 99 abstracts submitted for the 2003 National Syndromic Surveillance Conference



outside the typical workflow of the data provider. Ninety-five percent described systems designed to detect outbreak patterns in the data, with only 5% using syndromic surveillance for individual case detection (e.g., severe acute respiratory syndrome or West Nile encephalitis). Of the 35 abstracts that described system evaluation, 34% provided no data in the abstract, only describing the intent to present evaluation data. Nonetheless, 63% addressed the outbreak-detection experience in a prospective direction. Of the 35 abstracts that described system evaluation, 37% reported on the signaling of a system; 20% referred to one or more investigations; 57% addressed one or more outbreaks detected or missed; and 20% addressed timeliness in any fashion. None of the abstracts estimated the public health effect of early detection.

Discussion

The systems described in these conference abstracts are not a representative sample of jurisdictions conducting syndromic surveillance; rather, they are a synopsis from those jurisdictions willing to share their experiences at a national conference. Furthermore, certain presentations were invited talks for which abstracts were not submitted.

The diversity of data sources being used reflects the early stage of development of syndromic surveillance and the exploration of novel data sources (Table). The predominant focus, consistent with recommendations from the 2002 National Syndromic Surveillance Conference (*9*), is on data from emergency departments and other clinical sources. A substantial number of systems (30%) continue to rely on manual data collection at the data source. The sustainability of such a system has been questioned (*3,8–10,12*). Whether for routine data collection or for innovative surveillance

TABLE. Data sources for 60 abstracts on syndromic surveillance submitted to the 2003 National Syndromic Surveillance Conference

Data source	No. of abstracts*	%*
Emergency departments	29	48
Office or clinic visits	13	22
Hospital diagnoses	7	12
School absences	7	12
911 calls/EMS runs	6	10
Over-the-counter drug purchases	5	8
Poison control centers	5	8
Nurse advice lines	4	7
Veterinary clinics	3	5
Medical examiners	2	3
Pharmacy prescriptions	2	3
Laboratory results	2	3
Laboratory orders	1	2
Medical center parking-lot volume	1	2
Online obituaries	1	2
Subway-worker absences	1	2
School perception of an outbreak	1	2

* Certain systems used multiple data sources.

systems, automated data captured during the usual course of care (or business) is preferred to manual data collection when continuous, complete reporting is the goal. Manual data collection will continue to play a role in actual or threatened outbreak settings that have special data needs that cannot be filled by using existing electronic data (*3,7,9,10,12*).

A substantial number of abstracts (35%) focused on the evaluation of a system, although the rigor and methods of evaluation varied considerably. One third of abstracts that stated intent to present a system evaluation provided no data at all in the abstract regarding how effectively the system was working. However, approximately two thirds of the evaluation abstracts referred to tracking performance prospectively rather than simply analyzing historical data to identify known events. Not only is prospective identification of an outbreak a more substantial indicator of success, but it also offers benefits beyond identifying specific events (e.g., stronger relationships between clinicians and public health practitioners and higher quality surveillance data) (*4,13–15*).

To better understand the performance of outbreak-detection systems, basic measures of performance need to be counted. How often a system signals (i.e., how often it indicates that something worthy of further investigation is occurring) also needs to be reported. This applies to all the ways that health departments detect outbreaks (e.g., phone calls from the public), not just to syndromic surveillance. Every surveillance system should be able to report how many times in a given period (e.g., 1 month) it has triggered a follow-up investigation, yet only 37% of the evaluation abstracts gave any indication of system signals, much less a rate of signaling.

More information is needed about different responses to signals and the results of those responses. When a system signals, multiple responses can be made, from deciding not to act on the signal to launching a full investigation with staff participation and new data collection. Intermediate steps might include reviewing the data for errors, reviewing records manually within syndrome categories to search for patterns, conducting manual epidemiologic analysis for subgroup associations with the signal, examining data from other sources, and ensuring early submission of the next cycle of reports from affected locations. Although certain systems are potentially not signaling and therefore not instigating investigations, that only 20% of the systems presented in the evaluation abstracts have initiated investigations seems unlikely. Routine reporting of how often signals elicit a response and what those responses entail is essential.

Jurisdictions should report routinely both on outbreaks detected through syndromic surveillance and outbreaks missed. Practitioners should also report outbreaks detected through other methods to understand the relative value of syndromic surveillance. Of the 2003 evaluation abstracts, >50% addressed the detection or nondetection of outbreaks, but room for improvement remains.

Lastly, early detection is essential in syndromic surveillance, yet only 20% of the evaluation abstracts addressed timeliness. Measuring timeliness should be a routine part of reporting. The evaluation timeline in the Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks (*7*) provides milestones that should aid in the reporting of timeliness.

Conclusion and Next Steps

Evaluation requirements should be simplified and standardized to allow comparisons across systems and across outbreak-detection approaches. Simulations offer promise for testing and improving systems designed to detect rare events. The abstracts submitted to the 2003 conference reflect initial efforts to evaluate analytic methods in isolation with simulation exercises. Testing intact systems is needed to verify how well they might perform in practice at providing early warning of public health emergencies. Additional research is needed to validate the assumptions necessary for modeling disease outbreaks (e.g., the spread of disease in various scenarios, or the individual and community behavior patterns after onset of illness that might serve as early outbreak indicators).

Although detailed descriptions of systems would be a helpful step forward, the reporting burden could be heavy and additional experience is needed to determine the required system attributes and to standardize the descriptions. An interim approach might be to prioritize a limited number of measures of likely value now until experience is gained with other measures. A simplified version of the Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks (*7*) might focus on questions regarding timeliness, validity, and usefulness of an outbreak-detection system (Box 2). Such a framework could help standardize reporting of the different methods used by public health departments to detect outbreaks. Ultimately, the goal is to measure the effect of detection methods — how public health is improved by detection, and at what cost. The proposed framework could move the field forward incrementally by using readily available information and measures until additional information on metrics for outcomes and costs becomes available.

BOX 2. Priority evaluation questions for early outbreak-detection systems

1. How often does the system signal an event for further epidemiologic attention?
   a. What was the time period (e.g., 1 month)?
   b. What was the statistical threshold (e.g., p-value)?
   c. If the threshold has changed, explain why.
2. How were signals responded to?
   a. What percentage of signals were investigated through new data collection?
   b. What percentage caused increased reporting frequency from affected sites?
   c. What percentage conducted detailed manual analysis of any data available to the jurisdiction?
   d. What percentage conducted manual analysis of data from the system?
   e. What percentage were reviewed for data errors?
   f. What percentage of signals were ignored?
   g. What resources were directed to follow-up?
3. How many outbreaks were detected through the system?
   a. How timely was detection relative to other systems?
   b. How timely was detection relative to the stage of the outbreak?
   c. What were the agent, host population, and environmental conditions of the outbreak?
4. How many outbreaks were missed by the system?
   a. What were the agent, host, and environmental conditions?
   b. How was the outbreak detected?
5. What was the public health response to detection (e.g., no response, urgent communication to clinicians, or vaccination campaign)?

References

1. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis 2003. Available at http://www.cdc.gov/ncidod/EID/vol9no10/03-0231.htm.
2. Becker J. High-tech health watch draws cash, questions. Washington Post, November 23, 2003:A17.
3. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecur Bioterr 2003;1:1–5.
4. Sosin DM. Syndromic surveillance: the case for skillful investment. Biosecur Bioterr 2003;1:1–7.
5. Sosin DM. Draft framework for evaluating syndromic surveillance systems. J Urban Health 2003;80(2 Suppl 1):i8–13.
6. CDC. Updated guidelines for evaluating public health surveillance systems: recommendations from the guidelines working group. MMWR 2001;50(No. RR-13).
7. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5).
8. Henning KJ. Appendix B: syndromic surveillance. In: Smolinski MS, Hamburg MA, Lederberg J, eds. Microbial threats to health: emergence, detection, and response. Washington, DC: National Academies Press, 2003:281–312. Available at http://books.nap.edu/books/030908864X/html/281.html#pagetop.
9. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. J Urban Health 2003;80(2 Suppl 1):i1–7.
10. Pavlin JA, Mostashari F, Kortepeter MG, et al. Innovative surveillance methods for rapid detection of disease outbreaks and bioterrorism: results of an interagency workshop on health indicator surveillance. Am J Public Health 2003;93:1230–5.
11. Smolinski MS, Hamburg MA, Lederberg J, eds. Microbial threats to health: emergence, detection, and response. Washington, DC: National Academies Press, 2003. Available at http://books.nap.edu/books/030908864X/html/index.html.
12. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. J Urban Health 2003;80(2 Suppl 1):i76–88.
13. Brown K, Pavlin J, Mansfield J, Elbert E, Foster V, Kelley P. Identification and investigation of disease outbreaks by ESSENCE [Abstract]. J Urban Health 2003;80(2 Suppl 1):i119.
14. Johnson J, McClean C, Poggemeyer K, Ginsberg M. Application of bioterrorism surveillance methods in San Diego County [Abstract]. J Urban Health 2003;80(2 Suppl 1):i137.
15. Schumacher M, Nohre L, Santana S. Partial evaluation of a drop-in bioterrorism surveillance system in Phoenix, Arizona [Abstract]. J Urban Health 2003;80(2 Suppl 1):i118.

Measuring Outbreak-Detection Performance By Using Controlled Feature Set Simulations

Kenneth D. Mandl,[1,2] B. Reis,[1] C. Cassa[1,3]
[1]*Children's Hospital Boston, Boston, Massachusetts;* [2]*Harvard Medical School, Boston, Massachusetts;*
[3]*Massachusetts Institute of Technology, Cambridge, Massachusetts*

Corresponding author: Kenneth D. Mandl, Division of Emergency Medicine, Children's Hospital Boston, 300 Longwood Avenue, Boston, MA 02115. Telephone: 617-355-4145; E-mail: Kenneth.Mandl@childrens.harvard.edu.

Abstract

Introduction: *The outbreak-detection performance of a syndromic surveillance system can be measured in terms of its ability to detect signal (i.e., disease outbreak) against background noise (i.e., normally varying baseline disease in the region). Such benchmarking requires training and the use of validation data sets. Because only a limited number of persons have been infected with agents of biologic terrorism, data are generally unavailable, and simulation is necessary. An approach for evaluation of outbreak-detection algorithms was developed that uses semisynthetic data sets to provide real background (which effectively becomes the noise in the signal-to-noise problem) with artificially injected signal. The injected signal is defined by a controlled feature set of variable parameters, including size, shape, and duration.*

Objectives: *This report defines a flexible approach to evaluating public health surveillance systems for early detection of outbreaks and provides examples of its use.*

Methods: *The stages of outbreak detection are described, followed by the procedure for creating data sets for benchmarking performance. Approaches to setting parameters for simulated outbreaks by using controlled feature sets are detailed, and metrics for detection performance are proposed. Finally, a series of experiments using semisynthetic data sets with artificially introduced outbreaks defined with controlled feature sets is reviewed.*

Results: *These experiments indicate the flexibility of controlled feature set simulation for evaluating outbreak-detection sensitivity and specificity, optimizing attributes of detection algorithms (e.g., temporal windows), choosing approaches to syndrome groupings, and determining best strategies for integrating data from multiple sources.*

Conclusions: *The use of semisynthetic data sets containing authentic baseline and simulated outbreaks defined by a controlled feature set provides a valuable means for benchmarking the detection performance of syndromic surveillance systems.*

Introduction

Evaluation of surveillance systems for early detection of outbreaks is particularly challenging when the systems are designed to detect events for which minimal or no historic examples exist (*1*). Although infection by biologic agents is rare, exceptions have occurred. For example, in 1979, persons living in Sverdlovsk in the former Soviet Union were exposed to *Bacillus anthracis* during an unintentional release from a weapons plant (*2*), and a limited number of persons were exposed in Florida, New York, and the District of Columbia during 2001 when *B. anthracis* spores were released through the mail (*3*). However, absent sufficient real outbreak data, measuring a system's detection performance requires simulation. Simulated outbreaks must reflect the diversity of threats, both natural and man-made, that a surveillance system might reasonably be expected to encounter and detect. This paper describes a flexible approach to generating standardized simulated data sets for benchmarking surveillance systems and provides examples of its application. Rather than model all possible conditions and factors, the approach relies on simulated outbreaks characterized by a controlled feature set that systematically defines the magnitude, temporal progression, duration, and spatial characteristics of the simulated outbreaks on the basis of variable parameters.

Stages of Outbreak Detection

The goal of outbreak detection is to generate an alert whenever observed data depart sufficiently from an expected baseline (*4*). In other words, the system must be able to detect a signal (i.e., disease outbreak) against background noise (i.e., normally varying baseline disease in the region). Four basic methodologic stages are used to process data for outbreak detection: 1) the syndrome grouping stage, in which data acquired from different sources are used to assign each patient to a particular syndrome group (e.g., respiratory

infection or gastrointestinal infection); 2) the modeling stage, in which historic data, including data for patients during the past year(s), are analyzed to establish a model from observed temporal and spatial patient distributions; 3) the detection stage, in which the expected values (i.e., predicted daily frequencies of patients in each syndrome group) are compared with observed values to determine whether abnormal activity is occurring; and 4) the alert stage, in which thresholds are set to evaluate whether an unusual pattern warrants notifying public health authorities.

The first two stages can be accomplished by using historic data from a given region. Depending on the data source, different methods can be used to assign a case to syndrome group. For example, emergency department (ED) data can be categorized by chief complaint by using a naïve Bayesian classifier (*5*) or by a standardized grouping of *International Classification of Disease, Ninth Revision* (ICD-9) codes (*6*). Outbreaks are identified by comparing observations with the predictions generated by a model describing the expected baseline temporal or spatial pattern. Examples include time-series models (*7*), spatial scan statistics (*8,9*), and models of interpoint distance distributions (*10*).

At the detection stage, observed values must be compared with expected values; a signal containing outbreaks (hereafter referred to as an outbreak signal) is required to evaluate a system's detection performance. However, limited data are available concerning terrorism-related events, and none are available in the format used by existing syndromic surveillance systems.

Data Sets for Benchmarking Performance

Performance of outbreak-detection models can be measured by using authentic data, synthetic data, or combinations of the two (Table). Two kinds of purely authentic data sets are possible. One is genuine syndromic data contemporaneous with either a known large-scale local outbreak (e.g., a winter influenza surge) (*11*) or a more circumscribed event (e.g., a diarrheal outbreak) (*12*). The data set would contain the background of ordinary disease or symptom occurrence and the signal of the actual outbreak. A second type of authentic data set is a hybrid containing background from a regional surveillance system spiked with cases from a known outbreak. This approach was taken when over-the-counter medication-sales data were spiked with an outbreak based on the Sverdlosk incident (*13*). Alternatively, a hypothetical baseline can be constructed, and actual or simulated signals can be imposed and injected. Although this approach is valid, limited need exists to simulate background activity, given the abundance of readily available real-signal streams from surveillance systems.

The approach described in this paper superimposes a simulated signal onto an authentic baseline, permitting exploration of the effects of controlled variations of signal characteristics. Two main approaches can be taken to creating this simulated signal: 1) using multistage, multivariate mathematical models to produce the signal or 2) defining a series of parameters that enable generation of a controlled feature set simulated signal. For example, a complex mathematical model (*14*) might be based on a scenario in which a particular form of aerosolized *B. anthracis* is dispersed under a certain set of atmospheric conditions over a specific geographic region with a well-characterized population demographic. The number of susceptible persons might be estimated and their subsequent behaviors modeled. The resulting effect on the syndromic surveillance data set (e.g., retail sales, primary care visits, or ED visits) could be projected. However, this approach for evaluating outbreak-detection performance is labor-intensive, and the models are based on multiple assumptions. A more flexible approach is to use a set of variable parameters describing a particular outbreak. Defining feature sets of outbreaks (e.g., magnitude, shape, and duration) allows rapid determination of the limits of a system's ability to detect an outbreak under varying conditions.

Using Parameters To Specify Outbreak Characteristics

Background noise can be spiked with additional cases configured as spatial or temporal clusters, describable as a controlled feature set. Different adjustable parameters enable ready manipulation of the simulated outbreaks. Optimally, a training data set should be modeled, and the artificial outbreak signal should be injected into a validation data set. However, if suffi-

TABLE. Combinations of synthetic and authentic data to create semisynthetic data sets

		Signal	
		Authentic	**Synthetic**
Noise	**Authentic**	Sample containing outbreak, or signal and noise from two data sets	Authentic background spiked with simulated signal
	Synthetic	Signal from a naturally occurring outbreak superimposed on simulated noise	Simulated noise and signal

cient data are not available to do so, the artificial outbreak signal can be injected into the same data used for training.

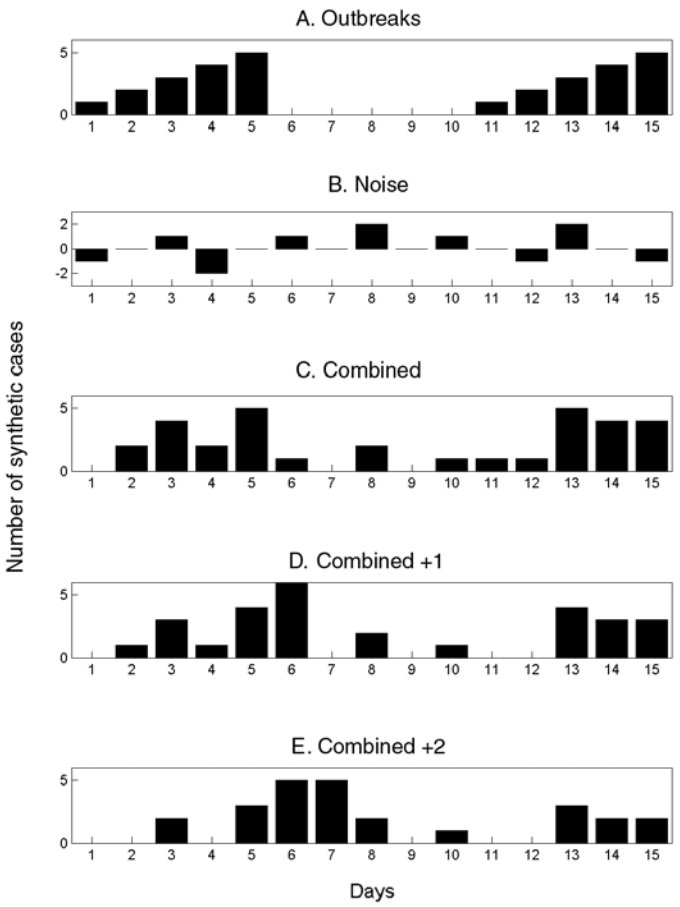
Outbreak Duration

A key parameter is the duration of the added outbreak signal. Executing simulations over a range of outbreak durations is useful, and various factors might influence the range chosen. Different agents can cause outbreaks of different lengths; for example, a surge in influenza activity lasts weeks or months, whereas a foodborne outbreak might last only 4–5 days. Furthermore, the temporal window used by the detection system might have a substantial effect on how outbreaks of different magnitudes are detected. If the detection window were based, for example, on a sliding moving average of 7 days, 2- or 3-day–long outbreaks would be smoothed out; under certain conditions, this smoothing might dilute the signal. Conversely, outbreaks gently trending upward in numbers might not be detected with a shorter sliding window.

Outbreak Spacing

One efficient way to measure outbreak-detection performance and the factors that influence it is to spike a data stream with a substantial number of individual outbreaks. The more outbreaks presented to a model-based system, the more accurately the system's detection performance can be characterized. To maximize the number of simulated outbreaks in the data set, one can introduce multiple nonoverlapping outbreaks in a single data set (e.g., a 5-day outbreak beginning on day 1, a second beginning on day 11, and a third on day 21). The outbreaks are then removed and replaced by a different set of nonoverlapping outbreaks and again presented to the system (e.g., days 2, 12, and 22). For measurement purposes, all individual outbreaks must be isolated temporally to ensure any response to the previous outbreak has been eliminated from the system before the next outbreak is encountered. For systems that analyze data by using a temporal window of >1 day, the spacing between outbreaks must be greater than that width to ensure independence. Although such temporal isolation is critical for accurate measurement of detection performance, it will not directly address the system's ability to detect overlapping outbreaks. Shifting the outbreaks in time ensures that outbreaks are affected by different regions of noise (Figure 1). Spacing outbreaks throughout the year also permits measuring the effect of seasonal changes in the background on outbreak detection. Understanding the effects of different regions of background noise cannot be accomplished without simulation.

FIGURE 1. Distorting effect of background noise on simulated outbreaks

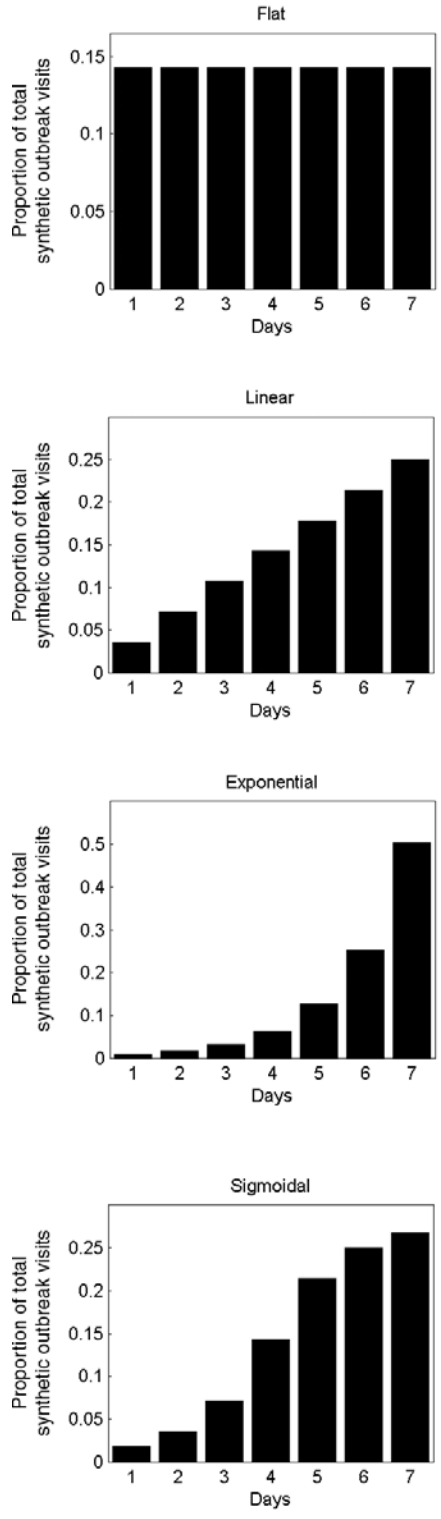


Note: Plot A depicts two simulated outbreaks spaced apart. Plot B depicts the background noise signal. Plot C illustrates the effect of noise distorting the outbreaks. Plot D demonstrates how the noise distorts the outbreaks differently when the outbreaks are shifted to the right by 1 day. Plot E demonstrates how the noise distorts the outbreaks when the outbreaks are shifted to the right by 2 days.

Outbreak Temporal Progression

The time course of an outbreak spreading through a population can follow multiple paths, effectively producing a signature shape related to the epidemic curve. For example, a highly infectious disease (e.g., smallpox) could spread exponentially over time, whereas a point-source exposure that is not contagious from person to person (e.g., a release of *B. anthracis*) would be unlikely to grow exponentially. Multiple canonical shapes of temporal progression (Figure 2) can be used in simulations to characterize the detection performance of surveillance systems. In a system monitoring daily ED visits, for example, flat outbreaks have a fixed number of extra visits/day for the duration of the outbreak (e.g., 10, 10, 10, 10, and 10 extra visits for a 5-day outbreak). Linear outbreaks have a linearly increasing number of extra visits/day

FIGURE 2. Four canonical shapes of temporal progression for simulated outbreaks



over the course of the outbreak (e.g., five, 10, 15, 20, and 25 extra visits for a 5-day outbreak). Exponential outbreaks have an exponentially increasing number of extra visits/day over the course of the outbreak (e.g., two, four, eight, 16, and 32 extra visits for a 5-day outbreak). Sigmoid-shaped outbreaks mirror epidemiologic phenomena in which the number of affected individuals increases exponentially at first, then slows down until it plateaus at a new fixed level (e.g., two, four, eight, 12, and 14 extra visits for a 5-day outbreak). Alternatively, a model of more complex shape described by a multinomial (e.g., the Sverdlosk [*2*] outbreak) might be desirable.

Outbreak Magnitude

Because the minimum detectable size of an outbreak is often of interest, outbreak-detection performance should be tested over a range of signal magnitudes; detection performance might vary substantially depending on these magnitudes. This variability is attributable primarily to the changes in signal-to-noise ratio that result from different outbreak sizes. For limited outbreaks that are at or near the "noise floor" of the model (i.e., the usual level of random variability in the model's predictions), the detection performance is typically poor because distinguishing outbreaks from the random noise of the model is difficult. As the relative size of an outbreak increases, identifying an outbreak in the presence of noise becomes easier. Once the outbreak magnitude is such that the noise does not effectively mask it, the outbreak-detection performance of the system typically plateaus at perfect or near perfect detection.

For identification of an appropriate range of outbreak magnitudes for simulations, the error or noise profile of the model should be characterized. The daily forecast errors of the model, defined as the forecast value minus the actual value for each day, must be calculated. The error profile can be visualized by plotting a histogram of these daily forecast errors and standard deviation of the error distribution. Outbreak magnitudes should range from near zero to at least twice the standard deviation of the forecast error. For example, in the case of a model of ED visits with mean of 140 visits/day and an error profile with a standard deviation of 20 visits, simulations of outbreaks ranging in magnitude from 0 to 40 visits/day should be run. This range can be sampled in intervals of five, yielding the following set of outbreak magnitudes: 0, 5, 10, 15, 20, 25, 30, 35, and 40.

The error profile of a model might vary during a year because of seasonal differences in signal variability. For example, respiratory-visit rates could vary more unpredictably in winter than in summer. In such cases, constructing separate error profiles for different seasons might be useful to tailor the detection test to each season.

Spatial Features

The outbreak cluster might describe the spatial relationship among the additional cases, which are represented as geocodes (i.e., latitude and longitude), possibly augmented by a time stamp. If so, the cluster can be described in terms of a maximum cluster radius, the distribution of cases within that radius, and the angle from a fixed point (e.g., a hospital). Simulating spatial clusters raises additional challenges, including the identification of realistic locations for simulated cases, based on the spatial features of a region (e.g., housing and of bodies of water).

Metrics for Detection Performance

Sensitivity and Specificity

A tradeoff always exists between sensitivity and specificity, and the ability to detect outbreaks must be balanced against the cost of false alerts (*1*). For evaluation purposes, holding sensitivity or specificity constant can be useful when plotting the other against another variable (e.g., outbreak magnitude or duration). For example, specificity might be held constant while plotting sensitivity versus outbreak magnitude. For each outbreak magnitude, the alert threshold should be tuned until the desired number of false alerts (and thus the desired specificity) is achieved. At this point, the resulting sensitivity under these conditions is measured. This process is repeated for each outbreak magnitude, ultimately yielding a plot of sensitivity versus outbreak magnitude with specificity fixed. The likelihood of not having an alert when no signal (specificity) exists can be measured simply by running the model on the baseline data without inserting artificial outbreaks.

Overall Outbreak Detection Versus Outbreak Day Number

Because outbreaks presented to the system typically will last >1 day, sensitivity and specificity can be measured either in terms of detection of specific outbreak days or of the overall outbreak. When the outbreak-day approach is used, each day is considered a separate, independent case; if a particular 5-day outbreak is detected on 3 days but missed on 2 days, three successes (true positives) and two failures (false negatives) are recorded. Similarly, each of the intervening nonoutbreak days is considered independently when false-positive and true-negative rates are calculated.

When the overall outbreak-detection approach is used, each outbreak is viewed as a single entity; if the outbreak is correctly detected on an outbreak day, the system has produced a true positive. An alternative criterion for a true positive is that the outbreak was correctly detected on a majority of the outbreak days. When the overall outbreak sensitivity is reported (e.g., "The system detected X% of all outbreaks presented to it"), full sensitivity and specificity statistics are reported by using the outbreak-days approach.

Receiver Operator Characteristic (ROC) Curves

The tradeoff between sensitivity and specificity is well-portrayed by ROC curves, which plot sensitivity versus one minus the specificity. For tests that have no diagnostic value, the ROC curve is a straight line along the diagonal of the plot. For plots of tests with higher diagnostic value, the line is curved away from the middle of the plot. The area under the ROC curve can thus be used as a measure of the diagnostic value of a test (*9*). The diagnostic value of two tests can be compared by comparing the areas under their respective ROC curves.

Controlled Feature Set Simulation Applications

A series of experiments was conducted by using semisynthetic data sets containing authentic background noise and controlled feature set simulated outbreaks. These experiments illustrate the flexibility of the approach. In all these experiments, the primary sources of data were ED chief complaints and ICD-9 codes from two urban academic teaching hospitals that share the same catchment area. The first experiments were performed to test the accuracy of the model used for the Automated Epidemiologic Geotemporal Integrated Surveillance (AEGIS) system, which was developed at Children's Hospital Boston and Harvard Medical School. This autoregressive integrated moving average (ARIMA) model was constructed on the basis of approximately a decade of historic data from a single ED. The model is run every 10 minutes on real-time data streams producing forecasts of ED volume over the next 24 hours. The system was presented with 7-day–long outbreaks of fixed size, spaced 15 days apart. Specificity was held constant at 97% to produce approximately one false alert/month. On average, 137 visits occurred each day. The results indicated a positive relationship between outbreak magnitude and system sensitivity at varying outbreak magnitudes (*7*).

For performance to be improved, a series of experiments was conducted in which the temporal detection window was widened from 1 day to 1 week, and a controlled feature set

simulation was used to measure the effects of temporal filters, differentially weighting the importance of each day for 1 week. The results demonstrated that the wider temporal window was able to more than double the detection sensitivity while holding the specificity fixed. The results also indicated that different temporal filter shapes provided complementary sets of benefits with regard to timeliness and overall sensitivity of detection (*15*).

Syndromic surveillance systems require data that allow the grouping of patients into syndromes or prodromes. Previous studies have examined the accuracy of different methods of syndrome grouping (*16–19*). This study assessed the effects of syndrome groupings on model accuracy, which is a key factor in outbreak-detection performance (*20*). Daily ED visit rates were analyzed from two urban academic tertiary-care hospitals. Three methods were used to group the visits into a daily respiratory-related syndrome category: chief complaint, diagnostic codes, and a combination of the two. These groupings were used to build historic models that were then tested for forecasting accuracy and sensitivity for detecting simulated outbreaks. For both hospitals, the data grouped according to chief complaint alone yielded the lowest model accuracy and the lowest detection sensitivity. Using diagnostic codes to group the data yielded better results. Smoothing of the data was demonstrated to improve sensitivity in all cases, although to varying degrees. Combining the two grouping methods yielded the best accuracy and sensitivity.

In the last set of experiments, the optimal method for integrating data from multiple regional EDs was determined (*21*). In one simulation, the synthetic outbreak was introduced evenly into both hospital data sets (aggregate model). In the second, the outbreak was introduced into only one or the other of the hospital data sets (local model). The aggregate model had a higher sensitivity for detecting outbreaks that were evenly distributed between the hospitals. However, for outbreaks that were localized to one facility, maintaining individual models for each location proved to be better. Given the complementary benefits offered by both approaches, the results suggested building a hybrid system that includes both individual models for each location and an aggregate model that combines all the data.

Limitations

This study is subject to at least four limitations. First, using simulated data for benchmarking syndromic surveillance systems carries the risk of evaluating performance under unrealistic conditions. Second, the controlled feature set simulation approach entails the explicit assumption that the historic data are pure noise and contain no signal. For terrorism-related events, this assumption is almost certainly true. However, detectable outbreaks of naturally occurring infection are likely contained within the historic data. Third, this approach does not account for processes occurring at the syndrome-grouping stage because artificial cases are injected directly into the data stream. A person with a case of true upper respiratory infection who reports to an ED might not be correctly assigned to the proper syndrome group on the basis of a chief complaint or ICD-9 code. However, the approach could be modified to introduce simulated cases earlier in the process, hypothetically presenting them to the syndrome classifier, enabling modeling of the accuracy of the syndrome grouping process. Finally, in live syndromic surveillance systems, records representing specific events for a given day might be transmitted from the data sources at different points in time. Such time delays could be incorporated into the controlled feature set simulations. In the experiments described, discrete parameter values are assigned. Another approach would be to use a method such as Monte Carlo simulation (*22*) to redefine the model parameters over a smoother distribution of values. Application of controlled feature set simulation to surveillance by using multivariate data streams requires explicit assumptions about the relationships among the signal features across data sets.

Conclusions

Use of semisynthetic data sets containing authentic background noise and outbreaks defined by a controlled feature set provides a valuable means for benchmarking the detection performance of syndromic surveillance systems.

Acknowledgments

This work was supported by the National Institutes of Health through a grant from the National Library of Medicine (R01LM07677-01), by contract 290-00-0020 from the Agency for Healthcare Research and Quality, and by the Alfred P. Sloan Foundation (Grant 2002-12-1).

References

1. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5):1–11.
2. Meselson M, Guillemin J, Hugh-Jones M, et al. The Sverdlovsk anthrax outbreak of 1979. Science 1994;266:1202–8.
3. Jernigan JA, Stephens DS, Ashford DA, et al. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg Infect Dis 2001;7:933–44.
4. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11:141–50.

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. US Department of Defense Global Emerging Infections System. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE). Available at http://www.geis.ha.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCEinstructions.asp.
7. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak 2003;3:2.
8. Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. Am J Public Health 1998;88:1377–80.
9. Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. J Urban Health 2003;80(2 Suppl 1):i57–65.
10. Bonetti M, Olson KL, Mandl KD, Pagano M. Parametric models for interpoint distances and their use in biosurveillance. Proceedings of the American Statistical Association, Biometrics Section [CDROM]. Alexandria, VA: American Statistical Association, 2003.
11. Lazarus R, Kleinman K, Dashevsky I, et al. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis 2002;8:753–60.
12. Hogan WR, Tsui F-C, Ivanov O, et al. Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. J Am Med Inform Assoc 2003;10:555–62.
13. Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proc Natl Acad Sci U S A 2002;99:5237–40.
14. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proc Natl Acad Sci U S A 2003;100:4346–51.
15. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci U S A 2003;100:1961–5.
16. Begier E, Sockwell D, Branch L, et al. The National Capitol Region's emergency department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? Emerg Infect Dis 2003;9:393–6.
17. 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.
18. Espino JU, Wagner MM. Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. Proc AMIA Symp 2001:164–8.
19. Mocny M, Cochrane DG, Allegra JR, et al. A comparison of two methods for biosurveillance of respiratory disease in the emergency department: chief complaint vs ICD-9 diagnosis code [Abstract]. Acad Emerg Med 2003;10:513.
20. Reis BY, Mandl KD. Syndromic surveillance: the effects of syndrome grouping on outbreak detection performance. Ann Emerg Med 2004 (in press).
21. Reis BY, Mandl KD. Integrating syndromic surveillance data across multiple locations: effects on outbreak detection performance. Proc AMIA Symp 2003:549–53.
22. Vose D. Quantitative risk analysis: a guide to Monte Carlo simulation modelling. Chichester, England: John Wiley & Sons, 1996.

Evaluation of Syndromic Surveillance Systems — Design of an Epidemic Simulation Model

David L. Buckeridge,[1,2] H. Burkom,[3] A. Moore,[4] J. Pavlin,[5*] P. Cutchis,[3] W. Hogan[6]
[1]*Palo Alto Veterans Health Care, Palo Alto, California;* [2]*Stanford University, Stanford, California;* [3]*Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland;* [4]*Carnegie Mellon University, Pittsburgh, Pennsylvania;* [5]*Walter Reed Army Institute of Research, Silver Spring, Maryland;* [6]*University of Pittsburgh, Pittsburgh, Pennsylvania*

Corresponding author: David L. Buckeridge, Stanford Medical Informatics, MSOB X-215, 251 Campus Dr., Stanford, CA 94305-5479.

Abstract

Introduction: *The paucity of outbreak data from biologic terrorism and emerging infectious diseases limits the evaluation of syndromic surveillance systems. Evaluation using naturally occurring outbreaks of proxy disease (e.g., influenza) is one alternative but does not allow for rigorous evaluation. Another approach is to inject simulated outbreaks into real background data, but existing simulation models generally do not account for such factors as spatial mobility and do not explicitly incorporate knowledge of the disease agent.*

Objective: *The objective of this analysis was to design a simulated anthrax epidemic injection model that accounts for the complexity of the background data and enables sensitivity analyses based on uncertain disease-agent characteristics.*

Model Requirements and Assumptions: *Model requirements are described and used to limit the scope of model development. Major assumptions used to limit model complexity are also described. Available literature on inhalational anthrax is reviewed to ensure that the level of model detail reflects available disease knowledge.*

Model Design: *The model is divided into four components: 1) agent dispersion, 2) infection, 3) disease and behavior, and 4) data source. The agent-dispersion component uses a Gaussian plume model to compute spore counts on a fine grid. The infection component uses a cohort approach to identify infected persons by residential zip code, accounting for demographic covariates and spatial mobility. The disease and behavior component uses a discrete-event approach to simulate progression through disease stages and health-services utilization. The data-source component generates records to insert into background data sources.*

Conclusions: *An epidemic simulation model was designed to enable evaluation of syndromic surveillance systems. The model addresses limitations of existing simulation approaches by accounting for such factors as spatial mobility and by explicitly modeling disease knowledge. Subsequent work entails software implementation and model validation.*

Introduction

Syndromic surveillance systems potentially allow rapid detection of outbreaks and enable prompt public health intervention (*1*). Although considerable effort and funding have been directed in recent years toward the development of systems and outbreak-detection algorithms, minimal evaluation of their performance in real surveillance environments has been conducted (*2,3*).

The conditions under which a syndromic surveillance system is likely to rapidly detect an outbreak need to be better understood. Public health decision-makers faced with funding decisions for terrorism preparedness should understand which types of disease agents and attack scenarios are likely to be detected by syndromic surveillance and which are not. Although speculations on this topic have been published (*4,5*), limited empirical research has been conducted. From a systems-development perspective, such evidence is required to ensure that developers understand which system configurations (especially which detection algorithms) are best suited to detecting specific attack scenarios and disease agents. The efficacy of algorithms in tightly controlled settings has been evaluated to an extent (*6,7*), but evaluation of outbreak-detection effectiveness in realistic settings has been minimal.

The ideal evaluation approach would assess system performance by using existing outbreaks of the type the system is intended to detect. However, for the majority of locations where systems are operating, essentially no previous data exist on outbreaks from agents of biologic terrorism. An alternative suggestion is to use data on seasonal outbreaks as a proxy signal for evaluation (*8*). This approach is useful but limited. Seasonal outbreaks are limited in number and might differ in

* The views expressed are those of the author and should not be construed as representing the position of the U.S. Department of Defense.

important ways from the type of outbreaks systems are intended to detect. Moreover, performing sensitivity analyses using real outbreak data is not usually possible.

Another alternative is to use simulated data for evaluation. Given the complexities of real data, evaluation should be based on real data injected with simulated outbreaks as opposed to relying on fully simulated data (*8*). To date, simulations have focused on injecting relatively simple signals with abstract characteristics into univariate time series (*6,9*) or on creating simple, abstract spatial signals (*7*). These simulation efforts are useful for understanding the general performance characteristics of detection algorithms, but they do not enable thorough evaluation of surveillance-system and detection-algorithm performance in realistic settings.

A limitation of existing simulation approaches is that they create signals with insufficient complexity to evaluate the effectiveness of certain algorithms in the scenarios and data environments for which they were designed. For example, algorithms used by syndromic surveillance systems often rely on spatial information (*10*) and on the joint distribution of multiple attributes (*7*). To evaluate the performance of a system that uses such algorithms, a simulation must be capable of producing a signal that accounts for such factors as the spatial mobility of persons among regions and the joint distributions of such variables as age and diagnosis. Another limitation of current simulation approaches (*6,7,9*) is that the disease agent responsible for the simulated signal is not explicitly modeled. Such explicit modeling is necessary to understand the plausible range of detection-performance results for a specified outbreak scenario. Different assumptions about disease-agent parameters (e.g., time spent in the incubation state) are required for a simulation model developed for system evaluation.

This paper describes the design for a simulation model intended to enable evaluation of the outbreak-detection characteristics of a syndromic surveillance system. The goal is to develop a model that 1) creates a realistic signal for injection into background data sources, 2) explicitly incorporates knowledge of disease, and 3) is as simple as possible. The aim is to design a model that can be generalized to multiple disease agents, geographic locations, and data sources. However, to focus model development, developers limited the model design to simulate exposure to aerosolized *Bacillus anthracis* spores in the Norfolk, Virginia, area and the resulting effect on outpatient clinical visits and pharmaceutical prescriptions.

Model Requirements and Assumptions

Model Requirements

Developing a simulation model requires simplifying reality in a way that sufficiently decreases complexity but still meets model requirements (*11*). The purpose of this model is to enable evaluation of the outbreak-detection characteristics of syndromic surveillance systems. Functionally, the model must simulate the effects of an epidemic in sufficient detail such that attributable cases can be plausibly injected into the background of authentic health-utilization records. The simulated records must account for such factors as the spatial mobility of the population and joint distributions of multiple attributes within and across data sources. From a design perspective, the model must explicitly incorporate knowledge of the disease agent in a way that enables analyses of the sensitivity of detection performance to key disease parameters.

Model Scope and Assumptions

Focusing on evaluation of timely outbreak detection provides a means of limiting the model's scope. This model assumes that outbreak detection by a surveillance system is successful only if it occurs 1) before the outbreak is evident because a sufficiently large number of persons seek care and 2) before a limited number of persons are diagnosed with a disease caused by a nonendemic Category A biologic agent (*12*). This assumption allows the scope of the model to be limited to the early stages of disease progression, up to hospital admission. However, it also requires that the model accurately reflect population and provider behavior before and after illness onset.

Another assumption is that before an epidemic is recognized, the behaviors of both health-care seekers and providers are reflected by historic data. This means that persons use health-care services, and health-care providers assign diagnoses and prescribe medications, according to historic patterns for persons with similar demographic characteristics and similar symptoms. Historic patterns for health-care consumers and providers can be determined empirically from background data, and this assumption substantially limits the need for quantitative data on health-care utilization in the early stages of an epidemic.

Parameters for Simulation of Inhalational Anthrax

To develop a model for inhalational anthrax, the investigators reviewed the literature on anthrax to ensure that the disease was modeled at a resolution appropriate to available knowledge. A limited number of studies have quantitatively modeled dispersion of anthrax spores, infection with inhalational anthrax, and disease progression among infected persons. A study of the Sverdlosk anthrax outbreak indicated that dispersion of an aerosol of *B. anthracis* is adequately described by a Gaussian plume model (*13*). Although it was based on incomplete data, that validation of the Gaussian plume model is the most complete analysis described in the literature. The Gaussian plume model seems to provide a reasonable first approximation in an urban setting, and others have used the model in an urban environment with essentially no modifications (*14*).

The estimates of an infectious dose of *B. anthracis* spores (ID_{10} = 1,135, ID_{50} = 8,940) proposed previously (*15*) are the geometric means of estimates from subject matter specialists, obtained before the U.S. anthrax cases in 2001. The age-specific estimates proposed (ID_{10} = 450–4,500, ID_{50} = 1,500–15,000) (*16*) were apparently based on the previous estimate (*15*) but were modified to account for knowledge derived from analysis of the 2001 exposures to mailborne *B. anthracis*. The revised age-specific values are consistent with the observation that the infectious doses in the 2001 cases were lower than previously thought necessary, and with observation from the Sverdlosk cases that children seem to require higher infectious doses (*13*). The probability of infection can be estimated from the number of spores inhaled (S) and the age category (n), by using functions described previously (*14,16*).

In terms of disease progression, one researcher (*14*) modeled five disease states for inhalational anthrax: uninfected, incubating, prodromal, fulminant, and dead. In addition, values determined from the Sverdlosk outbreak (*17*) were used to parameterize the lognormal distribution of duration in the incubation state (*14*). Parameters for the lognormal distributions of time in the prodromal and fulminant states from the 2001 cases in the United States (*18,19*) and an analysis of the time from exposure to death (*17*) were also estimated (*14*). These estimates in days are incubation (median = 10.95; dispersion = 2.04), prodromal (median = 12.18; dispersion = 1.41), and fulminant (median = 1.5; dispersion = 1.41), where the log of time in a state is normally distributed with mean μ and variance σ^2: $\log(t) \sim N(\mu, \sigma^2)$. Following other published work (*20*), the dispersion factor $d = \exp(\sigma)$. Approximately 68% of the cases in a state fall in the interval median/d to median * d, and roughly 95% of the cases fall in the interval median/d^2 to median * d^2. Human (*21*) and animal (*22*) evidence demonstrates that duration in the incubation state depends on the number of inhaled spores, although research indicates that the Sverdlosk data do not support this (*14*). In addition, animal evidence indicates that time from exposure to death is dose-dependent (*22*), although whether this is attributable only to a shortened incubation period and not also to a shortened duration of subsequent states is not clear.

Background Data and Simulation Region

Although the model is intended to be generalizable to other settings, our initial design focuses on two specific data sources drawn from the Norfolk, Virginia, region: ambulatory physician visit billing records and pharmaceutical prescription records for military personnel and their dependents. These types of data are used routinely by syndromic surveillance systems (*23–26*). Persons are uniquely identified with encrypted personal identifiers in a way that allows anonymous linkage of records for persons across the two data sources. The simulation region is defined as an area approximately 160 km by 200 km that encompasses 158 zip codes from two states. During July 2001–May 2003, a total of 115,732 persons from the simulation region made 231,116 clinical visits and 148,761 pharmacy visits. Within the region, clinical visits were made to 16 clinical facilities, and prescriptions were filled at 316 pharmacies. The fields in the background data sources are provided (Table 1).

Model Design

To facilitate overall model development and description, the model was divided into four components: a dispersion model,

TABLE 1. Fields in data sources used for simulation

Field	Physician visits	Pharmaceutical prescriptions
Scrambled subscriber identification number	*	*
Family member identification number	*	*
Facility type	*†	*†
Subscriber residential zip code	*†	*†
Encounter date	*†	
Date written		*†
Date filled		*
Facility identification number	*	*
Facility zip code	*†	*†
Facility type	*†	
Code	ICD-9†§	GC3†¶

* Field is present.
† Simulation model outputs field for simulated records.
§ *International Classification of Diseases, Ninth Revision.*
¶ Therapeutic class code.

an infection model, a disease and behavior model, and a data-source model (Figure 1). The dispersion model makes a calculation of the distribution of aerosolized spores over the study area. The infection model then takes the spatial distribution of spores, along with information on the covered population and inter-region travel information, to estimate the number of infected persons by home location. The disease and behavior model then determines progression of infected persons through disease stages and identifies the health-care–seeking behaviors of these persons over time. Finally, the data-source model converts the generic behaviors taken by persons into specific database records that can be combined with real background data.

The rest of this section describes each model component, focusing on the general structure, main assumptions, and parameters to be varied. Mathematical and technical details are available from the corresponding author and are not presented here.

Dispersion Model

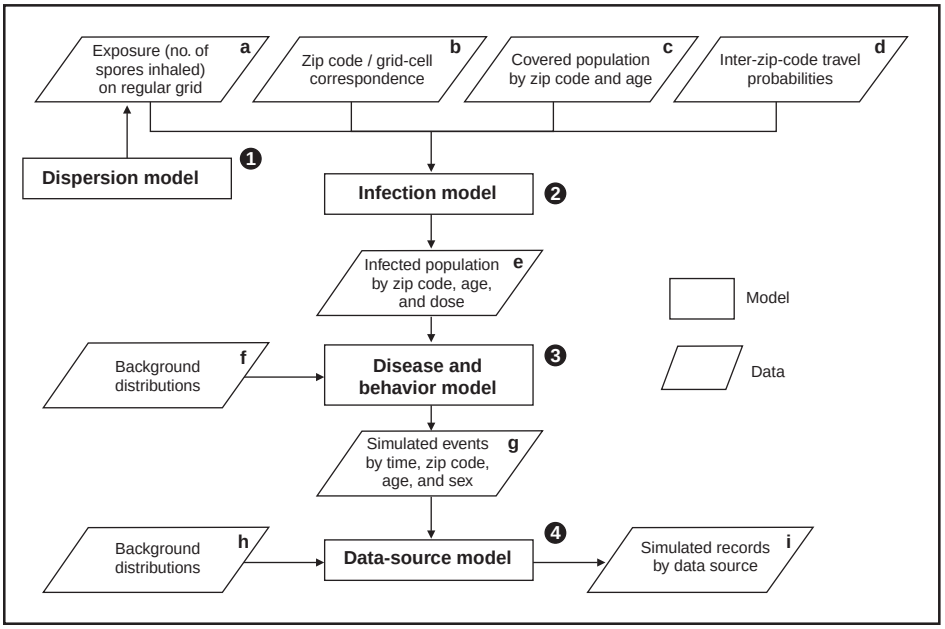
The dispersion model calculates the number of spores inhaled at point locations in the simulation region. A Gaussian plume model was used to simulate dispersion of spores over the region. Home locations of covered persons in the background data are available by zip code. Because considerable variation exists in the shape and size of zip codes within the region, the simplest approach of estimating spore exposure at a single point within each zip code was rejected. Instead, a regular grid over the simulation region, with at least one grid cell falling within each zip code, was defined. A cell size of 100 m is sufficient for this purpose in the Norfolk, Virginia, region. Therefore, each run of the dispersion model will take as input the release parameters (location, amount, and atmospheric conditions) and the grid description, and produce as output the number of spores inhaled at the center of each cell on the grid. The main parameters to vary within this model component are the amount of release, the location of the release, and the atmospheric conditions (wind direction and speed and atmospheric turbulence).

Infection Model

The infection model determines the number of infected persons from the covered population in each age/residential zip code/sex/spore-dose stratum. The covered population is defined as the set of unique persons represented in the background data sources. The average of the spore counts for the grid cells that fall within the zip code is used to determine the spore concentrations within each zip code. Correspondence of grid cells to zip codes is determined by overlaying the grid on the zip-code boundaries by using a geographic information system (GIS) and then using spatial topology to assign each grid cell the zip code that contains the centroid of the grid cell.

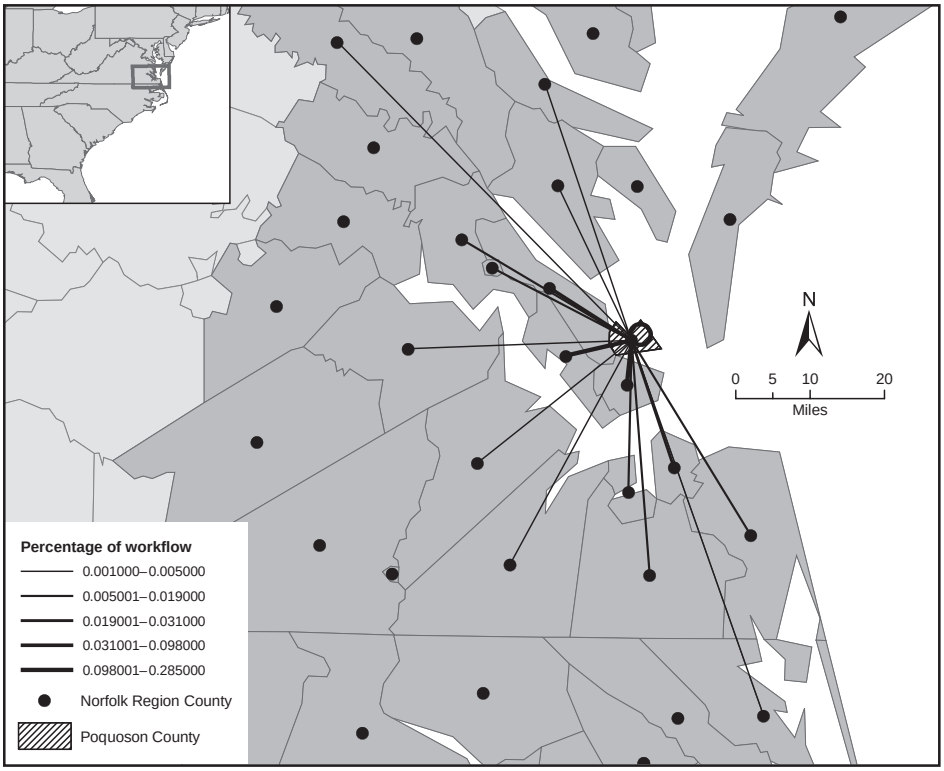
The geographic distribution of the covered population at the time of exposure is modeled as the probability of a person being in a zip code at a certain time given his or her residential zip code and age category. Time is divided into three categories (work/school, recreation, and home) on the basis of time of day and day of week, and three age groups are identified (young [0–18 years], middle-aged [19–64 years], and elderly [>64 years]. For the *work* time category and the *middle-aged* age group, probabilities are determined from U.S. Census workflow data (*27*) (Figure 2). For all other combinations of time categories and age groups, probabilities are determined by using inverse exponen-

FIGURE 1. Overview of an epidemic simulation model design illustrating the relation between model subcomponent and data sources



Note: A model run begins with calculation of spore distribution by the dispersion model (1) for a given release scenario. The infection model (2) then makes a stochastic calculation of infected persons. Disease course and health-seeking behaviors of infected persons are then simulated by the disease and behavior model (3). Finally, the data-source model (4) generates simulated records for insertion into background data sources.

FIGURE 2. Example of county-to-county workflow data used to simulate mobility between regions



Source: U.S. Census data, 2000.
Note: The proportion of workers leaving a county (only one origin county is provided here for clarity) to work in other counties is represented by the thickness of the arc between the origin county and the destination county.

tial driving distance between zip codes, with a distinct exponential weight for each time category. The weights are to be varied in sensitivity analyses and are chosen so that persons tend to be more widely dispersed during recreation times than during work or school times, and in turn more widely dispersed during work or school times than during home times.

The spore-concentration data, the geographic distribution of the covered population, and the probability of infection given dose and age (as described in Methods) are used to determine the probability of infection for each age/residential zip code/sex/spore dose stratum given the attack time. This probability is then used along with the number of persons in the covered population to sample the number infected in each stratum from a binomial distribution. Each run of the infection model will therefore take as input the time of the attack, the number of spores at each location on the grid (from the dispersion model), the covered population, grid cell-to-zip code correspondence, workflow mobility, inter-zip code driving distances, and distance weights. The output of this model will be the number of persons infected within each age/residential zip code/sex/spore-dose stratum. The main parameters to vary are the probability of infection given spore dose and the distance weights used to determine the geographic distribution of nonworking persons.

Disease and Behavior Model

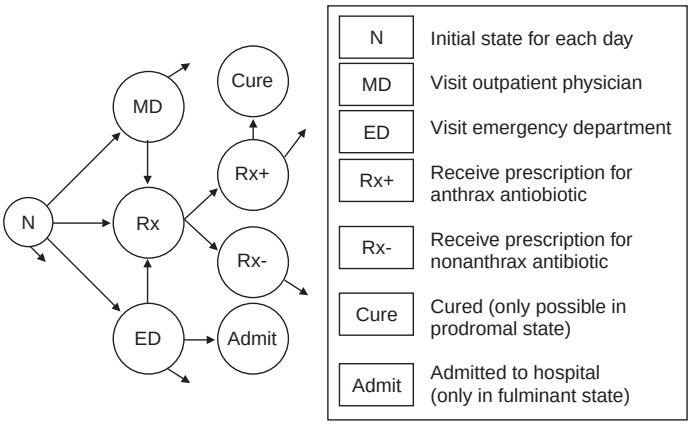
The disease and behavior model determines the progression of infected persons through disease states and the generic types of health-care–utilization behaviors of infected persons. Drawing on previous work in modeling anthrax (*14*), progression is modeled through three disease states: incubation, prodromal, and fulminant. The disease progression for each person is modeled as a semi-Markov process (*28*), with the transition time between states sampled from a log-normal distribution parameterized by the person's spore dose. Base case parameters are adapted from a previous simulation study (*14*) (Table 2). Each infected person begins in the incubation state and progresses to the prodromal state. Unless successfully treated with curative therapy while in the prodromal state, the illness progresses to the fulminant state and then exits the model after the simulated duration of the fulminant state.

For each day a person is in the prodromal or fulminant disease state, the person's health-care–utilization behaviors are simulated. The behaviors of persons are modeled as a Markov process (*28*), with the transition probabilities drawn primarily from the background data (Figure 3). The model is run

TABLE 2. Base case lognormal distribution parameters for dose-dependent duration of three inhalational anthrax disease states, by dose category and number of spores inhaled

Disease state	Dose category	Spores inhaled	Median (μ)	Dispersion (variance)
Incubation	High	>12,000	4 (1.4)	1.75 (0.31)
	Medium	4,000–12,000	10.95 (2.4)	1.75 (0.31)
	Low	<4,000	15 (2.7)	1.75 (0.31)
Prodromal	High	>12,000	1.0 (0)	1.25 (0.05)
	Medium	4,000–12,000	2.5 (0.9)	1.25 (0.05)
	Low	<4,000	4.0 (1.4)	1.25 (0.05)
Fulminant	High	>12,000	1.0 (0)	1.25 (0.05)
	Medium	4,000–12,000	1.5 (0.4)	1.25 (0.05)
	Low	<4,000	2.5 (0.9)	1.25 (0.05)

FIGURE 3. Disease-behavior model for persons in prodromal and fulminant disease states



Note: A person's path through the model is simulated for each day spent in the prodromal and fulminant state. Simulated behaviors on each day lead to the generation of corresponding records by the data-source model.

for each person on each day until the person exits the model. Each person begins in the initial state (N) from which he or she can seek care in one of three ways: 1) a physician visit (MD); 2) an emergency-department visit (ED); or 3) a prescription without a clinical visit (Rx), or not seek care and exit the behavior model for that day.

The first step in determining whether and how a person seeks care is to determine the daily background probability distribution of age/sex/diagnostic set for each care-seeking behavior. The diagnostic set is the set of *International Classification of Diseases, Ninth Revision* (ICD-9) diagnoses consistent with a person in the same disease state. Day-of-the-week variation in visit probability is taken into account when calculating background probabilities during the prodromal stage but not in the fulminant stage. In the fulminant stage, the assumption is that the only behavior that can be taken is to visit the ED, and that, at the first visit, the person is admitted and therefore leaves the simulation model. In the prodromal state, multiple visits can occur, and the background distribution of person-visit frequency is used to scale the probability of repeat visits. After the background probability of type of care by covariates has been determined, the next step in determining whether a person seeks care is to multiply the background probability by a scale factor unique to each disease state. These scale factors, to be varied in sensitivity analyses, account for the probability of not making a health-care visit for persons having symptoms consistent with the disease state. This cannot be estimated from the background data. Work is under way to identify these scale factors for classes of symptoms (e.g., lower respiratory, constitutional) through literature review and health-utilization surveys (*29*). After an individual care-seeking behavior is chosen, subsequent transition probabilities are determined directly from the background data for persons with the same age/sex/diagnostic set.

The disease component of the model is run once for each infected person, and the behavior component of the model is run once for each day an infected person is in the prodromal state and once for each day in the unhospitalized fulminant state. Input to the disease-behavior model is the number of infected persons in each age/sex/spore-dose stratum, the disease state transition parameters, the diagnostic sets for each disease state, and the scale factors for seeking care in each disease state. The output is a set of behavior records for each infected person with each record defining the date of health-care utilization, demographic information including residential zip code, and type of utilization. The main parameters to vary are the disease state transition parameters, and the diagnostics sets and care-seeking scale factors for each disease state.

Data-Source Model

The data-source model uses the behavior records from the disease and behavior model to generate records for injection into background data sources. The current model includes two data sources: clinical visits and pharmaceutical prescriptions. These data sources are described in the Methods section; a list of the fields in each data source is provided (Table 1). Creation of a data source record requires assigning a diagnostic (ICD-9) or pharmaceutical code (GC3) and facility (clinic, hospital, or pharmacy) to a behavior record and formatting the resulting information to match the background data structure. Facility location is chosen by sampling the background data distribution based on the historic use of facilities by persons from the same residential zip code with the same diagnostic set. Diagnostic and pharmaceutical codes are chosen by sampling historic data distributions for persons with similar demographic characteristics. The inputs to the data source model are the behavior records and the diagnostic sets for each disease state. The output is the records for injection into the background data sources. The only parameter to vary is the diagnostic sets.

Conclusions

This paper defines requirements and specifies a design for an injection simulation model that should enable evaluation of outbreak detection through syndromic surveillance. Although it is intended to be generalizable, the model is described in the form required to simulate an aerosol attack with *B. anthracis* spores in the Norfolk, Virginia, area. The model scope and complexity have been limited by making plausible assumptions regarding patient and health-care pro-

vider behavior. The model also demonstrates an approach to developing a sufficiently complex outbreak signal by incorporating spatial mobility and by relying on joint variable distributions in background data sources. Finally, a method for incorporating explicit models of disease and illness behavior into a simulation model was demonstrated. The degree of detail in the model should allow for sensitivity analyses based on uncertain disease and behavior parameters to determine their influence on detection performance.

References

1. Wagner MM, Tsui F-C, Espino JU, et al. The emerging science of very early detection of disease outbreaks. J Public Health Manag Pract 2001;7:51–9.
2. Becker J. High-tech health watch draws cash, questions. Washington Post, November 23, 2003;A:17.
3. Bravata DM, McDonald KM, Smith WM, et al. A critical evaluation of existing surveillance systems for illnesses and syndromes potentially related to bioterrorism. Ann Intern Med 2004;140:910–22.
4. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis 2003;9: 1197–204.
5. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 2003;1:77–81.
6. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. PNAS 2003;100:1961–5.
7. Wong W-K, Moore A, Cooper G, Wagner M. Rule-based anomaly pattern detection for detecting disease outbreaks. In: AAAI-02; Edmonton, Alberta: AAAI Press/The MIT Press;2002:217–23.
8. Sosin DM. Draft framework for evaluating syndromic surveillance systems. J Urban Health 2003;80(2 Suppl 1):8–13.
9. Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. PNAS 2002;99:5237–40.
10. Kulldorff M. A spatial scan statistic. Communications in Statistics. Theory and Methods. 1997;26:1481–96.
11. Law AM, Kelton WD. Simulation modeling and analysis. 3rd ed. Boston, MA: McGraw-Hill, 2000.
12. Darling RG, Catlett CL, Huebner KD, Jarrett DG. Threats in bioterrorism. I: CDC category A agents. Emerg Med Clin North Am 2002;20:273–309.
13. Meselson M, Guillemin J, Hugh-Jones M, et al. The Sverdlovsk anthrax outbreak of 1979. Science 1994;266:1202–8.
14. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proc Natl Acad Sci U S A 2003;100:4346–51.
15. Rickmeier GL, McClellan GE, Anno GA. Biological warfare human response modeling. Military Operations Research 2001;6:35–47.
16. Webb GF, Blaser MJ. Mailborne transmission of anthrax: modeling and implications. Proc Natl Acad Sci U S A 2002;99:7027–32.
17. Brookmeyer R, Blades N, Hugh-Jones M, Henderson DA. The statistical analysis of truncated data: application to the Sverdlovsk anthrax outbreak. Biostatistics 2001;2:233–47.
18. Jernigan JA, Stephens DS, Ashford DA, et al. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg Infect Dis 2001;7:933–44.
19. Jernigan DB, Raghunathan PL, Bell BP, et al. Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. Emerg Infect Dis 2002;8:1019–28.
20. Sartwell PE. The distribution of incubation periods of infectious diseases. American Journal of Hygiene 1950;51:310–318.
21. Brachman PS, Kaufmann AF, Dalldorf FG. Industrial inhalation anthrax. Bacteriological Reviews 1966;30:646–657.
22. Gleiser CA, Berdjis CC, Hartman HA, Gochenour WS. Pathology of experimental respiratory anthrax in *Macaca mulatta*. British Journal of Experimental Pathology 1963;64:416–26.
23. 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.
24. Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington DC area. Am J Prev Med 2002;23:180–6.
25. Lombardo J, Burkom H, Elbert E, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health 2003;80(2 Suppl 1):32–42.
26. 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.
27. US Census Bureau. County-to-county worker flow files. Washington, DC: US Census Bureau, Population Division, 2003. Available at http://www.census.gov/population/www/cen2000/commuting.html.
28. Shedler G. Regenerative stochastic simulation. New York, NY: Academic Press, 1992.
29. National Center for Health Statistics. Public use data tapes: National Ambulatory Medical Care Survey 2003. Hyattsville, MD: National Center for Health Statistics, 2003.

Benchmark Data and Power Calculations for Evaluating Disease Outbreak Detection Methods

Martin Kulldorff,[1] Z. Zhang,[2] J. Hartman,[3,4] R. Heffernan,[3] L. Huang,[2] F. Mostashari[3]
[1]*Harvard Medical School and Harvard Pilgrim Health Care, Boston, Massachusetts;* [2]*University of Connecticut, Storrs, Connecticut;*
[3]*New York City Department of Health and Mental Hygiene, New York, New York;* [4]*New York Academy of Medicine, New York, New York*

Corresponding author: Martin Kulldorff, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Avenue, 6th Floor, Boston, MA 02215. Telephone: 617-509-9757; Fax: 617-859-8112; E-mail: martin_kulldorff@hms.harvard.edu.

Abstract

Introduction: *Early detection of disease outbreaks enables public health officials to implement immediate disease control and prevention measures. Computer-based syndromic surveillance systems are being implemented to complement reporting by physicians and other health-care professionals to improve the timeliness of disease-outbreak detection. Space-time disease-surveillance methods have been proposed as a supplement to purely temporal statistical methods for outbreak detection to detect localized outbreaks before they spread to larger regions.*

Objective: *The aims of this study were twofold: 1) to design and make available benchmark data sets for evaluating the statistical power of space-time early detection methods and 2) to evaluate the power of the prospective purely temporal and space-time scan statistics by applying them to the benchmark data sets at different parameter settings.*

Methods: *Simulated data sets based on the geography and population of New York City were created, including effects of outbreaks of varying size and location. Data sets with no outbreak effects were also created. Scan statistics were then run on these data sets, and the resulting power performances were analyzed and compared.*

Results: *The prospective space-time scan statistic performs well for a spectrum of outbreak models. By comparison, the prospective purely temporal scan statistic has higher power for detecting citywide outbreaks but lower power for detecting geographically localized outbreaks.*

Conclusions: *The benchmark data sets created for this study can be used successfully for formal statistical power evaluations and comparisons. If an anomaly caused by an outbreak is local, purely temporal surveillance methods might be unable to detect it, in which case space-time methods would be necessary for early detection.*

Introduction

Early detection of disease outbreaks enables public health officials to implement disease control and prevention measures at the earliest possible time (*1–3*). For an infectious disease, improvement in detection timeliness by even 1 day might enable public health officials to control the disease before it becomes widespread. Real-time, geographic, early outbreak-detection systems have been used in New York City (NYC) (*4–8*), the greater Washington, D.C., area (*9*), Salt Lake City, Utah (*10*), and other locations (*11*). Because the onset of a disease outbreak is unpredictable, early detection methods need to continuously evaluate different incoming data streams (e.g., ambulance dispatches, emergency department [ED] visits, pharmacy sales, or health insurance claims). Furthermore, because early evidence of an outbreak might be localized, systems need to monitor multiple locations simultaneously because neither the extent nor geographic pattern of the outbreak is yet known.

The majority of traditional disease-surveillance methods are purely temporal in nature in that they seek anomalies in time-series data without using spatial information (*12*). Although temporal methods are important and can be used simultaneously for multiple areas, they do not take into account geographic location and might be unable to quickly detect localized outbreaks that do not conform to predefined areas. For this reason, different space-time early detection methods have been proposed (*13–17*). Research in this area is ongoing, and new or refined methods will likely be proposed soon. The effectiveness of these new methods will then have to be evaluated and compared with current methods.

When evaluating an outbreak-detection method, investigators should have knowledge of the method's ability to detect true outbreaks and the number of false alerts likely to result. The first aim of this study was to create simulated benchmark data sets that can be used for rigorous evaluation of the statistical power of early outbreak-detection methods, an important complement to other evaluations that use real data

sets with known outbreaks or real data sets with spiked outbreaks in which additional artificial cases are added to the real cases. The second aim was to estimate and compare the power of prospective purely temporal scan statistics with different versions of the prospective space-time scan statistics (*14*) that are used daily by the syndromic surveillance program of the NYC Department of Health and Mental Hygiene.

Methods

Benchmark Data

A collection of public benchmark data sets for statistical power comparisons was established to enable evaluation and comparison of early detection methods as they are developed. Geographic coordinates (representing the approximate center of each zip code) and population numbers for 176 NYC zip codes were used for these data sets.

A total of 134,977 benchmark data sets with a random number of cases of a hypothetical disease or syndrome were generated under either the null model or one of 35 alternative models, including a citywide outbreak with a relative risk of 1.5 and 34 geographically localized outbreaks in one of 17 different locations with either a high or modest excess risk. Three different sets of data sets were then generated under the null model and under each of the 35 alternative models, each with 31, 32, and 33 days, respectively. For each of the three null-model scenarios, with 31, 32, and 33 days, respectively, 9,999 random data sets were generated. For each of the 3 sets of 35 alternative models, defined by the number of days in the data and the location and relative risk of the outbreak, 1,000 random data sets were generated.

For each data set, the total number of randomly allocated cases was 100 times the number of days (i.e., 3,100 cases in the data sets containing 31 days, 3,200 cases in the data sets with 32 days, and 3,300 cases in the data sets with 33 days). The number 100 was chosen to reflect the occurrence rate of certain syndromes common to the NYC ED-based syndromic surveillance system.

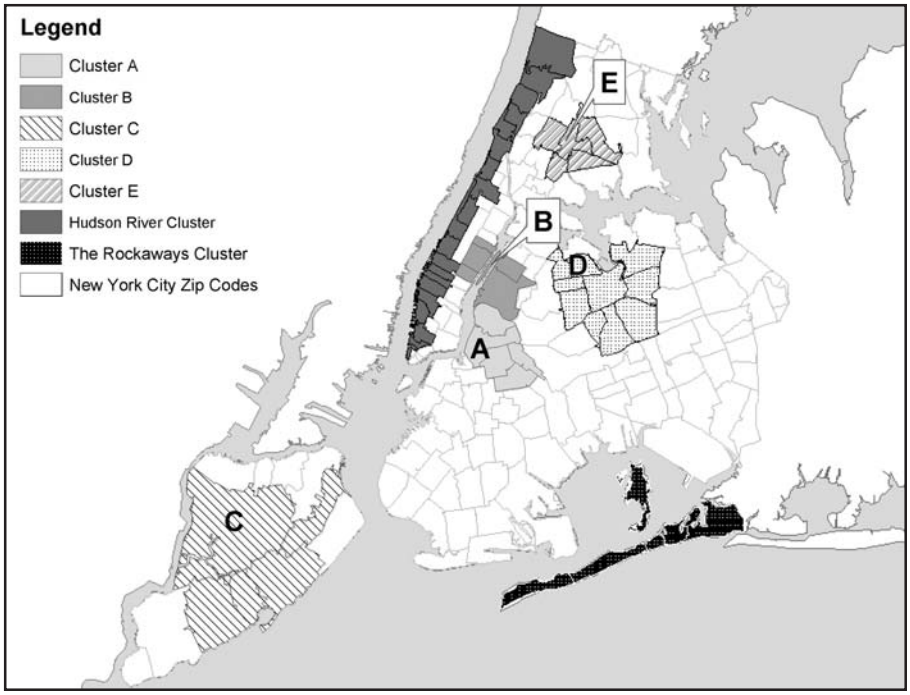
Under the null model, each person living in NYC is equally likely to contract the disease, and the time of each case is assigned with equal probability to any given day. Thus, each case was randomly assigned to zip code z and day d with probability proportional to $r_{zd} = pop_z$, where pop_z is the population of zip code z.

For the alternative models, one or more zip codes were assigned an increased risk on day 31 and, when applicable, on days 32 and 33 as well. For these zip code and day combinations, r_{zd} was multiplied by an assigned relative risk. For all other zip code and day combinations, r_{zd} did not change. Each case was then randomly assigned with probabilities proportional to the new set of r_{zd} to generate data under the alternative models.

Six alternative models in which the outbreak affected only one zip code were evaluated. The six zip codes varied in size and location. Next, six additional alternative models were considered, with the outbreak centered at the same six zip codes but also including four to nine neighboring areas. Seven additional alternative models, with outbreaks in the Rockaways region, along the Hudson River, and throughout each of the five NYC boroughs were also examined (Figure 1).

For each of the alternative models, the relative risk of the outbreak was assigned on the basis of the outbreak area's total population, with more populous areas assigned a lower relative risk. This was done so that the power was 99% to detect a signal at the $\alpha = 0.05$ level when a Poisson distribution was used to compare the observed relative risk within the outbreak area with the remaining zip codes by using only 1 day of data with a total of 100 cases. This approach permits evaluation of the relative strength of methods for detecting differ-

FIGURE 1. Location and size of simulated disease outbreaks — New York City



ent cluster types to be evaluated. For example, if a method has 85% power to detect an outbreak in one zip code and 80% power to detect a borough outbreak, the method is relatively more efficient at detecting smaller outbreaks. In addition to the relative risks created by using the 99% rule, a second group of data sets was created and evaluated by using the same rule but with 90% power. An alternative model with a citywide outbreak was also considered, with a relative risk of 1.5 in all zip codes during days 31, 32, and 33.

By using the same simulated data when comparing methods, the variance of the power-estimate differences is kept to a minimum (*18*). Availability of the simulated data sets (http://www.satscan.org/datasets) will enable new methods to be thoroughly evaluated and compared with minimal effort.

For statistical reasons, completely separate data sets with 31, 32, and 33 days, respectively (rather than one data set from which one could then use the desired number of days) were created to obtain proper power estimates. The majority of methods for conducting statistical evaluation of geographic clusters are based on Monte Carlo hypothesis testing (*19*), whereby the test statistic for the real data set is compared with the value of the test statistic for simulated data sets generated under the null hypothesis, after conditioning on the total number of cases observed. That is, the critical values for the likelihood ratio or any other type of test statistic are calculated conditioned on the total number of cases in the data set, so that only their geographic and temporal distribution are evaluated but not the total count observed. If only one data set was used for all three periods, the total number of cases during the first 31 days must be fixed to condition on the total number of cases in the 31-day analysis, but the total number of cases during the first 32 days must also be fixed to condition on the total number of cases in the 32-day analysis. However, if both of those are fixed, then the number of cases during day 32 is also fixed, which cannot be done because one should condition only on the total number of cases during the whole study period but not on individual days within that period.

Prospective Space-Time Scan Statistic

The benchmark data sets were used to estimate the power of the prospective space-time scan statistic (*14*). In brief, the prospective space-time scan statistic imposes a cylindrical window on the map and lets the center of the circular base move over the region, so that, at different positions, the window includes different sets of neighboring zip codes. For each circle center, the circle's radius is varied continuously from zero up to a maximum radius so that the window never includes, for example, >50% of the total population at risk. Thus, the window remains flexible, both in location and size. In addition, the height of the cylinder, representing time, is flexible such that the window might contain one or more days up to an upper limit. Hence, the window could cover a geographically small outbreak in a single zip code having lasted multiple days (a long and narrow cylinder), a geographically large outbreak affecting the entire city but present only during the last day (a short and fat cylinder), or any other combination of geographic size and temporal length. In total, the method creates thousands of distinct windows, each with a different set of neighboring zip codes and days within it, and each a possible candidate for containing a disease outbreak.

Only those cylinders that reach all the way to the end of the study period are considered. In mathematical notation, let $[B,E]$ represent the time interval for which data exist, and let s and t represent the start and end dates of the cylinder, respectively. All cylinders for which

$$B \le s \le t = E$$

are then considered. Different parameter options can be chosen in terms of the maximum geographic and temporal cluster size being considered; this study evaluated five different combinations.

Conditioning on the observed total number of cases, N, the definition of the space-time scan statistic S is the maximum likelihood ratio over all possible cylinders Z,

$$S = \frac{\max_Z L(Z)}{L_0} = \max_Z \frac{L(Z)}{L_0}$$

where L(Z) is the maximum likelihood for cylinder Z, expressing how likely the observed data are when allowing for different risk inside and outside the cylinder, and where L_0 is the likelihood function under the null model.

Let n_Z represent the number of cases in cylinder Z. Using a Poisson model for the observed number of counts, let $\mu(Z)$ be the expected number under the null model, so that $\mu(A) = N$ for A, the total region under study. Then,

$$\frac{L(Z)}{L_0} = \left(\frac{n_Z}{\mu(Z)}\right)^{n_Z} \left(\frac{N-n_Z}{N-\mu(Z)}\right)^{N-n_Z}$$

if $n_z > \mu(Z)$ and $L(Z)/L_0 = 1$. Details about the mathematical formulas, including derivations as likelihood ratio tests, have been published elsewhere (*20*). The cylinder for which this likelihood ratio is maximized identifies the most likely cluster. Its p-value is obtained through Monte Carlo hypothesis testing (*19*).

The prospective space-time scan statistic can be implemented by using different parameter options. As the standard analytic option, 50% of the population was used as the upper limit on

the geographic cluster size, and a period of 3 days was used as the upper limit on the temporal cluster size. The possibility of citywide outbreaks was also considered by including purely temporal clusters containing 100% of the population in addition to the 50% maximum size. No adjustment was made for the time-repeated analyses conducted daily. For selected alternative outbreak models, the power of the prospective space-time scan statistic was evaluated for the following changes in parameter options: 1) not including purely temporal clusters; 2) setting the maximum geographic cluster size at 5% of the population rather than 50%; 3) setting the maximum temporal cluster size at 1 and 7 days, respectively, rather than 3 days; and 4) adjusting for the multiple testing stemming from the repeated daily analyses such that only one false alert would be expected per year (*14*).

Purely Temporal Scan Statistic

The purely temporal scan statistic is mathematically a special case of the space-time scan statistic, in which counts from the entire surveillance area are aggregated so that no spatial information remains. Hence, the window is defined only by its temporal length, which could be one or more days. As with the prospective space-time scan statistic described previously, only those windows for which $B \leq s \leq t = E$ were considered. A period of 3 days was used as the maximum temporal length.

Power Estimations

The power estimations were conducted as follows. First, for the random data sets generated under the null model, the log likelihood ratio (LLR) was obtained for all cylindrical window locations and sizes, and its maximum noted, to obtain the maximum LLR for each simulated data set. A critical value corresponding to a 0.05 significance level was computed by identifying the 500th highest maximum LLR from among the 9,999 random data sets generated under the null model. Then, the estimated power for a particular alternative model was calculated as the percentage of the 1,000 random data sets for which the maximum LLR exceeds the critical value.

Separate critical values were obtained for each number of days considered (31, 32, and 33) and for each of the different analytic options used. However, as long as the number of days and the analytic options are the same, the same critical value can be used for different alternative outbreak models. All calculations were performed by writing additional routines for the SaTScan™ software (*21*).

Results

For the standard parameter options, the estimated powers for the different alternative models and different relative risks are provided (Tables 1 and 2). The power was good for both small and large outbreak areas. As expected, the power was higher when more days had elapsed since the start of the outbreak. The increase in power was rapid. The power was approximately the same for outbreaks of different sizes. The major exception was the Hudson River outbreak, for which the lower power was caused by using a circular geographic window to capture a long and narrow outbreak. The same loss of power was not seen in the similarly shaped Rockaways region, possibly because that region has fewer zip codes than the Hudson River outbreak region.

For selected alternative outbreak models, the estimated powers for each parameter option are provided, as well as for the purely temporal scan statistic (Table 3). Setting the maximum temporal cluster size to 1 day increased the power to detect the outbreak during the first day, at the expense of decreased power during subsequent days.

Adjusting for previous analyses reduces the power, a consequence of the unavoidable trade-off between power and the number of false positives. Hence, the choice of whether to adjust for previous analyses is similar to a choice of whether to use 0.01 instead of 0.05 as the α level. Both approaches will reduce the number of false alerts but also reduce the power to detect true outbreaks. The purely temporal scan statistic has considerably higher power for citywide outbreaks but does not perform well for localized outbreaks.

Certain power estimates were unexpected. For example, for an outbreak in a single zip code, the power would be expected to be higher with an upper limit of 5% rather than 50% on the geographic cluster size. However, for outbreak model A, the power is 0.86 in both cases (Table 3). The power is depicted as a function of the false-detection rate (α level) (Figure 2, top). The number of cases in the outbreak is always an integer, and if the outbreak area is limited, only a limited number of integer values are possible in the true outbreak area. Thus, the power function takes discrete jumps at certain α levels, and the location of the jump varies for different analytic options. Hence, for certain values of α, one method might be superior to another even though both methods have almost the same power at other α levels. The locations of these jump points are different for different single zip code outbreaks. As the number of zip codes in an outbreak area increases, this phenomenon disappears, such that the power functions are much smoother for model A with four neighbors (Figure 2, middle) and for Manhattan (Figure 2, bottom).

TABLE 1. Estimated power of the prospective space-time scan statistic for 17 different outbreak models with high excess risk, at different days of the outbreak

Outbreak area	No. of zip codes	Pop. %*	RR†	Expected/day§		Power on day		
				Null model	Alternative model	31	32	33
A. Williamsburg, Brooklyn	1	1.1	9.91	1.1	10.9	0.86	0.996	0.999
B. Roosevelt Island, Manhattan	1	0.1	57.08	0.1	5.7	0.92	0.996	1.000
C. Bulls Head, Staten Island	1	1.1	9.89	1.1	10.8	0.83	0.99	1.000
D. LaGuardia, Queens	1	0.5	18.63	0.5	9.3	0.85	0.998	1.000
E. West Farms, Bronx	1	0.7	13.76	0.7	9.6	0.83	0.997	1.000
A with 4 neighboring zip codes	5	4.0	4.47	4.0	17.8	0.85	0.996	1.000
B with 5 neighboring zip codes	6	3.2	5.02	3.2	16.0	0.82	0.996	1.000
C with 4 neighboring zip codes	5	3.3	4.93	3.3	16.2	0.83	0.99	1.000
D with 9 neighboring zip codes	10	8.2	3.24	8.2	26.4	0.88	0.996	1.000
E with 4 neighboring zip codes	5	3.7	4.62	3.7	17.0	0.86	0.99	1.000
Rockaways	5	1.3	8.48	1.3	11.0	0.84	0.997	1.000
Hudson River	20	10.3	2.97	10.3	30.4	0.66	0.96	0.996
Bronx	25	16.6	2.56	16.6	42.1	0.94	1.000	1.000
Brooklyn	37	30.8	2.25	30.8	68.4	0.98	1.000	1.000
Manhattan	40	19.0	2.47	19.0	46.5	0.92	1.000	1.000
Queens	62	28.0	2.28	28.0	63.1	0.98	1.000	1.000
Staten Island	12	5.5	3.82	5.5	20.9	0.87	1.000	1.000

* Pop. % = percentage of the city population represented by the outbreak area.
† RR = relative risk.
§ Expected/day = expected number of patients/day in the outbreak area under the null and alternative models, respectively.

TABLE 2. Estimated power of the prospective space-time scan statistic for 19 different outbreak models with medium excess risk, at different days of the outbreak

Outbreak area	No. of zip codes	Pop. %*	RR†	Expected/day§		Power on day		
				Null model	Alternative model	31	32	33
A. Williamsburg, Brooklyn	1	1.1	5.66	1.1	6.2	0.35	0.74	0.92
B. Roosevelt Island, Manhattan	1	0.1	24.19	0.1	2.4	0.37	0.73	0.93
C. Bulls Head, Staten Island	1	1.1	5.65	1.1	6.2	0.34	0.74	0.93
D. LaGuardia, Queens	1	0.5	9.42	0.5	4.7	0.32	0.67	0.91
E. West Farms, Bronx	1	0.7	7.36	0.7	5.1	0.29	0.72	0.90
A with 4 neighboring zip codes	5	4.0	3.06	4.0	12.2	0.42	0.79	0.94
B with 5 neighboring zip codes	6	3.2	3.33	3.2	10.6	0.40	0.77	0.95
C with 4 neighboring zip codes	5	3.3	3.29	3.3	10.8	0.33	0.77	0.94
D with 9 neighboring zip codes	10	8.2	2.39	8.2	19.5	0.42	0.85	0.97
E with 4 neighboring zip codes	5	3.7	3.13	3.7	11.6	0.43	0.79	0.96
Rockaways	5	1.3	5.01	1.3	6.5	0.34	0.76	0.91
Hudson River	20	10.3	2.24	10.3	23.0	0.33	0.65	0.82
Bronx	25	16.6	2.00	16.6	33.0	0.62	0.94	0.99
Brooklyn	37	30.8	1.82	30.8	55.6	0.79	0.98	0.999
Manhattan	40	19.0	1.95	19.0	36.8	0.57	0.90	0.99
Queens	62	28.0	1.84	28.0	51.1	0.73	0.97	0.998
Staten Island	12	5.5	2.71	5.5	14.9	0.43	0.81	0.97

* Pop. % = percentage of the city population represented by the outbreak area.
† RR = relative risk.
§ Expected/day = expected number of patients/day in the outbreak area under the null and alternative models, respectively.

TABLE 3. Powers of different analytic options for the prospective space-time scan statistic for selected outbreak models and at different days of the outbreak

Outbreak area	Maximum temporal size (days)	Maximum geographic size (days)	Analysis options: Include purely temporal	Analysis options: Adjustment for repeated analyses	Power on day 31	Power on day 32	Power on day 33
A: Williamsburg,	3	50	Yes	No	0.86	0.996	0.999
Brooklyn	3	50	No	No	0.86	0.996	0.999
1 zip code	3	5	No	No	0.86	0.995	0.999
(RR* = 9.91)	1	50	Yes	No	0.92	0.91	0.92
	7	50	Yes	No	0.86	0.996	0.999
	3	50	Yes	Yes	0.64	0.98	0.999
	3	N/A	Yes, only	No	0.19	0.30	0.42
A: Williamsburg,	3	50	Yes	No	0.85	0.996	1.000
Brooklyn	3	50	No	No	0.85	0.996	1.000
5 zip codes	3	5	No	No	0.86	0.99	1.000
(RR = 4.47)	1	50	Yes	No	0.90	0.91	0.89
	7	50	Yes	No	0.83	0.995	1.000
	3	50	Yes	Yes	0.64	0.98	0.999
	3	N/A	Yes, only	No	0.29	0.50	0.63
Manhattan	3	50	Yes	No	0.92	1.000	1.000
(RR = 2.47)	3	50	No	No	0.92	1.000	1.000
	3	5	No	No	0.77	0.98	1.000
	1	50	Yes	No	0.96	0.94	0.94
	7	50	Yes	No	0.90	0.999	1.000
	3	50	Yes	Yes	0.71	0.99	1.000
	3	N/A	Yes, only	No	0.75	0.96	0.99
Whole city	3	50	Yes	No	0.86	0.99	1.000
(RR = 1.5)	3	50	No	No	0.84	0.99	1.000
	3	5	No	No	0.40	0.69	0.81
	1	50	Yes	No	0.91	0.88	0.86
	7	50	Yes	No	0.81	0.99	1.000
	3	50	Yes	Yes	0.56	0.95	0.998
	3	N/A	Yes, only	No	0.996	1.000	1.000
No outbreak	3	50	Yes	No	0.05	0.06	0.05
(RR = 1.0)	3	50	No	No	0.05	0.06	0.05
	3	5	No	No	0.05	0.06	0.06
	1	50	Yes	No	0.05	0.06	0.05
	7	50	Yes	No	0.05	0.06	0.05
	3	50	Yes	Yes	0.002	0.0001	0.004
	3	N/A	Yes, only	No	0.05	0.06	0.04

*RR = relative risk.

Discussion

One goal of developing the benchmark data sets was to enable quick and simple comparison of new early detection methods with methods proposed previously. Inventors of new methods will hopefully make use of this opportunity. Pending evaluation of emerging methods, different parameter options of the prospective space-time scan statistic have been evaluated.

An important consideration when using the prospective space-time scan statistic is whether to include purely temporal cluster windows for detection of citywide outbreaks. Including this option increases the power for a citywide outbreak only marginally while minimally decreasing the power for the localized outbreak models (Table 1). In the majority of situations, purely temporal clusters should be included as an analysis option. For the same reason, using 50% as the upper limit in cluster size minimizes assumptions about the geographic cluster size.

The choice of maximum temporal-window size is less clear. Making the temporal window too small can substantially reduce the power to detect slowly emerging disease outbreaks. At the same time, these methods are meant for the rapid detection of disease outbreaks, and, depending on the disease, late detection of an outbreak might not provide any public health benefit. Compromise is needed and should be determined by the nature of the surveillance setting.

For the majority of these power evaluations, no adjustment was made for repeated analyzes performed daily. If such an adjustment were made, instead of keeping the false-alert rate at 5% for any given day (one false alert every 20 days), it could be set so that under the null model only one false alert/year (or per any other period) would be expected. As a result, the power would automatically decrease (Table 3). This decrease in power is attributable not to the method's strengths or weaknesses but to the ever-present trade-off between power and the number of false alerts. All power comparisons must use identical false-detection rates to be valid.

This study is subject to at least three limitations. First, the alternative outbreak models used for the benchmark data sets represent only a subset of the potential geographic and temporal features of actual disease outbreaks. As such, this study is a first step in creating different outbreak models for evaluating and comparing the statistical power of different outbreak-detection methods. For example, rather than a sudden increase in relative risk followed by a constant excess risk level in the outbreak area, one could construct outbreak models in which the relative risk increased gradually. Moreover, rather than simulating outbreaks that are geographically static

in time, an outbreak might first be spatially limited and then expand to neighboring zip codes, or it might start in one place and then gradually move to other areas of the city.

Second, this study examined data only for New York City. Simulated benchmark data sets for methods evaluation should be based on real geographic areas with realistic numbers for the underlying population at risk; NYC was selected for this study because it is where the investigators conduct outbreak surveillance. However, effective surveillance methods should work for different geographic areas and for different distributions of the population at risk. Evaluating outbreak-detection methods for geographic areas other than NYC would be valuable.

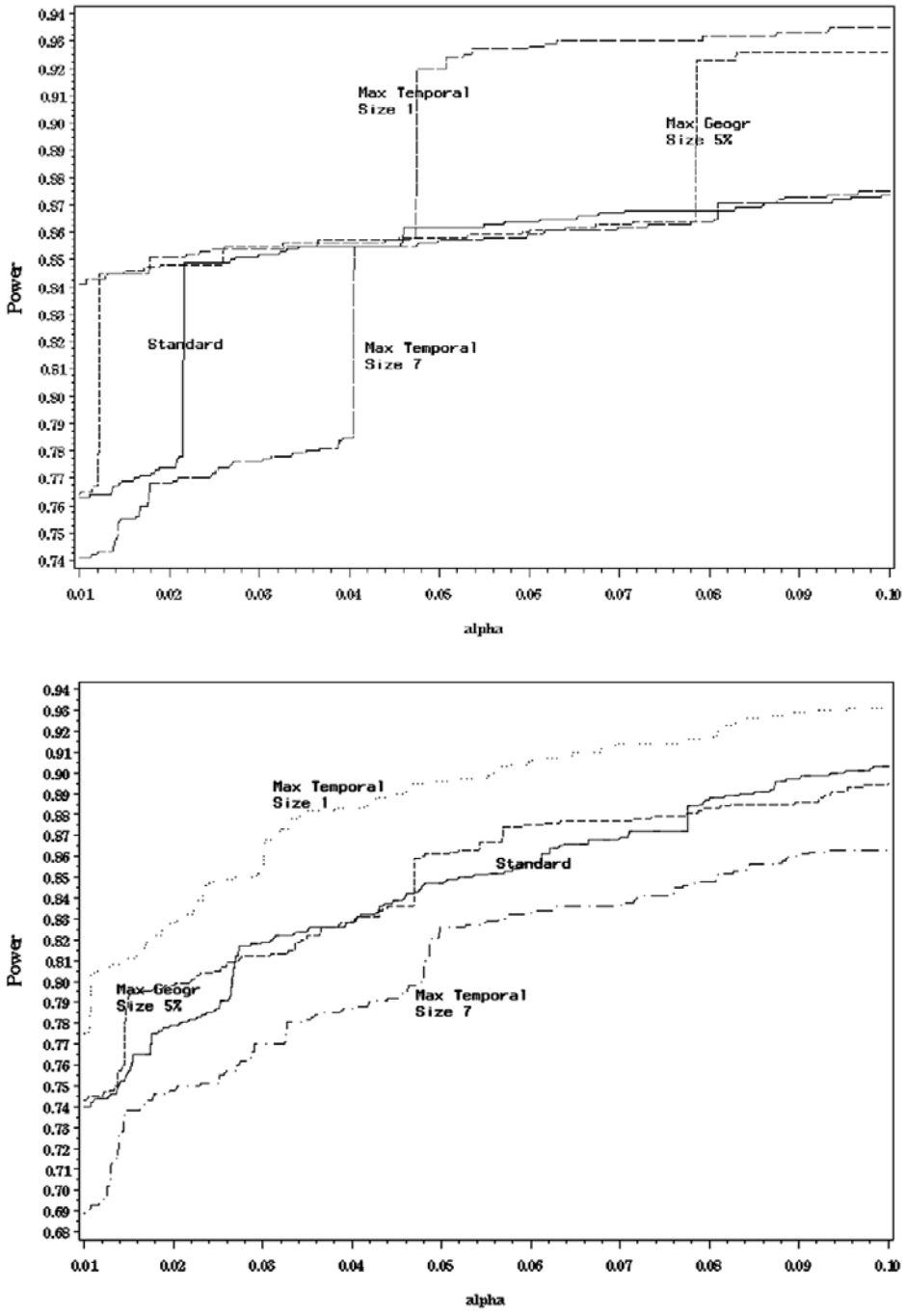
Finally, although these power estimates do capture the timeliness of a signal, they do not reflect its spatial accuracy. Only rarely will detected and true clusters coincide 100%, but the overlap might be better or worse for different methods.

Conclusions

The prospective space-time scan statistic performed well for all alternative models considered. Power was lowest for the Hudson River outbreak but remained surprisingly good considering that a circular window was used to detect a long and narrow cluster.

The low power of the purely temporal method to detect localized outbreaks provides a strong argument for using space-time surveillance methods for early outbreak detection, if the outbreak is expected to be localized. However, the purely temporal scan statistic performs substantially better at detecting a citywide outbreak, even when compared with a space-time method that includes the purely temporal outbreak as one parameter option. This is because less multiple testing needs adjustment when the multiple circles used to define localized outbreaks are not used, and an unavoidable trade-off exists between maximizing the power to detect localized versus citywide outbreaks.

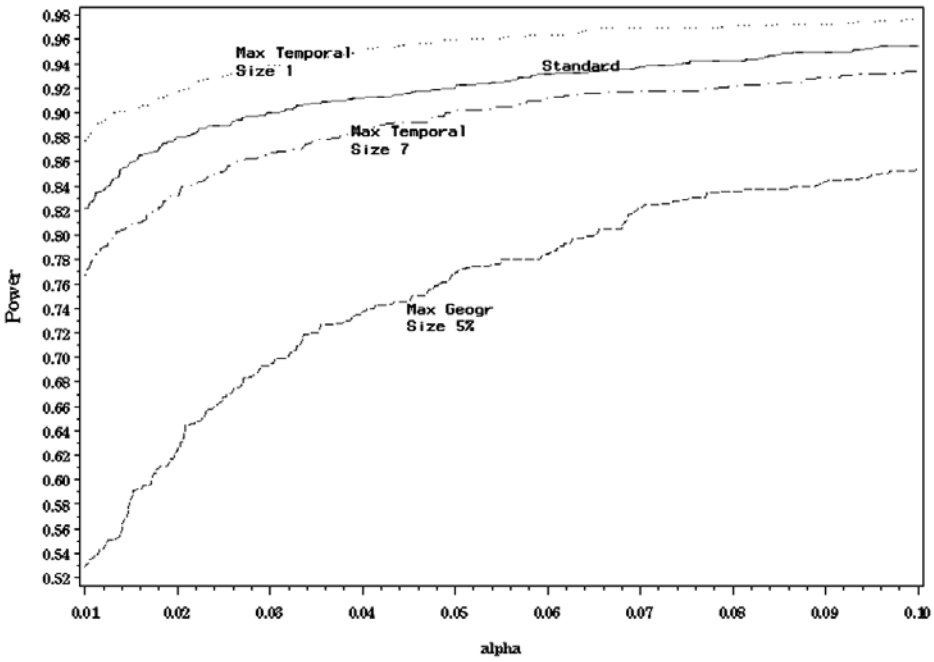
FIGURE 2. Power as a function of the false-detection rate (alpha), at day 31, for three different disease-outbreak models (A [top], A plus 4 neighbors [middle], and Manhattan [bottom])



Acknowledgments

This research was funded by the Alfred P. Sloan Foundation. The authors thank two anonymous reviewers for their valuable comments.

FIGURE 2. (*Continued*) Power as a function of the false-detection rate (alpha), at day 31, for three different disease-outbreak models (A [top], A plus 4 neighbors [middle], and Manhattan [bottom])



References

1. Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proc Natl Acad Sci U S A 2002;99:5237–40.
2. Hashimoto S, Murakami Y, Taniguchi K, Nagai M. Detection of epidemics in their early stage through infectious disease surveillance. Int J Epidemiol 2000;29:905–10.
3. 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.
4. Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection, New York City [Abstract]. MMWR 2004;53(Suppl):230.
5. Das D, Mostashari F, Weiss D, Balter S, Heffernan R. Monitoring over-the-counter pharmacy sales for early outbreak detection in New York City [Abstract]. MMWR 2004;53(Suppl):235.
6. Greenko J, Mostashari F, Fine A, Layton M. Clinical evaluation of the Emergency Medical Services (EMS) ambulance dispatch-based syndromic surveillance system, New York City. J Urban Health 2003;80(2 Suppl 1):i50–6.
7. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice: the New York City emergency department system. Emerg Infect Dis 2004;10:858–64.
8. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clustering: a potential early warning system for West Nile virus activity. Emerg Infect Dis 2003;9:641–6.
9. Lewis M, Pavlin J, Mansfield J, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC area. Am J Prev Med 2002;23:180–6.
10. 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.
11. Yih WK, Caldwell B, Kleinman K, et al. National Bioterrorism Syndromic Surveillance Demonstration Program. MMWR 2004;53(Suppl):43–6.
12. Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. J R Stat Soc A 2003; A166:5–21.
13. Järpe E. Surveillance of the interaction parameter of the Ising model. Communications in Statistics: Theory and Methods 1999;28:3009–27.
14. Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic. J R Stat Soc A 2001;164:61–72.
15. Rogerson PA. Surveillance systems for monitoring the development of spatial patterns. Stat Med 1997;16:2081–93.
16. Rogerson PA. Monitoring point patterns for the development of space-time clusters. J R Stat Soc A 2001;164:87–96.
17. Theophilides CN, Ahearn SC, Grady S, Merlino M. Identifying West Nile virus risk areas: the dynamic continuous-area space-time system. Am J Epidemiol 2003;157:843–54.
18. Kulldorff M, Tango T, Park P. Power comparisons for disease clustering tests. Computational Statistics and Data Analysis 2003;42:665–84.
19. Dwass M. Modified randomization tests for nonparametric hypotheses. Annals of Mathematical Statistics 1957;28:181–7.
20. Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods 1997;26:1481–96.
21. Kulldorff M, Information Management Services, Inc. SaTScan™ version 3.1: software for the spatial and space-time scan statistics, 2003. Available at http://www.satscan.org.

Bio-ALIRT Biosurveillance Detection Algorithm Evaluation

David Siegrist,[1] J. Pavlin[2]*
[1]Potomac Institute for Policy Studies, Arlington, Virginia; [2]Walter Reed Army Institute of Research, Silver Spring, Maryland

Corresponding author: David Siegrist, Potomac Institute for Policy Studies, 901 N. Stuart Street, Suite 200, Arlington, Virginia 22203. Telephone: 703-525-0770; Fax: 703-562-1000; E-mail: dsiegrist@potomacinstitute.org.

Abstract

Introduction: *Early detection of disease outbreaks by a medical biosurveillance system relies on two major components: 1) the contribution of early and reliable data sources and 2) the sensitivity, specificity, and timeliness of biosurveillance detection algorithms. This paper describes an effort to assess leading detection algorithms by arranging a common challenge problem and providing a common data set.*

Objectives: *The objectives of this study were to determine whether automated detection algorithms can reliably and quickly identify the onset of natural disease outbreaks that are surrogates for possible terrorist pathogen releases, and do so at acceptable false-alert rates (e.g., once every 2–6 weeks).*

Methods: *Historic de-identified data were obtained from five metropolitan areas over 23 months; these data included* International Classification of Diseases, Ninth Revision *(ICD-9) codes related to respiratory and gastrointestinal illness syndromes. An outbreak detection group identified and labeled two natural disease outbreaks in these data and provided them to analysts for training of detection algorithms. All outbreaks in the remaining test data were identified but not revealed to the detection groups until after their analyses. The algorithms established a probability of outbreak for each day's counts. The probability of outbreak was assessed as an "actual" alert for different false-alert rates.*

Results: *The best algorithms were able to detect all of the outbreaks at false-alert rates of one every 2–6 weeks. They were often able to detect for the same day human investigators had identified as the true start of the outbreak.*

Conclusions: *Because minimal data exists for an actual biologic attack, determining how quickly an algorithm might detect such an attack is difficult. However, application of these algorithms in combination with other data-analysis methods to historic outbreak data indicates that biosurveillance techniques for analyzing syndrome counts can rapidly detect seasonal respiratory and gastrointestinal illness outbreaks. Further research is needed to assess the value of electronic data sources for predictive detection. In addition, simulations need to be developed and implemented to better characterize the size and type of biologic attack that can be detected by current methods by challenging them under different projected operational conditions.*

Introduction

The Bio-Event Advanced Leading Indicator Recognition Technology (Bio-ALIRT) biosurveillance program was implemented during 2001–2004. The program's objective was to develop data sources, technologies, and prototypes for monitoring nontraditional data sources (e.g., animal sentinels, human behavioral indicators, and nondiagnostic medical data) that might enable public health authorities to detect terrorist release of a pathogen or toxin at the earliest possible moment. Technical challenges to the development of Bio-ALIRT have included 1) determining the value of each data source, alone and in combination with others, for earlier outbreak detection; 2) correlating and integrating information derived from heterogeneous data sources; 3) developing autonomous signal-detection algorithms with high sensitivity and low false alerts; and 4) maintaining privacy protection while correlating de-identified data sources.

Early detection of disease can be divided into two components: contributions made by the data, and contributions made by anomaly-detection algorithms. Bio-ALIRT investigators evaluated multiple data sources in comparison with standard data that indicated when an outbreak of influenza-like illness (ILI) or gastrointestinal illness (GI) actually occurred (as documented by de-identified insurance claims). The lead-time over those reference data and the confidence interval can then be calculated. (Additional information about the Bio-ALIRT data research is available from the corresponding author.)

This paper focuses on the evaluation of the detection algorithms as a component of a biosurveillance system. A common challenge problem and common data set are required to

* The opinions expressed in this paper are those of the authors and do not necessarily reflect the position of the U.S. Department of Defense.

evaluate detection algorithms; for the first Bio-ALIRT algorithm evaluation in August–September 2002, this was accomplished by using the BioWar simulation, which uses a software agent-based approach to simulate both a normal background and an outbreak signal (*1*). In 2003, to determine whether the algorithms could detect real disease outbreaks, the investigators used wholly authentic, de-identified, historic military and civilian data from five cities. An advantage of the evaluation approach used in 2003 is that it relied exclusively on real data and not on simulation, which might inadvertently introduce bias into the assessment. Also, by working with real data from cities of interest, the evaluators were able to hone their skills in a realistic environment that might also produce insights to further program goals. Limitations of the historic outbreak evaluation approach include uncertainty about the exact start dates and sizes of outbreaks and the inability to examine algorithm outbreak-detection capabilities under a substantial number and variety of conditions. Furthermore, pathogens likely to be used in a terrorist attack are presumed to have a different epidemiologic curve than an ILI outbreak. However, detecting slowly increasing seasonal respiratory outbreaks and more rapidly rising GI outbreaks across a metropolitan region were considered to be reasonable surrogates for detecting deliberate pathogen releases.

Bio-ALIRT was sponsored by the Defense Advanced Research Projects Agency (DARPA), a central U.S. Department of Defense (DoD) research and development agency, primarily to protect troops from biologic agents. Contract investigators included the Johns Hopkins University Applied Physics Laboratory in cooperation with the Walter Reed Army Institute of Research; the University of Pittsburgh/Carnegie Mellon University team; the General Dynamics Advanced Information Systems (formerly Veridian) team with the Stanford University Medical Informatics group; and IBM Corporation. The Potomac Institute performed an independent evaluation function. Both CDC and a municipal department of health also participated in the detection evaluation.

Methods

Data Sources

Authentic military and civilian data from five cities were analyzed. ILI and GI were used as surrogates for a biologic attack because these syndromes might mimic early symptoms of certain Class A pathogens on CDC's biologic terrorism threat list.[†] Naturally occurring historic outbreaks of ILI and GI were identified by using measurable phenomena (e.g., visits for symptomatic care to a health-care provider) that generated records (e.g., insurance claims from physicians' offices or hospital outpatient care) from which identifying information was removed.

Three data sources were obtained for the evaluation: military outpatient-visit records with *International Classification of Diseases, Ninth Revision* (ICD-9) codes, civilian ICD-9-coded outpatient visit records, and military outpatient prescription records. All data were stripped of identifying information to protect patient privacy. After geographic regions with overlap between available military and civilian populations were determined, five areas were selected for investigation: Norfolk, Virginia; Pensacola, Florida; Charleston, South Carolina; Seattle, Washington; and Louisville, Kentucky. DoD military treatment facility (MTF) coverage was approximately 100% for those five areas, whereas civilian coverage for the regions ranged from 15.9% to 32.7%, with a mean of 25.1%. All three data streams generated signals for the same disease outbreaks for approximately the same dates (Table 1), which increased the investigators' confidence in the overall quality of the data set.

The military data included ICD-9 codes from all MTF outpatient visits by active duty personnel, retirees, and dependent family members. These data included date of visit, ≤4 ICD-9 codes per visit, age, residential zip code, and MTF designator. Military pharmacy data captured all prescriptions paid for by the military health-care system and filled at either MTFs or civilian pharmacies. The evaluation data set included the pharmacy identification (ID) number; the date the prescription was written and filled; the drug name, generic drug classification, and therapeutic class identifier; whether the prescription was new or a refill; the number of refills; and the patient's age. Surveillance Data, Inc. (SDI) provided de-identified ICD-9 outpatient data from similar geographic regions, including the date of visit, ≤5 of the selected ICD-9 codes, and the patient's age and residential zip code.

Military outpatient ICD-9 information was captured electronically shortly after the outpatient visit. ICD-9 codes are added to the electronic record either by the provider or by a professional coder and sent with demographic and clinic information to a central repository. Pharmacy data were collected electronically at the time the prescription was filled. Over-the-counter drug prescriptions (e.g., decongestants and antidiarrheals) at MTFs were also included in the data.

All identifying information was removed from military outpatient and pharmacy data before their provision to the Walter Reed Army Institute of Research (WRAIR) and the teams. SDI data were generated from electronically transmitted

[†] Approximately 250 *International Classification of Diseases, Ninth Revision* (ICD-9) codes are closely associated with ILI and GI illness.

TABLE 1. Start date, date of estimated public health recognition, peak date, and end date of respiratory and gastrointestinal illness outbreaks in one metropolitan area, by data provider — Bio-ALIRT Biosurveillance Detection Algorithm Evaluation, 2003

Data provider	Start date of outbreak	Date of estimated public health recognition	Peak date of outbreak	End date of outbreak
Respiratory illness, February 10, 2003–April 29, 2003				
Ambulatory Data System (ADS)	Feb. 10	Feb. 24	Mar. 10	Apr. 21
Pharmacy Transaction Data Service (PDTS)	Feb. 10	Feb. 24	Mar. 10	Apr. 28
Surveillance Data, Inc. (SDI)	Feb. 10	Feb. 24	Mar. 10	Apr. 28
Final	**Feb. 10**	**Feb. 24**	**Mar. 10**	**Apr. 28**
Respiratory illness, September 16, 2002–April 28, 2003				
ADS	Sept. 16	Sept. 23	Mar. 10	Apr. 21
PDTS	Sept. 16	Sept. 18	Mar. 10	Apr. 28
SDI	Sept. 23	Sept. 30	Mar. 10	Apr. 28
Final	**Sept. 16**	**Sept. 23**	**Mar. 10**	**Apr. 28**
Gastrointestinal illness, October 21, 2002–February 10, 2003				
ADS	Oct. 21	Nov. 18	Jan. 6	Jan. 29
PDTS	Oct. 22	Nov. 25	Jan. 6	Jan. 29
SDI	Nov. 12	Dec. 10	Jan. 29	Feb. 10
Final	**Oct. 21**	**Nov. 25**	**Jan. 29**	**Feb. 10**
Gastrointestinal illness, February 24, 2002–March 13, 2003				
ADS	Feb. 25	Mar. 10	Mar. 10	Mar. 13
PDTS	Feb. 24	Mar. 3	Mar. 3	Mar. 12
SDI	N/A	N/A	N/A	N/A
Final	**Feb. 24**	**Mar. 10**	**Mar. 10**	**Mar. 13**

insurance claims for physician office services from a substantial sample of physicians across the United States. As claims were sent from the physicians to the insurers, identifying information was removed pursuant to the Health Insurance Portability and Accountability Act of 1996 (HIPAA), and data were transmitted to SDI and loaded into a data warehouse.

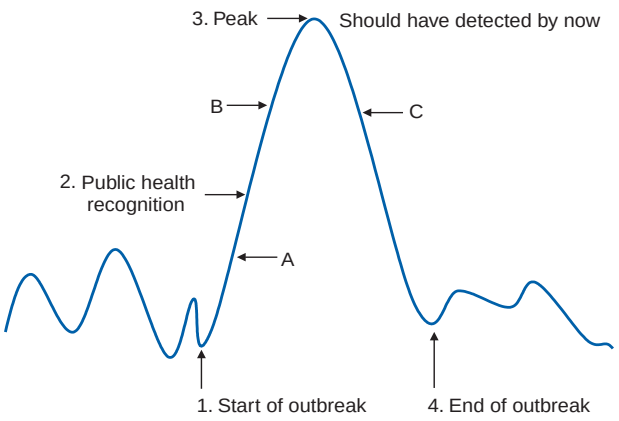
WRAIR uses military outpatient ICD-9 codes for an active disease surveillance system known as the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE). On the basis of previous experience and in conjunction with CDC, groups of ICD-9 codes and medications that best reflect respiratory and gastrointestinal illness were used for the evaluation. (A list of these ICD-9 codes is available from the corresponding author.) Only these drug categories and ICD-9 codes were provided to participants. However, participants were allowed to manipulate the syndrome categories and to delete or subgroup codes to improve their analysis. Data for July 2001–August 2002 were provided for training of the algorithms. The test data stream ran during September 2002–May 2003.

Outbreak Determination

An outbreak detection group (ODG) was formed to determine when natural outbreaks of gastrointestinal and respiratory illness took place in the selected areas and times. This group included medical specialists and epidemiologists from throughout the program; after joining the group, they were sequestered from participating in the detection portion of the evaluation. Using visual and statistical techniques, ODG found evidence of disease outbreaks in the data and also determined that the three data streams correlated effectively (Table 1). For convenience, a simple anomaly-detector algorithm was run over the data to assist in identification of outbreaks. Four dates were then determined for each of the agreed outbreaks: 1) start date, 2) date ODG expected that public health officials would declare prospectively that an outbreak was occurring, 3) peak date, and 4) end of outbreak (Figure). ODG included both broad, seasonal outbreaks and more concise disease-count elevations that occurred both inside and outside of seasonal fluctuations. Because the data were retrospective, outbreaks could not be confirmed in the majority of cases. However, because the algorithms being evaluated are intended to alert public health authorities to the likelihood of an outbreak, the presumptive standard was considered reasonable for the evaluation of the detection algorithms.

The data were divided into 14 months of training data and 9 months of test data. Two outbreaks were identified in the training data, and the dates of these outbreaks were provided to the teams. The dates of all outbreaks identified in the test data were withheld from the teams until after they submitted their detection results.

FIGURE. Detection of outbreaks on the epicurve



A Optimal time for detection — optimally, closer to the start
B Not a false alert if detection occurs at point B
C False alert — outbreak is still occurring, but counts are decreasing

Assessing Algorithm Performance

The sensitivity and timeliness of each outbreak-detection algorithm were assessed at false-positive rates of practical relevance for public health surveillance. Teams submitted detection results for ≤3 algorithms. The detection results for each algorithm consisted of two files, one for respiratory outbreaks and one for GI outbreaks. Each row in a file contained a date followed by five numbers, one for each city. The numbers were the algorithm output indicating the likelihood of an outbreak in a given city on a given day. For the majority of algorithms, the numbers were p-values, but the assessment method did not require this.

The three false-positive rates selected were one per 2 weeks, one per 4 weeks, and one per 6 weeks. Sensitivity and timeliness for respiratory and GI outbreak detection were calculated separately at each false-positive rate. This resulted in six estimates of sensitivity and timeliness for each algorithm. A false-positive rate for an algorithm corresponds to a threshold applied to the algorithm's numerical output. This threshold was determined separately for each type of outbreak (i.e., respiratory and GI) and for each city by examining the number of false alerts during nonoutbreak periods at each threshold. The numerator for sensitivity was defined as the number of outbreaks with ≥1 algorithm output over the threshold between the start date of the outbreak and the date public health authorities were expected to recognize the outbreak; the denominator for sensitivity was the number of outbreaks. The dates for respiratory and GI outbreaks in the five metropolitan areas are presented (Table 2).

Timeliness of outbreak detection was measured by using a variation of the activity monitor operating characteristic (AMOC) method (*2*). In practice, this entailed calculating the median time to outbreak detection for an algorithm at each false-positive rate. Median time to detection was used because mean time is problematic; outbreaks have different lengths, and no obvious way exists to penalize evaluation participants for an undetected outbreak. For example, if missed outbreaks are ignored, then a method that alerts late can have a larger mean than a method that does not alert at all. For a single outbreak, time to detection was defined as the number of days between the outbreak start date and the date the algorithm output first crossed the threshold. If the algorithm did not identify the outbreak before the date public health authorities were expected to recognize the outbreak, then an

TABLE 2. Start date, date of estimated public health recognition, peak date, and end date of respiratory and gastrointestinal illness (GI) outbreaks in five metropolitan areas — Bio-ALIRT Biosurveillance Detection Algorithm Evaluation, 2003

Outbreak type	Start date of outbreak	Date of estimated public health recognition	Peak date of outbreak	End date of outbreak	Detection difference (days)
Metropolitan area A					
Respiratory	Feb. 10, 2003	Feb. 24, 2003	Mar. 10, 2003	Apr. 28, 2003	14
Respiratory	Sept. 16, 2002	Sept. 23, 2002	Mar. 10, 2003	Apr. 28, 2003	7
GI	Oct. 21, 2002	Nov. 25, 2002	Jan. 29, 2003	Feb. 10, 2003	35
GI	Feb. 24, 2003	Mar. 10, 2003	Mar. 10, 2003	Mar. 13, 2003	14
Metropolitan area B					
Respiratory	Jan. 22, 2003	Feb. 18, 2003	Mar. 4, 2003	Mar. 31, 2003	27
GI	Feb. 16, 2003	Feb. 16, 2003	Feb. 17, 2003	Feb. 18, 2003	0
Metropolitan area C					
Respiratory	Jan. 27, 2003	Feb. 3, 2003	Feb. 10, 2003	Mar. 19, 2003	7
Respiratory	Oct. 17, 2002	Nov. 12, 2002	Dec. 9, 2002	Dec. 20, 2002	26
GI	Dec. 6, 2002	Dec. 18, 2002	Dec. 17, 2002	Jan. 28, 2003	12
Metropolitan area D					
Respiratory	Nov. 4, 2002	Dec. 10, 2002	Feb. 3, 2003	Apr. 14, 2003	37
Metropolitan area E					
Respiratory	Oct. 28, 2002	Nov. 18, 2002	Dec. 9, 2002	Feb. 3, 2003	21
Respiratory	Feb. 3, 2003	Feb. 18, 2003	Feb. 24, 2003	Apr. 15, 2003	15
GI	Nov. 14, 2002	Dec. 9, 2002	Jan. 29, 2003	Feb. 11, 2003	26
GI	Nov. 11, 2002	Dec. 9, 2002	Feb. 24, 2003	Apr. 16, 2003	29
GI	Feb. 22, 2003	Feb. 24, 2003	Feb. 24, 2003	Mar. 11, 2003	2

infinite time to detection for that outbreak was assigned to the algorithm. Assignment of an infinite time for a single outbreak does not unduly influence the calculation of the median the way it would influence calculation of the mean.

Charts for sensitivity and timeliness were calculated for each algorithm and used to compare performance of the various alerting algorithms on the 15 outbreaks identified by the ODG. The limited number of outbreaks precluded testing for statistically important differences in detection performance between algorithms.

Conducting the Test

Evaluation data were collected and distributed by WRAIR. Approximately 14 months of training data with two labeled outbreaks were released to the teams, including DoD ambulatory ICD-9 codes, DoD pharmacy data, and civilian medical-claims data for five cities. Unlabeled test data were released 6 weeks later. The teams were on the honor code to analyze the data prospectively (e.g., daily) as they were presented rather than identify peaks and then trace them back to their origins to determine the start of the outbreak. The processes used by the teams were asserted to be repeatable and thus verifiable. Two weeks after distribution of the test data, the algorithm-detection output was collected from participating teams, and software to score detection results was distributed to them. The software automatically computed sensitivity and timeliness. Desirable characteristics for the evaluation were high values for sensitivity (i.e., detecting that an outbreak occurred) and low values for timeliness (i.e., a slight delay in detecting the outbreak at the different false-alert rates).

Results

The best algorithms were able to detect all of the outbreaks, often for the same day the ODG had determined retrospectively that the outbreaks had begun, at a false-alert rate of one every 2 weeks (Tables 3 and 4). This study measured the number of days after the initial outbreak that the algorithms would detect the outbreak; therefore, detecting on day 1 is optimal. Compared with the human investigators, the algorithms detected the outbreaks "virtually prospectively." That is, the algorithms determined a probability of outbreak for a particular day as the date they were encountered, instead of when human investigators were projected to have detected the outbreak, leading to an average detection advantage of ≥18 days (Table 2). The detection advantage was more marked for seasonal respiratory outbreaks; the GI outbreaks peaked more rapidly and decisively. The leading detection algorithms included statistical process control methods applied to regression residuals, Bayesian change-point techniques, and wavelet methods. One of the analytic teams, instead of measuring raw syndrome counts, instead obtained good results by detecting variation in the total number of medical providers reporting and measuring the regression by using Hotelling's T^2 (*3*). A fuller description of the evaluation results and techniques will be forthcoming.

Conclusion

This paper has described a methodology and results for quantitatively evaluating the performance of outbreak detection algorithms used in biosurveillance. This methodology permits assessment of the performance of algorithms implemented by different research teams in detecting real outbreaks identified by expert opinion. Both timeliness and sensitivity were assessed at false-positive rates of practical relevance for public-health surveillance.

An advantage of the approach used is that it relied solely on actual data; no simulation was conducted that might inadvertently introduce bias into the assessment. Using real data from cities of interest enabled teams to hone their skills in a realistic environment that might also produce important insights that would further program goals. However, this approach has certain limitations, including uncertainty about the exact start date and size of outbreaks and inability to examine algorithm outbreak-detection capabilities under a substantial number of diverse conditions. In addition, the numbers of real outbreaks in the data set used in this evaluation were not sufficient to support statistical significance testing, which limited the precision of the results. Further, pathogens that would be used in a terrorist attack are presumed to have a somewhat different epidemiologic curve than a natural ILI outbreak, for instance. However, detecting slowly rising seasonal respiratory outbreaks, as well as more rapidly rising GI outbreaks, over a metropolitan region were considered to be reasonable surrogates for detecting deliberate pathogen releases.

The results of this analysis indicate that authentic historic data with real outbreaks can support evaluation across research teams by providing a common challenge problem and common data set. ODG members agreed on the number and dates of the outbreaks in all three parallel data streams for each of the five cities. The reliability of this agreement was not assessed quantitatively, but the general agreement indicates that the data were adequate to support the comparison. Epidemiologic investigators determined the dates of outbreaks on the basis of professional judgment. However, no further investigation was conducted to determine whether local public health authorities in these five metropolitan areas believed

TABLE 3. Performance of outbreak-detection algorithms at detecting respiratory illness — Bio-ALIRT Biosurveillance Detection Algorithm Evaluation, 2003

Team	False-alert rate	Best algorithm	Median timeliness (day)	Sensitivity: No. of outbreaks detected	Sensitivity: Total no. of outbreaks
RODS*	1 per 2 weeks	wav8ssmtwrf_sum†	1	8	8
	1 per 4 weeks	wav8ssmtwrf_sum	1	8	8
	1 per 6 weeks	wav8ssm_max§	1	7	8
ESSENCE¶	1 per 2 weeks	provReg/Hotel**	1	8	8
	1 per 4 weeks	EWMA†† C2	1	8	8
	1 per 6 weeks	EWMA C2	1	8	8
CDC	1 per 2 weeks	EARS§§ C3	3	7	8
	1 per 4 weeks		5.5	6	8
	1 per 6 weeks		18.5	5	8
General Dynamics	1 per 2 weeks	A	1.5	8	8
	1 per 4 weeks	A	2.5	8	8
	1 per 6 weeks	A	3.5	8	8
IBM	1 per 2 weeks	C	2.5	8	8
	1 per 4 weeks	C	2.5	8	8
	1 per 6 weeks	C	2.5	8	8

* Real-Time Outbreak Disease Surveillance.
† Wavelet algorithm to 8th power using all days of the week and summing results from all three data streams.
§ Wavelet algorithm to 8th level (2^8 days = 256 days); *ssm* refers to days of week modeled as their own time series (Saturday, Sunday, Monday); and *max* refers to reporting out the maximum standard deviation among the three individual data streams processed.
¶ Electronic Surveillance System for the Early Notification of Community-Based Epidemics.
** Provider count regression residuals, which are inputs to a multivariate Hotelling's T^2 algorithm.
†† Exponentially weighted moving average.
§§ Early Aberration Reporting System.

TABLE 4. Performance of outbreak detection algorithms at detecting gastrointestinal illness — Bio-ALIRT Biosurveillance Detection Algorithm Evaluation, 2003

Team	False-alert rate	Best algorithm	Median timeliness (day)	Sensitivity: No. of outbreaks detected	Sensitivity: Total no. of outbreaks
RODS*	1 per 2 weeks	wav8ssm_max†	1	7	7
	1 per 4 weeks	BCD§	1	6	7
	1 per 6 weeks	BCD	1	6	7
ESSENCE¶	1 per 2 weeks	provReg/Hotel**	1	6	7
	1 per 4 weeks	EWMA†† C2	1	6	7
	1 per 6 weeks	EWMA C2	1	6	7
CDC	1 per 2 weeks	Wavelet transform	1	6	7
	1 per 4 weeks	moving average	4	5	7
	1 per 6 weeks		9	5	7
General Dynamics	1 per 2 weeks	B, C	3	6	7
	1 per 4 weeks	B, C	3	6	7
	1 per 6 weeks	C	26	5	7
IBM	1 per 2 weeks	A	2	6	7
	1 per 4 weeks	A	3	6	7
	1 per 6 weeks	A	5	6	7

* Real-Time Outbreak Disease Surveillance.
† Wavelet algorithm to 8th level (2^8 days = 256 days); *ssm* refers to days of week modeled as their own time series (Saturday, Sunday, Monday); and *max* refers to reporting out the maximum standard deviation among the three individual data streams processed.
§ Biosurveillance using change-point detection.
¶ Electronic Surveillance System for the Early Notification of Community-Based Epidemics.
** Provider count regression residuals, which are inputs to a multivariate Hotelling's T^2 algorithm.
†† Exponentially weighted moving average.

that actual outbreaks took place at those times. Rather, the outbreak was determined on the basis of the fact that an unusual number of case counts were reported.

This evaluation provides a "snapshot" of the performance of certain algorithms and data-processing methods, in the hands of five teams, at detection of outbreaks identified by a panel of experts. Whether certain algorithms were better overall than others was not determined. The evaluation indicates that objective ways exist to compare critical aspects of biosurveillance systems by using authentic data from real outbreaks.

Acknowledgments

The following persons provided exceptional help in the conduct of the evaluations: Virginia Foster, WRAIR; David Buckeridge, Stanford University; Andrew Kress, Surveillance Data, Incorporated; Ted Senator, DARPA; Judith Brillman, University of New Mexico; Howard Burkom, Johns Hopkins University Applied Physics Laboratory; Murray Campbell, IBM; Karen Cheng, General Dynamics; Kathleen Carley, Carnegie Mellon University (CMU); Yevgeniy Elbert, WRAIR; Jay Mansfield, WRAIR; Andrew Moore, CMU; Doug Stetson, General Dynamics; Michael Wagner, University of Pittsburgh. The following persons provided helpful comments regarding earlier drafts: Howard Burkom; David Buckeridge; Virginia Foster; Bill Hogan, University of Pittsburgh; Henry Rolka, CDC.

References

1. Carley K, Fridsma D, Casman E, et al. BioWar: scalable multi-agent social and epidemiological simulation of bioterrorism events. Pittsburgh, PA: Center for Computational Analysis of Social and Organizational Systems, 2003. Available at http://www.casos.ece.cmu.edu/casos_working_paper/carley_2003_biowar.pdf.
2. Fawcett T, Provost F. Activity monitoring: noticing interesting changes in behavior. In: Proceedings of the Fifth ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY: ACM Press, 1999;53–62.
3. Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. J Biomed Inform (in press).

ESSENCE II and the Framework for Evaluating Syndromic Surveillance Systems*

Joseph S. Lombardo,[1] H. Burkom,[1] J. Pavlin[2]
[1]Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland; [2]Walter Reed Army Institute of Research, Silver Spring, Maryland

Corresponding author: Joseph S. Lombardo, Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Rd, Laurel, MD 20723. Telephone: 443-778-6287; Fax: 443-778-5026; E-mail: Joe.Lombardo@jhuapl.edu.

Abstract

Introduction: *The Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II) is a prototype syndromic surveillance system for capturing and analyzing public health indicators for early detection of disease outbreaks.*

Objectives: *This paper presents a preliminary evaluation of ESSENCE II according to a CDC framework for evaluating syndromic surveillance systems.*

Methods: *Each major topic of the framework is addressed in this assessment of ESSENCE II performance.*

Results: *ESSENCE captures data in multiple formats, parses text strings into syndrome groupings, and applies multiple temporal and spatio-temporal outbreak-detection algorithms. During a recent DARPA evaluation exercise, ESSENCE algorithms detected a set of health events with a median delay of 1 day after the earliest possible detection opportunity.*

Conclusions: *ESSENCE II has provided excellent performance with respect to the framework and has proven to be a useful and cost-effective approach for providing early detection of health events.*

Introduction

In response to the threat of biologic terrorism and the resurgence of virulent forms of infectious diseases, technologic advances are being applied to disease surveillance. Syndromic surveillance systems have emerged to capture and analyze health-indicator data to identify abnormal health conditions and enable early detection of outbreaks. Given the limited public health experience with biologic terrorism and the variety of possible terrorism scenarios, the research community is exploring the application of advanced detection technology to prediagnostic syndromic data. In 2003, CDC issued a draft framework for evaluating syndromic surveillance systems (*1*), which was later revised and published in *MMWR* (*2*). The CDC framework is designed for evaluation of relatively mature, fully operational syndromic surveillance systems. The technology to support syndromic surveillance is just maturing, with current operational experience gained from test-bed use. This paper applies the framework to the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE), a series of prototype systems developed by Johns Hopkins University Applied Physics Laboratory (JHU/APL) and the Division of Preventive Medicine at the Walter Reed Army Institute of Research.

* This research is sponsored by the Defense Advanced Research Projects Agency (DARPA) and managed under Naval Sea Systems Command (NAVSEA) contract N00024-98-D-8124. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of DARPA, NAVSEA, or the U.S. Government.

System Description

Purpose

Multiple versions of ESSENCE have been developed, each for different purposes. ESSENCE I provides worldwide surveillance for military personnel and their dependents at all military treatment facilities by using ambulatory records generated for TriCare, the military's health-care system. ESSENCE II is a regional system that supports advanced surveillance within the National Capital Region (NCR) test bed. The system is being developed by JHU/APL in collaboration with the Maryland Department of Health and Mental Hygiene, the District of Columbia Department of Health, and the Virginia Department of Health. Other versions of ESSENCE have been developed for military facilities and deployed forces. This description focuses on ESSENCE II only.

ESSENCE II is a test-bed system for 1) evaluating nontraditional health-care indicators, 2) developing and evaluating analytic techniques for early identification of abnormal disease patterns, and 3) providing an integrated view of NCR

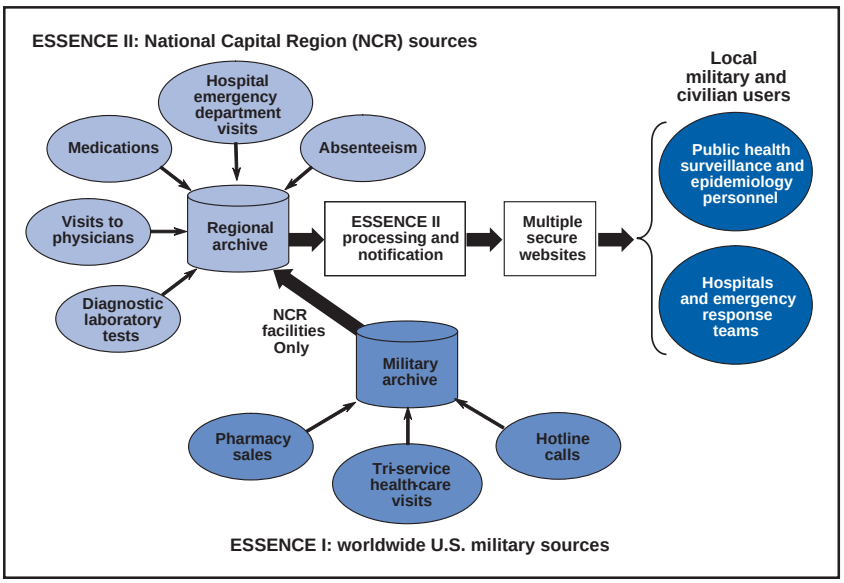
military and civilian health department data (*3*) (Figure 1). The system captures data on military ambulatory visits and prescription medications and merges them with civilian emergency department (ED) chief-complaint records, school-absenteeism data, over-the-counter (OTC) and prescription medication sales, civilian ambulatory visits, veterinary health records, and health department requests for influenza testing. All data are de-identified by their providers before being transferred to ESSENCE II, where they are archived, analyzed, and provided through secure Internet sites to local health departments and to hospitals that have data-sharing agreements with their health departments.

Stakeholders

NCR health departments conduct surveillance by using ED chief-complaint data from hospitals within and around the District of Columbia metropolitan area. ESSENCE II helps automate the processes of capturing hospital data, parsing chief-complaint text strings, and analyzing data for abnormalities.

ESSENCE technology is being used to form a regional collaborative disease-surveillance network. The network consists of four major nodes, one at each state and District of Columbia health department and a regional node for performing analysis across jurisdictional boundaries. The architecture permits fully identifiable information to be captured and archived at health departments for patients within their jurisdiction. The regional node negotiates the acquisition and distribution of data (e.g., military health-care data and OTC medication sales) across the region. The architecture also permits de-identification, aggregation, and sharing of information among the region's health departments while increasing the sensitivity for detection of abnormal health events occurring across jurisdictional boundaries.

FIGURE 1. Data sources for the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)



Operation

The data flow through an ESSENCE II node is illustrated (Figures 2 and 3). First, to expedite data collection and maintain confidentiality, the data providers create automated query software to extract recent data elements from their archives. These extractions are assembled into a de-identified update record, encrypted, and posted to a secure file transfer protocol (FTP) site. The query software automatically executes at a regular interval (e.g., daily at midnight or once every 8 hours) that can be changed easily. Although ESSENCE II can accept Health Level 7 (HL7) (*4*) data streams, the majority of data providers prefer the automated query approach. ESSENCE II polls the FTP sites to look for new entries, which are then ingested, cleaned, formatted, and archived in the primary system archive.

Data-sharing policies across the region have not been approved by all NCR health departments. After these policies are approved, selected data fields or aggregates of counts will be transmitted to other nodes in the network.

Chief-complaint data from hospital EDs 1) are received as text strings, which are of variable length; 2) include punctuation, misspellings, or abbreviations; and 3) can use varying syntax and vocabularies. A chief-complaint parsing algorithm developed for ESSENCE II converts text strings into syndrome groupings (*5*). The syndrome groupings agreed to by the NCR health departments are *death*, *gastrointestinal*, *neurologic*, *rash*, *respiratory*, *sepsis*, *unspecified*, and *other,* but the chief-complaint parsing algorithm can easily accommodate modifications. After ED data are entered into the primary archive, the parsing algorithm automatically converts the text strings into syndrome groupings. When the parser's performance is compared with that of human coders, the parser provides, on average, 97% sensitivity and 99% specificity. Whenever new hospital EDs are added to the system, the parser's performance is assessed to adjust for unfamiliar textual information. The algorithm provides approximately perfect conversion into syndrome groupings for the most prevalent syndromes (*respiratory* and *gastrointestinal*) and degraded performance for those less frequent (*neurologic*).

FIGURE 2. Data-acquisition flow for the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)

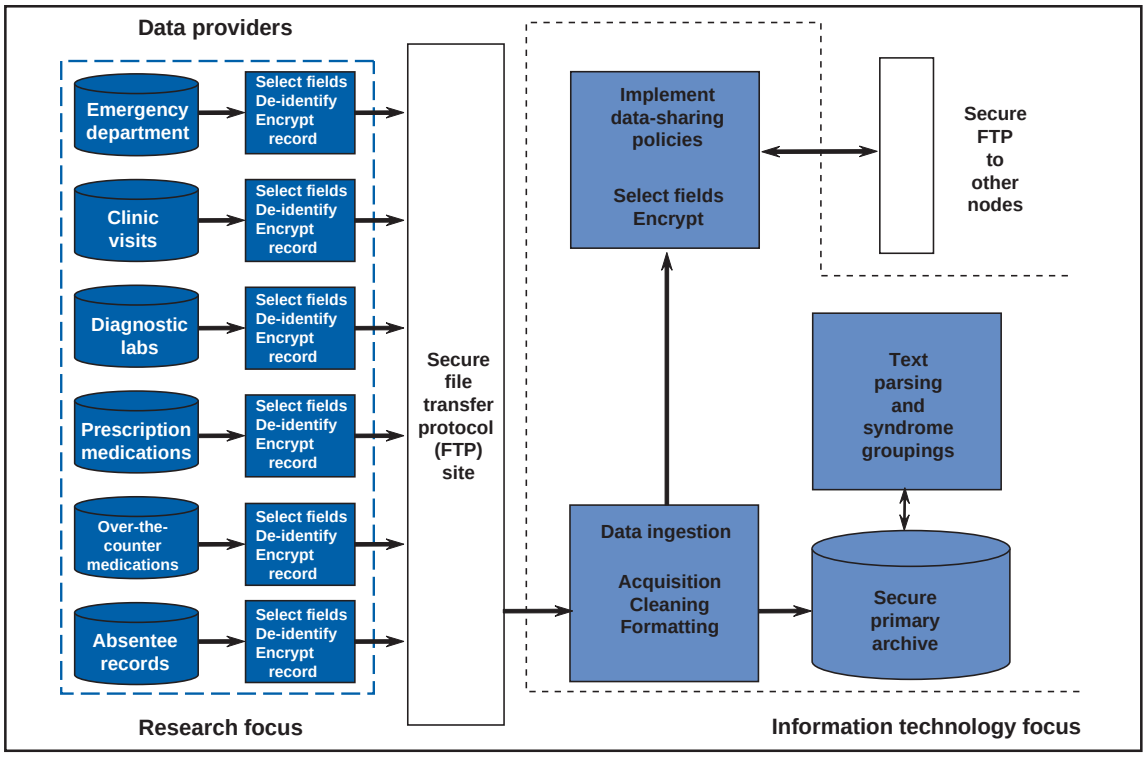
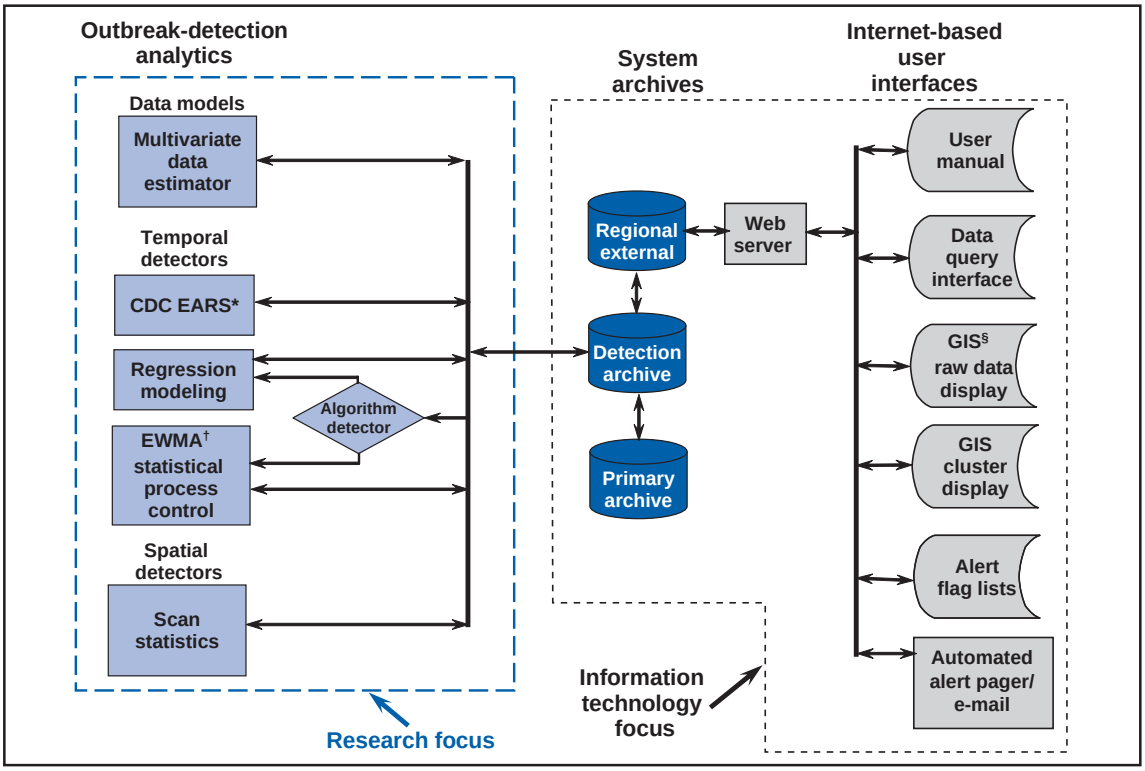


FIGURE 3. Processing and display flow for the Electronic Surveillance System for Early Notification of Community-Based Epidemics (ESSENCE)



* Early Aberration Reporting System.
† Exponentially weighted moving average.
§ Geographic information system.

In addition to ED chief-complaint information, ESSENCE II also receives data from physician-encounter claims in the form of *International Classification of Diseases, Ninth Revision* (ICD-9) codes and from retail merchants in the form of Universal Product Codes (UPCs) for OTC medications. These data are grouped into the same syndrome categories as the chief-complaint data to enable outbreak detection by syndrome.

Next, ESSENCE II applies outbreak-detection algorithms. These algorithms use a working archive known as the detection archive. New records are moved into the detection archive at the launching of the detection process. The detection algorithms are run every 4 hours, although this interval is adjustable. ESSENCE II can accommodate HL7 data streams if they are available from the hospital. Temporal and spatio-temporal algorithms are implemented in ESSENCE II to determine abnormalities. Also included are reference algorithms for assessing the performance enhancement provided by the ESSENCE II algorithms. CDC's Early Aberration Reporting System (*6*) algorithms were chosen as reference algorithms because they were already in use by regional health departments.

ESSENCE II uses two temporal algorithms: 1) an autoregressive modeling algorithm that predicts syndrome counts and looks for differences between actual counts and estimates and 2) the exponentially weighted moving average (EWMA), a statistical process control method. Details on these algorithms are published elsewhere (*7*). The autoregressive algorithm is based on a linear regression model that predicts a continually fluctuating daily expected count and threshold. The model bases its daily predictions on the previous 4 weeks of ESSENCE data, accounting for the day of the week and whether the day is a holiday or the day after a holiday. (The holiday function serves to explain artificial peaks in the data attributable to surges in patient visits after days when clinics are closed.) EWMA compares each observation to an average of past data that weights observations exponentially by time so that the most recent observations are most influential. Therefore, EWMA can be used when daily visit counts do not have the temporal structure required by a regression model. ESSENCE II uses a built-in goodness-of-fit statistic to determine whether the regression is useful in explaining the data; when this test fails, the automated checking process switches to EWMA.

A variant of the spatial scan statistic (*8*) is used to form clusters in time and space across the region by using zip codes as the smallest spatial resolution. The scan statistic has been modified to include multiple sources (*9*), which increases the sensitivity while controlling the false-alert rate.

ESSENCE II uses a secure website to transfer information to its users. Users must use individual passwords to access the website and can only access information for their respective jurisdictions. Four ESSENCE II portals enable users to view raw data and results from processed data:

- A map portal displays geographic distribution of raw data and clusters formed by scan statistics. The user can select data elements for geographic display and access details by clicking on the location of the data provider or the zip code(s) of interest. The details can be presented as tables or time graphs.
- The second portal provides alert lists for the output of the detection processes. These lists consist of color-coded flags to indicate algorithm outputs that are higher than expected. Upper confidence limits (UCLs) for the daily predictions are computed and used as alerting thresholds. If an observed count exceeds the 95% UCL but not the 99% UCL, a low-level (yellow) alert is generated; if it exceeds the 99% UCL, a high-level (red) flag results. The user can organize the lists to provide flags on data of interest, sort lists by elements of interest, and access data or link to the map portal to view the spatial distribution that resulted in the flag.
- The query portal enables a user interested in specific data to select from drop-down menus and view selected data elements over a selected timeframe as graphs or tables. All tabular information can be cut and pasted into a spreadsheet program for analysis offline.
- The fourth portal enables users to generate summary reports for export outside ESSENCE II. The user can select any data elements in the archive and view historic counts as well as upward or downward trends. This portal also contains tutorial material on operating ESSENCE II and a message board for making suggestions to developers or sharing thoughts with other users.

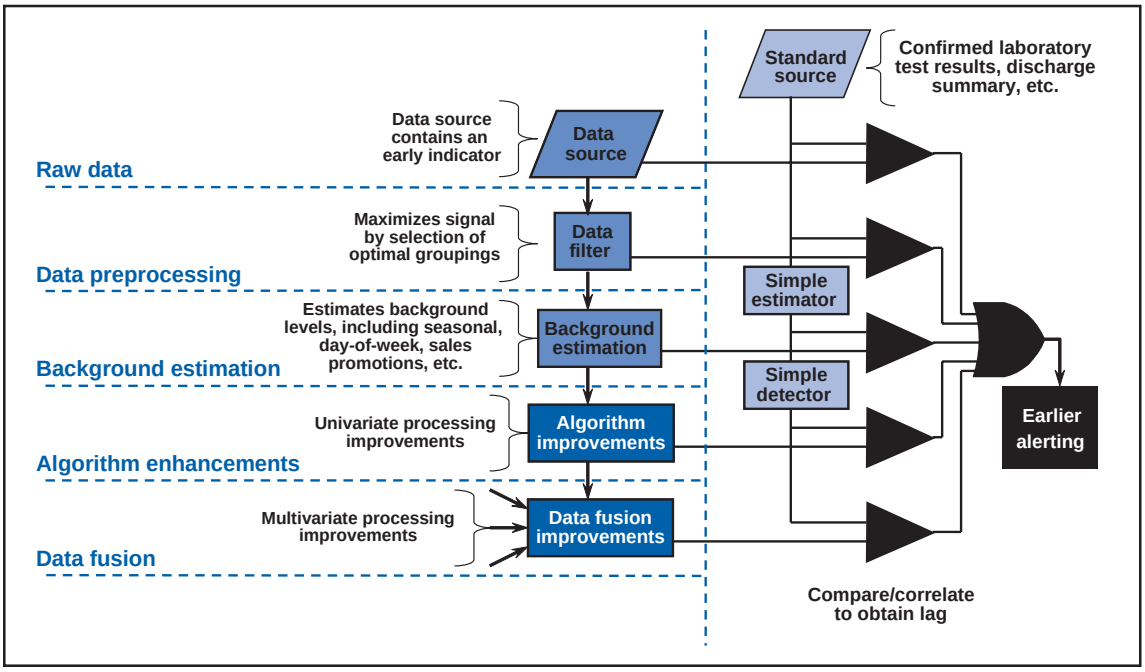
Outbreak Detection

Timeliness

The purpose of syndromic surveillance is to detect as early as possible abnormal disease patterns that could result in high mortality. This new technology should be evaluated and compared with traditional techniques to determine whether it improves upon detection timeliness. At least five layers of possible improvement exist (Figure 4). At each layer, the improvement is compared with a standard method to determine whether timelier notification is possible.

1. The first layer is the acquisition of a data source that contains an early indicator. For example, one promising data source is the nurse hotline service provided by certain health-care organizations.

FIGURE 4. Layers of possible improvement to outbreak-detection timeliness — Electronic Surveillance System for Early Notification of Community-Based Epidemics (ESSENCE)



2. The second layer involves filtering of the data stream to more closely match the population that exhibits early symptoms of disease. For example, because symptoms consistent with the release of a biologic agent at a facility (e.g., the Pentagon) would probably be observed among active-duty personnel at that facility, military data could be filtered by age to separate active-duty, retired, and dependent populations.
3. The third layer removes confounders from nontraditional data sources. For example, OTC medication sales are strongly influenced by sales promotions, seasonal effects, and day-of-week activity, as well as by the socioeconomic status of the community in which the sale occurred. ESSENCE II uses algorithms to model these confounders and remove their influence, thus allowing identification of the underlying pattern attributable solely to increases in disease.
4. The fourth layer addresses improvements to outbreak-detection algorithms that use a single data stream. Signal processing, regression modeling, and process control methods have been used to monitor single data streams.
5. The fifth layer addresses multivariate methods for gaining sensitivity needed for early recognition of an abnormality.

Improvements at any of the five layers or combination of layers can improve notification timeliness.

CDC's framework (*1,2*) provides a timeline, consisting of nine "anchor points," for measuring timeliness and performance of syndromic surveillance. The first three anchor points, *point-source exposure*, *symptom onset*, and *health-seeking behavior*, are independent of system performance; symptom onset is a function of the incubation period of the disease, and health-seeking behaviors depend on socioeconomic factors. The fourth anchor, *capture of the behavior in the record*, varies by data source, taking only seconds for scanning in OTC medications or hours to days for electronic claims. The fifth anchor point, *data source ready to share*, depends on the data provider and on system requirements for data updates. Data can be sent in real time (e.g., an HL7 feed from a hospital), hourly, daily, or at other predetermined intervals (e.g., ED chief-complaint data could be accumulated over 1 day and sent at midnight). ESSENCE II accepts both HL7 and ED chief-complaint data feeds. The data-ingestion module within ESSENCE II automates the *capture data into the system* process (anchor point six) within seconds. The seventh anchor point, *apply pattern-recognition tools/algorithms*, is also a function of the data-capture rate. If data are captured in real time, the detection algorithms must also operate in near real time. If data are captured daily, then the algorithms must be applied daily. ESSENCE II captures data throughout the day and applies the detection process every 4 hours but can alter the processing period when real-time data are received. After the detection process is complete, the *automated alert generation* process (anchor point eight) takes only seconds to minutes. The ninth anchor point, *initiate public health response*, depends upon policies and personnel at individual health departments and is independent of the syndromic surveillance system.

Validity

Algorithm performance can also be evaluated by detection of actual disease events within the community. In summer 2003, the ESSENCE II project participated in a blind evaluation conducted by the Defense Advanced Research Projects Agency (DARPA) Bio-ALIRT Program (*11*). This evaluation provided the opportunity for independent validation of results from the ESSENCE II outbreak-detection process and independent evaluation of participating syndromic surveillance systems. To conduct the evaluation, DARPA assembled an independent team of epidemiologists and physicians to identify respiratory and gastrointestinal events in data streams from five cities. The data included military and civilian ambulatory records and military prescription records. Team members identified eight respiratory and seven gastrointestinal events and, given only the raw data streams, were asked to estimate 1) start dates for the event, 2) date when a health department might recognize the event, 3) the peak of the event, and 4) the end of the event. Participants whose algorithms were being evaluated were provided only the raw data streams and asked to identify events.

Three ESSENCE II detection methods were selected for this evaluation (*10*): 1) a multivariate statistical process control algorithm applied to the residuals of a regression technique used to control for unexplained data dropouts, 2) a multiple univariate method based on the EWMA control chart, and 3) a Bayesian Belief Network applied to the outputs of the first two algorithms to optimize the decision for the two detectors. The results of these algorithms' detection performance and timeliness are provided as a function of false-alert rate, for rates of one false alert every 2 weeks, 4 weeks, or 6 weeks (Figure 5). In this context, a false alert does not imply the need for a laborious outbreak investigation but rather a more detailed review of the data and use of human judgment to dismiss alerting flags. For the highest false-alert rate, all three algorithms detected the eight respiratory events with a median detection time of 1 day after the start of the event (as determined by the epidemiology team). If the false-alert rate was constrained to once every 6 weeks, only the multiple univariate SPC method maintained its level of performance. For gastrointestinal events, only the Bayesian Belief Network successfully detected all seven events with a median delay of 1 day. Results might vary when the same algorithms are applied to other data streams and other seasons.

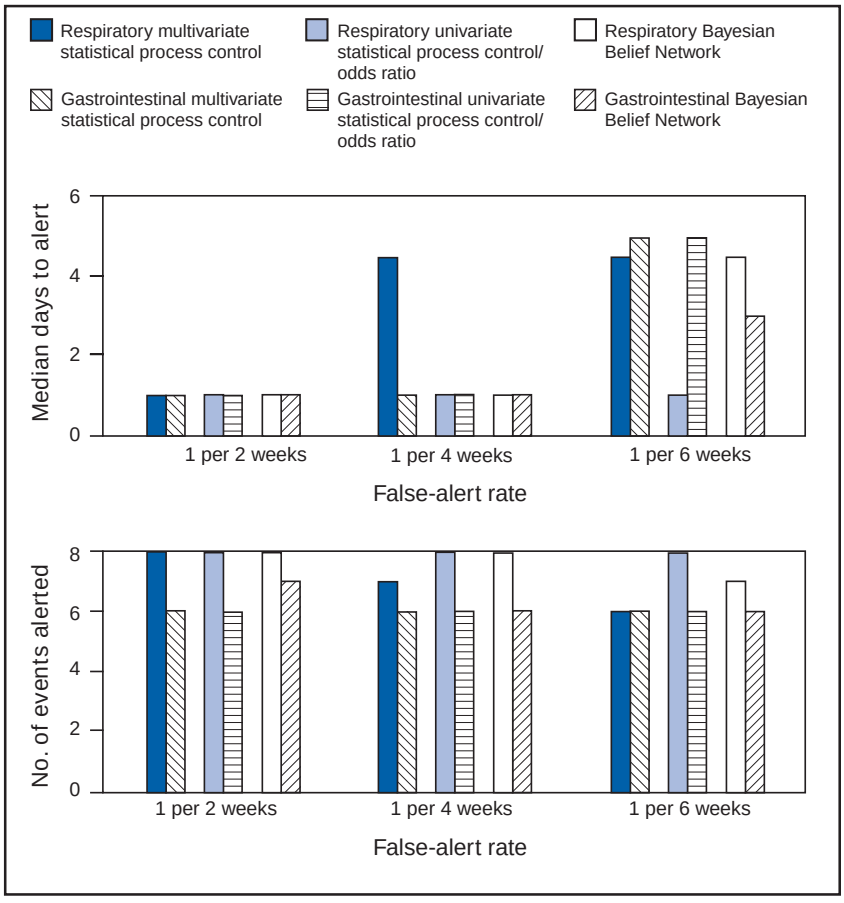
The majority of events used in the evaluation were seasonal epidemics attributable to colder weather, limited outdoor activity, and increased communicability during holiday gatherings; few, if any, of the cases comprising these events would result in death or were reportable diseases.

Experience

System Usefulness

ESSENCE II is used routinely by the Montgomery County (Maryland) Department of Health and Human Services for different purposes, including to accredit county hospitals for the capability to respond to mass casualties resulting from terrorism, to identify foodborne outbreaks, and to provide general knowledge of the county's health status. The department also requests changes to detection thresholds during high-profile events in the region that might affect public health in the county. The county health department continues to find new uses for ESSENCE II outputs; in 2004, it used the

FIGURE 5. Outbreak-detection performance of three algorithms in the Electronic Surveillance System for Early Notification of Community-Based Epidemics (ESSENCE)



system to determine when to initiate and cancel an influenza-vaccination program.

Flexibility and Portability

ESSENCE II acquires data feeds with minimal burden to data providers. The system accepts different data standards for acquisition and data sharing. Adding a new data source is more of a legal chore than a technical one because sources can be added with minimal hours of coordination or software development. ESSENCE is designed to enable persons with minimal programming skill to create new syndrome categories or change syndrome groupings in minutes. The system also allows users to access historic data to perform retrospective studies.

Multiple versions of ESSENCE II exist to accommodate different jurisdictions, data volumes, and data providers for both military preventive medicine and civilian health departments. ESSENCE II is also being provided to state and local health departments. Modifications are needed for local geographic shape files, zip codes, and data providers; these modifications can be performed by state health department IT staff.

System Acceptability

Acceptance by the majority of data providers has been exceptional. Currently, the test-bed version of ESSENCE II is used primarily when the level of risk increases. After the NCR network is fully implemented, usage levels are expected to increase. Full implementation is expected in 2004.

System Stability

Versions of ESSENCE II have been acquiring data since 1999 and have operated since then with minimal interruption. The system's size and complexity have expanded from the NCR military population and certain Maryland counties to include all of Maryland, Virginia, and the District of Columbia.

System Costs

System size and cost are a function of the jurisdiction's size, the number of data providers, and the size of the epidemiology department assigned to surveillance and follow-up. A minimum county-level configuration requires one or two computers, $15,000 for off-the-shelf software, one part-time epidemiologist, and one part-time IT professional. Cost-effectiveness depends upon the resources of the health department and the vulnerability of its population.

Conclusions

ESSENCE II is the first disease-surveillance system to incorporate both military and civilian data to improve the sensitivity and specificity of detecting abnormal disease occurrence. The design requires minimal resources from data providers, thus encouraging their participation. Research into algorithm improvements has been enhanced by operation of a test bed and by rapid upgrades to test improvements in an operational environment. Implementation of the NCR disease-surveillance network should provide operational insights for other jurisdictions considering collaborative surveillance systems.

CDC's framework for evaluating syndromic surveillance systems provides a needed reference for developers and health departments wishing to develop and implement new systems. Evaluation would be enhanced if CDC provided standard data sets to test the processes embedded within the systems and provide a benchmark for comparing system performance.

References

1. Sosin D. Draft framework for evaluating syndromic surveillance systems. J Urban Health 2003;80(2 Suppl 1):i8–13.
2. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5):1–11.
3. Lombardo J. A systems overview of the Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE II). J Urban Health 2003;80(2 Suppl 1):i32–42.
4. Health Level Seven, Inc. Health Level 7. Ann Arbor, MI: Health Level Seven, Inc., 2001. Available at http://www.hl7.org.
5. Sniegoski C. Methodology for categorizing emergency department chief records for syndromic surveillance. Presentation, Annual Meeting of the American Public Health Association, 2003. Available at http://apha.confex.com/apha/131am/techprogram/paper_68724.htm.
6. Hutwagner L. The bioterrorism preparedness and response Early Aberration Reporting System. J Urban Health 2003;80 (2 Suppl 1):i89–96.
7. Burkom H, Elbert YA. Development, adaptation, and assessment of alerting algorithms for biosurveillance. Johns Hopkins University Applied Physics Laboratory Technical Digest 2003;24:335–42.
8. Kulldorff M. A spatial scan statistic. Communications in Statistics—Theory and Methods 1997;26:1481–96.
9. Burkom H. Biosurveillance applying scan statistics with multiple disparate data sources. J Urban Health 2003;80(2 Suppl 1):i57–65.
10. Burkom H, Elbert YA. Role of data aggregation in biosurveillance with applications from ESSENCE. MMWR 2004;53(Suppl):67–73.
11. Siegrist D. Bio-ALIRT biosurveillance detection algorithm evaluation. MMWR 2004;53(Suppl):152–8.

Conducting Population Behavioral Health Surveillance by Using Automated Diagnostic and Pharmacy Data Systems*

Julie A. Pavlin,[1] P. Murdock,[2] E. Elbert,[1] C. Milliken,[1] S. Hakre,[1] J. Mansfield,[1] C. Hoge[1]
[1]*Walter Reed Army Institute of Research, Silver Spring, Maryland;* [2]*U.S. Army Health Facilities Planning Agency, Falls Church, Virginia*

Corresponding author: Julie A. Pavlin, Chief, Department of Field Studies, Walter Reed Army Institute of Research, 503 Robert Grant Ave., Silver Spring, MD 20910. Telephone: 301-319-9346; Fax: 301-319-9104; E-mail: julie.pavlin@us.army.mil.

Abstract

Introduction: *The Walter Reed Army Institute of Research used the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) to conduct population-based behavioral health surveillance among military-health–system beneficiaries. The study analyzed the effectiveness of using prescribing patterns of psychotropic medications to monitor changes in a community's behavioral health status.*

Objectives: *The objectives of this study were to 1) determine the feasibility of tracking psychiatric illnesses by monitoring prescriptions for psychiatric medications; 2) assess how often psychiatric medications are prescribed for patients with no record of psychiatric illness; 3) determine at what types of clinics these medications are prescribed most often and what other diagnoses are attributed to these patients; and 4) analyze data for potential changes in the population's mental health after high-stress events.*

Methods: *Correlation analysis and calculations of sensitivity and specificity were used to determine how well prescription medications correlate with outpatient diagnoses and how well they serve as proxies for outpatient diagnoses. A descriptive analysis was conducted of the types of clinics (e.g., primary care, behavioral health, or other specialty clinics) treating patients and the associated percentage of concurrence between prescriptions and diagnostic codes.*

Results: *In military treatment facilities, a diagnosis of depression or anxiety correlated significantly (r = 0.82) with antidepressant or anxiolytic prescriptions. Sensitivity of prescriptions when compared with outpatient visits was 0.76, and specificity was 0.94. Among those patients who visited a primary care clinic either the day before or the same day as an antidepressant or anxiolytic prescription was filled, 60.1% did not receive a diagnosis of any mental health disorder. Behavioral health clinics had the highest correlation between diagnoses and prescriptions; specialty clinics had the lowest.*

Conclusions: *Behavioral health trends in a population can be monitored by automated analysis of prescribing patterns alone. This method might be a rapid indicator of needed mental health interventions after acute stress-inducing events and be more sensitive than tracking diagnoses alone.*

Introduction

New approaches to public health surveillance that use automated, and often unconventional, data sources have focused on the threat of emerging infections and biologic terrorism. However, other uses for these technologies exist beyond traditional surveillance of infectious disease. Mental health surveillance using de-identified data has the potential to estimate the prevalence of certain mental illnesses, especially among persons who are sensitive to events that cause stress in their communities (e.g., natural or man-made disasters, regional unemployment, or deployments at military bases). These systems can support planning for more resource-intensive traditional mental health surveillance activities (*1*) and signal a need for community-based mental health interventions that emphasize normalization of responses to stress.

Automated public health surveillance systems are an inexpensive and timely augmentation to traditional health surveillance methodologies and can enhance provider alertness to public health threats (*2,3*). By using routinely collected electronic data (*4–6*) from different traditional and nontraditional sources (e.g., administrative, clinical, pharmacy, and retail databases, and school and work absenteeism data) (*4,7–11*), such systems can detect increases in the number of cases above that normally expected, with varying degrees of specificity.

* The views expressed are those of the contributors and do not reflect the position of the U.S. Army or the U.S. Department of Defense.

Certain mental illnesses might not require as urgent a response as infectious-disease outbreaks and therefore might not seem to justify use of automated data sources in surveillance. However, few, if any, active mental health surveillance tools exist, primarily because finding measures of mental health changes in a community is difficult. Thus, data on routine outpatient visits or pharmacy prescriptions might prove useful for determining a community's mental health status and assessing the effectiveness of interventions. Furthermore, communitywide increases in mental illness might require rapider intervention when illness could result in suicidal or homicidal behavior.

In 1997, the U.S. Department of Defense (DoD) instituted the Standard Ambulatory Data Record (SADR) to record demographic and diagnostic data on all military outpatient visits, including *International Classification of Diseases, Ninth Revision* (ICD-9) codes for each visit. In 1999, the Walter Reed Army Institute of Research (WRAIR) created the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) to detect and track infectious-disease outbreaks among military-health–system beneficiaries (*12,13*). Using the SADR database, ESSENCE automatically collects ICD-9 codes that potentially indicate infectious diseases and groups them into clinical diagnostic categories based on clusters of similar ICD-9 codes. Codes are grouped to reduce the variability and increase the sensitivity of administrative diagnostic data (*14*) and to improve baseline data-monitoring capability.

To study utilization of mental health services among military beneficiaries in the Washington, D.C., area after the September 11, 2001, attack on the Pentagon, WRAIR adapted the ESSENCE model to include psychiatric ICD-9 codes. Although no overall increase in utilization of mental health care was identified, the study did detect a significant change in the distribution of diagnoses, including relative increases in the median number of visits for adjustment reactions, anxiety, and acute-stress reactions during the first 5 months after the attack (*15*).

For the current study, groupings of mental health outpatient diagnostic data were correlated with pharmacy data and used to monitor changes in the mental health status of military communities. Diagnostic data based on ICD-9 codes might not be the best indicator of mental illness in a community, for multiple reasons. Anecdotal evidence published previously indicated that physicians might code only one diagnosis even when patients are seen for multiple conditions or make coding decisions based on codes most available or frequently used (*10*). In addition, diagnostic coding for mental health can be affected by stigma and employment culture. Stigma associated with a mental illness diagnosis is well-documented in the literature (*1,16–20*). In the military, a mental illness diagnosis can affect a service member's security clearance, flight status, and authorization to carry a weapon (*17,19*). A patient whose recorded diagnosis does not indicate a mental disorder might still receive a prescription for a psychiatric condition; therefore, prescriptions might be a better reflection of true mental health than the recorded diagnosis.

Pharmacy data provide insight into a clinician's treatment focus and might more accurately represent a patient's true condition. In addition, prescriptions are often renewed or refilled for chronic conditions regardless of the patient's primary complaint at the time of the visit (*10,21*). However, these data can be complicated by multiple indications for the same medication and are also sensitive to treatment setting (*6*). Because military patients, like other populations that have been studied systematically, receive a substantial percentage of their psychiatric care from primary care providers rather than mental health providers, measures of mental health treatment in primary care settings are needed.

This study's objectives were to 1) determine the feasibility of tracking psychiatric illnesses through the monitoring of psychiatric medication prescriptions by correlating diagnoses and the drugs prescribed; 2) assess how often psychiatric medications are prescribed to patients with no diagnostic record of psychiatric illness, particularly to estimate underreporting of psychiatric illnesses and determine whether pharmacy data might be a better indicator of mental health treatment; 3) determine at what types of clinics (i.e., primary care, specialty care, or behavioral health) psychiatric drugs are most often prescribed without corresponding mental health diagnoses, and identify what other diagnoses are attributed to these patients; and 4) evaluate whether any increases in anxiety or depression among family members of deployed military personnel could be detected.

Methods

Data were obtained for all outpatient visits at fixed military-treatment facilities (MTFs) and for prescriptions for all military beneficiaries during July 2001–August 2002. Approximately 8.8 million active-duty personnel, family members, and retirees are eligible for care of MTFs. Of this population, approximately 4.5 million are enrolled in the military's health-care system, Tricare Prime, which usually indicates they intend to receive care at MTFs (although some do access care outside of military hospitals and clinics). Those not enrolled in Tricare Prime can also receive treatment at MTFs on a different payment schedule.

The outpatient SADR database consists of ≤4 ICD-9 codes for every MTF visit. These codes are entered approximately at the time of the patient encounter, usually by the provider but also by professional coders at certain locations. The data include visits to all fixed MTFs worldwide but do not include deployed forces involved in military operations. All mental disorder ICD-9 codes in the 291–318 range, as well as related ICD-9 codes used by behavioral health clinics (e.g., mental health or substance abuse counseling; problems related to partner relationships, family circumstances, life circumstances, maltreatment, or abuse) were grouped according to established methods (*22,23*). In addition, a subset was created of all ICD-9 codes related to depression and anxiety, as those conditions would be more likely to increase during times of stress (Table 1).

TABLE 1. *International Classification of Diseases, Ninth Revision* (ICD-9) codes used to categorize depression or anxiety

ICD-9 code	Description
296.20	Major depressive disorder, single episode, unspecified
296.21	Major depressive disorder, single episode, mild
296.22	Major depressive disorder, single episode, moderate
296.23	Major depressive disorder, single episode, severe
296.24	Major depressive disorder, single episode, severe with psychotic behavior
296.25	Major depressive disorder, single episode, in partial or unspecified remission
296.26	Major depressive disorder, single episode, in full remission
296.30	Major depressive disorder, recurrent episode, unspecified
296.31	Major depressive disorder, recurrent episode, mild
296.32	Major depressive disorder, recurrent episode, moderate
296.33	Major depressive disorder, recurrent episode, severe
296.34	Major depressive disorder, recurrent episode, severe with psychotic behavior
296.35	Major depressive disorder, recurrent episode, in partial or unspecified remission
296.36	Major depressive disorder, recurrent episode, in full remission
300.00	Anxiety state, unspecified
300.01	Panic disorder
300.02	Generalized anxiety disorder
300.09	Other anxiety state
300.21	Agoraphobia with panic attacks
300.22	Agoraphobia without panic attacks
300.23	Social phobia
300.29	Other isolated or simple phobia
300.3	Obsessive-compulsive disorder
300.4	Neurotic depression
308.0	Acute reaction to stress, predominant emotional disturbance
308.3	Acute reaction to stress, other
308.4	Acute reaction to stress, mixed disorders
308.9	Acute reaction to stress, unspecified
309.0	Brief depressive reaction
309.1	Prolonged depressive reaction
309.81	Prolonged posttraumatic stress disorder
311	Depressive disorder, not elsewhere classified

The code for tension headache (307.81) was excluded from the analysis because this diagnosis is commonly used for headache unrelated to a mental disorder. The code for tobacco use disorder (305.1), which is included in the ICD-9 mental disorder category but not typically treated as a mental disorder, was also excluded. Deleting 305.1 also excluded use of certain antidepressants (e.g., bupropion) used as smoking cessation aids that could confound the analysis.

Outpatient pharmacy prescriptions at all MTFs and Tricare network pharmacies are collected in the Pharmacy Data Transaction Service database at the time they are filled (*24*). Prescriptions of medications used primarily to treat depression and anxiety (Table 2) were correlated with outpatient diagnoses. Certain medications were excluded to limit potential confounding factors. For example, trazodone, a potential antidepressant, is highly sedating and almost always used as a sleep aid. Hydroxizine has an anxiolytic indication but is almost always used for its antihistamine properties as an allergy medication. Amitriptyline is sedating and has cardiovascular side effects and is therefore rarely used as an antidepressant, although it is often used at low doses for pain conditions (e.g., headaches).

The strengths of correlations between antianxiety and antidepressant prescription medications and outpatient visits for mental health, anxiety, and depression were measured by using Pearson's correlation coefficient (*25*). Data were grouped by week to decrease the effect on the correlation of the usual weekly pattern of visits and prescriptions.

The two databases were then matched by using a code provided by Tricare that is uniquely assigned to each patient but does not allow patient identification. A match was determined for those patients who 1) had a new prescription written for one of the medications listed (Table 2) and 2) also had a recorded outpatient visit the day (or the day before) the prescription was written.

Prescriptions and outpatient visits were expected to have a correlation based on holiday and seasonal effects (e.g., fewer persons saw a health-care provider or were prescribed medications on holidays, compared with more persons during the winter influenza and seasonal affective disorder seasons); for this reason, the sensitivity, specificity, and positive predictive value of the prescription data were calculated by using outpatient visits both for depression and anxiety only and for all mental health concerns as the standard.

For those patients who were prescribed antidepressants or anxiolytics and who also had an outpatient diagnostic code from the same visit, the numbers of patients receiving depression or anxiety diagnoses, any mental health diagnoses, and all other diagnoses were calculated. The clinical setting was taken into account by grouping clinics into three categories:

TABLE 2. Antidepressants and anxiolytics

Class	Generic name
Antianxiety	Alprazolam
	Buspirone HCl
	Chlordiazepoxide
	Chlordiazepoxide HCl
	Clonazepam
	Clorazepate dipotassium
	Diazepam
	Halazepam
	Lorazepam
	Oxazepam
	Temazepam
Antidepressant	Amoxapine
	Bupropion HCl
	Citalopram hydrobromide
	Clomipramine HCl
	Desipramine HCl
	Doxepin HCl
	Escitalopram
	Fluoxetine HCl
	Fluvoxamine maleate
	Imipramine HCl
	Imipramine pamoate
	Isocarboxazid
	Maprotiline HCl
	Maprotiline
	Nefazodone HCl
	Nortriptyline HCl
	Paroxetine HCl
	Phenelzine sulfate
	Protriptyline HCl
	Sertraline HCl
	Tranylcypromine sulfate
	Trimipramine maleate
	Venlafaxine HCl
Combination drug	Amitriptyline HCl/chlordiazepoxide

1) mental health, 2) primary care (i.e., family practice, urgent-care clinics, emergency departments, internal medicine, and pediatrics), and 3) all other clinics (e.g., orthopedic, cardiology, or physical therapy). Prescription refills were not used in this analysis because the purpose of this surveillance was to detect acute psychiatric illness. However, certain prescriptions that appeared to be new and that were included might have represented dosage changes, brand changes, or renewals after all refills had been used (i.e., were not first-time prescriptions for the drug category).

Finally, outpatient visits and drug prescriptions among military beneficiaries were monitored to determine whether any increases in depression or anxiety had occurred; this was particularly relevant during 2003, when U.S. military deployments likely increased stresses on active-duty military and their families.

Although data for deployed forces were not available, data were examined from three installations from which substantial numbers of troops had been deployed to Iraq for Operation Iraqi Freedom (OIF). Trends in outpatient visits for anxiety and depression and filled prescriptions for antianxiety and antidepressant medications during July 2001–September 2003 were analyzed, by military beneficiary category, at all MTFs at the three installations. On any given day during surveillance, only the initial visit for anxiety and depression or the first prescription filled at the installation was included. To best reflect those who live at or near the installations, the analysis included only anxiolytics and antidepressants filled at pharmacies within a 50-mile radius of any of the three installations. The percentages of mental health visits and prescriptions for spouses (out of total outpatient visits or prescriptions) at that MTF were determined to decrease the effect of a changing population size. The Wilcoxon's rank-sum test (*25*) was used to test the alternative hypothesis that anxiety and depression visits and prescriptions differed significantly after January 9, 2003 (i.e., the date of deployment for OIF).

Results

During July 2001–August 2002, a total of 2,343,684 anxiolytic and antidepressant prescriptions were written for 894,922 unique patients. A total of 1,588,081 outpatient visits for 408,083 unique patients were given an ICD-9 code for any mental health disorder, and 675,564 (42.5%) of these visits, representing 224,459 unique patients, were for depression or anxiety, as defined previously (Table 1). Records containing the code for tension headache (2,712) or tobacco use disorder (44,828) were excluded.

The correlation coefficient was 0.82 when only new prescriptions for anxiolytics and antidepressants were compared with diagnoses of depression or anxiety. The coefficient was 0.85 when prescriptions were compared with all mental health diagnoses. Including prescription refills in the analysis increased the correlation coefficient to 0.85 for diagnoses of depression or anxiety and to 0.88 for all mental health diagnoses.

Of all antidepressant or anxiolytic prescriptions, 934,220 (40.0%) matched with a recorded outpatient visit. This number includes 650,100 patients who received one or more prescriptions and had a matching outpatient visit, with 87% of visits occurring the same day as and 13% the day before the prescription was written. Of those prescriptions that matched an outpatient visit, 37.4% were for refills and the remainder for new prescriptions. For prescriptions that did not match an outpatient visit, the percentage attributable to refills increased to 54%.

In the sensitivity and specificity analysis, prescription data were relatively sensitive (0.76) and highly specific (0.94) (Table 3). However, the positive predictive value was low

TABLE 3. Sensitivity and specificity analysis of antidepressants and anxiolytics when using matched outpatient visits for anxiety and depression as standard*

Anxiolytic or antidepressant prescribed	Outpatient visit for anxiety or depression		
	Yes	No	Total
Yes	243,476	690,744	**934,220**
No	75,575	10,542,493	**10,618,068**
Total	**319,051**	**11,233,237**	**11,552,288**

* Sensitivity = 0.76; specificity = 0.94; positive predictive value = 0.26.

(0.26). This result was expected because visits for anxiety or depression are relatively rare (8.0%) compared with all outpatient visits. The result also reflects the substantial number of prescriptions given without a corresponding diagnosis of depression or anxiety. If prescriptions without a corresponding diagnosis truly represent a mental illness, then prescriptions might be a better indicator of mental illness than the gold standard of outpatient visits. Because the outpatient codes chosen are broad and medications are more specific, these findings probably are conservative, and the correlation between prescriptions and mental health diagnoses might be stronger. When the gold standard is expanded to include all mental health visits, the sensitivity decreases (0.52) and the positive predictive value increases (0.31), indicating that antidepressants and anxiolytics are not as sensitive an indicator of any mental health condition.

Among patients with a matched prescription and visit, 62.4% who were prescribed anxiolytics or antidepressants in primary care clinics did not receive a diagnosis of any mental health disorder (Table 4). Behavioral health clinics had the highest correlation between prescription and diagnosis, and other specialty clinics had the lowest, with 11.5% and 91.2%, respectively, having no mental health diagnosis. This discrepancy might differ in civilian health-care settings that link diagnostic codes more closely to reimbursement and prescription justification.

The majority of the diagnostic codes for patients receiving medications were codes for common medical illnesses (e.g., hypertension and diabetes) or generic codes for counseling. In addition, certain diagnoses (e.g., insomnia, myalgia, and myositis) were not psychiatric but could justify the prescriptions given. Nonpsychiatric diagnoses for which an antidepressant or antianxiety drug could appropriately be used (e.g., back and joint pain and strains, urticaria and rash, headache and migraines, counseling, or insomnia) constituted approximately 19% of the codes.

ICD-9-code and prescription data were then used retrospectively for surveillance of military-health–system beneficiaries. Although use of psychotropic medications increased gradually during July 2001–September 2003, since the start of OIF deployments on January 9, 2003, or the start of OIF hostilities on March 19, 2003, no acute increases in outpatient visits for anxiety or depression or for prescriptions across the total military beneficiary population were determined. However, if the data are grouped by beneficiary category and if military installations with higher rates of deployment are isolated, certain trends become apparent. The percentage of total outpatient visits attributed to depression or anxiety and the percentage of total prescriptions for antidepressants or anxiolytics for spouses at the three installations that had high rates of OIF deployment were calculated (Figure). The rates of both outpatient visits and prescriptions during January 9, 2003–September 25, 2003, differed significantly ($p < 0.0001$ for both visits and prescriptions; Wilcoxon's rank-sum test) compared with the previous period of July 7, 2001–January 8, 2003.

Discussion

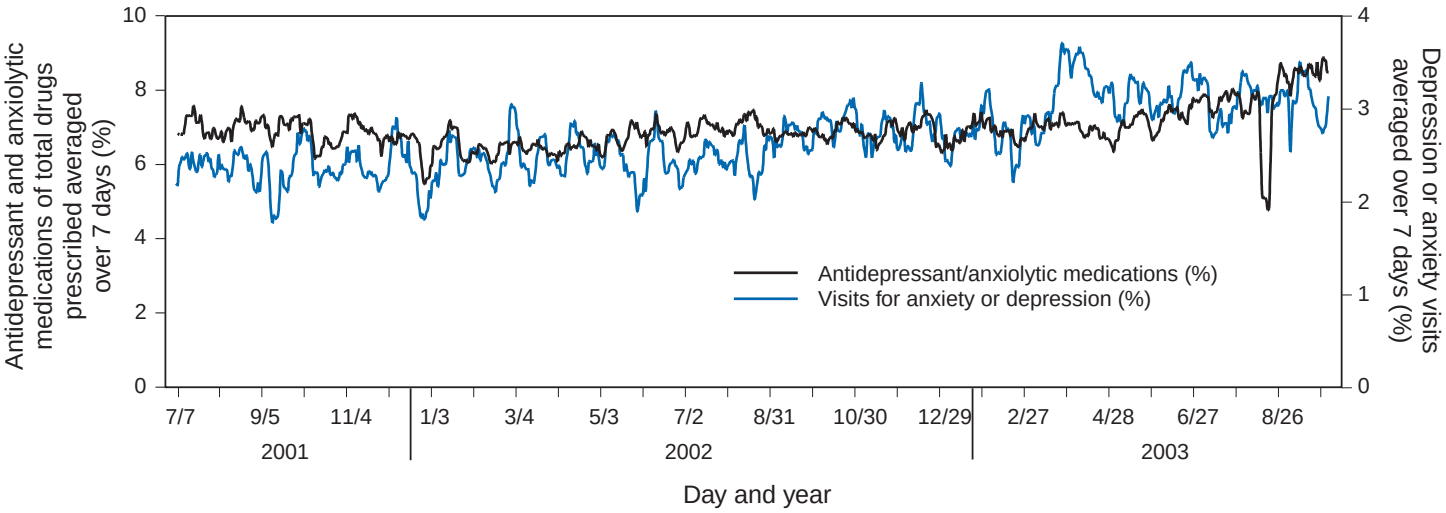
The analysis demonstrated a strong correlation between mental health outpatient diagnoses and prescription of antidepressants and anxiolytics. It also indicated that additional patients are prescribed these medications without a corresponding diagnosis for depression, anxiety, or any mental disorder. This result is similar to findings published previously that 55% of Medicaid beneficiaries who received psychotropic medication did not receive a mental health diagnosis (*21*). These results indicate that tracking outpatient medications might be a more sensitive means for detecting changes in the mental health of a population.

However, multiple potential confounders exist. First, 24% of patients who had an ICD-9 diagnosis for anxiety or depression were not prescribed psychotropic medications; therefore, if only prescriptions are surveyed, those patients will be overlooked. Second, in certain instances, the matching of outpatient visits

TABLE 4. Recorded diagnoses of military-health–system patients prescribed anxiolytics or antidepressants, by clinic type — July 2001–August 2002

Clinic type	No. of patients	Patients (%) with depression or anxiety diagnosis	Patients (%) with other mental health diagnosis	Patients (%) with other diagnosis
Behavioral health	129,454	89,878 (69.4)	46,351 (35.8)	14,942 (11.5)
Primary care	423,957	147,011 (34.7)	40,651 (9.6)	254,894 (60.1)
Specialty clinic	96,689	3,789 (3.9)	5,049 (5.2)	88,143 (91.2)
Total	**650,100**	**240,678 (37.0)**	**92,051 (14.2)**	**357,979 (55.1)**

FIGURE. Outpatient visits for anxiety or depression out of total visits and prescriptions for antidepressants or anxiolytics out of total prescriptions for spouses at three military installations with high rates of deployment



and prescriptions might not have been for the same encounter. For example, a patient could have had a telephone consultation or other provider interaction that resulted in a prescription; meanwhile, on the same day or the previous day, the same patient could have made an office visit for an entirely unrelated medical complaint. In such a case, the matching process would have linked the records and made the prescribed medication appear linked to an unrelated diagnosis. Such an error can make discrepancies between diagnosis and prescribing behavior appear greater than they are. However, such a situation also indicates that prescriptions might be a more sensitive surveillance tool, given that a prescription record exists despite no corresponding mental health diagnosis.

A third potentially confounding situation involves prescription of medications for disorders other than mental illness. Certain medical conditions (e.g., insomnia, pain conditions, urticaria, or migraine) can merit the prescription of antidepressants or anxiolytics. To adjust for this confounder, certain medications used more commonly for such conditions were removed and the analysis rerun. In that analysis, the percentage of patients receiving antidepressants or anxiolytics who had not been given a mental health-related diagnosis decreased by <1% in primary care clinics.

A fourth potentially confounding situation involves patients who take psychotropic medications for chronic conditions but who are treated for a different chief complaint during an office visit. A provider might code the visit accurately for the presenting complaint while also renewing the prescription for the chronic condition, which would make the number of coding errors appear greater and would decrease the specificity of using prescriptions for surveillance of acute events. Now that a longer historical record of patient visits and prescriptions is available, future studies will attempt to exclude from the analysis anyone who has ever received a medication in the anxiety or depression category.

Surveillance among military beneficiaries at three Army posts indicates that distress levels related to deployments might have increased in the population. Increases in mental health visits by military spouses were apparent. The increase in the rate of visits for anxiety or depression was greater than the increase in rate of psychotropic drug prescriptions. This finding highlights a potential limitation of relying on pharmacy data for mental health surveillance in a population. Deployment-related stress is common, and various counseling services that do not involve pharmacologic intervention are available to service members and families. Prescription-based indicators of distress might be less helpful in this context than they would be if a traumatic or terrorist event occurred in a population.

Conclusion

Automated analysis of prescribing patterns of psychotropic medications can be used to monitor behavioral health trends in a population. This surveillance method has potential to be a rapid and sensitive indicator of needed mental health interventions after acute stress-inducing events, especially in combination with surveillance of outpatient diagnoses. The importance of this surveillance is in its ability to react quickly to an increased need for mental health services. As with any other surveillance system that relies on data not originally gathered for surveillance purposes, any apparent increases in either prescribing behavior or outpatient visits should be veri-

fied by discussions with the provider or a review of records. If an increase in mental health needs is confirmed, early interventions can include community outreach and increased advertisement of available resources.

References

1. Hoge CW, Messer SC, Engel CC, et al. Priorities for psychiatric research in the US military: an epidemiological approach. Mil Med 2003;168:182–5.
2. CDC. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center. MMWR 2002;51(Special Issue):13–5.
3. US Department of Defense Global Emerging Infections Surveillance and Response System. ESSENCE: Electronic Surveillance System for the Early Notification of Community-Based Epidemics. Silver Spring, MD: Walter Reed Army Institute of Research, 2003. Available at http://www.geis.ha.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp.
4. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. J Urban Health 2003;80(2 Suppl 1) i1–7.
5. Shahar Y. A framework for knowledge-based temporal abstraction. Artificial Intelligence 1997;90:79–133.
6. Finnerty M, Altmansberger R, Bopp J, et al. Using state administrative and pharmacy databases to develop a clinical decision support tool for schizophrenia guidelines. Schizophr Bull 2002;28:85–94.
7. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5):1–11.
8. 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.
9. Buckeridge DL, Graham J, O'Connor MJ, Choy MK, Tu SW, Musen MA. Knowledge-based bioterrorism surveillance. Proc AMIA Symp 2002;76–80.
10. Sloan KL, Sales AE, Liu CF, et al. Construction and characteristics of the RxRisk-V: a VA-adapted pharmacy-based case-mix instrument. Med Care 2003;41:761–74.
11. Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proc Natl Acad Sci U S A 2002;99:5237–40.
12. Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC, area. Am J Prev Med 2002;23:180–6.
13. Lombardo J, Burkom H, Elbert E, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health. 2003;80(2 Suppl 1):i32–42.
14. Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc 2002;9:105–15.
15. Hoge CW, Pavlin JA, Milliken CS. Psychological sequelae of September 11. N Engl J Med 2002;347:443–5.
16. Rowan AB. Demographic, clinical, and military factors related to military mental health referral patterns. Mil Med 1996;161:324–8.
17. Pflanz S. Psychiatric illness and the workplace: perspectives for occupational medicine in the military. Mil Med 1999;164:401–6.
18. Fragala MR, McCaughey BG. Suicide following medical/physical evaluation boards: a complication unique to military psychiatry. Mil Med 1991;156:206–9.
19. Porter TL, Johnson WB. Psychiatric stigma in the military. Mil Med 1994;159:602–5.
20. Cooper LA, Gonzales JJ, Gallo JJ, et al. The acceptability of treatment for depression among African-American, Hispanic, and white primary care patients. Med Care 2003;41:479–89.
21. Gilmer T, Kronick R, Fishman P, Ganiats TG. The Medicaid Rx model: pharmacy-based risk adjustment for public programs. Med Care 2001;39:1188–202.
22. Hoge CW, Lesikar SE, Guevara R, et al. Mental disorders among US military personnel in the 1990s: association with high levels of health care utilization and early military attrition. Am J Psychiatry 2002;159:1576–83.
23. Garvey-Wilson AL, Hoge DW, Messer SC, Lesikar SE, Eaton KM. Diagnoses in behavioral health clinics: impact on perceived burden of mental health. In: Syllabus and scientific proceedings, annual meeting of the American Psychiatric Association. Washington, DC: American Psychiatric Association 2003;23.
24. US Department of Defense. Department of Defense Pharmacy Data Transaction Service. Available at http://www.pec.ha.osd.mil/pdts.htm.
25. Altman DG. Practical statistics for medical research. London, UK: Chapman & Hall, 1991.

Evaluation of an Electronic General-Practitioner–Based Syndromic Surveillance System — Auckland, New Zealand, 2000–2001

Nicholas F. Jones,[1] R. Marshall[2]
[1]*Auckland Regional Public Health Service, Auckland, New Zealand;* [2]*University of Auckland, Auckland, New Zealand*

Corresponding author: Nicholas F. Jones, Auckland Regional Public Health Service, Private Bag 92 605 Symonds Street, Auckland, New Zealand. Telephone: 64 9 262-1855; Fax: 64 9 623-4633; E-mail: nickj@adhb.govt.nz.

Abstract

Introduction: *During 2000 and 2001, Auckland Regional Public Health Service piloted a general-practitioner–based syndromic surveillance system (GPSURV).*

Objectives: *The pilot evaluated data capture, the method used to distinguish initial from follow-up visits, the definition of denominators, and the external validity of measured influenza-like illness trends.*

Methods: *GPSURV monitored three acute infectious-disease syndromes: gastroenteritis, influenza-like illness, and skin and subcutaneous tissue infection. Standardized terms were used to describe the syndromes. Data were uploaded daily from clinics and transferred to a database via a secure network after one-way encryption of patient identifiers. Records were matched to allow the distinction of follow-ups from first visits, based on between-visit intervals of ≤8 weeks. Denominator populations were based on counts of unique patients treated at participating clinics during the previous 2 years. Record completion was examined by using before-and-after surveys of self-assessed standardized-term recording. Between-visit intervals were counted for matching records and alternative denominators were calculated on the basis of different observation periods. Weekly influenza-like illness rates were compared with rates generated by an alternative system.*

Results: *Physicians' self-reported recording compliance was highest for skin and subcutaneous tissue infection (71%) and lowest for influenza-like illness (48%). Initial visits had 18%–19% greater compliance than follow-up visits. The number of physicians reporting increasing compliance during the pilot was greater than the number reporting decreases for all conditions. Comparison of data with an independent influenza-like illness surveillance system indicated a close agreement between the two data series.*

Conclusions: *These results indicate that incidence of acute syndromes can be monitored, at least as successfully as a manual system, by using standardized clinical-term data from selected general-practice clinics. The provision of feedback reports appears to have a limited but positive effect on data quality.*

Introduction

The potential to enhance public health surveillance by using general-practice data has been discussed by public health practitioners (*1*,*2*). Computerization of general practice records and increased emphasis on population health within primary care (*3*) have brought this potential closer to realization.

In New Zealand (NZ), electronic systems for physician reimbursement have contributed to widespread adoption of computerized family practice information systems. In 1995, an estimated 84% of NZ family physicians or general practitioners (GPs) used a computer for, at minimum, office management (*4*). A recent survey determined that 57% of NZ GPs use an electronic system to record and store clinical data; this figure was predicted to reach 89% by early 2004 (*5*). The potential for GP-based sentinel surveillance in NZ is also enhanced by virtually every GP clinic having, at minimum, dial-up connectivity to a secure wide area network.

These trends of increased information-system use among GPs created an opportunity for Auckland Regional Public Health Service (ARPHS) to develop a general-practitioner–based sentinel surveillance system (GPSURV). ARPHS provides public health surveillance for NZ's greater-Auckland region, which consists of seven districts or cities with a combined population of 1.29 million persons (*6*). GPSURV was designed to monitor community incidence of specified acute syndromes and rates of physician visits for common chronic conditions.

During 2000 and 2001, to test the feasibility of GPSURV, ARPHS undertook a pilot study with 27 volunteer GPs from nine clinics. After 3 months of system implementation, ARPHS evaluated the data collected to assess different aspects of internal validity, including data quality. External validity, or the degree to which observed trends were likely to represent communitywide trends, was examined after 12 months of data had been collected.

Objectives

This paper summarizes the evaluation of the GPSURV pilot with respect to acute syndrome surveillance. The evaluation assessed data capture, the validity of methods used to define illness episodes and denominator populations, the effect of physician participation on self-reported data-quality assessments, and the external validity of influenza-like–illness reporting.

Methods

CDC (*7*) and the World Health Organization (WHO) (*8*) have produced frameworks for evaluating established surveillance systems. The WHO protocol focuses on reviews of paper-based systems and therefore was not applicable to this study. The CDC framework accounts for the interchange of electronic data but is not intended to guide pilot studies, nor does it focus on the outbreak-detection capability of real-time surveillance systems. However, a recently published evaluation framework for evaluating syndromic surveillance systems (*9*) explicitly addresses evaluation of the outbreak-detection function of syndromic surveillance and guided the writing of this paper.

GPSURV Implementation

GPs were recruited from nine clinics whose physicians routinely used standardized terms to record patient assessments. Clinics were distributed across four cities, but locations were not random, and only one clinic was located in central Auckland. The combined population represented by the recruited clinics was 52,960 persons, or approximately 4.1% of the Auckland region's population.

GPSURV was designed to use standardized terms rather than free-text searches to identify patients with target conditions for three reasons. First, clinics were using different information systems, thereby necessitating use of a standard data-extract specification. Second, the project aimed to collect minimal data from clinic information systems with minimal disruption. Third, using standardized terms would likely enhance specificity and simplify analyses.

The standardized terminology used by participating physicians was the Read Codes, Version 2 (*10*). This terminology was widespread in NZ at the time of the pilot because the NZ Ministry of Health had promoted it as the national standard for electronic primary care records. The Read terminology incorporates a conceptual hierarchy within its coding system (*11*). Codes are used as shorthand for clinical terms, and variations of general terms use codes that incorporate the parent term code (e.g., the code for *viral gastroenteritis*, A07y0.00, includes the first two characters of the code for the parent term *intestinal infectious diseases*, A0.00).

Although not ideal for epidemiologic purposes, the Read hierarchy can be used to specify syndromes for surveillance. Three acute infectious clinical syndromes were chosen for the pilot: gastroenteritis, influenza-like illness, and skin infection. Physicians were provided case definitions and corresponding codes (Table 1).

Physicians were advised to record either the specified parent code or a more specific instance of the parent term or corresponding code, as clinically indicated. Data were uploaded daily from clinics via a secure network (Figure 1). A utility within each system enabled the physician or researchers to specify search terms or codes, thus ensuring the system had the flexibility to change conditions under surveillance. A unique patient identifier, the New Zealand National Health Index (NHI) was encrypted by an independent third party before data were transferred to the GPSURV database. Encryption enabled data for matching patients to be linked while maintaining patient privacy.

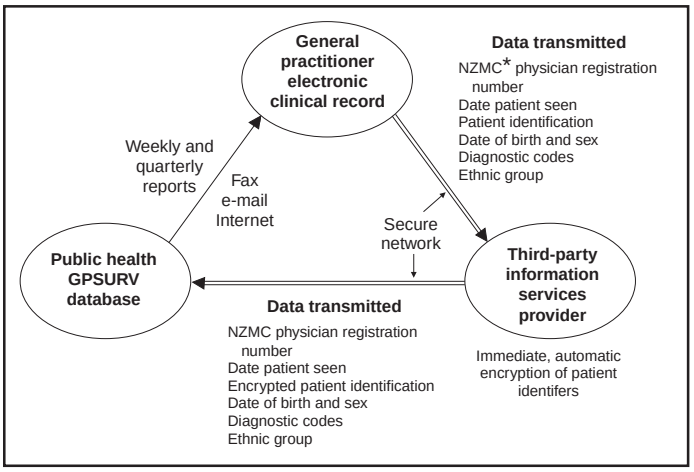
The electronic record system used by a majority of physicians did not allow physicians to distinguish an initial visit from follow-up visits for the same illness episode. Record linkage for this pilot allowed this distinction to be made by using

TABLE 1. Acute syndromes and codes tracked during pilot implementation of a general-practitioner–based syndromic surveillance system (GPSURV) — Auckland, New Zealand, 2000–2001

Syndrome	Read Code*	Definition
Gastroenteritis	A0.00	>3 loose stools/day or vomiting starting within last 5 days and not attributable to any noninfectious cause
Influenza-like illness	H27.00	Acute upper respiratory infection with abrupt onset and two or more of the following: fever, chills, headache, or myalgia
Skin and subcutaneous tissue infection	M0.00	Any presumptive bacterial skin infection, including superficial involvement (e.g., folliculitis) or deep involvement (e.g., cellulitis)

* **Source:** National Health Service Information Authority. The clinical terms version 3 (the Read Codes): incorporation of earlier versions of the Read Codes (the Superset). Birmingham, England: NHS Information Authority, 2000.

FIGURE 1. Information flow within a general-practitioner–based syndromic surveillance system (GPSURV) — Auckland, New Zealand



* New Zealand Medical Council.

an algorithm based on between-visit intervals. Visit records for the same patient and syndrome were categorized as follow-ups if the visit occurred within 8 weeks of a previous visit.

Because GPSURV aimed to compare disease occurrence among clinics and districts, denominators were required to calculate incidence rates. Unlike in the United Kingdom or the Netherlands where patients register with only one physician or clinic, NZ patients can visit as many GP clinics as they wish. This factor increased the difficulty of defining a denominator population. Alternative denominators have been recommended for countries in this situation (*12*). GPSURV defined denominators as active patients (*13*) and used counts of unique patients treated once or more by a participating physician during the previous 2 years. These counts were performed automatically by the clinic information system.

Physician-specific reports providing feedback on recorded illnesses and comparisons with regionwide trends were produced on a weekly and quarterly basis. Reports aggregating data to district and region levels were produced at the same time intervals. No statistical aberration-detection methods were used during the pilot because the focus was on assessing feasibility, data quality, and internal validity.

Data Quality

The sensitivity of GP-based surveillance systems is a function of diagnostic reliability and record completion or data capture. By defining the events under surveillance as conditions or problems identified by participating physicians, GP-based syndromic surveillance (e.g., GPSURV) is less concerned with diagnostic reliability than with record completion and data capture. Given the primary function of outbreak detection through detection of aberrations in time-series data, even incomplete data capture does not necessarily prevent such a system from fulfilling this function, provided data completion does not fluctuate over time. Nevertheless, the completion of recording and event data collection does affect system sensitivity.

Multiple approaches have been taken to assess the completion of term or code recording within electronic GP records. In the UK, where GPs have been required to retain both paper and electronic records, studies have measured completion by comparing those records (*14–16*). When clinics do not retain paper records, this approach is not possible. Direct inspection of electronic records would be possible but expensive and disruptive. Other approaches have included classifying physicians into adequate or inadequate recorders by comparing their incidence and prevalence rates with average values (*17,18*), and by using other data (e.g., diagnoses mentioned in hospital letters) as a proxy for prevalence (*18*). The proxy most commonly used has been data on prescribed medicines, obtained either directly from the clinic (*19*) or from centralized data collections (*20,21*). This method is useful only when medicines are prescribed exclusively for specified conditions.

Survey methods have demonstrated that GPs reliably self-report certain activities (e.g., asking patients about tobacco use [*22*]), and one study used a survey to examine electronic record-keeping within GP clinics in a UK network (*23*). No known studies have been published on the effect of individualized feedback on data quality in GP-based surveillance systems, although certain authors have reported that feedback is likely to have a positive effect (*21,24*).

For this study, a survey method was used to measure the completion of data recording for acute syndromes in the evaluation. Surveys of participating physicians were conducted before and after the first 3-month period of the pilot. For each surveillance condition and consultation type (i.e., initial and follow-up), respondents were asked to estimate the percentage of patient visits for which they recorded a standardized term or code (as opposed to free text).

To assess the effect on the denominator of changing the observation period, counts of active patients seen within previous 6-, 12-, and 18-month periods were compared with the denominator obtained by counting the number of patients attending during the previous 24 months. For evaluating the appropriateness of using an 8-week interval between consecutive visits to identify new illness episodes for the same health problem, distributions of between-visit intervals for matching patient records were examined.

External Validation

Generalizability of measured trends to the region's population would have depended on the geographic distribution of conditions under surveillance and the representativeness of disease events detected at the sentinel sites. A full evaluation of these concerns was beyond the scope of the pilot study. However, an attempt was made to examine external validity of observed trends by comparing data for one syndrome with data from an independent source. The age-sex structure of the study population was also compared with that of the region.

Results

Self-Reported Term-Recording Compliance

A total of 21 physicians completed a baseline survey, and 22 of 27 participating physicians completed a follow-up survey administered 3 months after the pilot began. Not all 22 of those completing the follow-up survey answered each question; nonrespondents for particular questions were removed from analysis. Compliance was defined as recording standardized terms for ≥90% of patient visits.

Of the acute syndromes studied, recording for skin and subcutaneous tissue infection had the greatest compliance (71% of physicians), and influenza-like illness had the least (48% of participants) (Table 2). For all conditions, physicians reported recording standardized terms for follow-up visits less frequently than for first visits.

Of the 21 physicians who had previously returned a baseline survey, 17 completed the follow-up survey. Before-and-after responses from these physicians were compared (Table 3). The number of physicians reporting a between-survey change for each diagnosis and visit type, based on a change of ≥10% in percentage of terms recorded, was determined. Although the number of participants was too limited to test any trends statistically, for all acute syndromes, more increases than decreases occurred.

TABLE 2. Number and percentage of doctors reporting ≥90% compliance in using standardized terms to record patient diagnoses, by syndrome and visit status — Auckland, New Zealand

	First visits		Follow-up visits	
Syndrome	No.	(%)	No.	(%)
Skin infection	15/21	(71)	11/21	(52)
Gastroenteritis	13/22	(59)	9/22	(41)
Influenza-like illness	10/21	(48)	6/20	(30)

TABLE 3. Changes of ≥10% from baseline survey to 3-month follow-up survey in physicians' (n = 17) self-assessed compliance in using standardized terms to record patient diagnoses, by syndrome and visit status — Auckland, New Zealand

	First visits		Follow-up visits	
Syndrome	Increases	Decreases	Increases	Decreases
Skin infection	6	2	7	0
Gastroenteritis	6	2	6	1
Influenza-like illness	7	3	6	2

Categorization of Follow-Up Visits

The percentage of visits for acute syndromes that were categorized as follow-ups (i.e., by using the 8-week between-visit interval) were as follows: 5% for influenza-like illness, 9% for gastroenteritis, and 25% for skin infections. Analysis of pairs of consecutive encounters for skin infections determined that 82% of follow-up visits occurred within 14 days of the previous matching encounter. Only three matching visits for any acute condition were recorded >8 weeks after the previous encounter; however, only 3 months of data were analyzed for matching pairs.

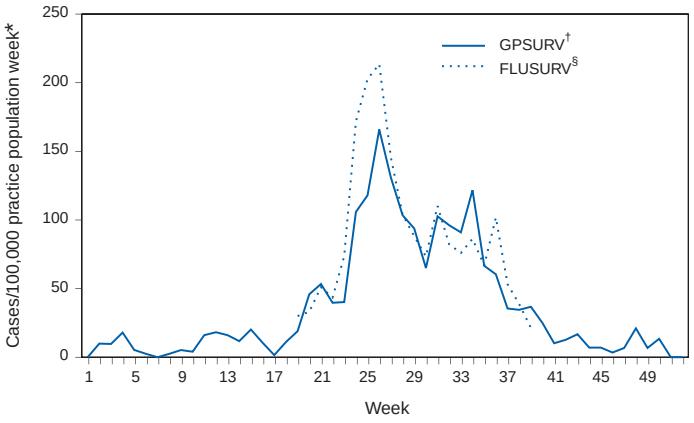
Denominator Populations

The size of the active-patient population increased with the period of observation, as would be expected. The number of active patients counted during a 6-month period was 60% of the 24-month count and 78% and 92% of the 24-month count for 12- and 18-month periods, respectively.

External Validation

Weekly ILI rates were compared with ILI rates as measured by a separate surveillance system. FLUSURV, a surveillance system for influenza and ILI, collects manually recorded data from approximately 40 volunteer GPs in the Auckland region. Participating FLUSURV clinics keep a written tally of patients meeting the WHO case criteria for ILI. Each week, a public health clerical staff member calls clinics to obtain data on the number of new cases. Denominator data for participating physicians are based on physician estimates of total patient population numbers. Only one clinic participated in both GPSURV and FLUSURV. The result of this comparison is illustrated (Figure 2). Although data are collected from a different network of clinics, incidence trends indicate statistical agreement ($t = 1.81$; $p = 0.085$ not significant, 20 degrees of freedom). The first peak of the season appears to be higher in the FLUSURV data, but incidence rates from GPSURV were age-standardized, which is likely to have reduced measured

FIGURE 2. Influenza-like illness incidence as measured by two surveillance systems — Auckland, New Zealand



*GPSURV data age-standardized to Auckland's population.
†General-practitioner–based syndromic surveillance system.
§FLUSURV is a surveillance system for influenza and influenza-like illness.

rates slightly. GPSURV appears to have detected the second substantial peak of the season earlier than FLUSURV.

Given the low self-reported compliance for recording of influenza-like illness, this result appears surprising. One possible explanation is that the initial 3-month pilot period was during late spring and early summer when ILI incidence was likely to be minimal. Thus, GPs might have been more likely to use alternative terms (e.g., *hay fever*) to record syndromes with upper respiratory symptoms.

Comparison of age-sex structures demonstrated that approximately all age-sex bands of the study population were within 2% of comparable percentages for the regional population. An exception was the <10 years age group; when compared with the regional age-sex distribution, this age group comprised 6% more of the study population for males and 5% more for females.

Discussion

This study examined the validity of disease-incidence measures based on the collection and analysis of clinical data routinely recorded by a network of volunteer family physicians. The study's findings indicate that, despite participant variability in data recording and problems with defining denominator populations, the incidence of common acute syndromes can be monitored at least as successfully by using standardized clinical-term data from selected GP clinics as by using manual methods. However, the sensitivity of this method will depend on the frequency of the syndrome under surveillance. For less common conditions, a larger sample of GPs would be required. Similarly, geographic variations in disease incidence probably would not be detected without increasing the geographic spread of participating clinics.

The study's findings indicate that the algorithm used to classify follow-up visits is probably working effectively. In the case of influenza-like illness, however, only 5% of visits were actually follow-ups. Thus, misclassification of these as first visits would have had minimal impact on measured rates. The evaluation indicated that approximately 80% of patients treated over a 2-year period would be counted over 12 months. The effect of changing observation period for defining the denominator would be more complicated given possible changes in age structure at different time periods. Nevertheless, if such a denominator were to be used for further surveillance, a 12-month observation period would probably suffice.

Clinic participation in the pilot appeared to have a limited but positive impact on data quality. This might have resulted from regular feedback provided to physicians in weekly and quarterly reports. Other aspects of participating in the project might also have contributed to improvements in data quality; for example; physicians might have gained an increased awareness of the public health benefits of providing valid data. However, observed fluctuations in the recording of standardized terms raise the possibility that this approach might be prone to artefactual aberrations in time-series data, and participating GPs would need to maintain consistency in their recording behavior for ongoing surveillance.

References

1. Adelstein AM. Policies of the Office of Population Censuses and Surveys: philosophy and constraints. Br J Prev Social Med 1976;30:1–10.
2. Pinsent RJ. The primary observer. Ecol Dis 1982;1:275–9.
3. Rigby M, Roberts R, Williams J, et al. Integrated record keeping as an essential aspect of a primary care led health service. BMJ 1998;317:579–82.
4. Thakurdas P, Coster G, Gurr E, Arrol B. New Zealand general practice computerisation; attitudes and reported behaviour. N Z Med J 1996;109:419–22.
5. New Zealand Ministry of Health. From strategy to reality—the WAVE project. Wellington, New Zealand: New Zealand Ministry of Health, 2002. Available at http://www.moh.govt.nz/moh.nsf/0/f34f8959738e992ccc256af400177998?OpenDocument.
6. Statistics New Zealand. Subnational population estimates (as at June 2003) — media release. Available at http://www.stats.govt.nz/domino/external/pasfull/pasfull.nsf/web/Media+Release+Subnational+Population+Estimates+As+at+30+June+2003?open.
7. CDC. Updated guidelines for evaluating public health surveillance systems: recommendations from the guidelines working group. MMWR 2001;50(No. RR-13).
8. World Health Organization (WHO). Protocol for the evaluation of epidemiological surveillance systems. Geneva, Switzerland: WHO, 1997. Available at http://www.who.int/emc-documents/surveillance/whoemcdis972c.html.

9. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR 2004;53(No. RR-5):1–11.
10. National Health Service Information Authority. The clinical terms version 3 (the Read Codes): incorporation of earlier versions of the Read Codes (the Superset). Birmingham, England: National Health Service Information Authority, 2000.
11. Britt H, Miller G. A critical review of the final report of the general practice coding jury. Sydney, Australia: Family Medicine Research Centre, University of Sydney, 2000. Available at http://www.fmrc.org.au/classifi.htm.
12. Schlaud M, Brenner MH, Hoopman M, Schwartz FW. Approaches to the denominator in practice-based epidemiology: a critical overview. J Epidemiol Community Health 1998;52:13S–19S.
13. De Loof J. Practice size: a fraction of the yearly attending group as practice size indicator. Gen Pract Int 1983;12:127–8.
14. Boydell L, Grandidier H, Rafferty C, McAteer C, Reilly P. General practice data retrieval: the Northern Ireland project. J Epidemiol Community Health 1995;49:22–5.
15. Neal RD, Heywood PL, Morley S. Real world data-retrieval and validation of consultation data from four general practices. Fam Pract 1996;13:455–61.
16. Pringle M, Ward P, Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. Br J Gen Pract 1995;45:537–41.
17. Bartelds AI. Validation of sentinel data. Gesundheitswesen 1993;55:3–7.
18. Jick H, Jick S, Derby L. Validation of information recorded on general practitioner based computerised data resources in the United Kingdom. BMJ 1991;302:766–8.
19. Vernon JG. Ensuring the accuracy and completeness of clinical data on general practice systems. J Inform Prim Care 1998;Nov:18–9.
20. Bruno G, Bargero G, Vuolo A, Pisu E, Pagano G. A population-based prevalence survey of known diabetes mellitus in northern Italy based upon multiple independent sources of ascertainment. Diabetologia 1992;35:851–6.
21. Gribben B, Coster G, Pringle M, Simon J. Non-invasive methods for measuring data quality in general practice. N Z Med J 2001;114:30–2.
22. Eccles M, Ford GA, Duggan S, Steen N. Are postal questionnaire surveys of reported activity valid? An exploration using general practitioner management of hypertension in older people. Br J Gen Pract 1999;49:35–8
23. Lawrenson RA, Coles G, Walton K, Farmer RDT. Characteristics of practices contributing to the MediPlus database and the implications for its use in epidemiological research. J Inform Prim Care 1998;Nov:14–18.
24. Chauvin P, Valleron AJ. Participation of French general practitioners in public health surveillance: a multidisciplinary approach. J Epidemiol Community Health 1998;52:2S–8S.

National Symptom Surveillance Using Calls to a Telephone Health Advice Service — United Kingdom, December 2001–February 2003

Duncan L. Cooper,[1] G. Smith,[1] M. Baker,[2] F. Chinemana,[3] N. Verlander,[4] E. Gerard,[3] V. Hollyoak,[5] R. Griffiths[6]
[1]*Health Protection Agency West Midlands, Birmingham, England;* [2]*Royal College of General Practitioners, London, England;*
[3]*NHS Direct Hampshire and the Isle of Wight, Southampton, England;* [4]*Health Protection Agency, Colindale, England;*
[5]*Health Protection Agency North East, Newcastle-Upon-Tyne, England;* [6]*Government Office for the West Midlands, Public Health, Birmingham, England*

Corresponding author: Duncan L. Cooper, Health Protection Agency West Midlands, Floor 2, Lincoln House, Heartlands Hospital, Birmingham, England B9 5SS. Telephone: 0121-773-7077; Fax: 0121-773-1407; E-mail: Duncan.Cooper@HPA.org.uk.

Abstract

Introduction: *Recent terrorist activity has highlighted the need to improve surveillance systems for the early detection of chemical or biologic attacks. A new national surveillance system in the United Kingdom (UK) examines symptoms reported to NHS Direct, a telephone health advice service.*

Objectives: *The aim of the surveillance system is to identify an increase in symptoms indicative of early stages of illness caused either by a deliberate release of a biologic or chemical agent or by common infections.*

Methods: *Data relating to 10 key syndromes (primarily respiratory and gastrointestinal) are received electronically from 23 call centers covering England and Wales. Data are analyzed daily and statistically significant excesses, termed* exceedances*, in calls are automatically highlighted and assessed by a multidisciplinary team.*

Results: *During December 2001–February 2003, a total of 1,811 exceedances occurred, of which 126 required further investigation and 16 resulted in alerts to local or national health-protection teams. Examples of these investigations are described.*

Conclusion: *Surveillance of call-center data has detected substantial levels of specific syndromes at both national and regional levels. Although no deliberate release of a biologic or chemical agent has been detected thus far by this or any other surveillance system in the UK, the NHS Direct surveillance system continues to be refined.*

Introduction

Recent terrorist activity has highlighted the need to improve surveillance systems for early detection of chemical or biologic attacks. A new United Kingdom (UK) surveillance system operated by the National Health Service (NHS) examines syndromes reported to NHS Direct, a national telephone health advice service (*1*). NHS Direct is a nurse-led helpline that provides the public with rapid access to professional health advice and information about health, illness, and NHS (*2*). NHS Direct is open 365 days/year and serves the entire population of England and Wales. NHS Direct nurses use clinical decision support software, the NHS Clinical Assessment System (NHS CAS), to respond to calls. NHS CAS contains >200 clinical algorithms that form tree-like structures of questions relating to the symptoms of the person about whom the call is made. The majority of calls result in a call outcome, either advice for self-care, a routine doctor referral, an urgent doctor referral, an emergency department (ED) referral, or a paramedic dispatch. Data derived from NHS Direct can be of value in disease surveillance.

When a deliberate release of a harmful agent causes an illness with an extended, mild, prodromal phase, certain persons are likely to contact NHS Direct before contacting any other health service. These contacts provide an opportunity to identify an increase in illness before it is identified by other primary- or secondary-care services. The aim of the surveillance system described here is to identify an increase in symptoms indicative of the early stages of illness caused by the deliberate release of a biologic or chemical agent, or by common infections. This project builds on existing surveillance of influenza-like-illness and gastrointestinal symptoms that uses NHS Direct call data (*3*–*5*).

Methods

Daily call data relating to 10 syndromes (cold/"flu," cough, diarrhea, difficulty breathing, double vision, eye problems, lumps, fever, rash, and vomiting) are received electronically by the Health Protection Agency (HPA) from all 23 NHS Direct sites in England and Wales. (Beginning April 2003,

eye problems replaced food poisoning as a syndrome category.) These data are analyzed daily by a surveillance team established in November 2001 and consisting of HPA and NHS Direct staff. The 10 syndromes were selected as indicative of the early stages of illnesses caused by biologic or chemical weapons. Data are categorized by NHS Direct site, symptom, age group, and call outcome. NHS Direct nurses triage rather than diagnose illness in callers.

Upper confidence limits (99.5% level) of calls for each syndrome, as a percentage of daily total calls, are constructed for each NHS Direct site. These confidence limits are derived from a standard formula for percentages (*6*) with the baseline numbers of total calls and symptom calls adjusted for seasonal effects (winter: December–February; spring: March–May; summer: June–August; autumn: September–October). A daily percentage of calls exceeding the 99.5% upper confidence limit is termed an *exceedance.*

In addition to confidence-interval analyses, control charts are constructed for five of the 10 syndromes (cold/"flu", cough, fever, diarrhea, and vomiting) at the 10 NHS Direct sites serving five major urban centers (London, Manchester, Leeds, Birmingham, and Newcastle). Baselines for the control charts are calculated by assuming that the number of syndromic calls follows a Poisson distribution with the total number of calls as an offset. A model is fitted to each site and each symptom separately, using data from December 2001. Each model always includes a public holiday and seasonal term. When necessary, a day of the week (weekday, Saturday, or Sunday) and a linear long-term trend factor are also fitted. Scaling is performed to account for overdispersion.

A normal approximation is not used to calculate the 99.5% upper control-chart limit of calls for each syndrome as it yields a greater percentage of exceedances than would be expected (i.e., >0.5%). Instead, a transformation to approximate normality with zero mean is performed and transformed back to the original scale. For control charts, the following formula for the 99.5% upper limit of syndromic calls is used:

$$\left(\sinh\left(\frac{z_{\alpha}/2+\sqrt{N-0.5}\,\sinh^{-1}\sqrt{p}}{\sqrt{N-0.5}}\right)\right)^2(N-0.75)-3/8$$

where N is given by the expected value divided by 1 less than the scale parameter; p is equal to the scale parameter minus 1; and z_{α} is the $100^{*}(1-\alpha)^{\text{th}}$ centile of the normal distribution. Ad-hoc choices of z are made to achieve the desired number of purely random exceedances (0.5%). The upper 99.5% control-chart limit of calls for each syndrome, as a percentage of total calls, is calculated daily.

Exceedances in calls for any of the 10 syndromes are automatically highlighted (for the confidence-interval and control-chart method) and assessed by the surveillance team (stage 1). If no reasonable explanation for the exceedance can be found, additional line listings of call details (including the call identification [ID] number and the caller's residential postcode) are requested for the date of the exceedance and for the current date (stage 2). The call ID number, which should be a unique number, is used to identify duplicate call records. Requesting calls for the current date (which will be complete up to the hour the request is made) is critical for monitoring what might be an evolving situation. If current call data indicate persistent statistical excesses (i.e., exceeding the 99.5% upper confidence limit) for a particular syndrome, a geographic information system can be used to map call data, although this procedure is not routine for all exceedances.

NHS Direct sites can export calls to other sites during periods of peak demand. A percentage of calls handled by NHS Direct sites (usually <10%) might therefore originate from outside their catchment areas. Catchment areas are based on local telephone area codes.

When the surveillance team determines that information provided by line listings necessitates further investigation (stage 2), the team generates an alert by passing call information to the relevant local or national public health teams for follow-up (stage 3). If the exceedance is suspected to represent a serious public health threat, the NHS Direct medical adviser can contact callers to obtain further clinical information. Weekly bulletins summarizing NHS Direct call activity are disseminated to relevant local and national health-protection colleagues.

Results

When the surveillance of 10 syndromes began in December 2001, call data were collected from approximately one-half of the total 23 NHS Direct sites. Subsequently, the mean number of NHS Direct sites providing daily call data increased from 12 sites in December 2001 to all 23 during October 2003 (Figure 1). A sudden decrease in the number of sites providing call data in July 2002 was attributable to surveillance staff absences. No constant differences in the level of data provision existed between the regions.

During December 2001–February 2003, a total of 1,811 confidence-limit exceedances occurred (stage 1), of which 126 (7%) required further investigation (stage 2) and 16 (1%) resulted in alerts (stage 3) (Table). Exceedance investigations

FIGURE 1. Number of NHS Direct sites providing daily call data for syndromic surveillance — England and Wales, December 2001–October 2003

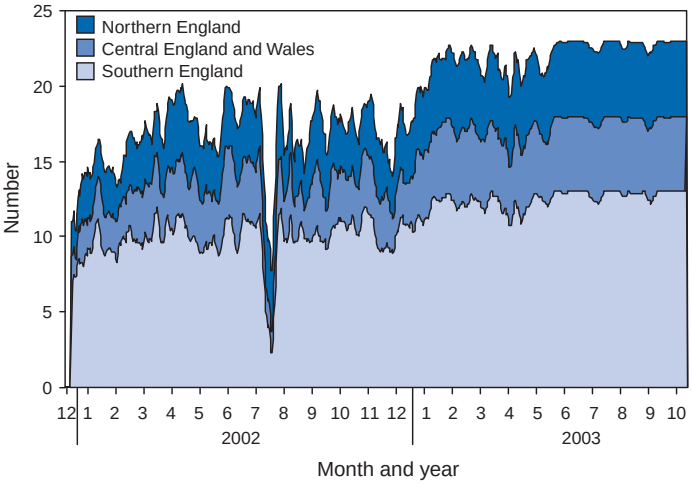


TABLE. Number of exceedances* (stage 1), exceedances investigated (stage 2), and alerts (stage 3), based on calls to 23 NHS Direct telephone advice line, by syndrome — England and Wales, December 2001–February 2003

	Stage 1	Stage 2	Stage 3
Syndrome	No. of exceedances	Exceedances investigated	Alerts†
Fever	328	23	4
Cough	279	4	0
Cold/"flu"	185	5	0
Vomiting	182	28	4
Double vision	180	2	0
Food poisoning	180	0	0
Lumps	142	14	1
Diarrhea	137	14	4
Difficulty breathing	123	22	2
Rash	75	13	1
Total	**1,811**	**126**	**16**

* An exceedance is a statistically significant excess of calls beyond the 99.5% upper confidence limit.
† Stage 3 exceedances have been described previously (*1*).

did not progress to alerts when 1) the observed increase in calls was a single-day exceedance only (46% of stage 2 investigations), 2) duplicate call records caused the exceedance (20%), or 3) the call data did not cluster geographically (15%).

An overview of the national daily numbers and percentages of calls for four syndromes is provided (Figures 2 and 3). As expected, a seasonal pattern of higher activity during the winter emerged for certain syndromes (e.g., cold/"flu" and vomiting), both in the numbers and percentages of calls. The numbers of calls for all 10 syndromes increased during weekends and on public holidays, when many routine primary-care services are closed. The percentage of calls regarding certain syndromes also increased during weekends (e.g., rash) and on public holidays (e.g., cough and vomiting).

During early August 2002, daily exceedances of callers reporting difficulty breathing occurred at eight of nine NHS Direct sites within the Thames basin and East Anglia. These exceedances accompanied a general increase in callers reporting difficulty breathing in eastern parts of Central and Southern England (Figure 4). This increase was preceded by elevated ozone levels and thunderstorms in this part of England. The timing and effect of these climatic and environmental conditions on call data are being analyzed. This detection of a sudden increase in calls has also generated new operational links between environmental health professionals in the Health Protection Agency and other central government departments.

In January 2003, traces of the chemical poison ricin were found in a North London apartment. In response, the surveillance team was asked to enhance symptom surveillance of call data collected from the five NHS Direct sites in London. Data were collected on four syndromes (Figure 5) and updated every 2 hours. Call data were also mapped by place of residence, as this might have provided the first clue that a deliberate release could have occurred at a particular location. NHS Direct data and other data sources have demonstrated no evidence thus far of any deliberate release of biologic or chemical agents within the UK.

Conclusions

This syndromic surveillance system is the only such system covering the entire population of England and Wales. Although the majority of exceedances do not result in subsequent investigation, when action is taken, health-protection teams are usually informed within 24–48 hours of calls being received by NHS Direct. Only 2 years of data have been collected, and the establishment of baselines and refinement of statistical methodology continue. Although no deliberate release of chemical or biologic agents has been detected, this surveillance system has detected elevated levels of activity in specific symptoms at both national and regional levels.

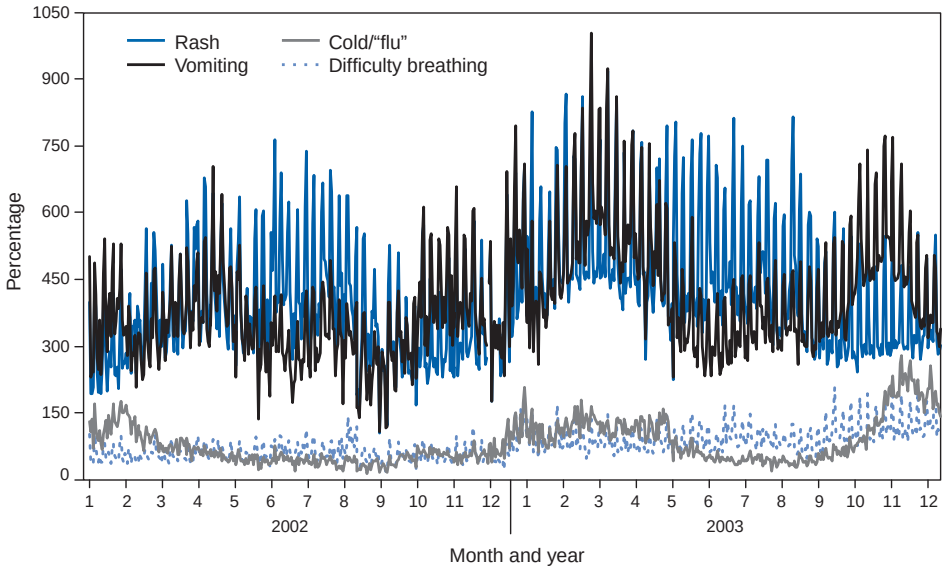
After an initial period in which duplicate call records led to investigation of exceedances that later proved spurious, data quality was improved. The surveillance now covers the entire population of England and Wales and is conducted daily. Although geographic locations of calls are available on request, the geographic resolution of the initial daily analysis (to identify exceedances) is at a site level. This means localized, subsite-level outbreaks might be overlooked. The surveillance team is investigating ways to collect and analyze call data by smaller geographic units.

Consistent and timely data returns have been achieved by concentrating on collecting routine NHS Direct data with minimal disruption to the data providers' work patterns and by ensuring continual feedback to all staff within the surveillance network. The annual operating cost of the surveillance system is <$200,000. Providing these surveillance data is now an integral part of NHS Direct's objectives and is a priority within the service; this was essential in January 2003 when real-time surveillance was needed to address a perceived threat of a ricin release. Analysis of the surveillance data in that instance helped to determine that no deliberate release had occurred.

A recent government strategy document announced a three-fold expansion of NHS Direct call-handling capacity over the next 3–4 years, with an equivalent increase in call volumes also expected (from 6 to 16 million calls/year in England) (*7*). This volume compares with approximately 14 million visits/year to EDs in England (*8*) and 190 million consultations with primary-care physicians (*9*). The increase in NHS Direct call volumes should improve the representativeness of the call data and the potential for early identification of disease outbreaks.

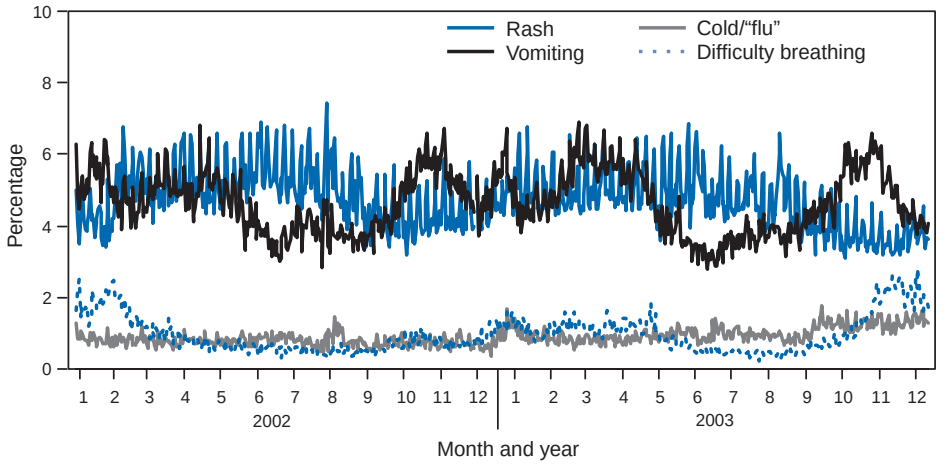
The value of surveillance of NHS Direct data in complementing existing surveillance for common infections (e.g., influenza) is being established (*1,3,4*). Whether the NHS Direct syndromic surveillance system will ultimately provide early warning of a chemical or biologic attack will only be demonstrated if such an event occurs.

FIGURE 2. National daily numbers* of NHS Direct calls for cold/"flu," difficulty breathing, vomiting, and rash



* Includes back data collected retrospectively for July 2002.

FIGURE 3. National daily percentages of NHS Direct calls for cold/"flu," difficulty breathing, vomiting, and rash



Acknowledgments

NHS Direct sites and the NHS Direct Health Intelligence Unit provided surveillance data for this research.

References

1. Baker M, Smith GE, Cooper DL, et al. Early warning and NHS Direct: a role in community surveillance? J Pub H Med 2003;24:362–8.
2. NHS Direct Online. Available at http://www.nhsdirect.nhs.uk.
3. Harcourt S, Smith GE, Hollyoak V, et al. Can calls to NHS Direct be used for syndromic surveillance? A pilot at three sites using influenza as an example. Commun Dis Public Health 2001;4:178–82.
4. Cooper DL, Smith GE, Hollyoak VA, Joseph CA, Johnson L, Chaloner R. Use of NHS Direct calls for the surveillance of influenza—a second year's experience. Commun Dis Public Health 2002;5:132–6.

FIGURE 4. Increase in the percentage of callers to NHS Direct sites who reported difficulty breathing — Eastern England, July–August 2002

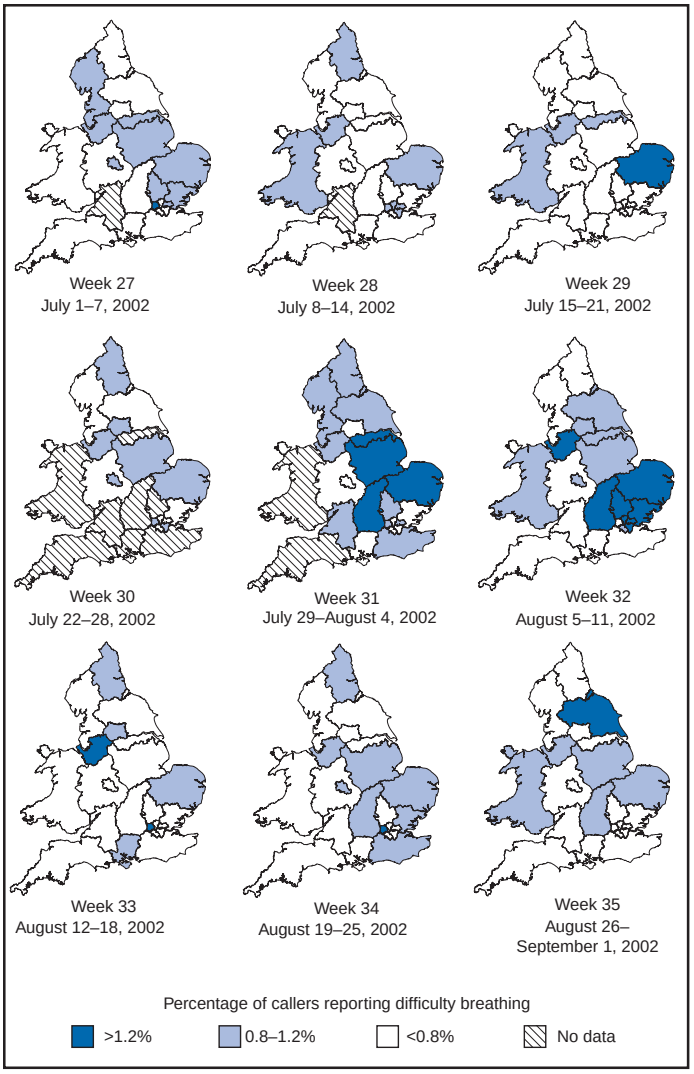
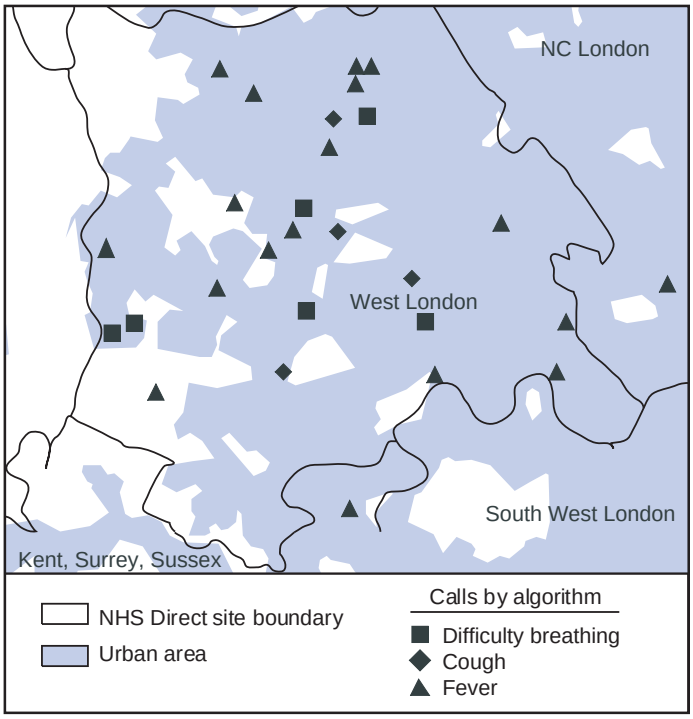


FIGURE 5. Calls to NHS Direct for difficulty breathing, cough, and fever, by residential postcode — West London, England, January 7, 2003



5. Cooper DL, Smith GE, O'Brien SJ, Hollyoak V, Long S. What can analysis of calls to NHS Direct tell us about the epidemiology of gastrointestinal infections in the community? J Infection 2003;46:101–5.
6. Armitage P, Berry G. Statistical method in medical research (2nd ed.). Oxford, England: Blackwell Scientific Productions, 1987.
7. Directorate of Access and Choice. Developing NHS Direct: a strategy document for the next three years. London, England: England Department of Health, 2003.
8. England Department of Health. Hospital activity statistics. Available at http://www.performance.doh.gov.uk/hospitalactivity/index.htm.
9. Rowlands S, Moser K. Consultation rates from the general practice research database. Br J Gen Pract 2002;52:658–60.

Field Investigations of Emergency Department Syndromic Surveillance Signals — New York City

Linda Steiner-Sichel, J. Greenko, R. Heffernan, M. Layton, D. Weiss
New York City Department of Health and Mental Hygiene, New York, New York

Corresponding author: Don Weiss, Director of Surveillance, Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, 125 Worth St, Box 22A, New York, NY 10013. Telephone: 212-442-5398; Fax: 212-676-6091; E-mail: DWeiss@health.nyc.gov.

Abstract

Introduction: *The New York City (NYC) Department of Health and Mental Hygiene (DOHMH) has operated a syndromic surveillance system based on emergency department (ED) chief-complaint data since November 2001. This system was created for early detection of infectious-disease outbreaks, either natural or intentional. However, limited documentation exists regarding epidemiologic field investigations conducted in response to syndromic surveillance signals.*

Objective: *DOHMH conducted field investigations to characterize syndromic surveillance signals by person, place, and time and to determine whether signals represented true infectious-disease outbreaks.*

Methods: *A DOHMH physician reviews ED-based syndromic surveillance results daily to look for signals. When necessary, field investigations are conducted and consist of a review of the patient line list, telephone interviews with hospital staff, chart reviews, interviews with patients, and collection and testing of specimens.*

Results: *In November 2002, a series of citywide signals for diarrhea and vomiting syndromes, which coincided with institutional outbreaks consistent with viral gastroenteritis, prompted DOHMH to send mass e-mail notification to NYC ED directors and institute collection of stool specimens. Three of four specimens collected were positive for norovirus. In December 2002, DOHMH investigated why an ED syndromic signal was not generated after 15 ill patients were transferred to a participating ED during a gastrointestinal outbreak at a nursing home. Field investigation revealed varying chief complaints, multiple dates of ED visits, and a coding error in a complementary DOHMH syndromic system, and confirmed a seasonal norovirus outbreak. During March 2003, the system generated a 4-day citywide respiratory signal and a simultaneous 1-day hospital-level fever signal in a predominantly Asian community. In those instances, epidemiologic investigation provided reassurance that severe acute respiratory syndrome was not present.*

Conclusion: *Detailed field investigations of syndromic signals can identify the etiology of signals and determine why a given syndromic surveillance system failed to detect an outbreak captured through traditional surveillance. Validation of the utility of syndromic surveillance to detect infectious-disease outbreaks is necessary to justify allocating resources for this new public health tool.*

Introduction

The New York City (NYC) Department of Health and Mental Hygiene (DOHMH) has operated a syndromic surveillance system based on emergency department (ED) chief-complaint data since November 2001. By November 2003, 44 of NYC's 67 EDs participated in this system, thereby capturing 80% of all NYC ED patient visits. This paper describes three investigations of ED syndromic signals that required in-depth fieldwork to characterize the syndromic signals by person, place, and time and to determine whether the signals represented true infectious-disease outbreaks.

Methods

Overview of the DOHMH ED Syndromic Surveillance System

The methods used for obtaining ED data for syndromic surveillance are described in detail elsewhere (*1*). Briefly, the DOHMH syndromic surveillance system receives ED data through daily electronic transmission of files containing free-text chief complaint, age, sex, residential zip code, and date and time of ED admission. A computer algorithm codes the free-text chief complaint into one of four syndromes: respiratory, fever (includes influenza-like illness), diarrhea, and vomiting. Daily statistical analyses evaluate citywide temporal trends and spatial clustering, by hospital and residential zip code, for respiratory and fever syndromes in persons aged ≥13 years and for diarrhea and vomiting syndromes in patients of

all ages. A signal is defined as a statistically significant increase in ED visits for a syndrome over a predetermined baseline. The temporal scan statistic (*2*), which compares the ratio of visits for one of the four syndromes to visits for other reasons (i.e., those that do not fall into an infectious-disease syndrome category) during the previous 1, 2, or 3 days to a 14-day baseline (*2*), is used for citywide analysis. A modified spatial scan statistic and a 14-day baseline are used for spatial analyses, adjusting for both purely temporal (e.g., a citywide increase in syndrome visits) and purely spatial variation (e.g., consistently higher syndrome visits within a particular zip code) (*3*). Significance is set at $p<0.01$, a level selected to manage the number of epidemiologic investigations while minimizing the probability of missing a real event. The extent of an investigation depends on the syndrome, the size and geography of the signal, its overlap with other syndromes, signals generated by complementary systems, and the current level of concern (e.g., during certain high profile events).

Steps in Spatial Signal Investigations

An analyst and a medical epidemiologist review the output daily and note any statistically significant citywide or spatial signals. Signal investigations are conducted according to the following priority: 1) spatial signals for fever and respiratory, 2) citywide signals for fever and respiratory, 3) spatial signals for diarrhea and vomiting, and 4) citywide signals for diarrhea and vomiting. Compared with a citywide signal, a spatial signal has limited geographic dimensions (i.e., ≥1 neighboring hospitals or zip codes) and a focused epidemiologic investigation. All spatial signals are investigated to varying degrees, depending on factors mentioned previously. A line list of ED visits captured in the signal is reviewed for duplicate entries and for typographic or coding errors. Descriptive statistics are generated on age, sex, residential zip code, and time of admission to examine patterns among the patients. Chief-complaint data are subcategorized to further uncover similarities among the patients.

Hospitals involved in the signal are then called to request interim data and to assess the volume and severity of patients visiting the ED during that time. ED staff are asked about the syndrome of interest and about any other severe or unusual clusters or similarities among patients. Speaking with hospital staff is a valuable component of a signal investigation; it provides a direct assessment of current ED activity and heightens the clinician's awareness of the specified syndrome. Speaking with physicians who worked the previous day is often helpful because they are usually more familiar with the ED visits responsible for the signal.

Data from complementary surveillance systems (e.g., ambulance calls and pharmaceutical sales) are reviewed for any signals occurring within the same syndrome category. Certain hospitals also provide an interim 12-hour ED chief-complaint log for the current day's data, which is coded and reviewed to evaluate whether the syndrome trend is continuing. If ongoing illness exists, DOHMH might ask ED physicians to lower their threshold for ordering certain diagnostic tests (e.g., blood cultures, stool cultures, chest radiographs, or rapid influenza tests). When necessary, patients are called at home to inquire about their condition. If evidence indicates that the outbreak is continuing, DOHMH staff are sent to EDs to interview patients (or their families), review charts of ED visits and hospital admissions by using a standardized chart-abstraction tool, and assist with collection and transport of specimens to the DOHMH public health laboratories (PHL).

Results

During November 15, 2001–November 14, 2003, a total of 142 citywide signals occurred on 111 surveillance days, including 22 respiratory syndrome signals and 33 fever syndrome signals during peak influenza season, and 25 diarrhea syndrome signals and 28 vomiting syndrome signals during the autumn and winter viral gastroenteritis seasons. Hospital-level signals included 51 signals for respiratory and fever syndromes and 58 signals for diarrhea and vomiting syndromes. At the zip-code level, 39 signals for respiratory and fever syndromes and 50 for diarrhea and vomiting syndromes occurred. The following section describes three in-depth epidemiologic field investigations conducted in response either to a syndromic surveillance signal or to the lack of a signal during an otherwise reported outbreak.

Investigation 1

Background. In October 2002, a series of citywide signals for diarrhea and vomiting syndromes coincided with institutional outbreaks clinically consistent with acute viral gastroenteritis. A hospital-level spatial signal for diarrhea syndrome involving two hospitals (A and B) occurred on both October 29 and 30 for both hospitals (Table 1).

Response. DOHMH sent an e-mail message to ED directors of hospitals participating in syndromic surveillance, alerting them to the citywide increase in gastrointestinal illness (GI) and asking them to lower their threshold for diagnostic testing, collect viral stool specimens, and identify common exposures or unusual circumstances among ED patients.

Hospitals A and B were involved in both days of this hospital-level spatial signal. Infection-control nurses at both hospitals

TABLE 1. Syndromic surveillance signals for diarrhea syndrome — New York City, October 29–30, 2002

Hospital	October 29 Observed cases	October 29 Expected cases	October 29 RR*	October 29 Excess	October 30 Observed cases	October 30 Expected cases	October 30 RR	October 30 Excess
A	10	4.3	2.3	6	9	4.0	2.3	5
B	12	5.2	2.3	7	7	2.7	2.6	4
C	1	0.3	3.3	1	—	—	—	—
D	—	—	—	—	3	2.1	1.4	1
E	—	—	—	—	4	1.8	2.2	2
F	—	—	—	—	7	4.2	1.7	3
Total	**23**	**9.8**	**2.3**	**14**	**30**	**14.8**	**2.0**	**15**

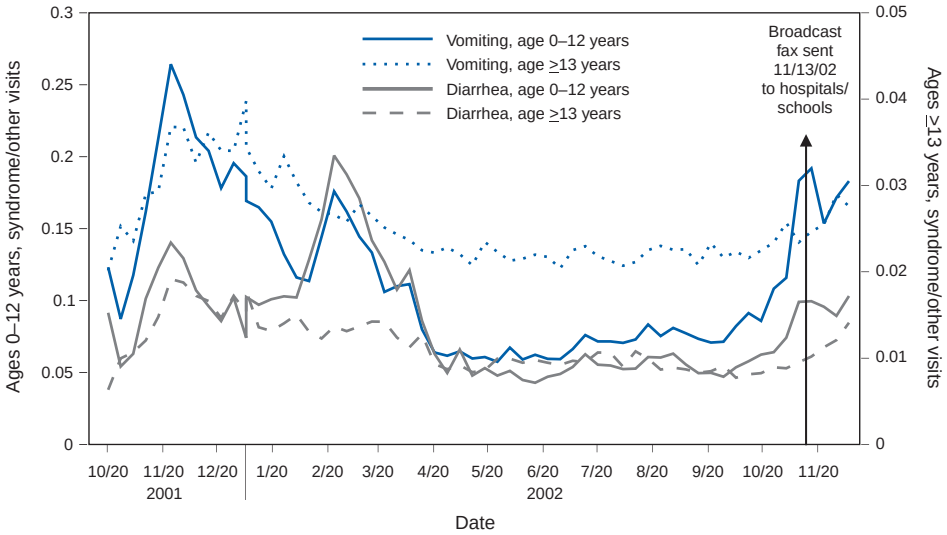
* Relative risk.

were contacted and patient line lists examined. Further analyses determined that the citywide increase in GI primarily affected young children (Figure). Staff were sent to Hospital A's pediatric ED to interview patients and collect stool specimens (for bacterial, ova and parasite, and viral testing) to determine an etiology. In addition, a health alert (http://www.nyc.gov/html/doh/html/cd/02md37.html) was sent to hospitals and schools citywide via broadcast fax and e-mail.

Findings. At Hospital A, two stool specimens were obtained on site and two collection kits were sent home with parents of ED patients and later retrieved and delivered to PHL through pre-arranged transportation. Three of the four specimens collected were positive for norovirus.

Norovirus was widespread throughout multiple parts of the United States, including New York City, during the winter of 2002. During November 2002–mid-January 2003, DOHMH received reports of 66 outbreaks of gastroenteritis epidemiologically consistent with norovirus infection affecting approximately 1,700 persons. Outbreak settings were primarily nursing homes and long-term care and rehabilitation facilities but also included hospitals, restaurants, a homeless shelter, and a school. In all, 29 stool specimens were tested, of which 19 (66%) were positive for norovirus (*4*). No epidemiologic links among patients presenting to Hospitals A and B were uncovered during the investigation. Thus, syndromic surveillance was an early indicator of citywide GI consistent with seasonal trends of norovirus.

FIGURE. Weekly emergency department visits for vomiting and diarrhea syndrome, by age group — New York City, October 2001–December 5, 2002



Investigation 2

Background. In December 2002, the DOHMH epidemiologist for foodborne illness received a call from a nursing-home director reporting 80 (of 320) residents with GI. On December 1, 2002, 25 nursing-home residents were transported by ambulance to four local hospital EDs. Although two of the four hospitals were participants in NYC's ED syndromic surveillance system, one of which (Hospital G) received 15 of the nursing home patients, NYC's ED syndromic surveillance system did not detect this GI cluster at Hospital G.

To investigate the system's failure to signal, DOHMH conducted a retrospective, age-specific (persons aged >60) spatial analysis, which did detect a GI cluster. Five hospitals, including Hospital G, were included in the cluster; however, none of the 15 cases transferred to Hospital G had been captured (Table 2).

Response. Hospital charts for the 15 patients reportedly transferred to Hospital G by ambulance on December 1, 2002, were requested for review, of which 13 (87%) were available. According to ED records, only nine of these 13 patients were treated at Hospital G's ED on December 1, 2002; the other four patients were brought in on December 2 or 3, 2002. Chief complaints for two of the nine patients were not for a gastrointestinal illness but for atrial fibrillation and syncope, which were either the primary or only reason for the ED visits. Chief complaints noted in the medical records of the remaining seven patients were consistent with gastroenteritis.

Because the nursing-home residents were transferred by ambulance to the EDs, DOHMH reviewed the emergency medical services (EMS)

TABLE 2. Retrospective spatial analysis of diarrhea syndrome among persons aged ≥60 years — New York City, December 1, 2002

Hospital	Observed cases	Expected cases	RR*	Excess
G	0	0.3	0.0	0
H	2	0.4	5.0	2
I	1	0.3	3.3	1
J	2	0.1	20.0	2
K	2	0.2	10.0	2
Total	**7**	**1.3**	**5.4**	**6**

* Relative risk.

ambulance-dispatch log, a complementary syndromic surveillance system. The data provided by the ambulance-dispatch log includes job number, date and time of call, call type, and the EMS chief complaint. This information was obtained from the ambulance call report, a copy of which was found in patient ED charts.

A detailed review of the EMS database indicated that eight of the nine patients were transferred by EMS from the nursing home to Hospital G's ED on December 1, 2003; the other patient was transported by private ambulance. Three EMS call types were documented for these eight patients: *sick*, *unconscious*, and *multiple casualty incident*. A single multiple casualty incident call accounted for five of the patients transported to the hospital. None of the EMS call types were related to GI.

Findings. The ED at Hospital G had incorrectly entered the nonspecific EMS call types indicated on the ambulance call report as the chief complaints instead of as the patients' subjective complaints given to EMS providers. Thus, these nonspecific call types, none of which indicated acute gastroenteritis, were received electronically by syndromic surveillance, instead of the chief complaints noted in the medical records. Therefore, critical information that would typically be captured and coded into a key syndrome was lost. Meanwhile, a concurrent foodborne-outbreak investigation determined that 10 of 11 stool specimens collected from ill nursing-home residents were positive for norovirus.

Investigation 3

Background. During March 2003, simultaneous citywide respiratory (4 days) and fever (3 days) signals occurred (Table 3). These signals coincided with the World Health Organization's global alert on March 12, 2003, about cases of atypical pneumonia, an outbreak later determined to be severe acute respiratory syndrome (SARS).

On March 16, 2003, a spatial signal for fever syndrome occurred in a predominantly Asian community for both the hospital (observed n = 23/expected n = 7.4; p = 0.001) and zip code (observed n = 9/expected n = 0.9; p = 0.002) analyses. Within the hospital signal, Hospital L (observed n = 20/expected n = 6.7) appeared to be driving the cluster, with 13 excess cases compared with Hospital M (observed n = 3/expected n = 0.7) (Table 4).

TABLE 3. Citywide signals for respiratory and fever syndromes — New York City, March 16–19, 2003

Date of signal	Observed cases	Expected cases	RR*	Excess
Respiratory				
3/16/03	364†	294	1.2	70
3/17/03	830§	707	1.2	123
3/18/03	1,289¶	1,084	1.2	205
3/19/03	1,259¶	1,155	1.1	104
Fever				
3/16/03	—**	—	—	—
3/17/03	452¶	381	1.2	71
3/18/03	490¶	408	1.2	82
3/19/03	508¶	438	1.2	70

* Relative risk.
† 1-day signal.
§ 2-day signal.
¶ 3-day signal.
** No fever syndrome signal occurred on March 16, 2003.

Response. DOHMH initiated an epidemiologic investigation on March 17, 2003. Patient line lists revealed that illness was distributed among all adult age groups and that chief complaints were consistent with influenza-like illness. ED staff were interviewed about concerning cases, unusual trends or clusters, and any travel histories, none of which were reported. The hospital infection-control practitioner collected contact information for patients with chief complaints consistent with fever and respiratory syndromes and identified patients admitted to the hospital. Patients treated in the ED for respiratory or fever syndromes on March 16, 2003, were contacted by telephone by a DOHMH physician on March 17, 2003. Sixteen patients were called and five patients were interviewed. All five reported improvement; one patient reported having traveled through Frankfurt Airport but denied having traveled to Asia, and one patient had visited the ED because of increased media reports on SARS. The remaining 11 patients

TABLE 4. Hospital- and zip-code–level spatial signals for fever syndrome — New York City, March 16, 2003

Signal location	Observed cases	Expected cases	RR*	Excess
Hospital				
L	20	6.7	3.0	13
M	3	0.7	4.3	2
Total	**23**	**7.4**	**3.1**	**15**
Zip code				
1	8	0.7	11.1	7
2	1	0.2	6.0	1
Total	**9**	**0.9**	**10.1**	**8**

* Relative risk.

either had incorrect or disconnected telephone numbers or did not respond after three attempts.

DOHMH staff visited Hospital L's ED on March 17 and 18, 2003, to review medical records and to interview staff and any patients (or their families) remaining in the ED who complained of fever or respiratory problems. Hospital staff reported no unusual clusters of illness or increase in patients complaining of fever or influenza-like illness. Two ED patients' families were interviewed, none of which reported recent travel to Asia. Sixteen medical records were reviewed, including those of patients admitted to the hospital.

Findings. This investigation of a hospital-level spatial signal for fever syndrome and concurrent citywide 3-day fever and 4-day respiratory syndrome signals did not uncover any features indicative of an outbreak or importation of SARS. No similarities in disease presentation, epidemiologic links, or etiologic agents were identified. These negative findings reassured the health department that no communitywide outbreak of febrile or respiratory illness related to SARS existed, particularly because the trend did not continue. Whether these signals represented an unusual statistical anomaly or focal community illness caused by one or more agents remains unknown.

Discussion

These field investigations illustrated both the difficulties of and resources required in identifying the cause of temporal and spatial aberrations in syndromic surveillance data. Syndromic data are nonspecific by nature. For illnesses that are self-limited and of short duration, resolving the syndrome into an etiologic diagnosis is not usually of direct benefit to the patient, not a priority for the clinician, and not always feasible with current technology. The advantage of using syndromic data for outbreak detection is timeliness. Experiences to date indicate that this advantage might only be theoretical. The time required to conduct investigations and retrieve diagnostic and epidemiologic information might negate the advantage of timely data acquisition. The absence of sustained syndromic signals is usually more reassuring that an outbreak does not exist than the information obtained by an immediate investigation.

Using ED syndromic surveillance for outbreak detection has certain limitations. Of the >40 spatial syndromic signals investigated by DOHMH during 2002–2003, none have been conclusively determined to be a discrete infectious-disease outbreak. Similarly, none of the localized outbreaks reported and investigated through traditional communicable disease surveillance (e.g., nosocomial- or foodborne-outbreak investigations) have yielded a simultaneous syndromic surveillance signal. This is a factor of both the difficulty of proving causality and the use of a sensitive but nonspecific detection system. Outbreaks reported through traditional means rarely involve sufficient ED visits and geographic localization to yield a syndromic signal. Even when both of these factors are present, the event might not be detected if complaint information is inaccurately recorded in the medical record, as evidenced by the second investigation described in this paper.

One advantage demonstrated by NYC's ED syndromic surveillance system has been its early detection of seasonal, widespread disease trends attributed to norovirus and influenza (*1*). These detections have enabled DOHMH to alert the medical community proactively and distribute prevention information to providers and the public. The effect of these measures has not yet been studied. Syndromic surveillance can also provide reassurance that a large-scale outbreak does not exist, as illustrated by the third investigation presented, in which cases of fever/influenza and respiratory illness were deemed unlikely to be SARS.

Using chief-complaint data instead of discharge diagnosis or information from the clinical evaluation might result in a more limited representation of patient illness; however, such clinical information is difficult to code and not timely. NYC's system relies on ED visits, which are uncommon for adults with mild or prodromal illness. More experience is needed with these systems, including an evaluation of the systems' performance in the presence of large outbreaks. Meanwhile, DOHMH has learned useful lessons for conducting future signal investigations (Box).

Conclusions

NYC's ED syndromic surveillance system provides rapid health information through timely electronic data collection and automated spatio-temporal analyses. The system receives data on 80% of daily ED visits citywide, which is representative of the population accessing care at city EDs.

Syndromic surveillance using ED chief complaint data has proved useful as an adjunct system to enhance traditional disease reporting methods at the DOHMH. It provides timely information on seasonal patterns of illness and disease trends citywide, which will allow for prompt epidemiologic investigation in the event of a significant deviation from baseline or a suspicious signal. After a citywide outbreak is detected, syndromic data might also provide information on the epidemic's pace and magnitude. However, the ability of syndromic surveillance to detect outbreaks that are either limited or result in mild disease is as yet unproven. Given the growing interest and investment in syndromic surveillance

BOX. Lessons learned in conducting syndromic signal investigations

Know the data

Knowing the expected range of values will help identify duplicate entries, syndrome miscodes, and patterns that can assist in signal investigations.

Be prepared for the site visit

Call in advance, plan emergency department (ED) visits to yield maximum information from chart reviews and patient interviews, and consider patient language and cultural factors.

Be flexible

Charts, log books, and review requests might be lost or delayed. Be prepared to interview staff and patients currently in the ED and review charts on admitted patients.

Plan for specimen collection

Specimen collection is time-consuming for health department staff and not usually a priority for EDs. Certain hospitals lack the ability to test for all pathogens, especially viral. A system for tracking specimens is necessary but can be difficult for EDs to implement. Although take-home kits can be useful for collecting stool specimens, they require preparation of the collection kit, laboratory slips, and language-sensitive instructions with health department contact information. Arranging transportation of specimens to the laboratory can increase the likelihood that samples are collected.

systems, continued evaluation of these systems is needed to determine the most useful data sources, analytic methods, and signal-investigation approaches.

References

1. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 2004;10:858–64.
2. Kulldorff M. Prospective time periodic geographic disease surveillance using a scan statistic. J R Stat Soc 2001;164:61–72.
3. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. Emerg Infect Dis 2003;9:641–6.
4. CDC. Norovirus activity—United States, 2002. MMWR 2003;52:41–5.

Should We Be Worried? Investigation of Signals Generated by an Electronic Syndromic Surveillance System — Westchester County, New York

William Terry, B. Ostrowsky, A. Huang
Westchester County Department of Health, New Rochelle, New York

Corresponding author: William Terry, Westchester County Department of Health, 145 Huguenot Street, 8th floor, New Rochelle, NY 10801. Telephone: 914-813-5203; Fax: 914-813-5182; E-mail: wwt1@westchestergov.com.

Abstract

Introduction: *In January 2003, the Westchester County Department of Health (WCDH) began conducting electronic syndromic surveillance of hospital emergency department (ED) chief complaints. Although methods for data collection and analysis used in syndromic surveillance have been described previously, minimal information exists regarding the responses to and investigations of signals detected by such systems. This paper describes WCDH's experience in responding to syndromic surveillance signals during the first 9 months after the system was implemented.*

Objectives: *The objectives of this analysis were to examine WCDH's responses to signals detected by the county's syndromic surveillance system. Specific goals were to 1) review the actual complaints reported by hospital EDs to determine whether complaint data were accurately identified and classified into syndrome categories, and provide feedback from this review to data collection and analysis staff to refine text terms or filters used to identify and classify chief complaints; 2) develop procedures and response algorithms for investigating signals; 3) determine whether signals correlated with reportable communicable diseases or other incidents of public health significance requiring investigation and intervention; and 4) quantify the staffing resources and time required to investigate signals.*

Methods: *During January 27–October 31, 2003, electronic files containing chief-complaint data from seven of the county's 13 EDs were collected daily. Complaints were classified into syndrome categories and analyzed for statistically significant increases. A line listing of each complaint comprising each signal detected was reviewed for exact complaint, number, location, patient demographics, and requirement for hospital admission.*

Results: *A total of 59 signals were detected in eight syndrome categories: fever/influenza (11), respiratory (6), vomiting (11), gastrointestinal illness/diarrhea (8), sepsis (7), rash (7), hemorrhagic events (3), and neurologic (6). Line-listing review indicated that complaints routinely were incorrectly identified and included in syndrome categories and that as few as three complaints could produce a signal. On the basis of hospital, geographic, age, or sex clustering of complaints, whether the complaint indicated a reportable condition (e.g., meningitis) or potentially represented an unusual medical event, and whether rates of hospital admission were consistent with medical conditions, 34 of 59 signals were determined to require further investigation (i.e., obtaining additional information from ED staff or medical providers). Investigation did not identify any reportable communicable disease or other incidents of public health significance that would have been missed by existing traditional surveillance systems. Nine staff members spent 3 hours/week collectively investigating signals detected by syndromic surveillance.*

Conclusions: *Standardized sets of text terms used to identify and classify hospital ED chief complaints into syndrome categories might require modification on the basis of hospital idiosyncrasies in recording chief complaints. Signal investigations could be reasonably conducted by using local health department resources. Although no communicable disease events were identified, the system provided baseline and timely objective data for hospital visits and improved communication among county health department and hospital ED staff.*

Introduction

Westchester County (2000 population: 923,459) is located directly north of New York City and is served by 13 acute-care hospitals with emergency departments (EDs). Existing communicable disease surveillance systems include passive surveillance based on notification by physicians or laboratories of reportable communicable diseases and reports from schools and health-care facilities. WCDH has routinely conducted active surveillance for specific diseases or situations (e.g., telephoning hospitals to identify possible cases of West Nile virus after the advent of this disease in 1999). Increasing concern about potential incidents of biologic terrorism has highlighted the need for surveillance systems to permit the

earliest possible detection of such an incident. Efforts to develop an electronic syndromic surveillance system in Westchester County were initiated before the September 11, 2001, terrorist attacks; the system was implemented in January 2003. Although methods for data collection and analysis used in syndromic surveillance systems have been described previously (*1,2*), minimal information exists about the responses to and investigations of signals detected by such surveillance systems. This paper describes responses by WCDH disease investigative staff to syndromic surveillance signals. The data collection and analysis methods used have been described previously (*1,3*).

Westchester County's Syndromic Surveillance System

On the basis of similar systems developed and implemented by other local health departments (*2,4*), WCDH implemented an electronic syndromic surveillance system in January 2003 in four of the county's 13 hospital EDs. By October 2003, the system had been expanded to include seven hospitals. Data from the seven EDs captured approximately 600 daily ED visits, which represented approximately 70% of total daily ED visits to all 13 county hospitals. Data collected on each patient included the chief complaint for which the patient was seeking medical attention, hospital name, patient age, sex, medical record number, municipality and zip code of residence, ED visit date, and whether the patient was subsequently discharged from the ED or admitted to the hospital. On the basis of text search terms and syndrome categories developed by other local health departments and CDC (*5*), chief complaints were classified into eight syndrome categories: 1) fever/influenza in patients aged >13 years, 2) respiratory complaints in patients aged >13 years, 3) vomiting, 4) gastrointestinal illness/diarrhea, 5) sepsis, 6) rash, 7) hemorrhagic events, and 8) neurologic events.

For each syndrome category, the number of complaints or visits for each category was analyzed to identify any statistically significant increases in visits. The cumulative sum method (CUSUM) was used for statistical analysis (*1*); three possible signal types (C1, C2, or C3) could be generated for each of the eight syndrome categories (*3*). A C1 signal was generated when the number of visits from the previous day exceeded the mean number of visits for the previous 7 days by 3 standard deviations. A C2 signal was generated when the number of visits from the previous day exceeded the mean number of visits for the 7 days beginning 9 days before the day being analyzed (excluding the mean number of visits for the 2 days immediately preceding the day being analyzed to smooth the data from any recent aberrations) by 3 standard deviations. A C3 signal was generated when an increase in the number of visits/day occurred on >1 of the 3 preceding consecutive days (*1*, L. Hutwagner, M.S., CDC, personal communication, 2004). Each time a signal was detected, WCDH disease investigative staff were notified and provided with a line listing of complaints comprising the signal and containing the data elements listed previously.

Objectives

The objective of this analysis was to examine WCDH's responses to signals detected by a syndromic surveillance system. Specific goals were as follows:

- review the actual complaints submitted by reporting hospital EDs to determine whether complaints were accurately identified and classified into syndrome categories;
- provide feedback from this review to data collection and analysis staff to refine text terms or filters used to identify and classify chief complaints into syndrome categories;
- develop procedures and response algorithms for investigating signals;
- determine whether signals correlated with reportable communicable disease or other incidents of public health significance requiring investigation and intervention; and
- quantify the staffing resources and time required to investigate signals.

Methods and Results

Signals Detected and Initial Response

During January 27–October 31, 2003, electronic files containing chief-complaint data from participating hospital EDs were collected daily (four EDs in January, expanding to six EDs in April and seven in July). On eight occasions, data transfers were not received from the hospitals but were transmitted the following day, and analyses were performed retrospectively. During the 277-day study period, 59 statistically significant increases or signals were detected on 57 separate days (two signals occurred on 2 days) in eight different syndrome categories (Table). The number of complaints or visits required to produce a signal varied by syndrome category. For the sepsis and neurologic categories, the number of complaints required to generate a signal ranged from three to 10. For other syndrome categories (e.g., gastrointestinal illness/diarrhea and fever/influenza), the number of complaints required to generate a signal ranged from 12 to 20.

TABLE. Number and types of signals generated by syndromic surveillance, by syndrome — Westchester County, New York, January 27–October 31, 2003

Syndrome	C1* only	C2† only	C1 and C2 combined	C3§	Total
Fever/influenza	2	2	6	1	**11**
Respiratory	0	2	4	0	**6**
Vomiting	4	2	5	0	**11**
Gastrointestinal/ diarrhea	1	0	5	2	**8**
Sepsis	3	2	2	0	**7**
Rash	1	3	3	0	**7**
Hemorrhagic	1	0	2	0	**3**
Neurologic	1	1	4	0	**6**
Total	**13**	**12**	**31**	**3**	**59**

Source: Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response early aberration reporting system (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96; L. Hutwagner, M.S., CDC, personal communication, 2004.

* C1 signals occurred when the number of visits from the previous day exceeded the mean number of visits for the previous 7 days by 3 standard deviations.

† C2 signals occurred when the number of visits from the previous day exceeded the mean number of visits for the 7 days beginning 9 days before the day being analyzed (excludes the mean number of visits for the 2 days immediately preceding the day being analyzed to smooth the data from any recent aberrations) by 3 standard deviations.

§ C3 signals occurred when an increase in the number of visits/day occurred on >1 of the 3 preceding consecutive days.

Each time a C1, C2, or C3 signal was generated in any of the eight syndrome categories, disease-investigation staff (including an infectious-diseases physician and a nonphysician epidemiologist) reviewed a line listing of the individual complaints comprising the signal. This line listing contained the absolute number of visits comprising the signal, the chief-complaint text for which the patient was seeking medical attention, hospital name, patient age, sex, municipality and zip code of residence, ED visit date, and whether the patient was subsequently discharged from the ED or admitted to the hospital. The number of complaints resulting in a signal and thus contained in line listings varied by syndrome (range: 3–103).

Terms Used To Identify and Classify Complaints into Syndrome Categories

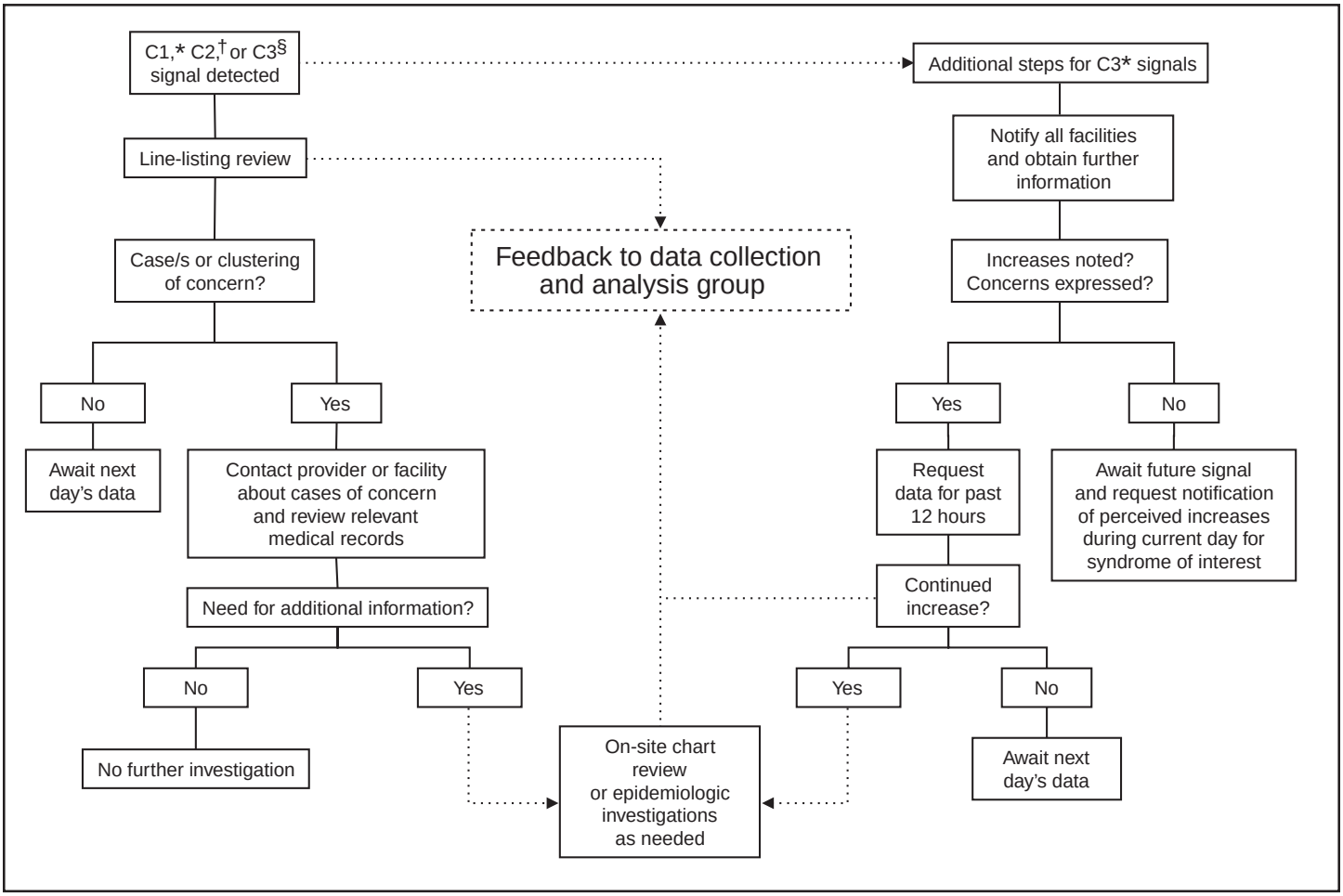
By using a system developed by the New York City Department of Health and Mental Hygiene (*2,4*), WCDH staff collected data files from hospital EDs containing fields of free text describing the patient's chief complaint. These text fields were searched for specific terms that were then used to classify complaints by syndrome category. For example, terms used to identify and classify ED visits into a fever/influenza syndrome category included *fever*, *temp*, *hot*, and *aches*, among others. Specific terms were also designated for exclusion from a syndrome category (e.g., chief complaints of nausea or vomiting including the terms *pregnant* or *pregnancy* were excluded from the vomiting syndrome category).

An infectious-diseases physician and a nonphysician epidemiologist compared text terms used to identify and classify complaints into syndrome categories with the actual text submitted by hospital EDs. Although they were not systematically quantified, the majority of terms used to identify and classify complaints into syndrome categories and terms used to exclude complaints from syndrome categories were determined to be correct. For example, the terms *temp* and *hot* correctly identified and classified the majority of complaints into a fever/influenza syndrome category, but such terms as *attempted suicide* and *gunshot* were also identified and included in the complaints for a fever/influenza syndrome category because they each contain the text of interest (*temp* and *hot*) within the larger text. Detection of the first 10–15 signals and subsequent line-listing reviews of complaints comprising these signals indicated that ≤3 complaints could result in a signal (C1 or C2, as described previously) for certain syndrome categories, meaning that incorrect identification of a limited number of complaints could result in a false-positive signal and trigger additional investigation. During reviews of line listings comprising signals, at least one term that could result in a complaint being classified incorrectly into one of the syndrome categories was identified in every line listing. Whenever this occurred, the data collection and analysis staff were instructed to exclude such terms to prevent future false-positive signals. Line-listing reviews also indicated that certain EDs recorded chief complaints by specifying complaints that patients did *not* have (e.g., *denies shortness of breath* or *denies fever*). As a result, certain chief-complaint terms detected by the system did not represent true cases of a particular syndrome. With repeated reviews, the number of complaints incorrectly identified or classified into syndrome categories decreased. On the basis of this limited experience, a standard set of text terms might not be universally applicable for syndromic surveillance systems but might require modifications based on idiosyncrasies in the text or words used to record chief complaints by individual hospital ED staff.

Investigation of Detected Signals

An algorithm was developed for responding to different types of signals detected by the syndromic surveillance system (Figure). In addition to determining whether patient visits might have been incorrectly included in a syndrome category in response to a C1, C2, or C3 signal, an infectious-diseases physician and a nonphysician epidemiologist reviewed the line listing of complaints comprising a signal to determine the need for additional investigation based on any clustering by hospi-

FIGURE. Syndromic surveillance signal response algorithm — Westchester County, New York



Source: Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96; L. Hutwagner, M.S., CDC, personal communication, 2004.

* C1 signals occurred when the number of visits from the previous day exceeded the mean number of visits for the previous 7 days by 3 standard deviations.

† C2 signals occurred when the number of visits from the previous day exceeded the mean number of visits for the 7 days beginning 9 days before the day being analyzed (excludes the mean number of visits for the 2 days immediately preceding the day being analyzed to smooth the data from any recent aberrations) by 3 standard deviations.

§ C3 signals occurred when an increase in the number of visits/day occurred on >1 of the 3 preceding consecutive days.

tal, patient municipality or zip code of residence, age, or sex, or whether the specific nature of a complaint was reportable in New York State. The complaints comprising a signal occurred in at least two to three hospitals and municipalities; in no instances did all of the complaints originate from a single hospital or municipality. To determine the need for further investigation, staff also assessed complaints for their potential to represent an unusual medical event and examined whether hospital-admission rates were consistent with the medical condition. For example, urosepsis in an elderly resident of a nursing home would be less an indication for further investigation than altered mental status in a young adult requiring hospital admission. Similarly, the percentage of visits requiring hospital admission varied depending on the complaint and the absolute number of visits. Hospital-admission rates not consistent with the medical condition were also an indication for further investigation. For example, three or four complaints or ED visits for seizures requiring hospital admission would be less of an indication for further investigation than 70–80 visits for diarrhea, of which 25% required hospital admission. No standard threshold percentage of visits requiring hospital admission could be used to determine the need for additional investigation; review and clinical judgment were required to make this assessment.

If the line-listing review identified no obvious cases or clusters of concern or cases potentially representing a reportable or unusual medical event, disease-investigation staff awaited results of the next day's data analysis to determine whether

the signal was sustained and whether investigation was needed. If any cases or clustering of concern were noted, disease investigative staff obtained further information. On the basis of the line-listing review, 29 of the 56 C1 or C2 signals detected by syndromic surveillance required further investigation. All C3 signals were investigated further. Follow-up investigation was conducted by telephone calls to hospital ED physicians, infection-control practitioners, or treating physicians and by requesting facsimiles of relevant laboratory and diagnostic tests results or medical records. The information obtained was sufficient to assess each situation, and no on-site hospital ED visits or chart reviews were conducted.

Because increases in ED visits in a given syndrome category on a single day could result in a C1 or C2 signal but an increase in ED visits in a given syndrome category on >1 of 3 consecutive days was required to generate a C3 signal, a C3 signal was believed to have an increased potential for an event of concern. Therefore, in response to the three C3 signals detected, all the investigative procedures described previously were followed, but further investigation was conducted regardless of the results of the line-listing review. WCDH staff contacted ED and infection-control staff at all 13 hospitals to notify them that the syndromic surveillance system had detected an increase in a particular syndrome (e.g., gastrointestinal illness/diarrhea) and to ask whether any increase in the syndrome of interest had been noted during or since the time encompassed by the most recent complaint data submitted, or whether any concern existed. If hospital ED or infection-control staff perceived no increases and expressed no concern, they were asked to report any perceived increase in ED visits for the syndrome of concern for the current day, and results of data analysis of complaints for the subsequent day were reviewed. Hospital staff perceived no increases or need for concern after any of the three C3 signals. Had such increases or concern been perceived, data collection and analysis staff were instructed to request, from all hospital EDs participating in syndromic surveillance, electronic files containing chief-complaint data encompassing the 12 hours subsequent to the last routine file transfer. None of the three C3 signals warranted this level of response.

Correlation of Signals with Reportable Communicable Disease or Other Incidents of Public Health Significance

After the response and investigation of syndromic surveillance signals, no events of concern or that were detected through other existing surveillance mechanisms were identified. On one occasion a complaint of encephalitis and on 11 occasions complaints of meningitis were noted on line listings comprising a signal. Because all types of meningitis and encephalitis are notifiable diseases in New York State and cases of meningococcal meningitis usually require intervention (e.g., postexposure prophylaxis of contacts), these cases were investigated by contacting the treating physician or hospital staff. In all cases, patients had received alternate diagnoses. Because line listings were reviewed only when a signal was detected, persons with meningitis might have reported to EDs on days on which no signal was detected and therefore would not have been detected through this mechanism. Although other clusters or reportable events were detected through telephone calls from medical providers or affected facilities during the 9-month period, the affected hospitals were not participating in syndromic surveillance, making correlation impossible. No cases of meningitis or other reportable diseases or events that had not been detected through otherwise existing surveillance mechanisms (typically telephone notification from hospital ED staff, infection-control staff, or treating physician, or by a New York State electronic laboratory reporting system) were detected through syndromic surveillance.

Efforts Required for Signal Follow-Up

Nine disease-investigation staff members spent a portion of their time responding to syndromic surveillance signals. An infectious-disease physician and a communicable-disease epidemiologist routinely reviewed the line listings, and seven public health nurses participated in follow-up investigations as described previously. The time and effort required for these activities varied depending on the number of signals received on a given day (1–2 signals/day) and the number of complaints on the line listing for each signal requiring additional investigation (range: 1–10 complaints/signal). Staff were asked to track for 1 month the time spent on follow-up. Signal and line-listing reviews typically required approximately 15 minutes and were performed by a physician and an epidemiologist. Telephone calls to medical providers and reviews of medical records received by facsimile for a single complaint typically required 30 minutes, including the time needed to reach and speak with a knowledgeable hospital staff member or for such staff to obtain relevant information. On average, disease-investigation staff collectively spent approximately 3 hours/week to investigate signals generated by the syndromic surveillance system, not including time required by data collection and analysis staff (*3*). Because information obtained through telephone calls and review of faxed medical records was sufficient to assess each of these situations, no on-site hospital ED visits or chart reviews were necessary.

Discussion and Conclusions

The information presented in this paper is primarily descriptive and encompasses only 9 months, during which time only four to seven of 13 hospital EDs in the county participated in syndromic surveillance. Despite these limitations, the research identified areas that might benefit from further evaluation, and Westchester County's experiences might be useful for others implementing syndromic surveillance.

Complaints identified by text terms developed for use in syndromic surveillance (*5*) routinely were incorrectly identified and classified into syndrome categories of interest. Standardized text terms to identify and classify hospital ED chief complaints into syndrome categories might not be broadly applicable but might require modification because of hospital idiosyncrasies in recording chief complaints. Assessment of signals by medical or clinical professionals was required to determine the need for further investigation.

The procedures used to assess and investigate syndromic surveillance signals could be reasonably conducted by using the resources of a local health department. No reportable or other disease events or events that required further investigation or intervention in addition to those detected by existing traditional surveillance systems were identified through the 59 syndromic surveillance signals detected and investigated during this 9-month period. Because ≤3 complaints were required to generate a signal, a limited number of incorrectly identified complaints could result in a signal and trigger additional investigation.

Further evaluation is required to establish the conditions in which syndromic surveillance is most useful. A jurisdiction of the size and complexity served by WCDH might represent the smaller end of the spectrum in which such systems are likely to be useful, and the disease events that occurred were not the type of events intended to be detected by syndromic surveillance.

Finally, the implementation of this system and investigation of detected signals provided additional benefits. Communications, working relationships, and personal familiarity among WCDH and hospital ED staff improved. ED staff awareness that WCDH staff were available 24 hours/day, 7 days/week as a resource increased. Physicians and hospital staff expressed appreciation for feedback provided by WCDH regarding potential disease activity of concern. A substantial number of the reportable or unusual events that occurred during the 9-month study period were detected through telephone calls from ED staff. This fact underscores the importance to disease surveillance of communication with local ED staff and indicates that syndromic surveillance should complement and not replace traditional reporting and surveillance systems. The system provided baseline and timely objective data for hospital visits and might provide a basis for future monitoring of other conditions of interest.

References

1. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96.
2. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. J Urban Health 2003;80(2 Suppl 1):i76–88.
3. Cho S, Hackmyer S, Li J, et al. Establishing an automated surveillance system [Abstract]. MMWR 2004;53(Suppl):232.
4. Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for community-wide influenza like illness, New York City. J Urban Health 2003;80(2 Suppl 1):i43–9.
5. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.bt.cdc.gov/surveillance/syndromedef/.

Public Health Practice

Public Health Information Network — Improving Early Detection by Using a Standards-Based Approach to Connecting Public Health and Clinical Medicine

Claire V. Broome,[1] J. Loonsk[2]
[1]*Office of the Director, CDC;* [2]*Information Resources Management Office, CDC*

Corresponding author: Claire V. Broome, Office of the Director, CDC, 1600 Clifton Rd., Mailstop D-68, Atlanta, GA 30333. Telephone: 404-639-7860; Fax: 404-639-7770; E-mail: cvb1@cdc.gov.

Abstract

Public health departments and their clinical partners are moving ahead rapidly to implement systems for early detection of disease outbreaks. In the urgency to develop useful early detection systems, information systems must adhere to certain standards to facilitate sustainable, real-time delivery of important data and to make data available to the public health partners who verify, investigate, and respond to outbreaks. To ensure this crucial interoperability, all information systems supported by federal funding for state and local preparedness capacity are required to adhere to the Public Health Information Network standards.

Introduction

The 2003 National Syndromic Surveillance Conference focused on design, development, and evaluation of systems that can rapidly detect terrorism-related outbreaks as well as naturally occurring epidemics. Public health departments and their clinical partners understand the urgency to have systems in place to support early detection and are moving ahead rapidly to implement systems that will provide early detection functionality. These systems obtain data from multiple sources, including traditional clinical-care delivery sites and clinical laboratories, as well as less traditional health-monitoring data sources (e.g., nurse call centers, over-the-counter retail sales, work and school absenteeism data, veterinary health data, or information from biologic-sensing devices). In their urgency to develop early detection systems, system developers should incorporate information-system standards to facilitate sustainable, real-time delivery of important data and to make data available to the public health partners who verify, investigate, and respond to outbreaks. To ensure this crucial interoperability, all information systems supported by federal funding for state and local preparedness capacity are required to use set information-system standards (*1*).

Standards-based system development is critical for three major reasons. First, the need for real-time information from multiple sources can best be accomplished by standards-based electronic messaging. Although individual custom interfaces can be created with the myriad potentially useful data sources, the cost of development would be prohibitive and the complexity of developing and managing such an array of custom interfaces would be formidable. The specification for standard Health Level 7 (HL7) (*2*) messages for early detection data permits health departments to leverage integration-broker technology and health-care delivery site information technology (IT) capacity for creation and processing of these standard HL7 electronic messages.

Second, the use of standards enables health departments to leverage previous investments in their IT infrastructures. Systems to support public health capacity for outbreak management, response, alerting, and information dissemination have been under development since Fiscal Year (FY) 1999 investments in the Health Alert Network and FY 2000 funding for the National Electronic Disease Surveillance System (NEDSS). A detection system is most valuable when it can communicate with those systems needed to investigate and respond to an epidemic. The availability of standards-based shareable directories, system security, and channels for bidirectional secure communication can support public health agencies' capacity to respond to outbreaks and provide key elements for early detection systems.

Finally, a consistent standards-based approach limits the burden on partners in the clinical-care delivery sector. Health-care providers and hospitals provide information to public health agencies for early detection and routine surveillance as part of their community responsibility. They are not compensated for the cost of providing that information. By using standard formats and electronic reporting, public health agencies can minimize the burden involved in reporting diseases and, ideally, use information that is already available in electronic format within the health-care delivery system.

Nationally, the importance of standards-based, interoperable electronic health records to support objectives for quality and safety within the health-care delivery system has been increasingly recognized. The National Committee on Health and Vital Statistics has recognized standards as an integral part of

the National Health Information Infrastructure (*3*). The critical role of standards has also been endorsed by the U.S. Department of Health and Human Services and the federal government through the Consolidated Health Informatics Initiative, a federal eGov initiative (*4*). Connecting for Health, a broad-based consortium of foundations, provider organizations, systems developers, and government organizations, is also pursuing this objective (*5*). These efforts have already identified and endorsed a number of relevant standards that can be used in early detection systems for the interchange of data between the clinical sector and public health.

To define how these broad standards can be implemented in surveillance systems that support the specific needs of public health practice, CDC and its state and local health department partners have identified key specifications and functions described as the Public Health Information Network (PHIN). By identifying standards for technology, data, vocabulary, and information security, PHIN is designed to enable the consistent exchange of health, disease-tracking, and response data among public health partners, to protect the security of these data, and to ensure the network's reliability in times of national crisis.

PHIN addresses five major functional areas — detection and monitoring, data analysis, knowledge management, alerting, and response. To support these public health functions, CDC and partners have developed specifications for nine IT functions, identifying the key vocabulary and technical standards relevant for creation of PHIN (*6*). These nine functions are as follows:

1. automated exchange of data between public health partners;
2. use of electronic clinical data for event detection;
3. manual data entry for event detection and management;
4. specimen and lab result information management and exchange;
5. management of possible case, contacts, and threat data;
6. analysis and visualization;
7. directories of public health and clinical personnel;
8. public health information dissemination and alerting; and
9. IT security and critical infrastructure protection.

Public Health Information Network — Functions and Specifications Relevant to Early Detection

Of the nine PHIN functions that should be incorporated into commercially or locally developed early detection systems, the following six functions have particular relevance to early detection:

- **Automated exchange of data between public health partners (No. 1) and use of the electronic clinical data for event detection (No. 2).** These standards address the use of electronic messages to transmit data from a clinical source over the Internet to the health department using secure encryption. These messages can be generated automatically on the basis of prior agreements by the trading partners regarding which data are potentially relevant for public health. The use of electronic messaging provides near real-time transmission of data needed to support early detection. The format standard used for messaging is HL7, one of the standards identified by the National Committee on Vital and Health Statistics, the Consolidated Health Informatics eGov Initiative, and Connecting for Health as the appropriate standard in this area (*2–5*). PHIN also provides a process for developing detailed specifications for early-detection message content.
- **Analysis and visualization (No. 6).** This standard governs the use of commercial applications for analysis and visualization, which use industry standards for accessing data from the database. This standard facilitates the use of a validated aberration-detection algorithm in multiple, diverse systems.
- **Directories of public health and clinical personnel (No. 7).** Such directories are critical tools, both for identifying the persons (or positions) who need to receive and transmit data, and to support role-based security to ensure appropriate access to data and secure data against unauthorized access. Because alerts frequently need to travel between jurisdictions, a standards-based directory (Lightweight Directory Access Protocol [LDAP], which uses a public health directory data model developed jointly by state, local, and federal partners as part of the PHIN process) can facilitate exchange of information among, for example, emergency-response personnel in adjacent local jurisdictions and public health personnel at the state level.
- **Public health information dissemination and alerting (No. 8).** This function is essential for communicating and responding to any outbreak identified by an early-detection system. Public health partners must be able to transmit and receive alerts in a timely fashion by appropriate mechanisms 24 hours/day, 7 days/week. The function might use e-mail or back-up modes (e.g., pagers and telephones) for notification. In addition, specifications are necessary to permit bidirectional, secure communications among health officials using PHIN-compatible directories and security so that sensitive information can be appropriately shared, discussed, and analyzed, An early-

detection system needs to address how it will interface with local and state secure-communications systems.

- **IT security and critical infrastructure protection (No. 9).** Security specifications are an essential element of early detection systems. Carefully planned approaches to protect system security and continuity of operations are needed to ensure that a system is available in the event of an emergency. A state's security strategy should be consistent with the state's approach to information-system security, rather than requiring an anomalous approach, such as implementation of two-factor authentication (i.e., use of two different modalities to ensure an individual is authenticated [e.g., password and secure token, or password and digital certificate]).

Implementing Systems Compliant with the Public Health Information Network

PHIN's specifications and functions are the building blocks for interoperable standards-based systems. However, considerable discussion has ensued about appropriate processes for turning these relatively high-level specifications into functioning systems. The CDC Information Council, the official governance body for CDC and its public health partners (including Association of State and Territorial Health Officials, National Association of County and City Health Officials, Council of State and Territorial Epidemiologists, Association of Public Health Laboratories, and National Association of Public Health Statistics and Information Systems) asked the Gartner Group, an experienced IT consulting firm, to recommend implementation approaches for PHIN specifications and functions, as well as processes for managing evolution of the architecture and data standards. In 2003, the Gartner Group issued a report addressing the PHIN functions and specifications and recommended approaches that might accelerate their implementation (*7*). The study team interviewed state and local health departments and examined documents and design specifications at CDC. The final report endorsed the PHIN standards and specifications as appropriate for use in public health. It also noted that CDC's public health partners universally agreed to the vision and overall direction of PHIN and emphasized that successful implementation of PHIN is critically dependent upon the commitment of CDC and its public health partners. The report also identified areas needing further clarification or expansion of the PHIN architecture.

For systems that are underway or still in development, the Gartner Group recommended an evolutionary approach toward PHIN compatibility. They recommended that application development teams focus first on compatibility of the data model with PHIN data standards and use of controlled medical vocabularies. Doing so would permit creation of data that can be easily aggregated at the national level by using extensible markup language (XML) schema. They also recommended use of HL7 messaging format for transport and security standards to share data securely between public health partners and CDC. A third recommendation was to focus on standards-based directory services (LDAP) to allow authorized and controlled access. Finally, they recommended that CDC provide tools (e.g., tools for secure message transport) built on PHIN standards that could be made available to states and their partners.

The Gartner study recommended that PHIN allow for multiple solutions, particularly for those components that are more technically challenging or new in the market (e.g., HL7 version 3.0, ebXML; http://www.hl7.org). However, they emphasized that the goal of a live network should be maintained even as different solutions are implemented. They recommended PHIN standards be required for investments of federal public health funding. Finally, they emphasized the importance of security at all levels of state public health infrastructure, recommending that states undertake independent verification and validation studies to provide an independent assessment of system security.

In addition to resources invested by states and local jurisdictions, additional funds are available to support PHIN in general and its use for early detection in particular. Since FY 1999, all 50 states have received funding through the Health Alert Network for continuous broadband internet connectivity among states and local health departments. Certain states have also used this funding to provide connections with clinical-care delivery partners and emergency-management partners. Since FY 2000, states have also received funding for standards-based surveillance systems through NEDSS, which implements the PHIN standards for clinical data exchange in the area of clinical laboratory data and nationally notifiable diseases. In FY 2002, the Public Health and Social Service Emergency Fund awarded >$1 billion for state and local public health preparedness capacity. A substantial portion of these funds have been directed to investments in IT systems; both CDC and the Health Resources Services Administration (HRSA) require that all IT investments use the PHIN specifications and functions (*1*). In September 2003, the second round of preparedness funding was awarded, which continued to require use of PHIN specifications and functions when funding IT investments. By September 2003, HRSA grants had increased to $498,000,000, directed toward enhancement of hospital surge capacity to deal with terrorist events. This

funding could be used, in part, to strengthen the communication and data interchange between hospital partners and public health.

Consistent with the Gartner Group recommendations, CDC has developed and made available tools to assist in developing PHIN-compliant systems. The PHIN Messaging System is a software program that supports standards-based, bidirectional, interinstitutional message transport using the ebXML standard with Public Key Infrastructure (PKI) encryption (*8*). It provides a message-transport tool for point-to-point messaging, thereby addressing the need for secure authentication and authorization between sender and receiver as well as handling encryption of the message payload.

In January 2004, CDC released a beta version of PHIN Vocabulary Services, which provides access to >80 key standard reference tables, as well as supporting version control and maintenance of those standard reference tables (*9*). This tool should facilitate using controlled vocabularies in local systems and support CDC-developed systems.

CDC has also published implementation guides that specify data standards for the message format for data exchange messages (e.g., those dealing with electronic laboratory reporting, test orders, and demographic information available from hospital admission discharge transfer [ADT] systems) (*10*).

Finally, CDC has collaborated with partners from the U.S. Department of Defense, U.S. Veterans Administration, the private sector, Harvard University, University of Pittsburgh, and state and local health departments to develop BioSense (*11*). BioSense is an Internet-accessible secure system that permits state or metropolitan-area users to visualize information about their locality from different early-detection data sources. It maps the data at a zip-code level and incorporates statistical analyses to identify possible aberrations warranting further investigation. Phase 1 of BioSense is in beta testing. It is intended to be complementary with local efforts. In Phase 2, BioSense will be able to incorporate local data-collection efforts that use PHIN standards to provide a more complete view of data sources relevant to a particular area.

Rapid detection of possible terrorist events is of considerable urgency. However, using a standards-based approach in surveillance is critical, both to accomplish the early detection objective and to facilitate rapid investigation of and response to multiple events of public health importance. Investing wisely by developing effective PHIN-compliant systems will have enormous benefits for the health of the public.

References

1. CDC. Continuation guidance for cooperative agreement on public health preparedness and response for bioterrorism — budget year four, Appendix 4: IT specifications and functions. Atlanta, GA: US Department of Health and Human Services, CDC, 2002. Available at http://www.bt.cdc.gov/planning/continuationguidance/pdf/appendix-4.pdf.
2. Health Level Seven, Inc. Health Level Seven. Ann Arbor, MI: Health Level Seven, Inc., 2001. Available at http://www.hl7.org.
3. US Department of Health and Human Services. Information for health: a strategy for building the national health information infrastructure — report and recommendations from the National Committee on Vital and Health Statistics, Washington, DC: US Department of Health and Human Services, 2001. Available at http://www.ncvhs.hhs.gov/nhiilayo.pdf.
4. eGov. Government to business portfolio: consolidated health informatics. Washington, DC: eGov, 2003. Available at http://www.whitehouse.gov/omb/egov/gtob/health_informatics.htm.
5. The Markle Foundation. Connecting for Health. New York: The Markle Foundation, 2003. Available at http://www.connectingforhealth.org.
6. CDC. Public Health Information Network standards, specifications, and functions. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/phin/architecture/index.htm.
7. Gartner, Inc. Technical review of issues related to version 1 of the Public Health Information Network functions and specifications. Stamford, CT: Gartner, Inc., 2003. Available at http://www.cdc.gov/phin/conference_presentations/05-13-03/.
8. CDC. Public Health Information Network (PHIN) messaging. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/phin/messaging/index.htm.
9. CDC. Public Health Information Network (PHIN) components. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/phin/components/index.htm.
10. CDC. Public Health Information Network (PHIN) messaging implementation guides/XML schemas/standards. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/phin/messaging/index.htm.
11. Loonsk JL. BioSense: a national initiative for early detection and quantification of public health events. MMWR 2004;53 (Suppl):53–5.

Information System Architectures for Syndromic Surveillance

William B. Lober, L. Trigg, B. Karras
University of Washington, Seattle, Washington

Corresponding author: William B. Lober, University of Washington, MS 357240, 1959 NE Pacific St., Seattle, Washington 98195. Telephone: 206-616-6685; Fax: 206-338-2455; E-mail: lober@u.washington.edu.

Abstract

Introduction: *Public health agencies are developing the capacity to automatically acquire, integrate, and analyze clinical information for disease surveillance. The design of such surveillance systems might benefit from the incorporation of advanced architectures developed for biomedical data integration. Data integration is not unique to public health, and both information technology and academic research should influence development of these systems.*

Objectives: *The goal of this paper is to describe the essential architectural components of a syndromic surveillance information system and discuss existing and potential architectural approaches to data integration.*

Methods: *This paper examines the role of data elements, vocabulary standards, data extraction, transport and security, transformation and normalization, and analysis data sets in developing disease-surveillance systems. It then discusses automated surveillance systems in the context of biomedical and computer science research in data integration, both to characterize existing systems and to indicate potential avenues of investigation to build systems that support public health practice.*

Results: *The Public Health Information Network (PHIN) identifies best practices for essential architectural components of a syndromic surveillance system. A schema for classifying biomedical data-integration software is useful for classifying present approaches to syndromic surveillance and for describing architectural variation.*

Conclusions: *Public health informatics and computer science research in data-integration systems can supplement approaches recommended by PHIN and provide information for future public health surveillance systems.*

Introduction

Automated acquisition of routine health-care data has enhanced public health surveillance capabilities. The 2003 National Syndromic Surveillance Conference featured model syndromic surveillance systems, including New York City's emergency department (ED)-based syndromic surveillance system (*1*), the Real-Time Outbreak Disease Surveillance system (RODS) (*2*), the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) (*3*), and other encounter-based systems. These systems use different data sources, including ED and primary care outpatient data (e.g., chief complaints or diagnoses), diagnosis-specific aggregate data (National Bioterrorism Syndromic Surveillance Demonstration Project [*4*]), and laboratory and radiology data, for early detection of disease outbreaks. Surveillance to detect clinical syndromes, whether inferred by secondary use of clinical data sources or directly coded by observers, is commonly called syndromic surveillance.

Despite increasingly widespread development of syndromic surveillance systems, continued efforts to understand different data-analysis strategies, and ongoing discussion of strategies to integrate syndromic surveillance into public health practice, the cost-benefit ratio of syndromic surveillance remains uncertain. Recommendations from the 2001 American Medical Informatics Association meeting stated that "public health informatics must create an information architecture that includes a longitudinal, person-based, integrated data repository…similar to the National Electronic Disease Surveillance System (NEDSS) model" (*5*). NEDSS has evolved into a prominent component of the Public Health Informatics Network (PHIN) initiative (*6,7*). A recent review of PHIN (*8*) concluded that "the PHIN vision must continue to broaden beyond the structured data obtained from surveillance systems and labs to include syndromic data from clinics, ERs, doctor's offices, pharmacies…," indicating that surveillance based on integration of heterogeneous data will become central to public health practice.

Implementing syndromic surveillance based on automated acquisition of clinical data requires both the development of secure, reliable information systems and the use of those systems in public health practice. The information technology (IT) activities include system design and integration and development of tools for data acquisition and analysis. Effective use of syndromic surveillance depends not only on IT activities but also on the system's integration with public health practices for outbreak detection, investigation, and response management.

Data modeling and data integration are integral IT components of syndromic surveillance information systems. *Data modeling* activities are those related to structure and content, and entail identifying relevant clinical variables; understanding both the vocabularies and coding schemes used to record these variables; and establishing procedures for clustering, re-coding, normalizing, or otherwise preparing data for analysis. *Data integration* activities are those related to movement and processing of data before their analysis or visualization, and entail acquiring, transforming, storing, and delivering information securely and reliably. Approaches to data modeling and integration and the trade-offs between different implementation technologies constrain the choice of system architectures.

This paper reviews these components in the context of basic and applied research in data integration, on the basis of an evolutionary model used to describe the development of biomedical informatics (*9*). This model provides a framework for reviewing architectures used for automated public health surveillance, both to classify them and to discuss the strengths, weaknesses, and roles of different research approaches.

Data Model Components

Limited development of syndromic surveillance systems, including the RODS and Syndromic Surveillance Information Collection (SSIC) systems (*10*), occurred before the anthrax outbreak in fall 2001. However, the 2001 terrorist attacks precipitated an increase in syndromic surveillance development, and implementation since then has balanced standardization with expediency. To implement systems rapidly before another terrorist attack, developers built systems tailored to readily available data. However, promulgation of national standards (e.g., PHIN) has emphasized the need for standardization of data types collected and of vocabularies used for individual data elements.

Data Elements

Two important data-element considerations are 1) the composition of the extracted data set and 2) the level of identification of the data. A 2001 review of data elements collected for surveillance by 10 different systems identified striking similarities (*11*). The majority of systems described at the 2003 NSSC continue to use data elements identified by that review. These systems collect data for patient ED or primary care visits and typically include age, sex, visit date and time, a measure of chief complaint and/or diagnosis, and a geographic measure; however, data elements and coding schemes vary among systems. Chief complaints or diagnoses, clustered into syndrome groupings, are used as variables for analysis, and both demographic and geographic variables are used to stratify the data. In contrast to the simple data model used by the majority of syndromic surveillance systems, the PHIN Logical Data Model provides a rich, detailed, object-oriented view of health-care data (*12*), encouraging both standardization and more granular data collection.

Public health agencies have legal authority to collect (the minimum necessary) data for surveillance, "without [patient] authorization, for the purpose of preventing or controlling disease, injury, or disability…" (*13*). However, certain barriers to provider data reporting have been identified, including regulatory issues, fit with business model, use of IT resources, public relations, accounting for public health disclosures, and release of competitive data (*14*). Despite certain states' legal authority to collect identified data, multiple system developers have chosen to collect either de-identified or minimally identified data to reduce these practical barriers. Although a masked or encrypted identifier can address these concerns while maintaining data quality, this approach was challenged by the final interpretation of the Health Information Portability and Accountability Act of 1996 (HIPAA) Privacy Rule (*15*). Concern has been expressed about the effect of this interpretation on medical and public health research (*16*).

Standard Vocabulary Usage

Standards for exchange of public health data arose from the need to combine heterogeneous data (i.e., comparable clinical information from different sources that is expressed by using different formats and coding schemes). PHIN specifies the ability to translate and manipulate Logical Observation Identifiers Names and Codes (LOINC®), Systematized Nomenclature of Medicine (SNOMED®), *International Classification of Diseases, Ninth Revision* (ICD-9), and current procedural terminology (CPT®) codes and to map local, legacy, or proprietary codes into these standards (Table). PHIN specifies LOINC as the vocabulary for laboratory reporting in conjunction with the PHIN notifiable-condition–mapping tables (*17*), which map LOINC and SNOMED codes to reportable conditions. Unlike laboratory reporting, syndromic surveillance systems might use local vocabularies and lack a fully developed transformation capability. Hospitals use different standard coding schemes, and transformation will become increasingly important as the scale of these systems increases.

Analysis Data Model

Aggregating data for analysis is also a challenge. Systems commonly use ICD-9 codes or chief-complaint data to categorize illnesses into syndrome groups. Different ICD-9 clus-

TABLE. Vocabularies referenced by the Public Health Information Network

Acronym	Title	Description
UMLS	National Library of Medicine's Unified Medical Language System®	Meta vocabulary collection that includes ICD-9, LOINC, CPT, and soon SNOMED
SNOMED	Systematized Nomenclature of Medicine — Clinical Terms (SNOMED CT®)	Nomenclature copyrighted by the College of American Pathologists; includes diseases, clinical findings, etiologies, procedures, and outcomes
ICD-9	*International Classification of Diseases, Ninth Revision*	Overlaps with SNOMED in diseases, events, and findings
LOINC®	Logical Observation Identifiers Names and Codes (LOINC)	Overlaps with SNOMED on findings and measures
CPT	Current Procedural Terminology	Overlaps with SNOMED in procedures/interventions concept; not as granular as LOINC; also designed for use in insurance data exchange
X12N 277	Claim status codes	Similar to CPT but focused for insurance claims; not specific enough for clinical reporting

tering schemes exist, including a collaborative effort of CDC and other agencies (*18*). Assigning chief complaints to syndrome groups has been implemented in different ways, including by Bayesian classification (*19*) and text substring searches (*2*), and is still being studied. Current algorithms and statistical approaches to detection have been implemented either by using standard statistical software packages or as part of the surveillance system. In either case, the information-system architecture should support preparation of an analysis data set by using a model appropriate to its intended use, the secure delivery of the data to the algorithm, and the data analysis and results presentation itself.

Data-Integration Components

Data integration is characterized by five functions: data extraction, secure data transport, transformation, normalization, and creation of an analysis data set or view. Systems use different approaches to perform these functions; PHIN cites multiple best practices.

Data Extraction

Data extraction refers to acquiring a data set from the source system. Query-based systems extract data through periodic execution of local queries or reports. IT staff responsible for the source system often develop these queries and run them automatically. In certain circumstances, queries against the source system are executed directly by the surveillance system. Message- or event-based systems send a message to the surveillance system whenever something of interest occurs in the source system. Typically, this stream of messages contains either the entire message set, or a filtered subset, of an electronic data interchange between hospital systems. These messages are commonly in Health Level 7 (HL7) format (*20*) and often can be rerouted by using the hospitals' HL7 interface engine or message switchboard. PHIN refers to a series of standards, including HL7 2.x and 3.0, to describe the appropriate formatting for data sent from a source system to public health authorities. However, both query-based and message-based data are consistent with PHIN.

Transport and Security

Public health surveillance data typically travel through the Internet. Although the chance of data either being intercepted or spoofed is low, certain techniques can ensure encryption of the message and protection of participants' identities (*21*). Files can be encrypted and signed by using a standard (e.g., Pretty Good Privacy [PGP]), transferred through a virtual private network (VPN), or transmitted by using a file transfer protocol (FTP) over a securely encrypted channel. PHIN specifies the PHIN Messaging System (PHINMS), which is based on ebXML standard for bidirectional data transport. Symmetric public key encryption (PKI), in which both parties use X.509 certificates, offers both high-quality channel encryption and authentication of both sides of the conversation and is used by PHINMS. PHIN also recommends annual security evaluation.

Transformation and Normalization

Data arriving from different source systems can be in different formats, and coding schemes used for individual data elements might need to be reconciled. Transformation of syntax and normalization of semantics must be organized and well-documented. The complexity of these steps is a direct result of the variance among the source systems. A trade-off exists between the complexity of programming needed to manage these transformations and the complexity of the human relationships needed to ensure that formats are synchronized among separate institutions. Certain systems represent data by using extensible markup language (XML), thus allowing data to be manipulated through standard transformation parsers. PHIN specifies use of XML and the need for a data-translation capability, without specifying software packages or platforms.

Analysis Set Creation and Delivery

Finally, integrated and normalized data need to be presented for analysis. The performance of detection algorithms is being researched, and the needs of different detection algorithms vary (*22*). Even when using a specific algorithm, users might not know whether to count each patient as a single data point or allow multiple data points for a patient who meets criteria for multiple syndromes. A flexible query system can present multiple analysis sets either to a human or to an automated detection algorithm. PHIN calls for the capability to analyze, display, report, and map data. These features are implemented in the model systems, but not always with comparable algorithms.

Architectures for Data Integration

The challenges of integrating data from heterogeneous sources into a single analysis set are not unique to public health surveillance. Decision support in business endeavors often depends on integrating and analyzing diverse data sets. Clinical practice increasingly requires this capability, as patient information is often widely distributed and patient care requires access to information at other institutions. This need to access distributed information is central to automated public health surveillance.

Three generations of data-integration techniques in biomedical informatics have been described (*10*). The simplest approach to data integration is to build a large-source system containing all data needed to satisfy a query. As data have multiplied, along with their diversity, uses, and ownership considerations, new integration approaches have been developed. Second-generation models integrate data from multiple sources at a central location. This technology is almost universal in public health surveillance systems. A third-generation approach is emerging that involves constructing relations between data sources so that they appear integrated to the surveillance user, even though the data remain at their original location, subject to the control of their original owner. Distinct models for this third-generation approach exist; research in this area has only recently been applied to public health surveillance.

First-Generation Integration

Surveillance based on first-generation systems is not practical unless a single information system contains sufficient data to represent the population of interest. A slight enhancement is the manual combination of data from multiple nonintegrated sources. This is often a first step in local public health surveillance. A health department might receive files containing surveillance reports and combine them manually by using a spreadsheet program, desktop database, or statistical data management package. This approach is straightforward but can result in data fields with cryptic, local meanings and in data elements represented in a combination of nonuniform coding schemes from different sources.

Second-Generation Integration

Second-generation integration has been characterized as the consolidation of data through enterprise information architecture (*10*). One sophisticated second-generation approach is *data warehousing*, characterized as "historical, summarized, and consolidated data… targeted for decision support" (*23*). Data warehousing systems are common in business and widely available in health care. Their characteristics closely match those desired for public health surveillance, although data are less timely than desired. Warehouse data are typically historic; although historic data can be useful for research and for developing event-detection algorithms, ongoing surveillance requires current data. At present, the common model for automated public health surveillance systems is a data warehouse with frequent updates (although the term *warehouse* is not typically used in syndromic surveillance literature).

Although multiple approaches to data warehousing exist, all approaches are implemented through construction of a centralized database that is optimized for resource-intensive queries against a substantial portion of data. Data from other sources are typically imported to the central warehouse database after a query is sent from the central database to the source database. A lag associated with periodic imports from the clinical database(s) into the warehouse is commonplace and has been noted in multiple query-based public health surveillance systems (*2,4,5,11*). Another limitation of this model is that a global schema, or data structure, is required, and a change in this schema typically requires changes in the import procedures from each source system.

One variant of this approach is for data sources to run local queries and transmit resulting data on a schedule. A second variant involves filtering the electronic data interchange messages used to transfer data between components of an enterprise clinical information system and storing data contained therein. These data are usually formatted according to the HL7 standard. This approach can improve data timeliness substantially, and the uniformity of HL7 encoding might simplify software development. However, substantial variation is permitted within the standard, and HL7-message decoding often requires customization. Moreover, this approach requires a consistent global schema and the mapping of that schema into multiple local variations. This approach is exemplified by RODS (*2*).

Third-Generation Integration

The need for future information systems to support automated information acquisition processes for decision-making activities was identified over a decade ago (*24*). To achieve this automation, an approach based on so-called *mediators* (i.e., software agents that translate a query from a global format to an appropriate local format for a specific database) was proposed. This model, in which queries are run against distributed, in-situ data, can be classified as a third-generation data-integration system.

Contemporary approaches to third-generation systems include federated databases, mediated query systems, and peer-to-peer data sharing. These approaches share an apparent integration of information that remains housed in multiple source systems. Although these systems might be more complex to build than earlier generation systems, they share the advantage that data are queried from their original location, which improves accuracy and decreases lag. Additionally, lightweight queries can be run routinely for surveillance purposes with less performance impact on the source system. More richly detailed underlying data might be equally available for focused investigations. However, early third-generation approaches expose the source system to performance degradation from additional queries and require that the source system be online to process a query. Ongoing research is aimed at minimizing these shortcomings and strengthening the approaches' advantages.

Federated databases are an association of independent databases that allow queries from a single source but have no common schema or organization (*25*). The lack of a common schema means that any application must contain the local schema for every database it wishes to query. This is efficient because the queries generated require no translation for the source systems, but each new data source added to the system might require a change in each application accessing data from the federation. A number of federated-database models exist; these differ in the locus and degree of centralized control over access to the system (*26*). The Kleisli system is one federated-database approach for integrating bioinformatics data, in which a set of drivers provides access to various heterogeneous databases (*27*).

One proposed mediator query model has been implemented in biomedical applications, which again provide integrated access to online genetic databases (*24*). Examples include the Biomediator (*28*) and transparent access to multiple bioinformatics information sources (TAMBIS) (*29*) systems. Mediated schema models offer real-time queries directly against source systems, combined with the single global schema of a data warehouse. This greatly simplifies application writing, as authors need to understand only the single common schema. This model has not yet been implemented in public health surveillance.

Perhaps the best-known applications of peer-to-peer communication are music- and file-sharing services (e.g., Napster and Gnutella). These services use somewhat different peer-to-peer models, using a common schema but maintaining their common index information in either a centralized or distributed fashion, respectively. This peer-to-peer file-sharing model, extended to include peered communication among intelligent data-sharing agents, has been described as a peer-data–management system (*30*).

Although third-generation systems are not widespread in public health surveillance, these models are promising. First, whether executed through a mediated schema or against a series of autonomous peer agents, the queries in these systems run directly against the source data. Timeliness and accuracy are ensured, and performance concerns can be mitigated by different strategies. These architectures are suitable to run against both modern and legacy databases, transparently presenting an integrated view of both. The intelligence built into each participant of a peer-data–management system lends itself to supporting queries that can dynamically configure themselves against the available data sources when they are run. Finally, local control over data sources, which is inherent in both peer-data–management systems and the mediated-schema approach, might enable owners of data at any level to provide access and detail appropriate to different stakeholders and in different situations while maintaining control of their own data.

Conclusions

In response to the threat of biologic terrorism, information-system–based public health surveillance has evolved rapidly. Second- and third-generation approaches offer the greatest utility for public health surveillance, and research is critical to the continued advancement of surveillance systems, especially third-generation systems.

Future research on surveillance architectures should explore combinations of methods. For example, a data warehouse that provides the data consolidation, rich historic record, comprehensible data structure, and ability to query the entire corpus of data at the local public health jurisdiction might be combined with a third-generation model for sharing of situation-dependent views of those data. The nature of this integration will be driven by issues of data ownership and privacy, as well as by an evolving understanding of the optimal data for various uses. Broad-scale application of these systems will also require policy development to address concerns of privacy and proprietary data. Public health agencies should partner with universities and research organizations to shape the agenda for data-integration research. At the same time, academics need

public health partners to ensure that research questions are grounded in relevant problems.

In addition, public health agencies must rely on proven technologies for their operational needs. This implies working with information system vendors to take advantage of data-integration solutions and to ensure those solutions meet public health needs.

As has been the case in clinical informatics, where data used for outcomes research are also useful for chronic disease management, quality assurance, health services research, and other purposes, surveillance systems likely will evolve to enhance public and environmental health practice and management. Public health leaders should pay attention to how these data-integration models scale in other domains; links with the research community will prove helpful. Although public health agencies must serve an immediate operational role in national security, aggressive research is required to extend the frontiers of data integration to ensure success.

Acknowledgments

Peter Mork, University of Washington Department of Computer Science and Engineering, provided a thorough review of the architecture section.

References

1. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. J Urban Health 2003;80(2 Suppl 1):i76–88.
2. 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.
3. Lombardo J, Burkom H, Elbert E, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health 2003;80(2 Suppl 1):i32–42.
4. Platt R, Bocchino C, Caldwell B, et al. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. J Urban Health 2003;80(2 Suppl 1):i25–31.
5. Yasnoff WA, Overhage JM, Humphreys BL, LaVenture M. A national agenda for public health informatics: summarized recommendations from the 2001 AMIA Spring Congress. J Am Med Inform Assoc 2001;8:535–45.
6. National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine. J Public Health Manag Pract 2001;7:43–50.
7. CDC. Public Health Information Network standards, functions, and specifications: Version 1.2. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/phin/architecture/index.htm.
8. Gartner, Inc. Technical review of issues related to version 1 of the Public Health Information Network functions and specifications. Stamford, CT: Gartner, Inc., 2003. Available at http://www.cdc.gov/phin/conference_presentations/05-13-03.
9. Stead WW, Miller RA, Musen MA, Hersh WR. Integration and beyond: linking information from disparate sources and into workflow. J Am Med Inform Assoc 2000;7:135–45.
10. Lober WB, Trigg LJ, Karras BT, et al. Syndromic surveillance using automated collection of computerized discharge diagnoses. J Urban Health 2003;80(2 Suppl 1):i97–106.
11. Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc 2002;9:105–15.
12. CDC. Public Health Information Network logical data model version 1.0 user guide. Atlanta, GA: US Department of Health and Human Services, CDC, 2004. Available at http://www.cdc.gov/phin/data_models/index.htm.
13. CDC. HIPAA privacy rule and public health. Guidance from CDC and the US Department of Health and Human Services. MMWR 2003;52(S-1).
14. Kress A. Data provider relationships: pros, cons, and considerations [Abstract]. MMWR 2004;53(Suppl):247.
15. 45 CFR § 164.514 (a)–(c).
16. Ehringhaus S. Testimony on behalf of the Association of American Medical Colleges before the National Committee on Vital and Health Statistics Subcommittee on Privacy. November 19, 2003. Available at http://www.ncvhs.hhs.gov/031120p2.pdf.
17. CDC. Vocabularies: PHIN notifiable condition mapping tables, data models. Atlanta, GA: US Department of Health and Human Services, 2003. Available at http://www.cdc.gov/phin/data_models.
18. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.bt.cdc.gov/surveillance/syndromedef/index.asp.
19. 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.
20. Health Level Seven, Inc. Health Level Seven. Ann Arbor, MI: Health Level Seven, Inc., 2003. Available at http://hl7.org/.
21. CDC. Function overview and supporting materials: IT security and critical infrastructure protection. Atlanta, GA: US Department of Health and Human Services, CDC, 2003. Available at http://www.cdc.gov/cic/functions-specs/function_9.htm.
22. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004; 11:141–50.
23. Chaudhuri S, Dayal U. Data warehousing and OLAP for decision support. ACM SIGMOD Record 1997;26:507–8.
24. Wiederhold G. Mediators in the architecture of future information systems. Computer 1992;25(3):38–49.
25. Bright M, Hutson A, Pakzad SH. A taxonomy and current issues in multidatabase systems. IEEE Computer Magazine 1992;25:50–60.
26. Sheth AP, Larson JA. Federated database systems for managing distributed, heterogeneous, and autonommous databases. ACM Computing Surveys 1990;22:183–236.
27. Wong L. Kleisli, a functional query system. J Funct Prog 2000;10:19–56.
28. Mork P, Halevy A, Tarczy-Hornoch P. A model for data integration systems of biomedical data applied to online genetic databases, Proc AMIA Symp 2001:473–7.
29. Baker PG, Brass A, Bechhofer S, Goble C, Paton N, Stevens R. TAMBIS: transparent access to multiple bioinformatics information sources. Proc Int Conf Intell Syst Mol Biol 1998;6:25-34.
30. Tatarinov I, Ives Z, Madhavan J, et al. The Piazza peer data management project. ACM SIGMOD Record 2003;32:47–52.

Perspective of an Emergency Physician Group as a Data Provider for Syndromic Surveillance

Dennis G. Cochrane
Emergency Medical Associates of New Jersey Research Foundation, Morristown, New Jersey; Morristown Memorial Hospital Residency in Emergency Medicine, Morristown, New Jersey

Corresponding author: Dennis G. Cochrane, Morristown Memorial Hospital Residence in Emergency Medicine, Morristown, NJ 07962. Telephone: 973-971-8919; E-mail: cochraned@verizon.net.

Abstract

The need for enhanced biologic surveillance has led to the search for new sources of data. Beginning in September 2001, Emergency Medical Associates (EMA) of New Jersey, an emergency physician group practice, undertook a series of surveillance projects in collaboration with state and federal agencies. This paper examines EMA's motivations and concerns and discusses the collaborative opportunities available to data suppliers for syndromic surveillance. Motivations for supplying data included altruism and public service, previous involvement in terrorism and disaster preparedness, academic research interests, and the opportunity to find added value in the group's existing information systems. Concerns and barriers included cost, maintaining patient confidentiality, and challenges in interacting with the public health community. The extensive and carefully maintained electronic medical record enabled EMA to conduct multiple studies in collaboration with state and federal agencies. The electronic medical record provides useful data that might be more sensitive and specific in detecting outbreaks than the patient-chief–complaint data more commonly used for surveillance. EMA's experience also indicates that opportunities exist for the public health community to work with emergency physicians and emergency physician groups as suppliers of data. Such collaborations not only are useful for syndromic surveillance systems but also can help build relations that might facilitate a response to an actual biologic attack.

Introduction

The terrorist attacks of September 11, 2001, and the subsequent release through the mail of *Bacillus anthracis* have increased awareness of the risk for biologic attack. The 2003 severe acute respiratory syndrome (SARS) outbreak also demonstrated the threat of emerging infectious diseases. Certain types of biologic attacks or emerging infectious disease outbreaks might initially present with nonspecific symptoms across a large population. At this stage of disease, a pathologic diagnosis might not be possible, although the symptoms might fall into a definable syndrome. Syndromic surveillance uses available data sources to detect such outbreaks at the earliest possible stage so early action can be taken to mitigate the effects and spread of disease.

Researchers are evaluating the early detection potential of such data sources as pharmacy sales, school and work absenteeism, and emergency department (ED) patient chief complaints. This paper discusses a less commonly used source of ED data — clinical data from an electronic medical record maintained by an emergency physician group practice. Such data can be made available in real time and can include detailed patient demographics, electronic versions of physicians' notes, physicians' choice of charting templates, laboratory test results, and clinical diagnoses. This paper discusses the motivations and concerns of an emergency medicine group as a data provider and examines opportunities for collaboration between the public health and emergency medical communities. It also describes how these data have been used for research in syndromic surveillance and how data from an electronic medical record might be used for enhanced real-time surveillance.

Practice Setting and Available Data Types

Emergency Medical Associates of New Jersey (EMA) is an emergency physician group practice that is fully owned by the practicing physicians and is constituted as a professional association. EMA contracts with hospitals to provide physician and physician-assistant coverage for 16 EDs in central and southern New Jersey and in New York State, with a combined volume of approximately 2,000 patients/day. The hospitals are a mixture of community hospitals and teaching hospitals, and group members function as faculty for two emergency medicine residencies. The practice receives an estimated one third of all ED visits in the northern half of New Jersey.

Patient visits are recorded by using the group's proprietary clinical software, EDIMS™ (Emergency Department Information Manager System). The software is integrated with the hospital's patient registration system and stores patient demographic information. It also tracks patient location and status during ED visits and records physicians' notes through a system of charting templates.

All data are uploaded electronically to EMA's central office in Livingston, New Jersey. Reports are generated by a proprietary reporting system, eMars™ (Emergency Medicine Analysis and Reporting System). These reports are routinely used to monitor billing and ED operations. All data are maintained in an Oracle™ database. Full clinical data, including physicians' electronic notes, are available from January 1996 to present. Billing data, including *International Classification of Diseases, Ninth Revision* (ICD-9) billing codes, are available from January 1988 to present.

Surveillance after the September 11, 2001, Terrorist Attacks

Before September 11, 2001, the primary research use of the eMars database had been for epidemiologic studies of emergency medicine conducted by the group's physicians (*1–4*), who had minimal interest in biologic surveillance. Any interest in disaster management and multiple casualty incidents was concentrated on internal and external disaster plans.

This changed dramatically after the terrorist attacks on the World Trade Center (WTC) in downtown Manhattan on September 11, 2001. On that day, the group's emergency physicians waited at their EDs or at disaster staging sites near the WTC for a potential onslaught of patients that never materialized. Because the threat of an associated biologic attack seemed real, physicians at each ED scrambled to prepare their decontamination equipment and gather information about illnesses that might result from such an attack. They understood that daily life had changed fundamentally and that emergency physicians needed to rethink aspects of disaster preparedness, especially the need to detect and respond to a biologic attack.

Although the WTC attack did not include a release of a biologic agent, it was soon followed by the mailborne release of *B. anthracis*. Over the following months, a substantial number of patients reported to EDs to "get checked for anthrax." EMA's 16 hospitals treated as many as 62 patients/day (representing 3.5% of all visits groupwide) requesting a test for exposure to *B. anthracis* and often requesting prophylactic medications. These patients were expecting expert, reliable advice. The ED physician's sense of responsibility was reinforced when an emergency physician was sued for failing to detect one of the first cases of anthrax. Emergency physicians already knew that the ED needed to be prepared to respond to a mass-casualty biologic attack and now realized that they could be held legally liable for not detecting an attack in its earliest phases.

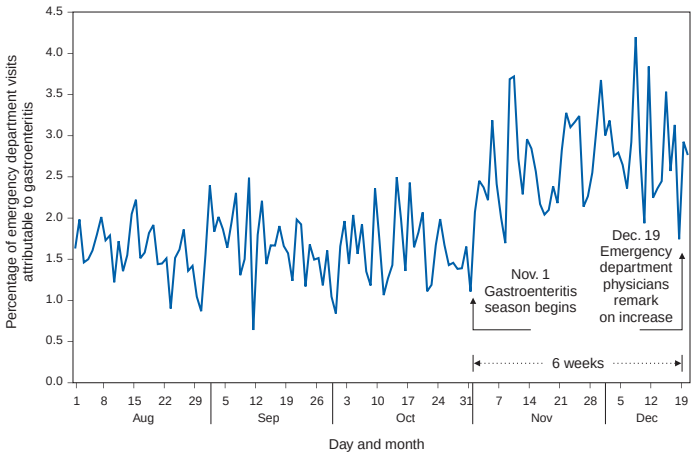
Difficulty of Detecting Changes in Illness Patterns

Surveillance for sentinel cases would rely on astute observation by the ED physician. The New Jersey Department of Health and CDC websites were helpful in establishing diagnostic criteria and reporting mechanisms. The majority of emergency physicians would likely identify a sentinel case of anthrax if the features were typical. However, physicians also realized that in a biologic attack, a person might report initially in a nonspecific way. In addition to looking for a sentinel case, physicians were also advised by CDC to look for "illness patterns and diagnostic clues that might indicate an unusual infectious disease outbreak associated with intentional release of a biologic agent" (*5*).

The individual emergency physician, working in isolation, might have difficulty detecting a subtle increase in patients reporting with a given nonspecific symptom. Emergency physicians see patients with a diverse group of illnesses whose incidence varies widely. On any given day, emergency physicians expect a greater than usual disease incidence of one or more conditions on the basis of chance alone. For example, at the end of a work shift, a physician might not report seeing three cases of diarrheal illness during that shift even though the average is only one case. A substantial change in case mix over a 24-hour period that would be obvious from examining aggregate data from multiple physicians might appear as random variation to an individual physician seeing only a subset of those patients.

Individual physicians face difficulties in identifying outbreaks. For example, in December 2002, two EMA physicians examined EMA's ED volume data to determine whether the data indicated a seasonal gastroenteritis outbreak, which they believed had started 2 weeks earlier. The data revealed that the outbreak had actually begun 6 weeks earlier (Figure). An ED physician might need to work multiple shifts over a week or more to notice an aberration (e.g., a doubling or tripling of the average number of gastroenteritis cases). The difficulty of outbreak detection is even greater when an individual physician is looking for multiple disease patterns simultaneously.

FIGURE. Disparity between physicians' perceptions of the start of a gastroenteritis outbreak and the actual start of the outbreak



Syndromic surveillance of aggregate visit data is an important component of preparedness. A biologic attack might cause a sudden increase in the volume of patients with a specific set of symptoms that would be invisible to an individual physician but apparent to analysts using combined data in real time. This might activate heightened surveillance for sentinel cases. In addition, even if an attack was first detected by other means, having rapid access to information about patient volumes could help determine the appropriate response and allocation of resources.

Motivations for Participating in Syndromic Surveillance

EMA was motivated to become involved in syndromic surveillance for multiple reasons. An initial motivator was that syndromic surveillance represented an opportunity for doing research needed to validate its effectiveness. EMA's academic physicians had conducted epidemiologic research by using the billing and clinical databases for >15 years; the same methods could be adapted for research into syndromic surveillance. In particular, EMA's well-maintained and clinically rich database and substantial patient volume would facilitate the study of questions difficult to research in other settings. By collaborating with other agencies, especially public health, EMA physicians might be able to make a contribution to this new field.

The opportunity for public service was another motivator. EMA hospitals cover approximately one third of all ED visits in central and northern New Jersey. Therefore, the group might be able to contribute directly to syndromic surveillance efforts locally. The availability of real-time clinical information from the electronic medical record might offer a unique ability to track and respond to outbreaks.

Another motivator for the group's administrators was that involvement in syndromic surveillance might enhance the group's image in the marketplace. By participating in important public health efforts, EMA's physician group might be perceived as being at the forefront of the specialty in this new area. Such projects might also be a way to demonstrate the added value of the group's information management systems.

Finally, emergency physicians have a personal vested interest in early detection of outbreaks. As illustrated by the 2003 severe acute respiratory syndrome (SARS) epidemic, emergency and hospital personnel can become infected at the initial stages of an outbreak, and health-care personnel can be disproportionately affected overall. Any advance warning could help emergency physicians augment infection-control procedures at the earliest possible time.

Potential Barriers to Participating in Biologic Surveillance

One difficulty with implementing any project in an ED setting is the environment's unpredictable and often chaotic nature. However, the clinical systems in place gather data that can be obtained passively without making further demands on personnel.

Costs were also a concern. To an extent, EMA's robust reporting system, used to produce regular reports for billing, financial management, and operations management and to track physician productivity, could be readily adapted to syndromic surveillance. Because the needed data were already being gathered for other purposes, the expense of generating reports for research would be minimal. The larger expense would come from improving system infrastructure to accommodate the real-time gathering of data. Data are collected daily, but certain data reporting is delayed up to 3 days to ensure its completion on site. Implementing real-time reporting would require system enhancements to enable the necessary fields to be uploaded immediately. Another cost might be the need to reformat data in a standardized format to share with local, state, and national agencies.

Initial costs for generating reports for research purposes were accommodated through the EMA Research Foundation. Improvements in real-time gathering of data were included in an upgrade of EMA's data collection systems. Recently, EMA initiated Internet-based reporting of syndromic trends to EMA physicians as part of a program to facilitate communications and operations within the group using Internet-based technology.

Patient confidentiality and compliance with the Health Insurance Portability and Accountability Act of 1996 (HIPAA)

was another concern. Fortunately, EMA's billing and information systems personnel are well-versed in HIPAA requirements and experienced in sharing de-identified subsets of data for billing and reporting purposes. However, large-scale biologic surveillance of the ED population could be perceived as invasive of privacy. This problem might be reduced in a group medical practice in which clinical follow-up and quality control activities are part of everyday operations. Patients are often contacted the day after an ED visit to ensure follow-up with physicians' instructions and to monitor patient outcomes. Biologic surveillance for unusual clusters or trends is a reasonable extension of ongoing medical services. All individual identifying information stays within the practice.

A more substantial barrier is potential resistance within the public health community. Syndromic surveillance is a new field that requires research and validation. Responding to alerts from a syndromic surveillance system might burden the public health infrastructure. Pursuing syndromic surveillance would be futile without the interest of the public health community. Ultimately, EMA identified ample opportunities to collaborate with public health agencies.

Research Projects and Collaborations

EMA's initial research effort into syndromic surveillance was to determine whether the existing database could track known seasonal disease outbreaks. A set of nine ICD-9 code syndrome groupings were developed and used to filter the database. This enabled creation of time-series graphs for each syndrome group over a period of years (*6*). The data were encouraging in that they depicted seasonal variation for nearly all of the syndrome groups. The seasonal influenza epidemic was identified, as were annual spikes appearing to correlate with the seasonal rotavirus epidemic in children. Seasonal variations in asthma were also identified.

EMA first collaborated with the New York State Department of Health (NYSDOH) to study biologic surveillance methods based on patient chief complaints. By applying methods adapted from the New York City Department of Health and Mental Hygiene to the EMA database, the group was able to demonstrate key seasonal illness patterns, particularly the influenza season, by using a chief-complaint methodology (*7*). The system's ability to track the influenza season lent credence to its ability to detect other types of outbreaks.

As part of a syndromic surveillance working group, EMA also supplied data for a study of syndromic definitions and ICD-9 code groupings (*8*). The working group included members of the U.S. Department of Defense's Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) project, CDC, Harvard-Pilgrim Health Care, and EMA. EMA's contribution was to supply ED data that could be used to test different choices of ICD-9 groupings. In addition to providing raw data, participating EMA personnel were able to interpret the data in the ED setting. The data allowed the working group to identify ICD-9 codes commonly used in the ED and differentiate them from codes that are less commonly used but might be better markers of biologic terrorism. This provided a rationale for stratifying codes within a syndrome, so that, if desired, the more common but less specific codes could easily be removed to search for a signal among the less common but more specific codes.

Having studied the existing ICD-9 and chief-complaint methods, EMA and NYSDOH were then able to compare the sensitivity and specificity of the two methods by using a single database (*9*). This study examined the chief-complaint method for respiratory syndrome by using the ICD-9 method as the criterion standard (*9*). Two results emerged. First, a substantial difference between the syndrome definitions used for the two methods was noted; although the study initially found poor sensitivity (31%) for chief complaints as compared with ICD-9 codes, sensitivity improved substantially when the methods were adjusted to more closely reflect similar syndrome definitions (sensitivity: 53%). Second, a difference existed between the information captured in the chief complaint and the information captured in the ICD-9 code that could not be resolved. For example, a patient with a chief complaint in the respiratory syndrome (e.g., cough) might easily be assigned an ICD-9 code in a different syndrome (e.g., fever), and vice versa.

These studies were facilitated by the fact that the existing corporate database, originally developed for billing and clinical purposes, was able to provide a large data set with consistent capture of ICD-9 codes and clinical information. Also, the existing data-analysis methods, originally used for corporate analysis, proved a good match for the needs of syndromic surveillance.

Unique data sources within EMA's electronic medical record were also explored. For example, in EMA's system, the physician chooses one of approximately 450 charting templates at the time he or she sees the patient. Thus, the physician's choice of charting template is available in real time before the patient leaves the ED. Because the choice of charting template embodies the physician's clinical judgment, a high level of agreement can exist between the physician's choice of charting template and the final ICD-9 coding of the patient. Comparison of the two methods using the Kappa statistic

determined approximately perfect agreement for the *asthma*, *chest pain*, and *headache* filters; excellent agreement for the *skin*, *any gastrointestinal*, and *diarrhea* filters; and moderate agreement for the *respiratory* and *fever* filters. (*10*). These results indicate that the physician's choice of charting template might be useful for real-time biologic surveillance when available in an electronic medical record system.

Ongoing Role for Biologic Surveillance Within EMA's Group Practice

Biologic surveillance reports are now included with other daily, weekly, and monthly reports generated by eMars. These reports provide statistics on the incidence of illness within the practice, categorized by syndrome (e.g., gastrointestinal, respiratory, or febrile illness). The statistics are based on chief-complaints, ICD-9 codes, and physician's choice of charting templates. Reports are circulated to the practice's physicians through group e-mail and posted on the group's website.

Anecdotal feedback from physicians indicates that they appreciate these biologic surveillance reports because they provide early warning of disease outbreaks. For example, by knowing that the influenza season had begun, physicians were able to apprise themselves of the latest recommendations and options for influenza treatment and prophylaxis. A lively discussion ensued about the possible use of neuraminidase inhibitors for patients reporting to the ED with influenza-like symptoms and for their caretakers. In another example, when reports revealed that the annual pediatric gastroenteritis epidemic had begun, EMA's pediatric emergency physicians were able to adjust their treatment of affected children. For children whose pattern of illness matched the pattern expected for the seasonal epidemic, physicians felt more comfortable proceeding with fewer laboratory tests and trusting their clinical impressions. This reduced time, expense, and patient discomfort.

Conclusion

EMA has successfully used its corporate database for collaborative studies with public health agencies. Such efforts represent only a limited portion of similar projects completed or underway in the emergency medicine community, as multiple publications have documented (*11–20*). Emergency physicians are active in disaster management and terrorism preparedness locally as well as at county, state, and federal levels. Emergency physicians are not only involved in passive surveillance but also have participated in active surveillance (e.g., the drop-in SARS surveillance system implemented recently in Milwaukee [*21*]).

The partnerships that result from collaborative biologic surveillance projects might be more important than the projects themselves. Because the nature of a future terrorist attack cannot be anticipated, developing collaborative relationships now will enhance the ability of public health authorities to respond flexibly and effectively should such an attack occur.

EMA's experience indicates that opportunities exist for the public health community to work with emergency physician groups as data providers. Such collaborations are useful not only for syndromic surveillance but can also help build relations that might be useful when responding to an actual biologic attack.

Acknowledgments

The author appreciates the assistance of the New York State Department of Health (Carolyn Stetson, Melissa Mocny, and Trang Nguyen), the New York City Department of Health and Mental Hygiene (Richard Heffernan and Farzad Mostashari), the ESSENCE Project (Julie Pavlin and Virginia Foster), CDC (Dan Sosin and Samuel Groseclose), and the Emergency Medical Associates Research Foundation (John Allegra and Jonathan Rothman). Ed Barthell also provided valuable insight.

References

1. Cody RP, Weisel CP, Birnbaum G, Lioy PJ. The effect of ozone associated with summertime photochemical smog on the frequency of asthma visits to hospital emergency departments. Environmental Research 1992;58:184–94.
2. Allegra JR, Cochrane DG, Biglow R. Monthly, weekly, and daily patterns in the incidence of congestive heart failure. Acad Emerg Med 2001;8:682–5.
3. Allegra JR, Cochrane DG, Allegra E, Cable G. Calendar patterns in the occurrence of cardiac arrest. Am J Emerg Med 2002;20:513–7.
4. Alves DW, Allegra JR, Cochrane DG, Cable G. Effect of lunar cycle on temporal variation in cardiopulmonary arrest in seven emergency departments during 11 years. Eur J Emerg Med 2003;10:225–8.
5. CDC. Recognition of illness associated with the intentional release of a biologic agent. MMWR 2001;50:893–7.
6. Cochrane D, Allegra J, Rothman J. Real-time biologic surveillance using an existing emergency department electronic medical record database [Abstract]. J Urban Health 2003;80(2 Suppl 1):i120.
7. Shannon R, Davisson M, Trang N, Stetson C, Jones K. The use of hospital emergency department chief complaint data as a "near" real-time marker for assessing public health risk of infectious disease outbreak [Abstract]. J Urban Health 2003;80(2 Suppl 1):i121.
8. Foster V. Syndrome definitions for disease associated with critical agents. Presented at the 2003 National Syndromic Surveillance Conference, New York Academy of Medicine, New York, New York, October 24, 2003.
9. Mocny M, Cochrane DG, Allegra JR, et al. A comparison of two methods for biologic surveillance of respiratory disease in the emergency department: chief complaint versus ICD-9 diagnosis code [Abstract]. Acad Emerg Med 2003;10:513.

10. Cochrane DG, Allegra JR, Rothman J. Comparison of physician's choice of charting template to ICD-9 codes for biologic surveillance using an emergency department electronic medical records database [Abstract]. Acad Emerg Med 2003;10:525.
11. Barthell EN, Cordell WH, Moorhead JC, et al. The Frontlines of Medicine Project: a proposal for the standardized communication of emergency department data for public health uses including syndromic surveillance for biological and chemical terrorism. Ann Emerg Med 2002;39:422–9.
12. Moran GJ, Kyriacoud DN, Newdow MA, Talan D. Emergency department sentinel surveillance for emerging infectious diseases. Ann Emerg Med 1995;26:351–4.
13. EMERGEncy ID NET Study Group. EMERGEncy ID NET: an emergency department-based emerging infections sentinel network. Ann Emerg Med 1998;32:703–11.
14. Cordell WH, Overhage JM, Waeckerle J. Strategies for improving information management in emergency medicine to meet clinical, research, and administrative needs. Ann Emerg Med 1998;31:2.
15. Richards C, Brustein J, Waeckerle J, Hutson HR. Emergency physicians and biological terrorism. Ann Emerg Med 1999;34:183–90.
16. Hirshon JM. The rationale for developing public health surveillance systems based on emergency department data. Acad Emerg Med 2000;7:1428–32.
17. Pollock DA, Lowery DW. Emergency medicine and public health: new steps in old directions. Ann Emerg Med: 2001;38:675–84.
18. 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.
19. Suyama J, Sztajnkrycer M, Lindsell C, Otten E, Daniels J, Kressel A. Surveillance of infectious disease occurrences in the community: an analysis of symptom presentation in the emergency department. Acad Emerg Med 2003;10:753–63.
20. Begier E, Sockwell D, Branch L, et al. The national capitol region's emergency department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? Emerg Infect Dis 2003;9:393–6.
21. Foldy SL. Linking better surveillance to better outcomes. MMWR 2004;53(Suppl):12–7.

SARS Surveillance Project — Internet-Enabled Multiregion Surveillance for Rapidly Emerging Disease

Seth L. Foldy,[1,2] E. Barthell,[2,3] J. Silva,[4] P. Biedrzycki,[1] D. Howe,[1] M. Erme,[5] B. Keaton,[6] C. Hamilton,[7] L. Brewer,[7] G. Miller,[8] E. Eby,[9] R. Coles,[10] K. Pemble,[3,11] C. Felton[3]

[1]Milwaukee Health Department, Milwaukee, Wisconsin; [2]Medical College of Wisconsin, Milwaukee, Wisconsin; [3]Infinity Healthcare, Inc., Milwaukee, Wisconsin; [4]Silva Consulting Services, Eldersburg, Maryland; [5]Akron Health Department, Akron, Ohio; [6]Summa Health System, Akron, Ohio; [7]Tarrant County Public Health, Fort Worth, Texas; [8]University of Colorado Health Science Center, Denver, Colorado; [9]Emergency Physicians at Porter Hospitals, Denver, Colorado; [10]Colorado Department of Public Health and Environment, Denver, Colorado; [11]University of Wisconsin at Milwaukee, Milwaukee, Wisconsin.

Corresponding author: Seth L. Foldy, 3061 N. Marietta Avenue, Milwaukee WI 53211. Telephone: 414-906-0036; Fax: 414-906-9667; E-mail: sfoldy@sbcglobal.net.

Abstract

Introduction: *On March 15, 2003, CDC requested health-care and public health agencies to conduct surveillance for severe acute respiratory syndrome (SARS). The SARS Surveillance Project (SARS-SP) was established to rapidly implement multiregional SARS surveillance in emergency departments (EDs) by using existing Internet-based tools.*

Objectives: *The objectives of SARS-SP were to 1) disseminate and update SARS screening forms for ED triage, 2) establish surveillance for SARS syndrome elements by using Regional Emergency Medicine Internet (REMI), 3) expand surveillance to multiple regions, and 4) evaluate the usefulness of Internet tools for agile surveillance during a rapidly emerging global epidemic.*

Methods: *SARS-SP developed, distributed, and updated an Internet-based triage form to identify patients for infection control and public health reporting. EDs then were invited to report visit frequencies with various SARS syndrome elements to local public health authorities by using the REMI Internet application (first in one metropolitan area, and later in four). After pilot-testing in one metropolitan area, the surveillance system was implemented in three others.*

Results: *Active syndromic surveillance was established by health departments in Milwaukee, Wisconsin; Denver, Colorado; Akron, Ohio; and Fort Worth, Texas. A total of 27 EDs reported syndrome frequencies from >146,000 patient encounters.*

Conclusions: *ED and public health partners reported being satisfied with the system, confirming the usefulness of Internet tools in the rapid establishment of multiregion syndromic surveillance during an emerging global epidemic.*

Introduction

On March 15, 2003, CDC urgently requested health-care and public health agencies to conduct surveillance for severe acute respiratory syndrome (SARS) (*1*), a pneumonia later attributed to a newly discovered coronavirus (SARS-CoV). SARS had spread rapidly by air travel to three continents and appeared to be highly infectious to health-care workers and patients in health-care settings (*1*). The cause of SARS was then unknown, and diagnostic tests were lacking. Basic epidemiologic facts (e.g., the range of clinical symptomatology, whether persons with mild or asymptomatic infection could transmit disease, and the range of possible routes of infection) were unknown. Minimal assurance could be given that SARS was not already circulating in the United States. As a result, public health systems had to deploy complex, rapidly changing measures to protect health-care facilities and to take an agile approach to surveillance.

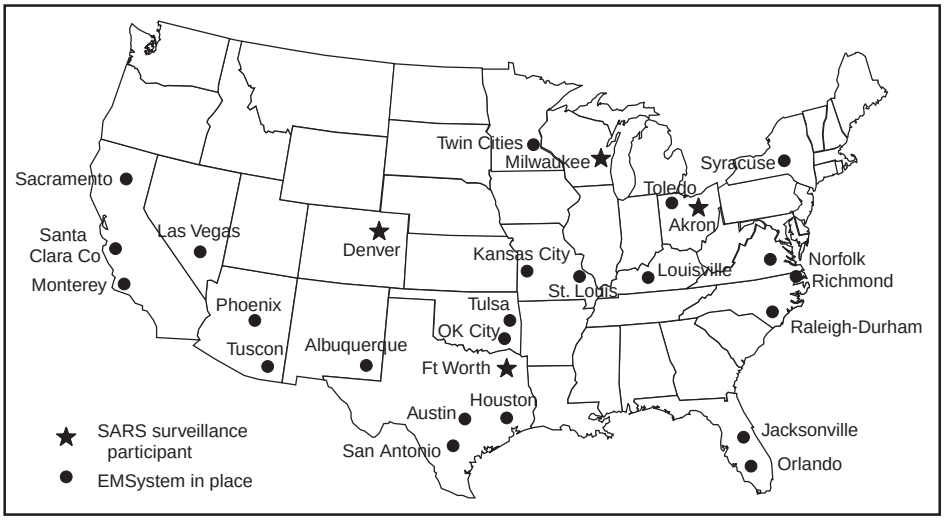
Frontlines of Medicine (http://www.frontlinesmed.org) is a collaborative of emergency medicine, public health, and informatics professionals organized to enable better public health surveillance of emergency department (ED) information (*2*). Frontlines of Medicine created the SARS Surveillance Project (SARS-SP) workgroup to develop, disseminate, and update a practical screening (case-finding) form for potential SARS patients in EDs. The form was used to measure daily ED volumes of SARS syndrome elements. These counts were transmitted and assembled regionally by using EMSystem® Regional Emergency Medicine Internet (REMI).* Because EMSystem was in use in 26 cities (Figure 1), syndromic surveillance developed in one city was presumed

* EMSystem is an Internet-served REMI that allows restricted viewing of Internet screens protected by standard Secure Sockets Layer (SSL) with 128-bit encryption, and can alert participants using text mail messages. EMSystem and similar networked REMI applications were developed to improve situational awareness of emergency departments regarding ambulance diversions, mass casualty events, and other emergency medical services system changes. They have since been used for other functions including public health alerting, monitoring health-care utilization and readiness, and syndromic surveillance (*3*).

to be portable to multiple urban areas quickly and inexpensively.

The objectives of SARS-SP were to 1) create, disseminate, and update SARS screening forms for ED triage, 2) conduct SARS surveillance by using REMI, 3) expand surveillance to multiple regions, and 4) evaluate the usefulness of Internet tools for agile surveillance during a rapidly emerging global epidemic. SARS triage forms and surveillance were field-tested in Milwaukee, Wisconsin. The form was then distributed over the Internet, and three other urban regions initiated surveillance.

FIGURE 1. EMSystem® and SARS Surveillance Project sites — United States, 2003



Methods

Case-Finding Triage Forms

A single-page screening form was created for ED triage personnel (available at http://www.frontlinesmed.org/SARS-SP). The form was designed to 1) identify patients requiring immediate infection control and public health notification (case finding) and 2) facilitate counting and reporting to public health officials the number of daily visits featuring SARS syndrome elements for time-trend surveillance. Three check boxes recorded the presence or absence of the following elements of the SARS case definition (hereafter referred to as "SARS elements"): fever (history or finding of temperature >38ºC); respiratory findings (i.e., cough, shortness of breath, difficulty breathing, pneumonia, or respiratory distress syndrome); and either recent travel to locations associated with SARS transmission or contact with a suspected SARS patient (hereafter referred to as "SARS risks"). Pulse oximetry <95% was recorded separately.

Screening was originally recommended only for patients with fever; later, after CDC recommended assessing patients for possible SARS on the basis of either fever or respiratory symptoms, triage personnel were instructed to screen patients with either complaint. The screening form encouraged ED staff to telephone the local public health authority immediately for any patient with the triad of fever, respiratory findings, and SARS risks.

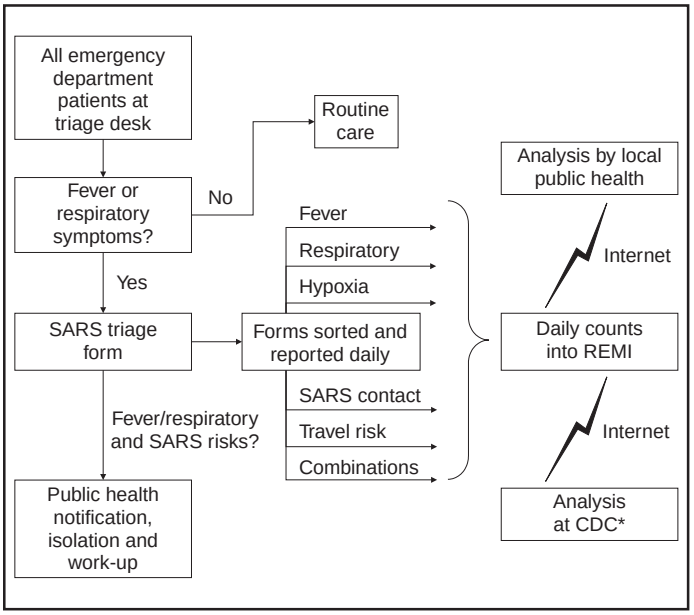
On March 17, 2003, forms were distributed to Milwaukee EDs via REMI. On March 30, revised forms were posted online and the national membership of the American College of Emergency Physicians was notified by e-mail of the screening form website. Persons downloading forms were invited to enter an e-mail address to receive notification of updated forms and to participate in the voluntary syndromic surveillance effort. Screening forms were revised twice (and registered users notified) to matching changing CDC recommendations.

Syndromic Surveillance

The Milwaukee Health Department (MHD) invited local EDs to report daily visit totals and the numbers of screened patients sorted by mutually exclusive combinations of SARS elements (e.g., fever only, fever with respiratory findings only, respiratory findings with SARS risks only, etc.). Because little was then known of the clinical spectrum of SARS infection, surveillance was performed for each clinical element so that health authorities could be alerted to rising rates of febrile or respiratory illness even if patients failed to meet CDC criteria for SARS diagnosis (Figure 2). The reporting system was similar to that employed in Milwaukee the previous summer during the 2002 Major League Baseball All-Star Game using EMSystem (*4,5*).

Detailed instructions were e-mailed to ED managers and mounted on REMI for reference, with a follow-up conference call. MHD staff provided assistance as needed. Designated ED staff collected all screening forms for 24-hour periods and sorted them into mutually exclusive sets of SARS elements. REMI automatically reminded EDs daily to enter the previous day's totals on a screen designed for that purpose. Only authorized staff could enter or view surveillance data.

FIGURE 2. Workflow for the SARS Surveillance Project



* CDC analysis conducted for short-term proof of concept in Milwaukee only.

Only visit counts were entered into REMI; no personally identifiable health information was transmitted. Triage personnel stamped each form with patient identification and retained completed forms in case public health investigation of a particular patient was needed. If REMI reports included visits with the triad of fever, respiratory illness, and SARS risks, public health officials could ask the ED to identify the patient.

During nationwide dissemination, those who downloaded screening forms were asked if they would conduct syndromic surveillance. If ED staff expressed interest, the state or local public health agency offered assistance. If EMSystem was not already in use in the area, local interface screens, log-on accounts, server accounts, data storage, and 24-hour/day technical assistance were offered at no charge to EDs and health departments, using existing EMSystem infrastructure.

Participating public health staff used password-protected accounts to download daily jurisdiction-specific data from REMI as a tab-delimited spreadsheet. Each health department had exclusive access to its local data and controlled how it was analyzed and acted on. Milwaukee data were also downloaded remotely at CDC for analysis with the Early Aberration Reporting System (EARS) to test the feasibility of remote analysis (*6*).

Surveys

Participating health department surveillance coordinators provided summary statistics and impressions of the project. In July 2003, surveys were also sent to nurse managers at the 13 participating Milwaukee-area EDs.

Results

During May–September 2003, a total of >500 SARS-SP website hits were logged, and 257 persons requested e-mail notification of screening-form changes. Much smaller numbers visited the site after receiving e-mail notification of revised forms. The total number of EDs or clinics that used the screening form is not known.

During March 19–June 25, 2003, a total of 13 Milwaukee-area EDs participated in syndromic surveillance of 105,669 visits. Three other metropolitan areas (Denver, Colorado; Akron, Ohio; and Fort Worth, Texas) established ED syndromic surveillance with reporting to health authorities. During April 23–May 31, 2003, nine EDs in Denver, Colorado, that already used REMI sent surveillance information on 16,997 encounters to the Colorado Department of Public Health and Environment (CDPHE). During May 1–June 1, 2003, three EDs in Akron, Ohio, reported information from 12,939 encounters to the Akron Health Department (AHD). Neither the hospitals nor AHD had previously used REMI. During May 12–October 12, 2003, two hospitals in Fort Worth, Texas, that already used REMI reported on 10,941 encounters to Tarrant County Public Health (TCPH), with surveillance continuing beyond October. EDs in eight other cities expressed interest in daily syndromic surveillance, but efforts to recruit a public health agency failed in seven. The eighth city initiated a surveillance pilot in fall 2003.

Only one person in all four cities ultimately met the CDC criteria for possible SARS, and no confirmed cases were reported. Thus, neither case-finding sensitivity nor specificity can be measured. During March 15–October 1, 2003, three of the four jurisdictions investigated 42 potential SARS cases, of which 22 (52%) were prompted by the triage form. In Milwaukee, five investigations originated from telephone calls about positive ED triage forms; four originated from REMI electronic reports; and five originated outside EDs. All 13 investigations by CDPHE began with REMI reports. All 15 TCPH investigations began before initiation of SARS-SP surveillance and originated from nonmedical settings (e.g. from airlines). No patient investigated for possible SARS visited a participating ED but failed detection by the screening form.

The median percentage of surveillance period days for which participating EDs reported syndrome frequencies electronically by using REMI was 89% (range: 52%–100%). The most common data-quality problems cited by public health surveillance coordinators were nonreporting, reports lacking total ED visit census, and errors in the date of surveillance; telephone calls were sufficient to resolve these concerns. In Milwaukee, questions and data-quality concerns required fre-

quent calls (7–9 daily) to and from EDs early in the project but only 1–2 calls by the end.

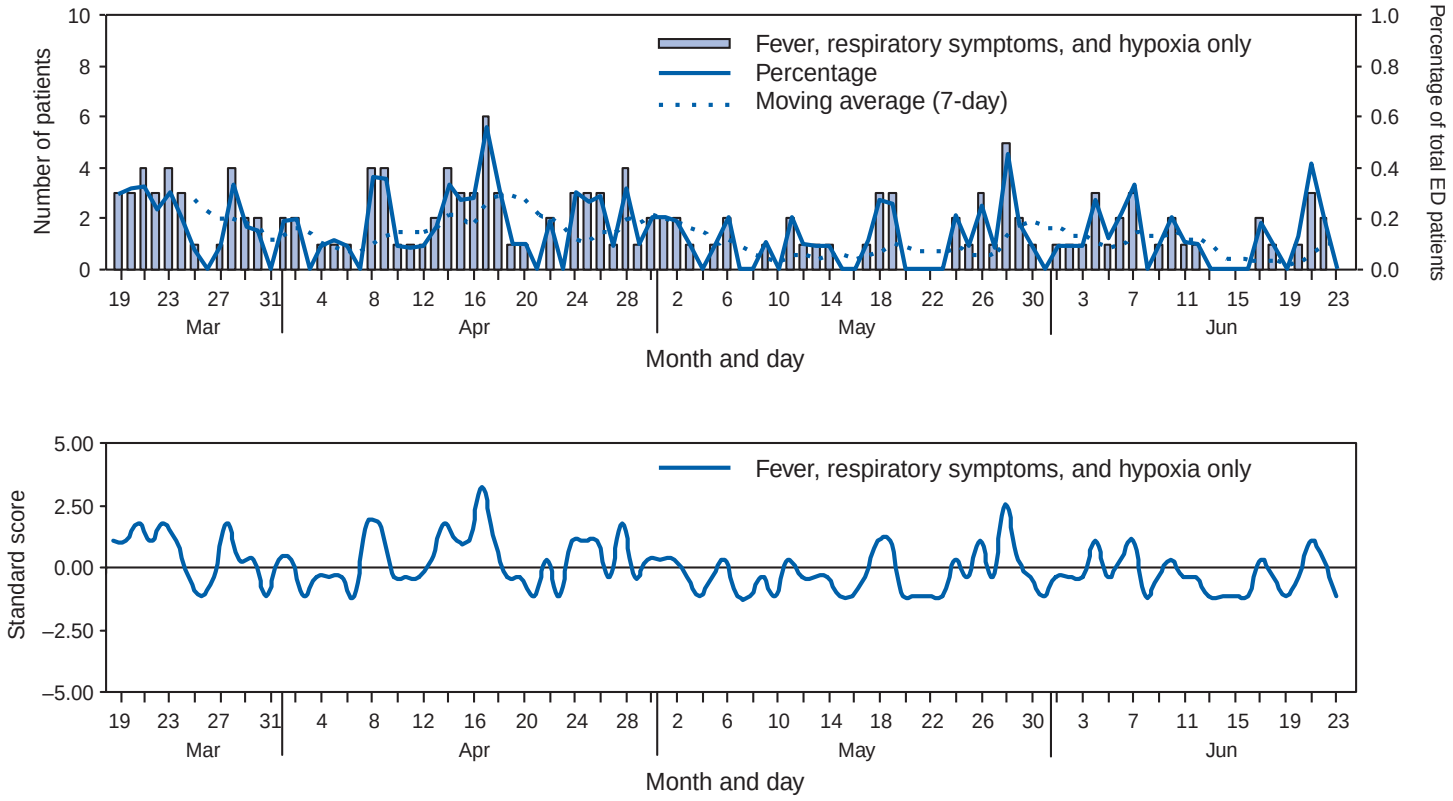
Resources did not permit on-site chart review to validate the accuracy of SARS element frequency reporting. Also, the standard ED record would not necessarily collect SARS risk history (travel or contact) and thus is not an ideal standard for comparison.

Each city performed its own analyses of syndromic time-series data. Cross-city analysis was not performed. In Milwaukee, staff graphed time series of SARS elements as crude counts, proportions of total ED census, and standard scores (i.e., the difference of daily counts from the cumulative mean, divided by the standard deviation) to display significant aberrations from the mean (Figure 3). The overall incidence rate of ED visits with each SARS element varied widely between cities, which is not surprising given the different geographic areas and date ranges of surveillance. Local surveillance-period incidence rates of ED patients reporting fever plus respiratory illness ranged from 0.33% in Akron to 1.4% in Denver. Two cities (Milwaukee and Fort Worth) investigated increasing syndrome trends; in both cases, telephone queries and record reviews by ED staff proved sufficient to exclude SARS as the cause.

During March 22–April 20, 2003, CDC easily downloaded daily Milwaukee data for EARS analysis, but these files did not include corrections made by local public health staff after telephone contact with EDs. Permitting online correction of data files on REMI would enable more accurate remote analysis.

Six of 13 participating Milwaukee ED managers returned nonanonymous surveys. Four of six believed SARS screening was performed as requested during all shifts. On a five-point scale ("strongly agree," "somewhat agree," "neutral," "somewhat disagree," and "strongly disagree,") five of six managers at least somewhat agreed they felt more secure knowing screening was being performed and also that screening increased the index of suspicion for SARS in their ED (one response to each item was neutral). Four at least somewhat agreed that data tabulation and data entry were easy (with one respon-

FIGURE 3. Daily emergency department (ED) visits by patients with three severe acute respiratory syndrome (SARS) elements (fever plus respiratory symptoms plus hypoxia) — Milwaukee, Wisconsin, 2003



Note: The upper graph displays crude counts, counts as a percentage of all ED visits, and counts displayed as a 7-day moving average. The lower graph displays standard scores of the crude counts (cumulative mean minus daily count divided by the standard deviation).

dent neutral and the other somewhat disagreeing to both items). The average estimate for the time to complete the form at triage was 2.6 minutes (range: 1–5 minutes; median: 3 minutes), and the average estimate for daily tabulation and reporting was 17 minutes (range: 5–45 minutes; median: 15 minutes). These compared favorably with estimated time spent on syndromic surveillance during the 2002 All-Star Game project, further validating that surveillance from the controlled confines of the triage desk was more manageable. Two managers had participated in syndromic surveillance during the previous summer; both strongly agreed that triage-based surveillance was superior, and both at least somewhat agreed that prior experience with REMI surveillance facilitated the rapid start-up of SARS surveillance.

The four public health surveillance coordinators all reported that they were glad they had participated and were interested in similar surveillance opportunities. Queried on ways to improve the system, two coordinators stated that they wished they had recruited additional EDs to participate, and two stated that they desired better communications between public health agencies and ED staff.

Discussion

SARS traveled extremely quickly, and new information about the disease evolved at a similar pace. SARS-SP, a rapidly organized, voluntary response, leveraged three capabilities to help clinicians and health officials keep pace: 1) interdisciplinary collaboration between emergency medicine, public health, and informatics; 2) an always-on, secure REMI network used in >24 metropolitan areas, and 3) rapid Internet information dissemination to clinicians. These were applied to two critical tasks: 1) helping ED staff detect possible SARS cases (case-finding) so they could protect patients, staff, and the community and 2) establishing syndromic surveillance to warn local health officials if illness consistent with SARS was increasing in their communities. The latter was deployed because CDC's surveillance focused on identifying known or suspected SARS risks but might not alert authorities to illness from unsuspected SARS contact (e.g., from asymptomatic transmission or unreported cases).

Ready-to-use screening forms helped busy ED staff to consistently meet complex, rapidly changing CDC guidance. ED triage (through which every patient passes early in an ED visit) was selected for case-finding and syndromic surveillance on the basis of ED workflow and previous experience. The 2002 All-Star Game surveillance project determined that relying on treating staff to record syndrome data produced poor-quality surveillance data and substantial staff-time demands (*4,5*). In contrast, triage nurses equipped with a well-crafted case-finding form could consistently "Screen—Isolate—Call Public Health." Although the sample size was limited, ED managers in Milwaukee reported higher satisfaction, greater confidence in data collection, and more reasonable time demands from triage-based surveillance than from the earlier 2002 All-Star Game surveillance program.

Paper-based forms have important limitations. Manual data check-off, tabulation, and entry each multiply the risk of data error and consume staff time. However, surveillance methods relying exclusively on mined data from existing registration, discharge, or other routine data sets would miss relevant information (e.g., recent travel), and they would not provide a real-time alert to ED personnel to implement infection control, diagnostic testing, and public health reporting. Therefore, data mining alone does not replace intelligent tools at the point of service for agile surveillance and response. Ideally, future triage information systems could be modified rapidly to collect and analyze newly important information (e.g., travel) alongside other routinely collected data (e.g., chief complaints) as part of routine workflow. The right combinations of data would automatically alert staff and public health authorities of a potential case while data for ongoing syndromic surveillance are collected with no additional human effort. Intelligent, programmable, and interoperable electronic medical record systems, linked through clinical networks such as REMI, could result in automated yet agile surveillance.

Milwaukee had used REMI previously to facilitate drop-in ED surveillance. Resulting experience and relationships helped MHD rapidly implement SARS surveillance. EDs in other cities appeared more prone to participate when they already used REMI in their day-to-day work (as was the case in 24 of the 27 participating EDs). Staff used the same application for surveillance that they used daily for other purposes, eliminating the need for new hardware and simplifying training. By contrast, public health agencies that were unfamiliar with the REMI application appeared more reluctant to participate.

Existing experience, servers, and 24-hour technical assistance capability that already supported the REMI system were leveraged to support rapid, multiregional surveillance. The project demonstrated that remote CDC specialists could use aberration analysis on remote REMI data. Ideally, such data should be quality-checked locally before analysis.

Rapid dissemination and updating of the screening form was enabled by ACEP's membership e-mail list and Internet tools. Because SARS-SP anticipated rapid evolution of case definitions, clinicians were encouraged to subscribe for

updates. However, not surprisingly, busy clinicians often failed to return for updated forms after downloading the original form. Ideally, REMI-networked clinical information systems would automatically incorporate updates and eliminate outdated tools from the point of service.

EDs in 12 urban areas expressed willingness to submit syndromic surveillance information to public health authorities, but only four health departments participated. The Council of State and Territorial Epidemiologists and the National Association of County and City Health Officials did not promote the project among their members because it lacked formal CDC endorsement. Such endorsement might be a precondition to participation, particularly in a fast-moving emergency with competing time demands.

Although this was a successful proof of concept of multiregional REMI-enabled surveillance, it had limitations. First, sensitivity and specificity of the triage screening and reporting cannot be calculated without SARS cases. Second, data were not validated by chart review. Third, ED records do not routinely record all information (e.g., travel) solicited. Finally, the system emphasized sensitivity over specificity.

With sufficient proportion of EDs involved, a sharp or sustained increase in community incidence of febrile and respiratory illness would likely be detected. Stamping and storing complete screening forms simplified rapid public health investigation. Because all four health departments reported being satisfied that they had participated in the surveillance project, it appears that a low positive predictive value for SARS was nevertheless practically manageable. Surveillance did not exhaust the patience of either EDs or public health agencies in springtime, but the outcome might have been different if the incidence rate of influenza and other common respiratory viruses were rising rather than falling.

Conclusion

SARS syndromic surveillance was rapidly established under emergency conditions by a loose network of collaborators using the tools available. It was handicapped by the lack of a legal or practical framework for sharing surveillance information across jurisdictions, and resources did not allow rigorous evaluation of the system's performance. Nevertheless, the ability to share surveillance tools across communities in a rapidly evolving outbreak illustrates how networked tools (e.g., REMI), which now reach >18% of the nation's EDs, have become practical instruments for agile surveillance across multiple regions. This is enhanced when clinicians and public health agencies are familiar with the applications from regular use. State and federal public health involvement might elicit participation by more agencies and could exploit untapped potential of these applications, such as integrating data across multiple regions and employing more sophisticated aberration algorithms.

Acknowledgments

Lori Hutwagner, National Center for Infectious Diseases, CDC, provided advice and EARS analysis; American College of Emergency Physicians publicized the screening tool and promoted participation in surveillance; Infinity HealthCare, Inc., developed and deployed new surveillance screens, accounts, and technical assistance without charge; and the staffs of participating hospitals and health departments established local procedures to transmit and review surveillance data. Milwaukee Health Department staff were supported in part by Lynde Uihlein through the Milwaukee Center for Emergency Public Health Preparedness, and by the Wisconsin Division of Public Health/CDC Cooperative Agreement #U90/CCU517002-03-02.

References

1. CDC. Outbreak of severe acute respiratory syndrome—worldwide, 2003. MMWR 2003;52:226–82.
2. Barthell EN, Cordell WH, Moorhead JC, et al. The Frontlines of Medicine Project: a proposal for the standardized communication of emergency department data for public health uses including syndromic surveillance for biological and chemical terrorism. Ann Emerg Med 2002;39:422–9.
3. Barthell EN, Foldy SL, Pemble KR, et al. Assuring community emergency capacity with collaborative Internet tools: the Milwaukee experience. J Public Health Manag Pract 2003;9:35–42.
4. Foldy S, Biedryzcki P, Barthell E, et al. The Milwaukee Biosurveillance Project: real time syndromic surveillance using secure regional Internet. [Abstract]. J Urban Health 2003;80(2 Suppl 1):i126.
5. Foldy S, Biedrzycki P, Barthell E, et al. Syndromic surveillance using Regional Emergency Medicine Internet. Ann Emerg Med 2004 (in press).
6. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80(2 Suppl 1):i89–96.

Health Information Privacy and Syndromic Surveillance Systems

Daniel Drociuk,[1] J. Gibson,[1] J. Hodge, Jr.,[2]
[1]South Carolina Department of Health and Environmental Control, Columbia, South Carolina;
[2]Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Corresponding author: Daniel Drociuk, 1751 Calhoun Street, Columbia, SC 29201. Telephone: 803-898-0289; Fax: 803-898-0897; E-mail: drociukd@dhec.sc.gov.

Abstract

The development of syndromic surveillance systems to detect potential terrorist-related outbreaks has the potential to be a useful public health surveillance activity. However, the perception of how the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule applies to the disclosure of certain public health information might affect the ability of state and local health departments to implement syndromic surveillance systems within their jurisdictions. To assess this effect, a multiple-question survey asked respondents to share their experiences regarding patient confidentiality and HIPAA Privacy Rule requirements when implementing syndromic surveillance systems. This assessment summarizes the results of a national survey of state terrorism-preparedness coordinators and state epidemiologists and reflects the authors' and others' experiences with implementation.

Introduction

State and local public health authorities use reports of diagnosed diseases or clinical syndromes to monitor disease or condition patterns (*1*). Syndromic surveillance refers to the systematic gathering and analysis of prediagnostic health data to rapidly detect clusters of symptoms and health complaints that might indicate an infectious-disease outbreak or other public health threat (*2*). Examples include electronic monitoring of routinely collected syndromic data (e.g., fever, gastrointestinal illness, or respiratory complaints in emergency departments) and time-sensitive data collection at regional hospitals before, during, or after major public events (e.g., Super Bowl, Salt Lake Winter Olympic Games, or political conventions).

Although certain syndromic surveillance activities use nonidentifiable data, the majority of systems require individually identifiable health data to permit rapid and efficient investigation of signals and follow-up with affected persons. As a result, syndromic surveillance systems often require medical providers or others to disclose identifiable health information to state, tribal, or local public health agencies; these data are typically shared with federal public health authorities in a nonidentifiable format. Multiple legal concerns arise from such data practices, including questions about systems' underlying legal authority and the relevance of health information privacy regulations pursuant to the HIPAA Privacy Rule or other health information privacy laws.

These legal concerns include the questions of 1) whether state statutory authorization for disease reporting applies to syndromic data; 2) the perceived effect of HIPAA's requirement that covered entities account for disclosures to public health agencies; 3) the effect of HIPAA requirements on investigating signals of possible outbreaks; and 4) the cost to reporting organizations or public health agencies of establishing a flow of syndromic data (*3*). In particular, public health professionals are concerned about whether potential reporting organizations might be incorrectly citing the HIPAA Privacy Rule to justify their refusal to disclose syndromic health data to public health agencies.

To examine these concerns, researchers e-mailed a survey to state epidemiologists and terrorism-preparedness coordinators, asking about the effect of privacy concerns on their ability to establish and conduct syndromic surveillance. Because little has been published on this topic, the survey instrument was designed to be exploratory and hypothesis-generating rather than hypothesis-testing. After asking initial questions to determine the status of a state's or city's syndromic surveillance system, the instrument used open-ended questions to capture anecdotal remarks and perceptions regarding barriers to the implementation of syndromic surveillance systems. The survey targeted the 50 states, four localities, and eight territories that are current recipients of CDC cooperative agreements on public health response and terrorism preparedness (*4*). Responses from county-level entities with syndromic surveillance systems were also accepted and incorporated into respective state-level responses.

Methods

Data Sources

A survey instrument was developed to assess the impact of the HIPAA Privacy Rule on syndromic surveillance within

states. After hypothesis-generating conversations with representatives from two states and one city (Connecticut, Kentucky, and New York City) that have implemented syndromic surveillance systems and with staff of the CDC Division of Public Health Surveillance and Informatics, a multiple-question survey was designed. With funding from CDC cooperative agreements for terrorism preparedness and response, multiple states have explored the implementation of syndromic surveillance systems under the surveillance and epidemiology capacity section (Focus Area B) of these cooperative agreements (*4*). All states, the District of Columbia, the three largest U.S. municipalities (Los Angeles County, New York City, and Chicago), and eight U.S. territories have received Focus Area B funding. Persons identified as Focus Area B leaders are tasked with operational oversight and implementation of critical capacities and benchmarks pertaining to surveillance and epidemiologic initiatives. The survey was distributed electronically on October 15, 2003, to each identified Focus Area B leader (N = 58) in each jurisdiction awarded resources under this cooperative agreement. The survey was also e-mailed to all 50 state epidemiologists to gather information from the four states without an identified Focus Area B leader, as well as to gain additional perspectives from others who might be involved in state-level syndromic surveillance activities. Responses were requested by October 21, 2003, to provide preliminary data for the National Syndromic Surveillance Conference at the New York Academy of Medicine on October 24, 2003.

Statistical Analysis

The survey design provided nine categorical questions with open-ended response options allowing for anecdotal remarks. Two investigators coded each response. If classification of a given response was questionable, the investigators evaluated the response separately and then compared results. In the event of a disparity, a third party evaluated the response and determined its final category. All analyses were performed by using data exported to S-PLUS® 2000 statistical software (*5*).

Because the sampling frame was primarily intended to be the recipients of CDC terrorism-preparedness cooperative agreements (50 states, four localities, and eight territories), the most relevant response rate seemed to be the percentage of grantees who had a syndromic surveillance system either under development or in operation. However, the denominator for that rate was unknown, because the total number of CDC terrorism-preparedness grantees that already had or were developing a syndromic system was not known. To estimate that denominator, knowledgeable consultants (two terrorism consultants from the Council of State and Territorial Epidemiologists and a CDC surveillance program staff member) were asked to help generate a list of known states with such systems. The resulting estimate of 40 states, localities, and territories with syndromic surveillance systems (either under development or in operation) provided the denominator used to determine the survey coverage rate.

Results

Of the 48 Focus Area B leaders who received the survey, as verified by documented successful transmission of the e-mail, 33 responses were received from 32 states, cities, and counties and one territory (Table). Of the 32 responses from states, cities, and counties, two states reported a state-level perspective along with a city-level perspective from a jurisdiction performing a pilot project. One state provided three separate county-level perspectives reflecting three distinct syndromic surveillance projects; these were combined with the state response to form a single response for each state when tabulating the coverage rate. Thus, total responses from the 62 states, localities, and territories with CDC terrorism-preparedness cooperative agreements were 29. Because not all states, localities, or territories have active syndromic surveillance systems, the consensus estimate of the total state, locality, and territory grantees with active syndromic systems was 40, which yields a coverage rate for this survey of 74.4%. Each respondent was given the option to report anonymously or be identified by jurisdiction. Of the 33 responses received, a majority of respondents (54.6%) requested anonymity in reporting.

To capture the nature of the responses to each question, examples are provided here. Responses to specific questions often addressed additional concerns to those raised by that question; to avoid subjectively imposing the investigators' views, such responses are reported here in conjunction with the question with which they initially appeared, even if the response seemed more relevant to another question.

When compared with those respondents who identified "no problems" in the implementation of a syndromic surveillance system, more than one half (54.2%) reported either "some" or "substantial" problems caused by real or perceived patient-confidentiality concerns and HIPAA Privacy Rule requirements. For example, one respondent stated, "Even our routine investigations encounter roadblocks. Many people in the trenches don't know enough about HIPAA and do not give information beyond the minimum necessary. This hampers all disease surveillance activities." Another reported, "Almost every hospital we approached raised issues of compliance with HIPAA. Discussions led to what is a minimum data set." Of

TABLE. Results of survey examining the effect of privacy regulations on jurisdictions' ability to establish and conduct syndromic surveillance — state, city, and territorial terrorism-preparedness coordinators and state epidemiologists, 2003

Question	No. of respondents	%
Status of a syndromic surveillance system?	33	
Active system	16	48.5
Currently implementing (i.e., recruiting reporting partners)	4	12.1
Considering or planning a system	10	30.3
Not considering implementing a system	3	9.1
Experience in dealing with patient confidentiality concerns and HIPAA requirements?	24	
Substantial problems	3	12.5
Some problems	10	41.7
No problems	11	45.8
Issues arising from the perception that syndromic reporting is not the same as disease reporting and therefore might not be mandated by state statute?	33	
Yes	17	51.5
No	15	45.5
Other	1	3.0
Has consideration been given to adding syndromic surveillance indicators to state reporting statutes or regulations?	33	
Yes	11	33.3
No	22	66.7
Regarding syndromic surveillance systems, have concerns originated from the HIPAA requirement to account for (i.e., track) disclosures to public health agencies?	31	
Yes	7	22.6
No	24	77.4
Issues arising from the need to investigate signals generated by the syndromic data flow?	30	
Yes	13	43.3
No	17	56.7
Issues around providers' fears that participating in syndromic surveillance would cause a negative public perception (i.e., marked as higher risk)?	32	
Yes	0	0
No	32	100.0
Issues regarding providers' costs incurred by providing syndromic data to public health?	32	
Yes	11	34.4
No	21	65.6
How are the costs of participation in syndromic surveillance addressed?	11*	
Provider responsibility (to use grant-funded sources)	6	54.5
Public health responsibility (to use grant-funded sources)	2	18.2
Joint responsibility	3	27.3
Concerns about adequate security of data transmission or storage at the health department?	32	
Yes	5	15.6
No	27	84.4

* Of those reporting active syndromic surveillance systems (n = 16).

those respondents who indicated "no problems" in the implementation, responses included the following: "Our syndromic surveillance system collects aggregate data, so this has not been as much of an issue for us as it has been for many other states. In fact, of the more than 600 sites participating in our syndromic surveillance system last year, less than two dozen commented or asked about confidentiality concerns and HIPAA requirements."

One survey question asked whether any concerns had arisen from the perception that syndromic reporting is not the same as disease reporting and might not be mandated by state statute. A similar percentage of respondents reported that such concerns were present (51.5%) or nonexistent (45.5%). Concerns included the following: "As we create more and more notifiable disease conditions, the medical community is increasingly resistant to accept without question." The question of whether syndromic surveillance was legally mandated yielded such responses as, "Reporting of diseases was both mandated and considered to be a more accurate surveillance tool, which caused many to feel it was unnecessary to ask people to also report syndromic information."

Respondents were then asked whether they had considered adding syndromic surveillance indicators to state reporting statutes. A majority (66.7%) indicated this was not being considered. Among the rationales given was that adding syndromic surveillance to a mandated reporting list would be problematic because generally accepted methods and content for syndromic surveillance systems have not yet been established

and because its effectiveness is still in question. Of respondents who reported that adding syndromic surveillance indicators was not being considered, 33% indicated that until clearer indicators are identified, the use of currently mandated reporting of clinical criteria (e.g., "clusters of extraordinary occurrence of illness" and "clusters of unusual illness") could apply to syndromic surveillance indicators. In addition, 54% of state-level respondents noted that, in the event of a recognized threat, their state health director is authorized to request that syndromic surveillance be conducted for a renewable period of time on the basis of an identified clinical presentation (e.g., severe acute respiratory syndrome [SARS]).

The survey also asked whether the HIPAA Privacy Rule requirement to account for disclosures to public health agencies was an obstacle to conducting syndromic surveillance. Of those responding, 22.6% reported concerns originating from this requirement. A majority of respondents indicated that their data were exchanged in limited data sets or aggregate form (e.g., emergency department [ED] visits or admission numbers), which do not require an accounting of disclosures. Twenty-three percent of respondents replied that the accounting requirement would be a concern if more detailed data were to be obtained. One respondent's jurisdiction had decided to collect only a limited data set, partly because of patient confidentiality concerns, and partly because the jurisdiction "rarely identified notifiable conditions via syndromic surveillance, and would end up having to call the hospital back for patient follow-up." Respondents expressed concerns about their syndromic surveillance system's ability to provide meaningful data when using only a limited data set. "Providers are cautious and it is not clear what we could provide to reassure them that a general accounting rather than transaction-specific accounting would work."

The burden on local health departments of investigating signals generated by syndromic surveillance systems was also explored. A majority of respondents (56.7%) indicated that, in cases where signals were identified, facility staff did not raise concerns about cost or feasibility. However, 43.3% of respondents reported concerns regarding the investigation of a signal. Responses included the following: "Since we are no longer collecting patient identifiers, the only way we can follow up with hospitals is by asking them to trace back the patient(s) who were seen at the date(s) and time(s) of interest. Perhaps because we rarely make such requests, we have not had problems with compliance."

The benefit of collecting individual patient records was evident from five respondents, who reported that the investigation of signals quickly showed that individually identifiable patient records are needed to trace information. As one respondent indicated, "We now have these and it is an immense improvement, saving considerable time for both investigators and providers."

The survey also asked whether providers feared participation in syndromic surveillance would harm their public perception or increase their vulnerability to a terrorist attack. All respondents indicated that providers had a positive outlook towards participating in syndromic surveillance. A total of 30% noted that they enhance community security by looking for unusual patterns that might not normally be observed. The majority of states reported that their hospital staff view participation in syndromic surveillance positively and as an indication of readiness. One state respondent indicated that staff at certain health-care facilities view syndromic surveillance as helping them meet internal requirements. One respondent reported, "Most of our hospitals want to participate and say it is helpful for them to see a summary of who has been to the ED the previous day."

Two survey questions examined concerns regarding the costs of syndromic surveillance. In the first question, a majority of respondents (65.6%) reported no issues associated with providers' costs of providing data. Thirty-seven percent of respondents identified initial problems regarding cost and have taken steps to reduce the burden on health-care facilities, including applying for federal grants for rural facilities and providing computers and Internet service by contracting with the facilities for data access and programmer time. Twenty-one percent reported using resources from the CDC terrorism-preparedness cooperative agreements and the Health Resource Services Administration (HRSA) Bioterrorism Hospital Preparedness Program for these activities. However, 14% expressed concern about setting a precedent of paying hospitals for their participation in the syndromic surveillance system. Certain respondents also acknowledged that not assisting hospitals with implementation might place an undue burden on facilities that do not have the information technology (IT) capacity to provide needed data. A total of 21% are establishing stipends to compensate hospital's IT departments for expenses incurred on behalf of their participation.

Of those states that reported an active syndromic surveillance system (n = 16), 54% require the hospital to cover the costs of participation, whereas 45% either pay for initial costs outright or share costs with health-care facilities. The majority of respondents recommended federal sources for covering the costs of program initiation.

The survey's final question addressed concerns about the security of data once they arrive at the health department. A substantial majority (84.4%) of respondents reported no security concerns from syndromic surveillance system partici-

pants, in part because of measures taken to ensure secure transmission (e.g., virtual private networks from a secure state file transfer protocol (FTP) site to the data sources) and off-system data-archiving protocols. For those collecting data without name-specific or identifiable data, security was of little concern to source participants.

Discussion

This was the first known survey on this topic targeted to state terrorism-preparedness managers and state epidemiologists. The survey attempted to assess the effect of the HIPAA Privacy Rule on the implementation of syndromic surveillance systems. Weaknesses of this survey include the limited response rate, the uncertain representativeness of the 33 responding jurisdictions (28 state or city and one territory cooperative-agreement recipients), and the limited time allotted for respondents to poll staff about problems with confidentiality or data access. Denominator data to account for all syndromic surveillance systems in the United States are difficult to quantify; therefore, the denominator estimate might not be accurate.

The study is based on qualitative judgments of senior managers responsible for syndromic surveillance systems, many of which were initiated by using CDC funding. However, substantial attempts were made to notify Focus Area B leaders and state epidemiologists for all states and terrorism-preparedness–funded localities to act as an information-gathering conduit for this survey. The state points-of-contact are likely to be closely involved with development of syndromic surveillance systems, on the basis of their involvement in implementing routine disease surveillance systems and the initiatives supported by CDC terrorism-preparedness grants. Accordingly, the authors believe that few managers of large syndromic surveillance systems were unaware of the survey. As the only known attempt to gather representative data on these issues, this study provided new information on matters of importance and identified areas for future research.

Ten percent of survey respondents also indicated that they request only limited data sets to more easily obtain permission and participation from covered entities. This can result in delays in investigating signals and, in certain cases, outright refusals of access to data on patient visits generating the signals. When signal investigations are substantially delayed, the syndromic surveillance system's value decreases. The added burden of retracing data to determine a signal's origin might hinder a timely response to an emerging situation.

One source of covered entities' reported reluctance to provide data appears to be the perceived requirement that they account for all disclosures for public health purposes under the Privacy Rule; this concern was reported by 22.6% of respondents. This problem persists despite favorable interpretations on the use of simplified "routine accounting" processes under the Rule, as discussed in CDC guidance and through the U.S. Department of Health and Human Services, Office of Civil Rights (OCR). Respondents from two states indicated that attorneys or risk managers for health departments and certain covered entities do not deem OCR authoritative enough to require a change in their data-release policies. In addition, perception about the scope of HIPAA was reported to be a substantial source of concern to state and local participants at the National Center for Vital and Health Statistics Privacy Subcommittee's hearings on the impact of the HIPAA Privacy Rule on public health (*6*). Incorrect interpretations of health departments' existing legal authority and the HIPAA Privacy Rule might cause substantial delays, extra work, and obstacles to obtaining necessary data for various surveillance systems, including syndromic surveillance.

OCR should disseminate clearer statements detailing how covered entities can use simplified accounting methods for routine disclosures to public health agencies and their contractual partners, pursuant to syndromic surveillance reporting requirements. The narrow perception of the accounting requirement and the view that it places an intolerable burden on covered entities might negatively impact the performance of syndromic surveillance systems within the United States, while accomplishing little to protect individual privacy where the existence of the disclosures underlying these public health practices are readily known.

References

1. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. J Urban Health 2003;80(2 Suppl 1):i1–7.
2. CDC. Syndromic surveillance: an applied approach to outbreak detection. Atlanta, GA: US Department of Health and Human Services, CDC, Division of Public Health Surveillance and Informatics. Available at http://www.cdc.gov/epo/dphsi/syndromic.htm.
3. Broome CV, Horton HH, Tress D, Lucido SJ, Koo D. Statutory basis for public health reporting beyond specific diseases. J Urban Health 2003;80(2 Suppl 1):i14–22.
4. CDC. Continuation guidelines for cooperative agreements on public health response and preparedness for bioterrorism. Atlanta, GA: US Department of Health and Human Services, CDC. Available at http://www.bt.cdc.gov/planning/continuationguidance/pdf/guidance_intro.pdf.
5. Mathsoft, Inc. S-PLUS® 2000 statistical software, professional release 1. [Software]. Seattle, WA: Mathsoft, Inc., 2000.
6. National Center for Vital and Health Statistics, Subcommittee on Privacy and Confidentiality. Hearings on the effect of the HIPAA Privacy Rule on public health practice, November 19, 2003, Silver Spring, MD.

Abstracts

Monitoring System for Detecting Starts and Declines of Influenza Epidemics

Eva Andersson
Goteborg University, Goteborg, Sweden

Corresponding author: Eva Andersson, Statistical Research Unit, Goteborg University, P.O. Box 660, SE 405 30, Goteborg, Sweden. Telephone: 0046 (0)31 773 12 64; Fax: 0046 (0)31 773 12 74; E-mail: eva.andersson@statistics.gu.se.

Abstract

Introduction: A nonparametric surveillance system was constructed for early detection of influenza outbreaks. The system uses weekly data on the number of influenza cases.

Objectives: For this analysis, a nonparametric method of surveillance was compared with the likelihood ratio method, which is optimal because it yields a minimal expected delay for a fixed false-alert probability. The evaluation was conducted by using probability of successful detection within a specified time and predictive value at different time points. The optimal surveillance method requires knowledge of the parametric model for the given process (i.e., the influenza cycle). Influenza cycles differ in shape and amplitude from one season to the next. Therefore, finding a parametric model based on influenza data from previous seasons is difficult. Also, using data from previous seasons might lead to misspecification of the cycles.

Methods: In the nonparametric method, the influenza cycles were estimated under monotonicity restrictions (i.e., monotonically increasing during the outbreak and monotonically decreasing during the outbreak's decline). The surveillance system was evaluated in a theoretical simulation study. The performance of the nonparametric method was compared with that of the optimal method. The effect of a misspecification of the parametric model was also studied.

Results: For most surveillance methods, the probability of successful detection of an influenza outbreak within 1 week depends on when the outbreak began relative to the start of the surveillance. The predictive value depends on when the alert is generated (Table).

Conclusions: The nonparametric method has lower detection probability then the optimal method when the outbreak begins immediately after surveillance is started. However, the nonparametric method avoids misspecifications. A parametric method with a misspecification results in poor detection probability for early outbreaks and low predictive value for late alerts.

TABLE. Probability of successful detection of an outbreak within 1 week (by start date of outbreak) and predictive value (by time of alert) for three surveillance methods

	Probability of successful detection		Predictive value	
	Start of oubreak (week)		Time of alert (week)	
Method	2	15	2	15
Optimal	0.64	0.64	0.69	0.76
Nonparametric	0.23	0.48	0.53	0.76
Parametric, misspecified	0.09	0.83	0.98	0.59

Evaluation of School Absenteeism Data for Early Outbreak Detection — New York City, 2001–2002

Melanie Besculides, R. Heffernan, F. Mostashari, D. Weiss
New York City Department of Health and Mental Hygiene, New York, New York

Corresponding author: Melanie Besculides, Mathematica Policy Research, 955 Massachusetts Avenue, Suite 801, Cambridge, MA 02139. Telephone: 617-491-7900, extension 273; Fax: 617-491-8044; E-mail: mbesculides@mathematica-mpr.com.

Abstract

Introduction: School absenteeism data might serve as an early indicator of disease outbreaks. However, before resources are committed to prospective surveillance, absenteeism data should be evaluated.

Objectives: This study evaluated the usefulness of school absenteeism data for early outbreak detection.

Methods: Data obtained from the New York City Department of Education on 1.2 million students (1,160 schools) for the 2001–02 academic year consisted of the number of students registered and absent by grade, school, and day. Reason for absence is not routinely collected. Citywide trends were examined separately for elementary and secondary students. Linear regression models predicted the expected percentage absent after controlling for day of week and pre- or post-holidays. Geographic clustering was assessed by the spatial scan statistic.

Results: Average daily absenteeism was higher among secondary students (13.7%) than elementary students (7.6%). No sustained increase in absenteeism was associated with the peak of the 2001–02 influenza A season (this period overlapped with winter break). A 2-week increase in absenteeism in March among elementary school children corresponded with peak influenza B season. Spatial analysis detected 790 clusters of absenteeism at $p<0.01$ (where only two clusters would have been expected by chance alone). Two of these clusters occurred during a previously reported gastrointestinal outbreak at one school.

Conclusions: A multiday, citywide increase in absenteeism among elementary students coincided with peak influenza B activity, but school absenteeism data were not useful for detecting the influenza A season. Although the system was able to detect one known localized gastrointestinal outbreak, this cluster did not stand out among other major clusters. Information on reason for absence and improved analytic methods might make absenteeism data more useful for early outbreak detection.

System To Generate Semisynthetic Data Sets of Outbreak Clusters for Evaluation of Outbreak-Detection Performance

Christopher A. Cassa,[1,2] K. Olson,[1,3] K. Mandl[1,3]
[1]*Children's Hospital Boston, Boston, Massachusetts;* [2]*Massachusetts Institute of Technology, Cambridge, Massachusetts;*
[3]*Harvard Medical School, Boston, Massachusetts*

Corresponding author: Christopher Cassa, Massachusetts Institute of Technology, 77 Massachusetts Ave., Rm. E25-519, Cambridge, MA 02139. Telephone: 617-355-2930; Fax: 617-730-0921; E-mail: cassa@mit.edu.

Abstract

Introduction: The outbreak detection performance of a syndromic surveillance system can be measured in terms of its ability to detect signal (disease outbreak) against background noise (normal variation of baseline disease within a region). However, because a limited number of persons have been infected with agents of biologic terrorism, such data are virtually nonexistent. Therefore, simulation is necessary. One approach to evaluation is to present detection algorithms with semisynthetic data sets. These data sets contain simulated signal superimposed on real background noise.

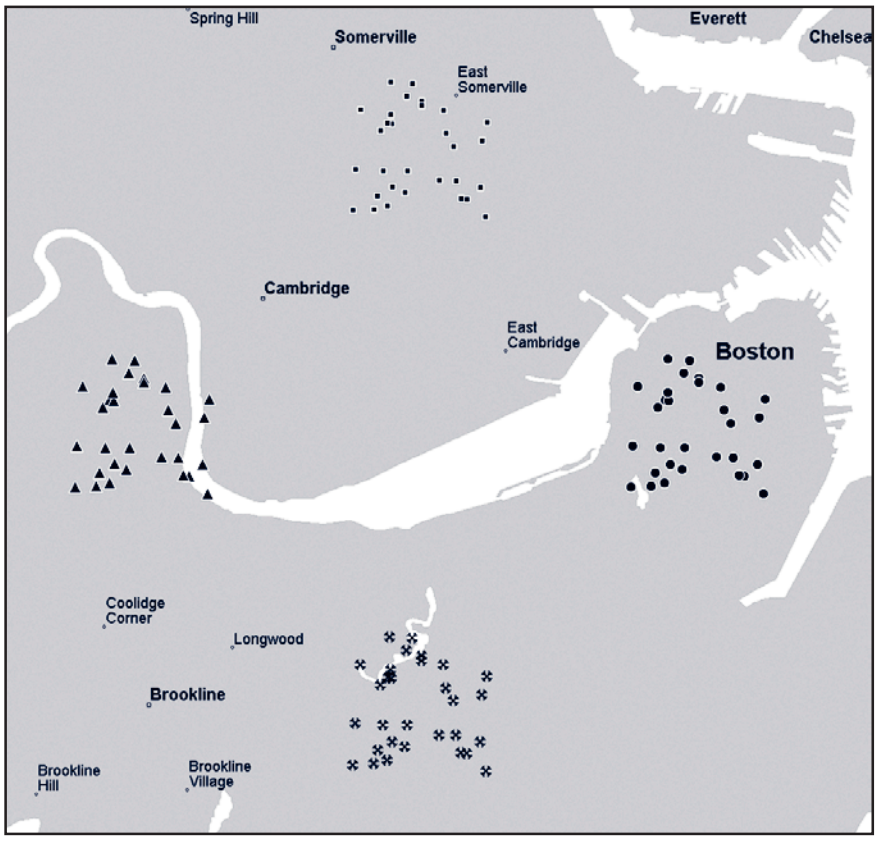
Objectives: The Children's Hospital Informatics Program (CHIP) Cluster Generator automates the creation of spatio-temporal patient cluster data to help evaluate epidemic-detection software. The spatio-temporal data can then be used to analyze the sensitivity and specificity of spatial or temporal detection algorithms.

Methods: A software tool (available at http://www.chip.org/biosurv/resources.htm) was created to generate artificial outbreaks of spatially clustered cases and inject them into background noise. Each cluster is defined by a controlled feature set. Parameters (e.g., outbreak magnitude, duration, temporal progression, and location) can be varied by the user.

Results: The open-source program accepts a valid set of patient test cluster parameters and creates geospatial patient test data for a single cluster or a series of clusters. The tool automates the creation of valid patient data sets for rigorous testing of outbreak-detection algorithms. The tool outputs either single-patient clusters or series of patient clusters as files containing patient longitude and latitude coordinates. When used with geographic information system software, these clusters can be displayed on a map (Figure). In testing, all generated clusters were properly created within the parameters set at program execution. The cluster generator is in use for rigorous testing of outbreak-detection algorithms.

Conclusions: Automated generation of semisynthetic data sets facilitates evaluation of public health surveillance systems for early detection of outbreaks.

FIGURE. Creation of a series of four system-generated outbreak clusters centered in Cambridge, Massachusetts (with the angle varied)



Establishing an Automated Surveillance System

Stanley S. Cho,[1] S. Hackmyer,[2] J. Li,[1] J. Lipsman,[1] O. Meruelo,[2] D. Mottola,[3] K. Mulvey,[2] M. Pereira,[2] R. Recchia[1]
[1]Westchester County Department of Health, New Rochelle, New York; [2]Westchester County Department of Information Technology, White Plains, New York; [3]Applied Systems Group, Inc., Tuxedo, New York

Corresponding author: Stanley S. Cho, 145 Huguenot St., 8th Floor, New Rochelle, NY 10801. Telephone: 914-813-5979; Fax: 914-813-5044; E-mail: ssc1@westchestergov.com.

Abstract

Introduction: In January 2003, Westchester County Department of Health (WCDOH) launched its Community Health Electronic Surveillance System (CHESS). CHESS receives daily data electronically from multiple hospital information systems, automatically analyzes data to detect elevated levels in each syndrome category, and generates electronic reports of results.

Objectives: This article describes the construction and implementation of an automated syndromic surveillance system in Westchester County.

Methods: WCDOH and multiple health-care providers reached agreement for daily acquisition, encryption, and transmission of data files. Providers were not required to use a standard file format. When files are not received by a specified time, the system automatically e-mails reminders to providers. Files of varying formats, based on scripts written individually for each provider, are automatically detected, decrypted, and loaded into the main database. CHESS was adapted from the syndromic surveillance methods developed by the New York City Department of Health and Mental Hygiene and CDC.

Results: CHESS analyzes data from a majority of the county's 12 emergency departments. Analysis and reporting are scheduled at given daily intervals and results are automatically e-mailed to WCDOH staff for appropriate action.

Conclusions: CHESS is advanced in the surveillance arena for its flexibility in accepting data from providers in varying file formats and its automation of internal processing and communication of results, allowing for ongoing system refinements. WCDOH has demonstrated the possibility of creating a local syndromic surveillance system that minimizes reporting burden on providers and maximizes use of internal resources and technical support.

Change-Point Detection Using Directional Derivatives

Allan B. Clark, A. Lawson
University of South Carolina School of Public Health, Columbia, SC

Corresponding author: Andrew B. Lawson, Department of Epidemiology and Biostatistics, Norman J. Arnold School of Public Health, 800 Sumter St., Columbia, SC 29208. Telephone: 803-777-6647; Fax: 803-777-2524; E-mail: alawson@gwm.sc.edu.

Abstract

Introduction: Individual-level disease maps, which estimate the risk for disease across a geographic region, usually are based on observing a set of spatial locations of cases and controls. This study examined the extension of where cases and controls form a space-time point process (locations and dates) and focused on assessing whether the intensity of cases changed across time.

Objectives: The objectives of the study were to develop a method for detecting changes in individual-level disease maps. The method was applied to a data set of birth abnormalities in the United Kingdom, which included locations and times of all the live (singleton) births and abnormalities during a 5-year period.

Methods: The change in intensity across time was measured through the directional derivative, with respect to time, of the geotemporal surface. This can be estimated nonparametrically and is used to check for both sudden and gradual changes over time.

Results: The results do not demonstrate the descriptive ability of an approach that relies on a map of changes in risk. The directional derivative was computed at 10 update points, corresponding to when data were available, and summary statistics were produced (Table). Isolated departures from constant risk were indicated, but the measure aggregated over a map did not demonstrate any change.

Conclusions: A directional derivative approach might yield optimal answers and is worthy of further research. The example data (Table) demonstrate that isolated changes are occurring, but data aggregated over a map did not indicate any change. Therefore, the geography of the problem should be considered, but an analysis that is aggregated over geography should not be performed.

TABLE. Summary statistics of directional derivative at each time point

	Update point									
Statistic	1	2	3	4	5	6	7	8	9	10
Minimum	–0.02	–0.08	–0.05	–0.01	–0.02	–0.02	–0.06	–0.05	0	–0.07
Mean	0	0.01	0	0	–0.01	0	0	0	0.01	–0.01
Median	0	0.01	0.01	0	–0.01	0	0	0	0.01	–0.01
Maximum	0.13	0.02	0.02	0.01	0.06	0.08	0.03	0.02	0.04	0.01

Physician's Choice of Charting Template Versus ICD-9 Code — Agreement Between Two Syndromic Surveillance Methods Using Emergency Department Electronic Medical Records

Dennis G. Cochrane,[1,2] J. Allegra,[1,2] J. Rothman[2]
[1]*Morristown Memorial Hospital Residency in Emergency Medicine, Morristown, New Jersey;*
[2]*Emergency Medical Associates of New Jersey Research Foundation, Livingston, New Jersey*

Corresponding author: Dennis G. Cochrane, Morristown Memorial Hospital Residence in Emergency Medicine, Morristown, NJ 07962. Telephone: 973-971-8919; E-mail: cochraned@verizon.net.

Abstract

Introduction: Although syndromic surveillance is often performed by tracking patterns of *International Classification of Diseases, Ninth Revision* (ICD-9) codes, ICD-9 codes are frequently not available in real time. In certain practice settings, the physician's choice of charting template (PCCT) is available in real time and therefore might have an advantage for use in syndromic surveillance.

Objectives: This study quantified the level of overlap among patients selected by PCCT and ICD-9 code.

Methods: A retrospective analysis was conducted of a database of patient visits in 15 New Jersey emergency departments during January 1999–October 2002. Two investigators reviewed all ICD-9 codes and PCCTs used during this period and chose by consensus those relevant to each of nine syndromes. For each syndrome, counts were generated of patient visits selected by ICD-9 code and by PCCT. The kappa statistic was then used to characterize the level of agreement between the two techniques. Sensitivity and specificity of the PCCT method were calculated by using the ICD-9 code as a criterion standard.

Results: The database contained 1,729,866 patient visits. Kappa calculations indicated near perfect agreement for *asthma* (0.82), *chest pain* (0.81), and *headache* (0.82) syndromes (Table). Excellent agreement was determined for *skin* (0.6), *any gastrointestinal* (0.74), and *diarrhea* (0.69) syndromes. Calculations indicated moderate agreement for *respiratory* (0.52) and *fever* (0.49) syndromes and only fair agreement for *weak* (0.34) syndrome.

Conclusions: Moderate to near perfect agreement between ICD-9 code and PCCT was determined for eight of the nine syndromes examined. PCCT might be useful for real-time syndromic surveillance using electronic medical records.

TABLE. Agreement (kappa), sensitivity, and specificity for physician's choice of charting template versus ICD-9 code in 15 emergency department databases, by syndrome — New Jersey, January 1999–October 2002

Syndrome	Kappa	Interpretation of kappa*	Sensitivity†	Specificity†
Headache	0.82	Near perfect	0.80	1.00
Asthma	0.82	Near perfect	0.81	1.00
Chest pain	0.81	Near perfect	0.83	0.99
Any gastrointestinal	0.74	Excellent	0.79	0.97
Diarrhea	0.69	Excellent	0.81	0.99
Skin	0.60	Excellent	0.60	0.99
Respiratory	0.52	Moderate	0.47	0.97
Fever	0.49	Moderate	0.44	0.98
Weak	0.34	Fair	0.40	0.98

* Based on a commonly used interpretation of kappa.
† Sensitivity and specificity of physician's choice of charting template using ICD-9 method as the criterion standard.

Monitoring Over-the-Counter Pharmacy Sales for Early Outbreak Detection — New York City, August 2001–September 2003

Debjani Das, F. Mostashari, D. Weiss, S. Balter, R. Heffernan
New York City Department of Health and Mental Hygiene, New York, New York

Corresponding author: Debjani Das, New York City Department of Health and Mental Hygiene, 125 Worth Street, Room 318, CN#22A, New York, NY 10013. E-mail: ddas@health.nyc.gov.

Abstract

Introduction: Because over-the-counter medications (OTCs) are commonly taken before patients seek medical care, OTC sales data might serve as an early indicator of communitywide illness. Since August 2002, the New York City Department of Health and Mental Hygiene (DOHMH) has tracked OTC sales from New York City pharmacies to enhance detection of natural and intentional infectious disease outbreaks.

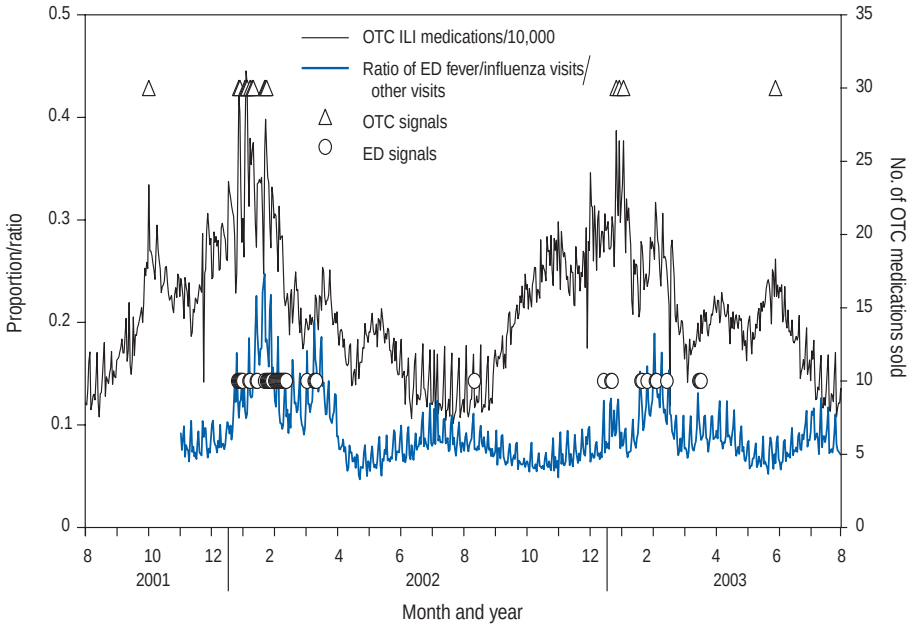
Objectives: First-year surveillance results on OTC sales were summarized and compared with results from an emergency department (ED) syndromic surveillance system.

Methods: A file containing the number of OTC units sold the previous day, by drug name and retail store, is transmitted to DOHMH daily from a central pharmacy database. The influenza-like illness (ILI) drug category includes cough and influenza medications whose sales correlate strongly with annual influenza epidemics. The antidiarrheal drug category includes generic and brand-name loperamide. Citywide trends are evaluated by using a linear regression model, controlling for seasonality, day of week, promotional sales, and temperature and are compared with ED data. Spatial clustering by store is evaluated by using the spatial scan statistic (SaTScan™ software, available at http://www.satscan.org).

Results: Citywide ILI drug sales were highest during annual influenza epidemics and elevated during the spring and fall allergy seasons, similar to trends in the ED system (Figure). Loperamide sales peaked during influenza season but did not increase substantially during the November 2002 viral gastroenteritis season. A spike in loperamide sales occurred after the August 2003 New York City blackout. Spatial signals for ILI sales occurred on 277 of 365 days.

Conclusions: The effect of allergies and asthma on respiratory illness should be considered when interpreting trends in OTC sales for ILI. Loperamide sales were not a useful indicator of a large 2002 norovirus outbreak detected by ED surveillance. Spatial cluster analysis was sensitive to variability in sales data by store and has not proven useful. OTC sales indicate correlation with other syndromic surveillance systems, but methods require refinement.

FIGURE. Sales of over-the-counter (OTC) influenza-like illness (ILI) medications per 10,000 population and ratio of emergency department (ED) fever/influenza visits to other visits, with their respective signals — New York City, August 1, 2001–August 1, 2003



Death Certificate Surveillance — New Hampshire

Kim K. Fallon, D. Boone
New Hampshire Department of Health and Human Services, Concord, New Hampshire

Corresponding author: Kim K. Fallon, New Hampshire Department of Health and Human Services, 29 Hazen Dr., Concord, NH 03301. Telephone: 603-271-4820; Fax: 603-271-0545; E-mail: KFallon@dhhs.state.nh.us.

Abstract

Introduction: New Hampshire is one of the only states in the United States that uses Vital Records Vision 2000, a system in which death certificates are filed electronically with the Division of Vital Records within 24 hours of being signed by a physician. The average time between date of death until certificates are filed with the state is 2.37 days. A surveillance coordinator reviews death certificates daily.

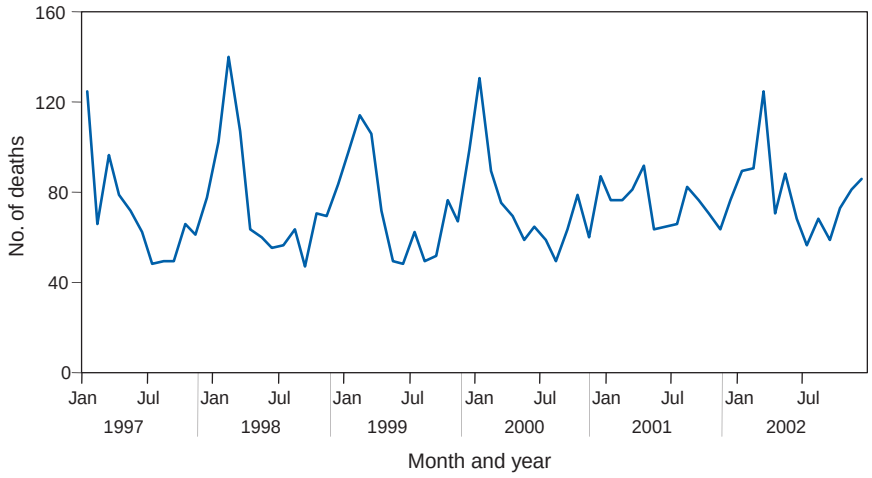
Objectives: This surveillance system is designed to detect clusters of deaths, deaths considered unusual, and deaths relevant to public health.

Methods: A query was developed that details >50 illnesses potentially related to terrorism. When an unusual death or cluster of deaths is found, the surveillance coordinator contacts the health-care provider to obtain more information. The state's communicable disease control unit investigates if warranted.

Results: Three unusual deaths were identified in 2003. None had been reported to public health authorities. Two previously healthy young persons were hospitalized with undiagnosed pulmonary infections, one for 7 days and the other for 11 days, before death. Specimens from both patients were retrieved and sent to CDC for further testing. In addition, infectious encephalitis was listed as the cause of death for an older patient suspected of having West Nile virus; specimens were obtained and sent to the New Hampshire Public Health Laboratory, where West Nile virus was ruled out. In addition, a review of death-record data for the period 1997–2002 demonstrates a consistent trend in pneumonia deaths over time (Figure).

Conclusions: Death certificate surveillance is able to 1) identify deaths that should have been, but were not, reported to public health agencies; 2) confirm the presence or absence of cluster deaths; 3) provide timely information on deaths statewide; and 4) provide information on seasonal trends in disease and death.

FIGURE. Pneumonia deaths — New Hampshire, 1997–2002



Correlation of West Nile Virus Infection with Emergency Department Chief Complaints by Using a Passive Syndromic Surveillance Model

John Flaherty, M. Gillam
Evanston Northwest Healthcare, Evanston, Illinois; Northwestern University, Feinberg School of Medicine, Chicago, Illinois

Corresponding author: John Flaherty, Evanstone Northwestern Healthcare, 2650 Ridge Rd., Evanston, IL 60201. Telephone: 847-570-2114; Fax: 847-634-1693; E-mail: jjflaherty@northwestern.edu.

Abstract

Introduction: West Nile virus infection appeared diffusely in Illinois in 2002, with >800 cases and 63 deaths. This number of confirmed cases was the highest in the nation and resulted in triple the number of deaths of any other state.

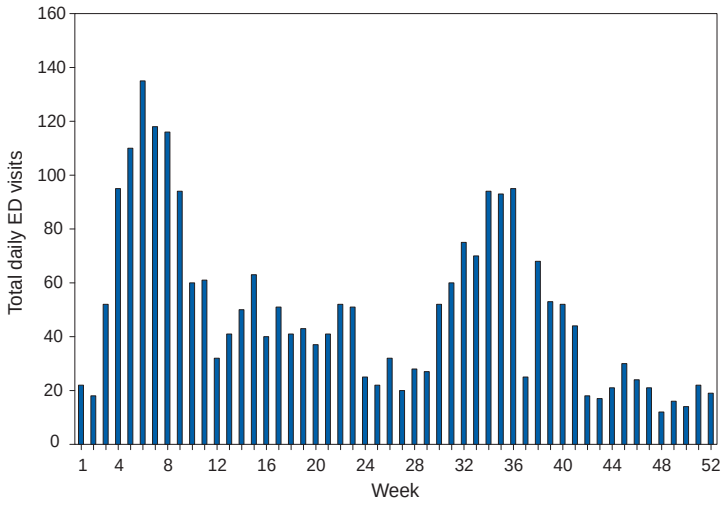
Objectives: This study used a passive syndromic surveillance model to analyze emergency department (ED) patient chief complaints of fever and headache, influenza-related symptoms, and viral syndrome, and correlate these data with known West Nile virus cases and with the epidemic curve of confirmed cases in northern Illinois.

Methods: A passive syndromic surveillance system using a computerized patient log was implemented. A retrospective cohort study used structured query language (SQL) queries to search for patient chief complaints of fever and headache, influenza-related symptoms, or viral syndrome. Positive matches were compiled in a graphical and geographic database.

Results: SQL queries revealed a biphasic distribution, with a first peak corresponding to influenza cases during the second week of February and a second unexpected peak during the second week of September 2002 (Figure). Geocoding and frequency analysis matched the confirmed outbreak. A majority of these patients were discharged, and no deaths occurred. IgM serology was positive in 5% of cases. Statistical analysis determined no significant differences in distribution and a coefficient of determination of 0.67.

Conclusion: Passive syndromic surveillance systems can retrospectively detect West Nile virus infection. The system was able to detect an increase in syndromic cases in the ED during a confirmed outbreak of West Nile virus. Further study is needed to quantify this effect. Serologic confirmation will also aid in validation.

FIGURE. Emergency department (ED) visits for viral syndrome, by week — one health-care system, Evanston, Illinois, 2002



Note: The first peak (weeks 4–10) in ED visits for viral syndrome was attributable to an anticipated increase in influenza cases. The second peak (weeks 30–42) was attributable to West Nile virus.

Evaluation and Validity of Chief Complaints and Discharge Diagnoses in a Drop-In Syndromic Surveillance System

Aaron T. Fleischauer,[1] B. Silk,[2] M. Schumacher,[3] K. Komatsu,[4] S. Santana,[3] V. Vaz,[4] M. Wolfe,[5] L. Hutwagner,[1] J. Cono,[1] R. Berkelman,[2] T. Treadwell[1]

[1]*Bioterrorism Preparedness and Response Program, National Center for Infectious Diseases, CDC;* [2]*Emory University, Atlanta, Georgia;* [3]*Maricopa County Department of Public Health, Phoenix, Arizona;* [4]*Arizona Department of Health Services, Phoenix, Arizona;* [5]*Division of HIV/AIDS Prevention, National Center for HIV, STD, and TB Prevention, CDC*

Corresponding author: Aaron T. Fleischauer, Bioterrorism Preparedness and Response Program, National Center for Infectious Diseases, CDC, 1600 Clifton Rd, MS C-18, Atlanta, GA 30308. Telephone: 404-634-3995; Fax: 404-639-0382; E-mail: alf6@cdc.gov.

Abstract

Introduction: Syndromic surveillance systems are being explored to determine their capacity to detect outbreaks, including those caused by biologic or chemical terrorism. However, few systems have been validated.

Objectives: This study evaluated a syndromic surveillance system by comparing syndrome categorization in the emergency department (ED) with medical chart review.

Methods: During October 27–November 18, 2001, a surveillance form was completed for each ED visit at 15 participating Arizona hospitals. One of 10 clinical syndromes or "none" was selected per patient to best represent the patient's primary condition. Medical records were reviewed for a weighted, random sample of 16,886 available forms. ED chief complaints and discharge diagnoses were abstracted as standards to compare with surveillance forms. Clinicians assessed concordance between the selected syndromes and standards.

Results: Of 1,956 patient records from six selected hospitals, 1,646 (85%) indicated either one syndrome or none, and 313 (15%) were blank. Overall, system concordance was 71% and 85% when using chief complaint and ED discharge diagnosis, respectively. Discharge diagnosis outperformed chief complaint in the overall system (+14%) and within syndromes (range: 0%–65%). Concordance of *respiratory tract infection with fever* for chief complaint was low (27%) compared with its concordance with ED discharge diagnosis (83%). Similarly, concordance of chief complaint was low for *sepsis* (6%), *rash with fever* (24%), and *myalgia with fever* (40%).

Conclusions: This ED-based syndromic surveillance system was able to classify patients into an appropriate syndrome category rapidly and with accuracy. However, syndromic surveillance systems might perform better when based on ED discharge diagnosis in addition to or instead of chief complaint.

Using the Toxic Exposure Surveillance System To Detect Potential Chemical Terrorism Events

Amy B. Funk,[1] J. Schier[1], M. Belson,[1] M. Patel,[1] C. Rubin,[1] W. Watson,[2] T. Litovitz,[2] E. Kilbourne[3]
[1]Division of Environmental Hazards and Health Effects, National Center for Environmental Health, CDC; [2]American Association of Poison Control Centers, Washington, D.C.; [3]National Center for Environmental Health/Agency for Toxic Substances and Disease Registry, CDC

Corresponding author: Amy B. Funk, CDC, 4770 Buford Hwy., MS F-46, Atlanta, GA 30341-3717. Telephone: 770-488-3402; Fax: 770-488-3450; E-mail: ajf9@cdc.gov.

Abstract

Introduction: CDC and the American Association of Poison Control Centers are using the Toxic Exposure Surveillance System (TESS) to improve public health surveillance of health hazards associated with chemical exposures. TESS is a national real-time surveillance database that records all human exposures to potentially toxic substances reported to U.S. poison control centers.

Objectives: TESS is used to facilitate early detection of illness associated with a chemical release by monitoring daily clinical effects reported to the database.

Methods: Computer-generated surveillance is conducted daily on each clinical effect (n = 131). The frequency of each clinical effect during a 24-hour interval is compared with a historic baseline. The historic baseline is defined as the mean frequency for each clinical effect during the 2-week period surrounding the 247-hour interval, during the preceding 3 years. An aberration is identified when the observed number of cases with a given clinical effect exceeds the expected limit (historic baseline plus 2 standard deviations). Cases identified through this system are evaluated, and respective poison control centers are contacted when unusual patterns in location, substance, or outcome are noted.

Results: Aberrations have identified clusters of clinical effects occurring within a 24-hour period. Further investigation has identified clusters with a single etiology (e.g., 16 cases of severe gastrointestinal illness from intentional tampering of coffee with arsenic at a church picnic).

Conclusions: Detection of these aberrations indicates that conducting surveillance by using TESS can identify illnesses resulting from intentional or unintentional chemical releases that occur at a single site or, potentially, across multiple locations.

Making Syndromes Reportable Diseases — Authorizing, Mandating, or Both? A Perspective on the Legal Basis for Syndromic Surveillance

Per H. Gesteland,[1,2] R. Rolfs[2,3,4]
[1]*Intermountain Health Care Institute for Health Care Delivery Research, Salt Lake City, Utah;* [2]*University of Utah, Salt Lake City, Utah;* [3]*Utah Department of Health, Salt Lake City, Utah;* [4]*National Center for Chronic Disease Prevention and Health Promotion, CDC*

Corresponding author: Per H. Gesteland, Division of Inpatient Medicine, Primary Children's Medical Center Department of Pediatrics, University of Utah School of Medicine, 100 North Medical Drive, Salt Lake City, Utah 84113. Telephone: 801-588-3813; Fax: 801-588-3814; E-mail: per.gesteland@hsc.utah.edu.

Abstract

In preparation for the Salt Lake 2002 Olympic Winter Games, Utah established legal authority for syndromic surveillance by enacting an administrative rule based on current communicable disease reporting authority. That rule required designated emergency centers to report data on patients seen the previous day for whom diagnostic information indicated the presence of ≥1 of 11 tracked syndromes. Data could be reported by emergency centers or collected by public health personnel.

Concurrently, the Detection of Public Health Emergencies Act was passed during Utah's 2002 legislative session. That Act gave the Utah Department of Health (UDOH) authority to designate diseases, conditions, or syndromes as "reportable emergency illness and health condition(s)" under subsequent administrative rule. UDOH is working to enact administrative rules that specify details of syndromic reporting based on that authority.

The Act authorizes voluntary reporting under normal circumstances and mandatory reporting upon declaration of a public health emergency. That approach was chosen to avoid imposing an unacceptable burden on facilities that lack technical infrastructure to report electronically. However, voluntary reporting poses the risk that providers will not participate for fear of being exposed to legal and public relations problems. Furthermore, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule has led certain providers to require a specific legal mandate to report.

Other challenges under this approach are that current Utah law does not authorize collection of protected health information for patients not determined to have one of the defined syndromes, data that are needed to permit normalization for statistical analysis. Another concern is whether records should be processed to identify syndromes at the health-care facility, necessitating greater technical investment at each facility, or at the public health entity, requiring at least temporary disclosure to the public health entity of records not meeting the syndrome-reporting criteria.

New Twist on Old Methods — Simple Schemas for Disease and Nonbattle Injury Surveillance in Deployment Settings

J. Kevin Grayson, P. Gould
Air Force Institute for Operational Health, San Antonio, Texas

Corresponding author: Kevin Grayson, 60 MDSS/SGSE, 101 Bodin Circle, Travis AFB, CA 94535. E-mail: kevin.grayson@travis.af.mil.

Abstract

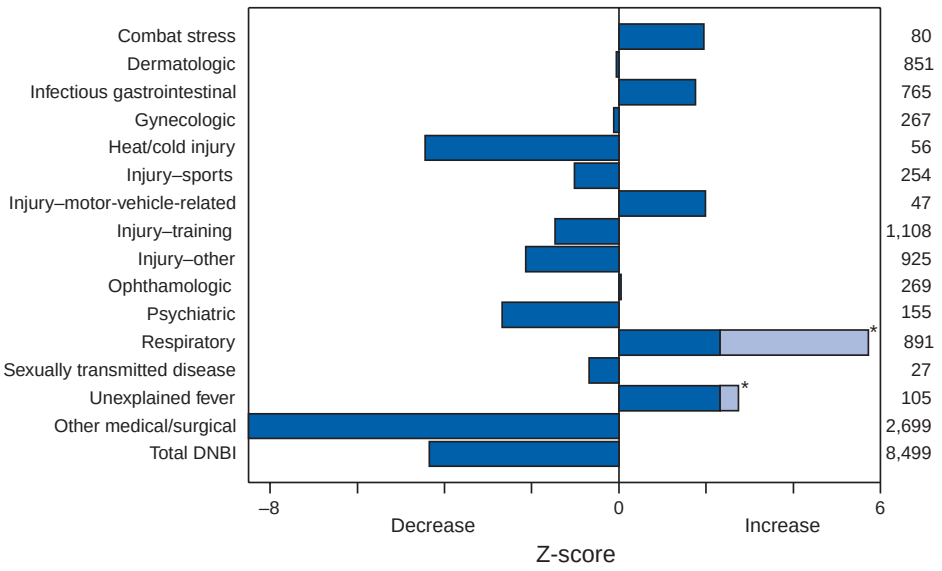
Introduction: Medics deployed with U.S. troops routinely collect disease and nonbattle injury (DNBI) data to provide early outbreak detection and identify adverse trends. Robust statistics are available to analyze surveillance data; however, these methods often require extensive historic data, are computationally difficult, and can be confusing to inexperienced users.

Methods: A modified current-past experience graph (CPEG) and statistical process control charts (SPCCs) were developed to track DNBI trends among deployed service members. The CPEG method compares weekly counts for 16 DNBI categories with expected values from the previous 4 weeks by using the Poisson function normal approximation. These are transformed to z-scores and charted with color codes to indicate when a value exceeds threshold limits, corresponding to the 99th percentile. The u-bar method (i.e., a statistical process control method that also relies on Poisson approximation) is used to produce SPCC, comparing observed rates with the average from the previous 20 weeks for each DNBI category. Although the necessary calculations could be performed by hand, spreadsheet templates were produced for field use. Stata® statistical software is used routinely to automate the process and provide graphs to customers over the Internet.

Results: These charts have been used to monitor DNBI reports from Operation Enduring Freedom and Operation Iraqi Freedom since their inception. A typical CPEG chart shows significant increases in the respiratory and unexplained fever categories (Figure). The CPEG and SPCC methods are complementary. CPEG summarizes all data on a single chart and is highly sensitive. SPCC provides more detail and underscores long-term trends.

Conclusions: Customers find CPEG and SPCC useful because they summarize a substantial amount of information and are readily understood by nonmedical commanders.

FIGURE. Observed versus expected case counts of 16 categories of disease and nonbattle injury (DNBI) among deployed U.S. military service members, as depicted by a Current-Past Experience Graph — Southwest Asia, week of September 20, 2003



* Statistically significant excesses.

Dual-Model Approach to Syndromic Surveillance Using Hospital Emergency Department Data

Karen Green,[1] B. Miller,[1] M. Hadidi,[1] M. Zimmerman,[2] R. Danila[1]
[1]*Minnesota Department of Health, Minneapolis, Minnesota;* [2]*Hennepin County Medical Center, Minneapolis, Minnesota*

Corresponding author: Karen Green, 717 Delaware Street SE, Minneapolis, MN 55414. Telephone: 612-676-5414; Fax: 612-676-5743; E-mail: karen.1.green@health.state.mn.us.

Abstract

Introduction: The initial symptoms of diseases resulting from biologic terrorism are likely to appear as respiratory illness (RI) or gastrointestinal illness (GI). Increased counts of RI- or GI-related *International Classification of Diseases, Ninth Revision (*ICD-9) codes in a syndromic surveillance system might indicate an outbreak. Only a limited number of syndromic surveillance systems analyze data for temporal and geographic disease clustering simultaneously.

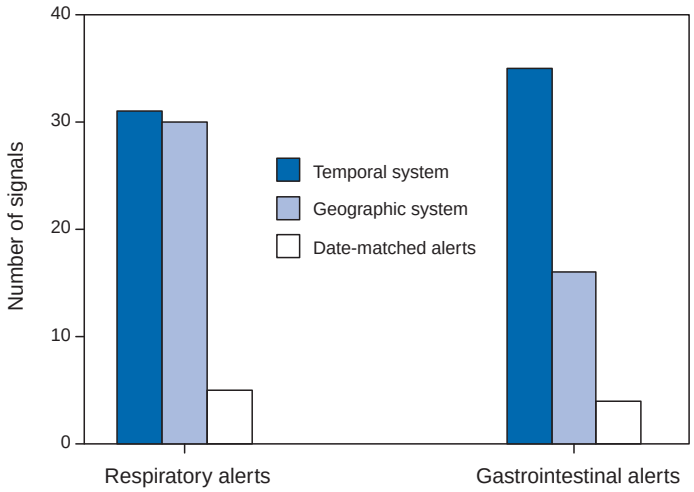
Objectives: A comprehensive syndromic surveillance system was created through analysis of emergency department (ED) data by using both time-series (TSS) and geographically based syndromic surveillance (GSS) models.

Methods: Minnesota Department of Health receives patient-encounter data from the Hennepin County Medical Center (HCMC) ED via secure file transfer protocol (FTP) file daily. TSS uses a regression model adjusted for day-of-week and seasonal effects. Autocorrelation and cumulative sum analysis of predictive residuals detects unexpected increases of ICD-9 counts. The GSS model is an adapted mixed models approach. Daily counts are compared with historic data by using the binomial probability mass function. Analyses were performed for HCMC ED patients reporting during January 2001–August 2003 (32 months). The ED treats approximately 100,000 patients annually.

Results: RI counts exceeded threshold 31 times under TSS and 30 times under GSS, matching on five dates (9%). GI counts exceeded threshold 35 times under TSS and 16 times under GSS, matching on four dates (9%) (Figure).

Conclusions: Unmatched dates resulted from the differing statistical approaches of each model. Signals detected under TSS indicate temporal clustering; signals detected under GSS indicate spatial clustering. These combined analyses allow observation of disease patterns by examining concurrent temporal and geographic effects. A signal detected by using TSS or GSS can initiate further examination of encounter data, including chart reviews by medical facility staff. Concurrent signals detected from both TSS and GSS warrant rapid follow-up.

FIGURE. Unique signals detected under temporal and geographic syndromic surveillance models — Hennepin County Medical Center Emergency Department, Minneapolis, Minnesota, January 1, 2001–August 31, 2003



Technique for Rapid Detection and Localization of Attacks with Biologic Agents*

Ronald Hoffeld, L. Candell
Massachusetts Institute of Technology Lincoln Laboratory, Lexington, Massachusetts

Corresponding author: Ronald Hoffeld, MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02420. Telephone: 781-981-2785; Fax: 781-981-6873; E-mail: hoffeld@ll.mit.edu.

Abstract

Introduction: A technique is presented for simultaneously detecting, localizing, and estimating time of attacks with biologic agents or other infectious sources that have a distinct spatial-temporal point pattern. The proposed technique uses high-quality individual location histories coupled with self-reported health status to search for areas where a high density of currently ill persons had congregated in the past. The increased infection rate associated with this detection is indicative of a possible infectious outbreak.

Objective: The system, named BACTrack (Biological Attack Correlation Tracker), was assessed through simulation and analysis to determine achievable sensitivity relative to attack size, infection rate, and participating population.

Method: A sample cohort of the general population was simulated to continuously record their location histories and to report the onset of illness. Developments in location-based cellular phone services enable simplified automation of these functions. Detection was performed by dividing the surveillance area into space-time regions, determining the ratio of ill persons to total population within each region, and flagging regions that exceeded an adaptive threshold on the basis of the statistical variation of the background illness.

Results: A *Bacillus anthracis* attack affecting 1,100 persons in a city of 150,000 population was simulated. Detection, location, and time of attack were determined with 90% probability 1.5 days after appearance of initial symptoms. The simulation was conducted with health status data with that reflected only whether the person was healthy or ill.

Conclusion: The results demonstrate an ability to operate on poor-quality symptom information. Early detection is made possible by using early diffuse symptoms, efficient data collection, and the signal-processing gain that results from performing location correlation at the time of the attack.

* This work was sponsored under Air Force Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Air Force.

From Data Sources to Event Detection — Summary of the Southern California Regional Surveillance Summit

Jeffrey M. Johnson, L. Gresham, D. Browner, C. McClean, M. Ginsberg, S. Wood
County of San Diego Health and Human Services Agency, San Diego, California

Corresponding author: Jeffrey M. Johnson, 1700 Pacific Highway, MS P511C-A, San Diego, CA 92101; Telephone: 619-531-4945; Email: jeffrey.johnson@sdcounty.ca.gov.

Abstract

Introduction: Approximately 20 million persons reside within southern California. Each county and city health department in southern California has approached syndromic surveillance somewhat independently and is at different stages of development.

Objectives: The Southern California Regional Surveillance Summit was held June 16, 2003, to enable professionals to share capacities and best practices and to assess the potential for regional collaboration.

Methods: Using terrorism-preparedness funds, San Diego County sponsored a summit of county and city health department representatives. Selected counties presented their syndromic surveillance efforts. Roundtable discussions were held regarding data sources, aberration-detection algorithms, model syndromic surveillance systems and information technology interfaces, and evaluation of signals and alerts. Roundtable discussions were summarized and next steps explored. A compendium was developed for all participants.

Results: With representation from 12 California counties, the state of California, U.S./Mexico Border Health, and the U.S. Navy, all participants described a level of syndromic surveillance. Potential data sources were prioritized, meaningful methods identified, and the potential for regional collaboration outlined. Across southern California, syndromic surveillance capacity varied substantially, with certain regions using real-time data-capture systems and state-of-the-art aberration-detection methods, whereas others face shortages in staffing, insufficient access to data sources, or lack of formal evaluation techniques.

Conclusions: The summit enabled professionals from county and city health departments to exchange information, highlight lessons learned, and explore the potential for future collaboration. It was an important step toward a multijurisdictional effort of using surveillance for early disease detection.

Syndromic Surveillance of Infectious Diseases in Taiwan — Before and After the Challenges of Severe Acute Respiratory Syndrome (SARS)

Chwan-Chuen King,[1] F-Y. Shih,[1,8] M-Y. Yen,[2] F-C. Hu,[1] J-S. Wu,[1,3] F-K. Chang,[3] L-W. Lin,[3] J-Y. Yang, H-Y. Chen,[3] T-S. Wu,[1] D-J. Wang,[3] K-T. Chen,[3] H-T. Yu,[4] C-A. Hsiung,[4] S-W. Lu,[3] C-M. Chang,[3] S-T. Lin,[2] C-J. Fu,[6] C-M. Huang,[2] M-S. Ho,[5] H. Chang,[7] J-H. Chou,[3] S-J. Twu,[4] I-J. Su,[3] M. Marx,[9] H. Sobel[10]

[1]National Taiwan University, Taipei, Taiwan; [2]Taipei Municipal Jen-Ai Hospital, Taipei, Taiwan; [3]Center for Disease Control, Taipei, Taiwan; [4]National Health Research Institute, Taipei, Taiwan; [5]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; [6]Taipei Municipal WanFang Hospital, Taipei, Taiwan; [7]Taipei City Government Department of Health, Taipei, Taiwan; [8]National Taiwan University Hospital, Taipei, Taiwan; [9]New York City Department of Health and Mental Hygiene, New York, New York; [10]World Health Organization, Geneva, Switzerland

Corresponding author: Chwan-Chuen King, Institute of Epidemiology, College of Public Health, National Taiwan University, 1 Jen-Ai Road Section 1, Taipei, Taiwan (100). E-mail: a1234567@ccms.ntu.edu.tw.

Abstract

Introduction: Timely and sensitive outbreak detection is a high priority of syndromic surveillance. Early detection enables officials to allocate limited public health resources to contain outbreaks and thereby decrease morbidity and mortality.

Objectives: This study retrospectively evaluated Taiwan's respiratory syndromic surveillance system (RSSS), established in July 2000, for its ability to detect severe acute respiratory syndrome (SARS).

Methods: Reporting through RSSS was encouraged for patients aged >5 years with unexplained cough, respiratory distress, pulmonary edema, or other severe symptoms. Their specimens were collected for laboratory testing of suspected etiologic agents.

Results: Among 112 reported acute respiratory syndrome cases during January 1–August 5, 2003, etiologic agents were identified for 26 cases, and only four SARS cases and one case co-infected with SARS-associated coronavirus (SARS Co-V) and *Mycoplasma* were detected. Only five (0.75%) of 664 probable SARS cases were captured through RSSS. The first SARS case, reported on March 14, 2003, was not detected by RSSS, reflecting the system's low sensitivity. RSSS did not detect a SARS case until March 17, 2003, after awareness had been raised by media reports.

Conclusions: Because RSSS was both insensitive and rarely used before the SARS outbreak, and because public health administrators urgently needed daily updated case numbers and laboratory results, Taiwan instituted an Internet-based reporting and a day-to-day medical follow-up form immediately after the peak of hospital-associated SARS. Emergency department-based syndromic surveillance was established in July 2003, and different hospital data sets are being integrated into the system to facilitate detection of future outbreaks of emerging infectious diseases.

Syndromic Tracking and Reporting System — Overview and Example

Jylmarie Kintz,[1] Eliot Gregos,[1] David Atrubin,[1,2] Jeff Sanchez[1]
[1]*Hillsborough County Health Department, Tampa, Florida;* [2]*Florida Department of Health Epidemic Intelligence Service, Tallahassee, Florida*

Corresponding author: Jylmarie Kintz, Hillsborough County Health Department, 1105 East Kennedy Boulevard, P.O. Box 5135, Tampa, FL 33675-5135. Telephone: 813-307-8010; Fax: 813-276-2981; E-mail: jylmarie_kintz@doh.state.fl.us.

Abstract

Introduction: In cooperation with CDC and the Florida Department of Health's Bureau of Epidemiology, the Hillsborough County Health Department (HCHD) first participated in syndromic surveillance during the 2001 Super Bowl. Ongoing syndromic surveillance was implemented in November 2001. Nine hospital emergency departments (EDs) in the county report syndromes daily.

Objectives: The Syndromic Tracking and Reporting System (STARS) augments HCHD's traditional disease reporting by acquiring near real-time syndromic data from hospital EDs. STARS is designed to detect terrorism-related and naturally occurring outbreaks in which affected persons seek ED care.

Methods: Seven different syndromes are monitored by ED physicians. ED staff then enter limited patient information and the appropriate syndrome into an Internet-based system. The data are housed at HCHD and analyzed by using CDC's Early Aberration Reporting System (EARS) software, which detects statistical aberrations. A decision matrix is used to decide which aberrations require follow-up by HCHD's epidemiology staff.

Results: Statistical aberrations have been investigated periodically. On March 24, 2003, STARS detected 20 reported cases of diarrhea/gastroenteritis syndrome from one hospital. This spike in the data was flagged by EARS statistical aberration software. Follow-up investigation revealed that 14 of 20 affected persons had chronic conditions that were not of infectious disease concern, and no outbreak was determined to have occurred.

Conclusions: The system worked as intended. Studies are under way to evaluate data quality and assess the validity and sensitivity of STARS.

Addressing the Concerns of Data Providers — Lessons from the Private Sector

Andrew Kress, G. Hess
Surveillance Data, Inc. Plymouth Meeting, Pennsylvania

Corresponding author: Andrew Kress, Surveillance Data, Inc., 220 W. Germantown Pike, Suite 140, Plymouth Meeting, PA 19462. Telephone: 610-834-0800; Fax: 610-834-8817; E-mail: akress@survdata.com.

Abstract

Introduction: Emphasis on development of syndromic surveillance programs by public health administrators has resulted in proliferation of public-private partnerships for data provision. However, serious concerns arise from these partnerships that require consideration of the motivations and concerns of providers and an understanding of the challenges stemming from working with data from these sources.

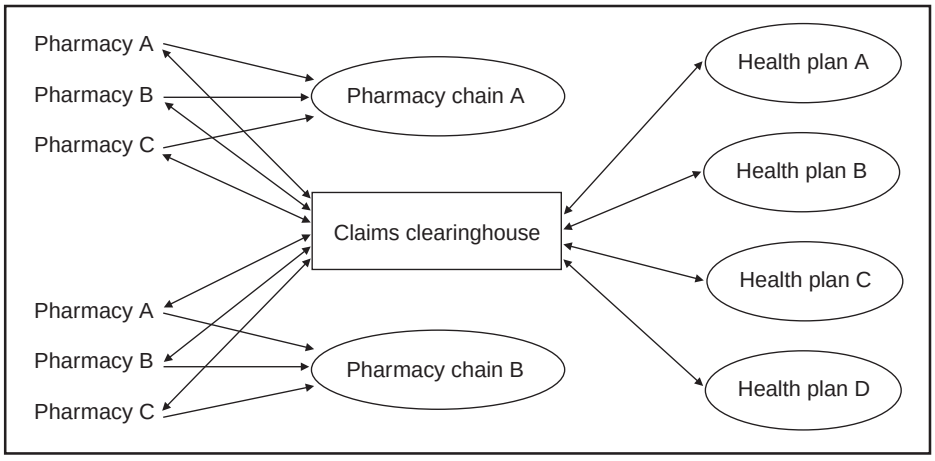
Objectives: The paper provides an overview of selected data-source types, concerns relating to partnerships with data providers, and challenges of working with shared data.

Methods: The authors conducted a study based on their experience in working with private-sector data providers, of different data types and provider partnerships. The study focused on the benefits of working with data providers, concerns and motivations of data providers, reasons for participating in data-sharing partnerships, and technical and legal problems of data sharing.

Results: Benefits of working with data providers include substantial-sized samples, broad geographic coverage, timeliness of data, and passive data collection. Problems arising from working with data providers include complexity of data extraction, need to protect patient privacy and confidentiality, resource requirements, limited financial benefits, concerns about public opinion, and duplicative data requests. Reasons for provider participation include commercial benefit, limited resource requirements, and corporate goodwill. Other challenges for data recipients include data processing, quality control, storage requirements, and lack of available, proven analytic techniques for data interpretation.

Conclusions: Substantial sample size, timeliness, and passive collection can be gained by using certain types of data. However, to attain these advantages, end-users should be prepared to address the concerns of data providers and cope with the methodologic and technical complexities involved in undertaking such partnerships (Figure).

FIGURE. Potential data access points within a network



From Implementation to Automation — A Step-by-Step Approach to Developing Syndromic Surveillance Systems from a Public Health Perspective

Brian M. Lawson,[1] E. Fitzhugh,[1] S. Hall,[1] L. Hutwagner,[2] G. Seeman[2]
[1]*Knox County Health Department, Knoxville, Tennessee;* [2]*National Center for Infectious Diseases, CDC*

Corresponding author: Brian M. Lawson, Knox County Health Department, 140 Dameron Ave., Knoxville, TN 37917. Telephone: 865-215-5095; Fax: 865-584-4604; E-mail: brian.lawson@knoxcounty.org.

Abstract

Introduction: Implementing new surveillance for biologic terrorism is becoming an essential function of local public health departments. Although syndromic surveillance systems can be implemented by multiple methods (including commercially available products), one model might be particularly well-suited to public health. CDC's Early Aberration Reporting System (EARS) is a syndromic surveillance system that uses aberration-detection models to identify deviations in current data when compared with a historic mean. The Knox County Tennessee Health Department (KCHD) is using a 7-day seamless surveillance system based on the EARS program that incorporates multiple data sources, automated data transfer via file transfer protocol (FTP), scheduled batch analysis, and remote access to surveillance data.

Objectives: KCHD developed a 10-step process for designing a syndromic surveillance system, from implementation to automation.

Methods: The steps are as follows:

1. Contact CDC staff to discuss acquisition of EARS programs.
2. Assess infrastructure to implement EARS.
3. Engage stakeholders.
4. Identify staff and assign specific tasks.
5. Select syndromes or symptoms to monitor.
6. Establish daily data exchange.
7. Develop automation routines for data transfer via FTP and for importing data into SAS.
8. Schedule EARS analysis programs as a batch job.
9. Establish a review and response protocol.
10. Develop plans for long-term collaboration and system expansion, including evaluation.

Results: By following these 10 steps, KCHD has made substantial progress toward implementing a multifaceted, seamless, 7-day syndromic surveillance system.

Conclusions: Public health departments can use these 10 steps as a framework for developing local syndromic surveillance systems.

Site-Based Biosurveillance*

Chung-Sheng Li,[1] C. Aggarwal,[1] M. Campbell,[1] Y. Chang,[1] M. Hill,[1] V. Iyengar,[1] M. Naphade,[1] J. Smith,[1] M. Wang,[1] K. Wu,[1] P. Yu,[1] A. Kress[2]
[1]IBM T.J. Watson Research Center, Yorktown Heights, New York; [2]Surveillance Data, Inc., Plymouth Meeting, Pennsylvania

Corresponding author: Murray Campbell, IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598. Telephone: 914-945-2942; Fax: 914-945-4077; E-mail: mcam@us.ibm.com.

Abstract

Introduction: Site-based biosurveillance presents multiple opportunities for data collection that might not be available in less permissive environments.

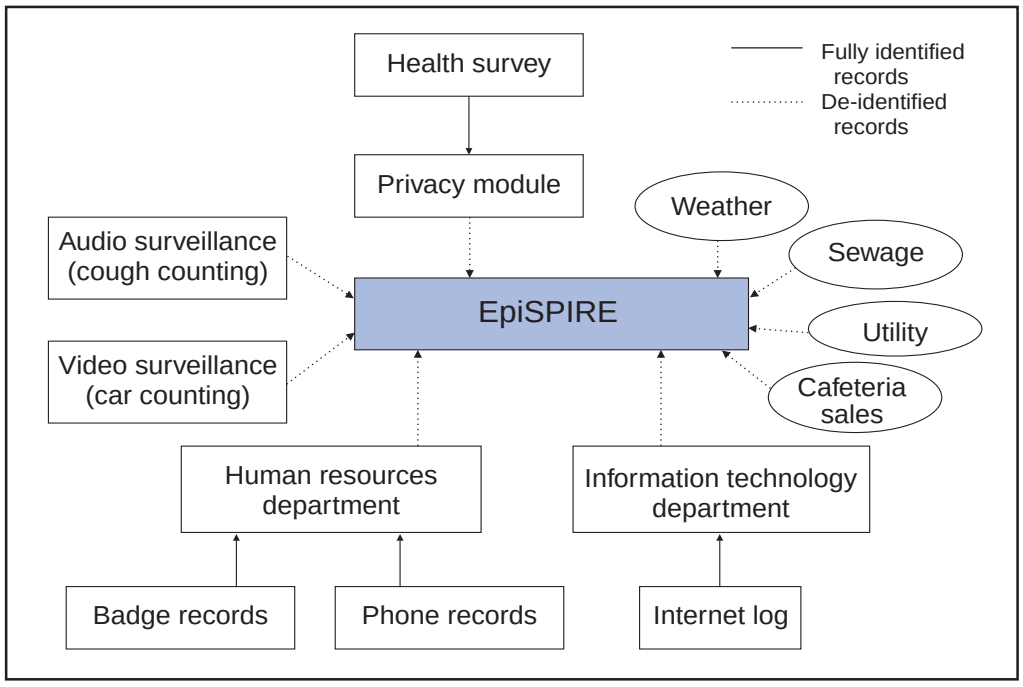
Objectives: This study examined the potential for using site-based biosurveillance — the monitoring of a geographically contained site (e.g., work site, university campus, or military base) — to detect disease outbreaks.

Methods: Available data sources were catalogued, and an initial characterization of those data sources with respect to their value for disease surveillance was performed. A system (EpiSPIRE) for managing surveillance data (both site and regional) and outbreak-detection algorithms was also developed (Figure). The study was conducted at the IBM T.J. Watson Research Center, which is located at two sites 10 miles apart: Yorktown Heights, New York, and Hawthorne, New York. Data collection started in late 2001. Physician office-visit data for respiratory illness in the Westchester County area was supplied by Surveillance Data, Inc., for use in evaluating the site data sources.

Results: Two site data sources were identified as most promising: 1) a survey of self-assessed health and 2) phone calls to medically related phone numbers. Absenteeism, Internet queries, cafeteria sales, and traffic data, though less promising, are worthy of further study. Cough counting and utility usage appear to have less value for site surveillance.

* This work is supported by the Air Force Research Laboratory (AFRL)/Defense Advanced Research Projects Agency (DARPA) under AFRL Contract No. F30602–01–C–0184. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the AFRL or DARPA.

FIGURE. Data sources examined by the EpiSPIRE site-based biosurveillance system — IBM T.J. Watson Research Center, Yorktown Heights and Hawthorne, New York



Expansion of ESSENCE for Use in Joint Military and Civilian Surveillance in Nine Cities*

Nicola Marsden-Haug,[1] J. Pavlin,[1] V. Foster,[1,2] S. Rechter,[1] J. Lombardo,[2] S. Lewis[2]
[1]*Walter Reed Army Institute of Research, Silver Spring, Maryland;* [2]*Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland*

Corresponding author: Nicola Marsden-Haug, Nicola Marsden-Haug, Walter Reed Army Institute of Research, Division of Preventive Medicine, 503 Robert Grant Avenue, Silver Spring, MD 20190. Telephone: 301-319-7355; Fax: 301-319-9104; E-mail: nicola.marsden-haug@na.amedd.army.mil

Abstract

Introduction: Syndromic surveillance systems are increasingly commonplace, as multiple states and CDC have begun using them for potentially timelier and more sensitive outbreak detection. Although different nontraditional indicators are being used to achieve earlier detection, optimally sensitive systems should capture data from civilian, military, and veteran populations.

Objectives: Walter Reed Army Institute of Research and Johns Hopkins University Applied Physics Laboratory are participating in the U.S. Department of Defense Joint Services Installation Pilot Project (JSIPP). This project targets nine military installations as model sites for integrated surveillance, protection, and response. Under the force-protection component, sites will acquire chemical and biologic detection capabilities and emergency-response equipment. Sites will also receive an upgraded version of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE IV).

Methods: ESSENCE IV was developed for pilot testing at JSIPP sites. Military outpatient and prescription data will be integrated with civilian *International Classification of Diseases, Ninth Revision* (ICD-9) claims, emergency department chief complaints, and outpatient Veterans Affairs data from surrounding communities. Other enhancements include a new user interface, a geographic information system for mapping disease distribution and spatial clusters, and new temporal signal detection methods.

Results: The challenge of integrating military and civilian data is engaging appropriate personnel from both jurisdictions. For JSIPP, military preventive medicine and civilian public health will jointly define data-sharing agreements and standard operating procedures. Workshops will be held to establish alert-response protocols.

Conclusions: This program can serve as an example for establishing joint surveillance across military and civilian borders.

* The views expressed are those of the authors and do not reflect the position of the U.S Army or the U.S. Department of Defense.

Investigation of Diarrheal Illness Detected Through Syndromic Surveillance After a Massive Blackout — New York City, August 2003

Melissa A. Marx,[1,2] C. Rodriguez,[1] J. Greenko,[1] D. Das,[1] F. Mostashari,[1] S. Balter,[1] R. Heffernan,[1] M. Layton,[1] D. Weiss[1]
[1]New York City Department of Health and Mental Hygiene, New York, New York; [2]Epidemiology Program Office, CDC

Corresponding author: Melissa A. Marx, Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, 125 Worth St., Box 22A, New York, NY 10013. Telephone: 212-442-9072; E-mail: mmarx@health.nyc.gov.

Abstract

Introduction: After a massive blackout in New York City on August 14, 2003, a larger number of patients than expected visited city emergency departments (EDs) for diarrhea.

Objective: New York City Department of Health and Mental Hygiene conducted a case-control study to determine risk factors for diarrheal illness among patients who visited EDs after the blackout.

Methods: Subjects were selected from patients who visited EDs participating in syndromic surveillance during August 16–18, 2003. All persons with diarrhea syndrome were designated case-patients. Control patients were a stratified random sample of patients with other syndromes. Structured telephone interviews were used to collect information about exposures between the blackout and symptom onset. Patients whose symptom onset occurred before the blackout were excluded.

Results: Of 759 subjects selected, 287 (38%) were reached and eligible, agreed to participate, and reported their age. Approximately 68% of study participants reported consuming chicken, meat, seafood, dairy products, or deli meat between the time of the blackout and symptom onset. Although case-patients (n = 58) and control patients (n = 100) aged <13 years indicated no differences in food consumption, more case-patients (n = 58) than control patients (n = 71) aged ≥13 years ate seafood (odds ratio [OR] = 4.8; 95% confidence interval [CI] = 1.6–14.1) or meat (OR = 2.7, 95% CI = 1.2–6.1) after the blackout. No differences existed in percentage of case- and control patients who discarded foods after the blackout. Overall, 67% of patients heard messages recommending the discarding of food; the most common sources for those messages were television (35%) and radio (28%).

Conclusions: Without refrigeration, meat and seafood spoil quickly. Diarrheal illness among adults in this study was associated with consumption of meat and seafood and might have been associated with food spoilage after the blackout. Syndromic surveillance was essential for detecting the increase in diarrhea after the blackout and for framing the study to investigate this increase.

Semantic Approach to Public Health Situation Awareness — Design and Methodology

Parsa Mirhaji, S. Lillibridge, R. Richesson, J. Zhang, J. Smith
University of Texas Health Science Center at Houston, Houston, Texas

Corresponding author: Parsa Mirhaji, University of Texas Health Science Center at Houston, Houston, Texas 77030. Telephone: 713-500-3157; Fax: 713-500-0370; E-mail; pmirhaji@uth.tmc.edu.

Abstract

Introduction: A public health situation awareness system is proposed that 1) uses explicit representation of surveillance concepts based on the user's cognitive model, and 2) is optimized for efficacy of performance and relevance to the process and task, rather than for ontic accuracy of syndrome definitions.

Objectives: The goal of this effort is to develop a prototype knowledge-based system that demonstrates the utility of knowledge-intensive approaches in 1) integrating heterogeneous information (e.g., patient triage data, pharmacy sales data, and school absenteeism data); 2) eliminating the effects of incomplete and poor-quality surveillance data; 3) reducing uncertainty in syndrome and aberration detection; and 4) enabling visualization of complex information structures in surveillance settings, particularly in the context of biologic terrorism preparedness.

Methods: For this approach, explicit domain knowledge is the foundation for interpreting public health data, as opposed to conventional systems for which statistical methods are central. The system uses the Resource Definition Framework (i.e., a framework for representing information that enables machines and humans to communicate) and expressive language (i.e., Web Ontology Language [OWL]) to explicate human knowledge into machine-interpretable and computable problem-solving modules that can guide users and computer systems in sifting through relevant data to detect outbreaks.

Results: A prototype knowledge-based system for early detection of outbreaks of influenza, which has a complex natural pattern and is a potential agent for biologic terrorism, is being developed. A model has been developed (using OWL ontology language) to enable case detection for respiratory illness syndromes caused by weaponized influenza. A knowledge-based system to integrate relevant health data from nine community hospitals has also been developed.

Conclusions: Preliminary data from this effort will evaluate the utility of knowledge-based approaches in information integration, syndrome and aberration detection, information visualization, and cross-domain investigation of root causes of events.

Improving Agreement Between Two Algorithms for Biosurveillance of Respiratory Disease in the Emergency Department — Chief Complaint and ICD-9 Code

Melissa Mocny,[1] D. Cochrane,[2,4] J. Allegra,[2,4] T. Nguyen,[1] J. Pavlin,[3] J. Rothman,[4] R. Heffernan,[5]
[1]*New York State Department of Health, Albany, New York;* [2]*Morristown Memorial Hospital Residency in Emergency Medicine, Morristown, New Jersey;*
[3]*Walter Reed Army Institute of Research, Silver Springs, Maryland;* [4]*Emergency Medical Associates of New Jersey Research Foundation, Livingston, New Jersey;*
[5]*New York City Department of Health and Mental Hygiene, New York, New York*

Corresponding author: Dennis G. Cochrane, Morristown Memorial Hospital Residence in Emergency Medicine, Morristown, NJ 07962. Telephone: 973-971-8919; E-mail: cochraned@verizon.net.

Abstract

Introduction: *International Classification of Diseases, Ninth Revision* (ICD-9) codes and patient chief complaints have both been advocated for use in syndromic surveillance of emergency department (ED) visits.

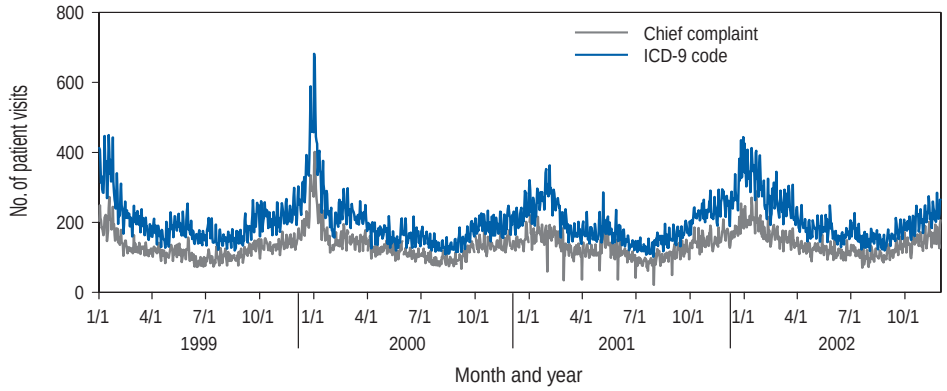
Objectives: The objective of this analysis was to determine whether two algorithms, one based on ICD-9 codes and the other on patient chief complaints, identified similar patterns and patient populations for respiratory illness. An attempt was also made to improve agreement by equalizing and expanding syndrome definitions.

Methods: Retrospective analysis was performed for consecutive visits to 15 New Jersey EDs. The Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) project supplied a then-current version of its ICD-9 algorithm. The New York State Department of Health extended a chief-complaint algorithm originally developed by the New York City Department of Health and Mental Hygiene and also made modifications to ESSENCE ICD-9 code groupings. A time-series graph was generated, and a correlation coefficient was calculated. Agreement between the two algorithms was examined in three stages: 1) initial chief complaint and ICD-9 algorithms, 2) after modifying the algorithms to match more closely, and 3) after expanding both algorithms to include *fever*.

Results: A total of 2,250,922 visits were used to compare seasonal variations as measured by the two methods (Figure). High correlation existed between the two algorithms ($r = 0.90$; $p<0.01$). A subset of 174,520 visits was examined; for stages 1, 2, and 3, respectively, agreement by kappa statistic was 0.28 (fair), 0.42 (moderate), and 0.56 (moderate); sensitivity was 0.31, 0.53, and 0.71; and specificity was 0.94, 0.91, and 0.90.

Conclusions: ICD-9 and chief-complaint algorithms for *respiratory* syndrome identified similar patterns of illness. The level of agreement was improved both by equalizing and by expanding the syndrome definitions.

FIGURE. Patient visits to 15 emergency departments for respiratory syndrome, by ICD-9 code and chief complaint — New Jersey, January 1, 1999–January 1, 2003



Syndromic Surveillance Using Chief Complaints from Urgent-Care Facilities During the Salt Lake 2002 Olympic Winter Games

Michael B. Mundorff,[1] P. Gesteland,[1,2] M. Haddad,[3,4] R. Rolfs[2,3,5]
[1]*Intermountain Health Care, Salt Lake City, Utah;* [2]*University of Utah, Salt Lake City, Utah;* [3]*Utah Department of Health, Salt Lake City, Utah;* [4]*Epidemiology Program Office, CDC;* [5]*National Center for Chronic Disease Prevention and Health Promotion, CDC*

Corresponding author: Michael B. Mundorff, Intermountain Health Care, 36 South State St., Suite 2100, Salt Lake City, UT 84111. Telephone: 801-442-2896; Fax: 801-442-3486; E-mail: michael.mundorff@ihc.com.

Abstract

Introduction: During the period surrounding the Salt Lake 2002 Olympic Winter Games, Utah's health departments conducted syndromic surveillance for potential early identification of natural or terrorist-introduced communicable disease outbreaks.

Objectives: For minimal onsite intrusion and protection of patient confidentiality, electronic data from 19 urgent-care facilities were routed to public health authorities by using a computer program that mapped free-text chief complaints from patient registrations into selected syndromes. After the Games, the system's usefulness for determining five syndrome categories was evaluated.

Methods: During January 15–March 23, 2002, syndromes were monitored as daily counts and proportions of total visits. Changes in occurrence over time were tracked by statistical process control charts. Findings were compared with other public health surveillance streams (e.g., influenza surveillance). Classification validity was examined by comparing five syndromes initially identified through chief-complaint data with a reference standard of *International Classification of Diseases, Ninth Revision* (ICD-9) discharge diagnoses. Sensitivity, specificity, predictive values, and likelihood ratios were measured.

Results: *Respiratory* syndrome changes paralleled influenza seasonality. Occurrence of other syndromes remained relatively constant. The predictive value of classification by chief complaint varied substantially among syndromes (Table).

Conclusions: Syndromic surveillance findings provided reassurance that no unexpected communicable disease outbreaks occurred. Validity measures for *respiratory*, *gastrointestinal*, and *rash* syndromes appear sufficiently promising to warrant additional investigation of this approach's value for detecting outbreaks manifesting as these syndromes. Effective syndrome classification by free-text chief complaint requires knowledge of disease presentations, local information systems, and linguistic conventions used by registration clerks.

TABLE. Comparison of keyword-based chief-complaint (CC) classification system with ICD-9* discharge diagnosis classification system for patient visits to 19 urgent-care facilities during the period surrounding the Salt Lake 2002 Olympic Winter Games — Utah, January 15–March 23, 2002

	Syndrome (n = 59,404)				
Measure	**Respiratory**	**Gastroenteritis**	**Rash**	**Neurologic**	**Botulinic**
CC total counts	15,514	2,293	1,721	697	143
Proportion of visits (%)	26.1	3.9	2.9	1.2	0.24
ICD-9 total counts	30,061	946	1,056	7	62
Proportion of visits (%)	50.6	1.6	1.8	0.01	0.10
Sensitivity	0.42	0.46	0.54	0.14	0.097
Specificity	0.90	0.97	0.98	0.99	0.998
Predictive value positive	0.81	0.19	0.33	0.001	0.04
Predictive value negative	0.60	0.99	0.99	0.9999	0.999
Positive likelihood ratio	4	15	27	12	42

* *International Classification of Diseases, Ninth Revision.*

Fast Grid-Based Scan Statistic for Detection of Significant Spatial Disease Clusters

Daniel B. Neill, A. Moore
Carnegie Mellon University, Pittsburgh, Pennsylvania

Corresponding author: Daniel B. Neill, Department of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213. Telephone: 412-621-2650; E-mail: neill@cs.cmu.edu.

Abstract

Introduction: The spatial scan statistic is a commonly used statistical test for detecting significant disease clusters. However, the time needed to compute the scan statistic increases as the square of the number of data points M, making the test computationally infeasible for large data sets (M >100,000). One solution is to aggregate data points to a uniform grid — when the grid is dense, the scan statistic can be computed substantially faster, with complexity $O(M\sqrt{M})$ instead of $O(M^2)$. However, even this approach can require multiple days to compute when M is large. Because disease clusters must be found in minutes rather than days for real-time detection, a more efficient algorithm is needed.

Objectives: Given a grid of squares, where each square has an associated count (number of disease cases) and underlying population, the goal is to quickly find the region with the maximum value of the scan statistic (the most significant disease cluster).

Methods: A multiresolution algorithm is proposed that partitions the grid into overlapping regions, bounds the maximum score of each region, and prunes regions that cannot contain the most significant cluster. This method enables users to search across all possible regions while examining only a fraction of the regions. This reduces complexity to $O(M)$ for dense test regions. As in the original scan statistic, randomization testing is used to calculate the statistical significance (p-value) of the detected cluster. (For additional details, see the full paper at http://www.cs.cmu.edu/~neill/papers/sss-techreport.pdf.)

Results: The algorithm was tested on seven data sets ($M \approx$ 200,000), including western Pennsylvania emergency department data. The algorithm identified the most significant disease clusters in 20–130 minutes, 20–150 times faster than exhaustive search (Table).

Conclusions: The algorithm results in substantial speedups as compared with exhaustive search, making real-time detection of disease clusters computationally feasible. This algorithm is being applied toward automatic real-time detection of outbreaks.

TABLE. Performance of a multiresolution algorithm for detection of spatial disease clusters, as compared with exhaustive search

Data set	Time (1,000 replications)	Speedup versus exhaustive*
Standard, large test region	17 minutes, 3 seconds	154x
Standard, small test region	29 minutes, 51 seconds	88x
City, large test region	21 minutes, 26 seconds	122x
City, small test region	131 minutes, 44 seconds	20x
High variance, large region	17 minutes, 21 seconds	151x
High variance, small region	34 minutes, 55 seconds	75x
Emergency department	46 minutes, 42 seconds	85x

* Speedup is defined as the run time of exhaustive search divided by the run time of the algorithm. For example, a 10x speedup means that the algorithm finds the most significant disease cluster in 1/10 the time of exhaustive search.

Population-Adjusted Stable Geospatial Baseline for Outbreak Detection in Syndromic Surveillance

Karen L. Olson,[1,2] M. Bonetti,[3] M. Pagano,[3] K. Mandl[1,2]
[1]*Children's Hospital Boston, Boston, Massachusetts;* [2]*Harvard Medical School, Boston, Massachusetts;* [3]*Harvard School of Public Health, Boston, Massachusetts*

Corresponding author: Karen L. Olson, Children's Hospital Informatics Program, Enders 5, Children's Hospital Boston, 300 Longwood Ave., Boston, MA 02115. Telephone: 617-355-6718; Fax: 617-730-0921; E-mail: karen.olson@childrens.harvard.edu.

Abstract

Introduction: Surveillance systems should detect outbreaks that become evident when cases cluster geographically. Identifying an illness with an abnormal spatial pattern of disease requires a stable model of what is normal, adjusting for underlying population density.

Objectives: Observations indicate that the distribution of all pairwise interpoint distances among patients in the catchment area of a hospital is stable over time. This study sought to demonstrate that baseline spatial distributions can be established.

Methods: Emergency department visits made during 2 years at two urban academic medical centers (one a pediatric hospital) were classified into syndromes according to chief complaints and *International Classification of Diseases*, *Ninth Revision* codes. Distances between all pairs of patient addresses were calculated. The number of visits and the distance distributions for respiratory and gastrointestinal syndrome at each hospital, by season, were determined.

Results: For respiratory syndrome at one hospital, the number of visits ranged from a summer low of 1,932 to a winter high of 4,457 (mean: 3,203; standard deviation: 795). Variability and seasonal effects were present. By contrast, the interpoint-distance distributions were characterized by remarkable similarity over time without seasonal effects. When individual distance distributions for each season for 3 years are plotted, they overlap to substantially, demonstrating their stability. This same pattern of results was identified for respiratory visits at one hospital and gastrointestinal visits at both hospitals.

Conclusions: Empirical and parametric methods that rely on detecting differences between interpoint-distance distributions have been described previously. Although the number of cases varies substantially over time, a stable geographic baseline can be established against which clusters can be detected. Therefore, syndromic surveillance is enhanced when location is incorporated into a system that can detect outbreaks in space, even when the number of cases is too small to generate alerts on the basis of frequency.

Improving Outbreak Detection by Signal Integration*

Paola Sebastiani,[1,2] L. Wang,[1,2] K.D. Mandl,[2,3] M. Ramoni[2,3]
[1]*Boston University, Boston, Massachusetts;* [2]*Center for Biopreparedness at Children's Hospital, Boston, Massachusetts;* [3]*Harvard Medical School, Boston, Massachusetts*

Corresponding author: Paola Sebastiani, Department of Biostatistics, Boston University School of Public Health, 715 Albany Street, Boston MA 02118. Telephone: 617-638-5877; Fax: 617-638-6484; E-mail: sebas@bu.edu.

Abstract

Introduction: A critical problem of surveillance systems is the trade-off between true and false detections. Integration of different monitors and information from exogenous sources can increase the true-detection rate by limiting the false-detection rate.

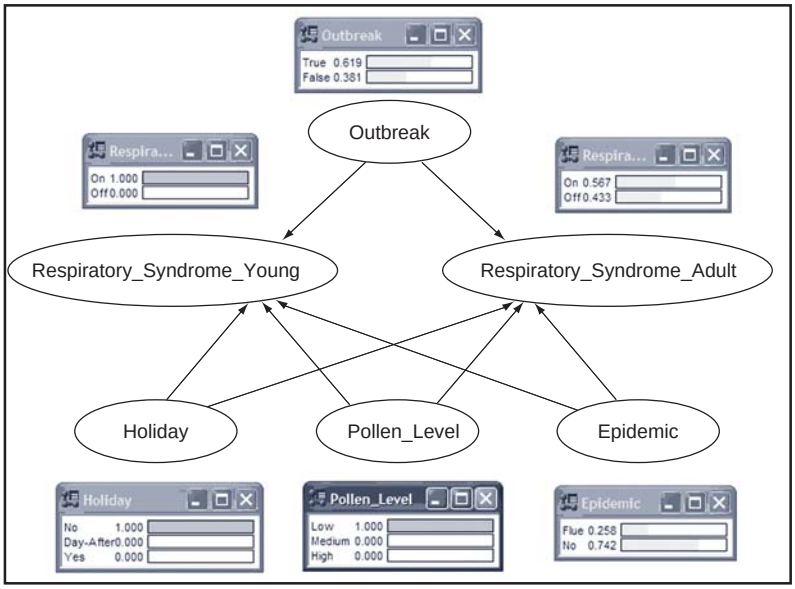
Objective: The authors introduce a probabilistic architecture able to achieve a substantial detection rate while keeping false detections low.

Methods: The architecture is a Bayesian network that encodes probabilistic information through a directed graph. The nodes and arrows represent variables and stochastic dependencies quantified by probability distributions. The integration of two systems for syndromic surveillance at a pediatric and adult hospital is illustrated by using a respiratory illness outbreak (Figure). Empirical evaluations have demonstrated that true and false-alert rates are affected by influenza epidemics, by air quality as measured by pollen level, and by whether the alert day is a holiday. The network integrates the sources of information to compute the probability of an outbreak (given that one or both systems generate alerts) and what is known about the other variables. The probability tables quantifying the network were obtained from data contaminated with different simulated outbreaks. The integrator was validated on 84 simulated outbreaks.

Results and Conclusions: This study indicates that the integration of the two monitoring systems with exogenous information has a 73% true-detection rate with an 8% false-detection rate in limited outbreaks (i.e., an average of four ill persons/day), and 97% true-detection rate with 10% false-detection rate in more substantial outbreaks (i.e., an average of eight ill persons/day).

* This work was supported by the Alfred P. Sloan Foundation (Grant 2002-12-1).

FIGURE. Integration of two monitoring systems of syndromic data (nodes Respiratory_Syndrome_Young and Respiratory_Syndrome_Adult) with exogenous variables that provide information about an influenza epidemic (node Epidemic), air quality (node Pollen_Level), and whether the alert day is a holiday (node Holiday).



Note: The figure illustrates that if one of the two systems generates an alert during a normal work day with good air quality, the probability of an outbreak is 61.9%.

Evaluation of Taiwan's Syndromic Surveillance System after the Severe Acute Respiratory Syndrome — Taiwan, 2003

Frank Fuh-Yuan Shih,[1,2] M-Y. Yen,[3] F-K. Chang,[4] L-W. Lin,[4] J-S. Wu,[4] C-A. Hsiung,[5] M-S. Ho,[6] I. Su,[3] M. Marx,[7] H. Sobel,[8] C-C. King[1]
[1]College of Public Health, National Taiwan University, Taipei, Taiwan; [2]National Taiwan University Hospital, Taipei, Taiwan; [3]Jen-Ai Municipal Hospital, Taipei, Taiwan; [4]Center for Disease Control, Taipei, Taiwan; [5]National Health Research Institute, Taipei, Taiwan; [6]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; [7]New York City Department of Health and Mental Hygiene, New York, New York; [8]World Health Organization, Geneva, Switzerland

Corresponding author: Chwan-Chuen King, Institute of Epidemiology, College of Public Health, National Taiwan University, 1 Jen-Ai Road, Section 1, Taipei, Taiwan (100). E-mail: a1234567@ccms.ntu.edu.tw.

Abstract

Introduction: Taiwan's clinical syndromic surveillance system faced substantial challenges during the 2003 outbreak of severe acute respiratory syndrome (SARS).

Objectives: This study aimed to evaluate the feasibility of syndromic surveillance for health-care workers and delineate obstacles to the reporting process.

Methods: Six months after the SARS outbreak, self-administered, structured questionnaires were mailed to 270 Taiwan health-care workers at medical centers, community hospitals, and other health-care facilities. The questionnaire gathered information about demographics, difficulties in reporting, reasons for delayed reporting or underreporting, and types of information health-care workers expected for feedback. Chi-square and paired t-tests were used for data analysis.

Results: A total of 229 completed questionnaires (84.8%) were analyzed. Respondents cited the following problems in reporting SARS cases: waiting for laboratory data (48%), ambiguous clinical presentations (45%), and protection of patient privacy (45%). Health-care workers in medical centers expressed greater concern about rigorous control from hospital authorities but had less difficulty in arranging consultations and were less influenced by mass media. By contrast, health-care workers in community hospitals waited longer for treatment responses, had more consultation regarding confusing laboratory results, and experienced more pressure from patients and their relatives not to report their illnesses. Respondents cited a need for improved guidelines, recommendations, standard operating procedures, and the effectiveness of prevention and control measures.

Conclusions: Future SARS surveillance in Taiwan requires simplified case definitions with different levels of confirmation, built-in mechanisms to prevent release of confidential information, enhanced infection-control training, timely communication of appropriate feedback information; and enhanced use of information technology to simplify the reporting process and integrate different data sets.

Temporal Correlation of Nontraditional and Traditional Evidence of a Natural Outbreak

Douglas Stetson, R. Bloom, D. Crary, K. Cheng, G. McClellan
General Dynamics Advanced Information Systems, Arlington, Virginia

Corresponding author: Douglas Stetson, 1400 Key Blvd., Ste. 100, Arlington, VA 22209. Telephone: 757-377-5283; E-mail: douglas.stetson@gd-ais.com.

Abstract

Introduction: Although syndromic surveillance typically involves monitoring of traditional clinical data sources (e.g., emergency department visits), monitoring nontraditional sources might also provide information about community health. This study demonstrated that parking use data from a medical center parking facility reflected an unusual increase in regional outpatient visits for respiratory illness associated with a well-publicized public health event.

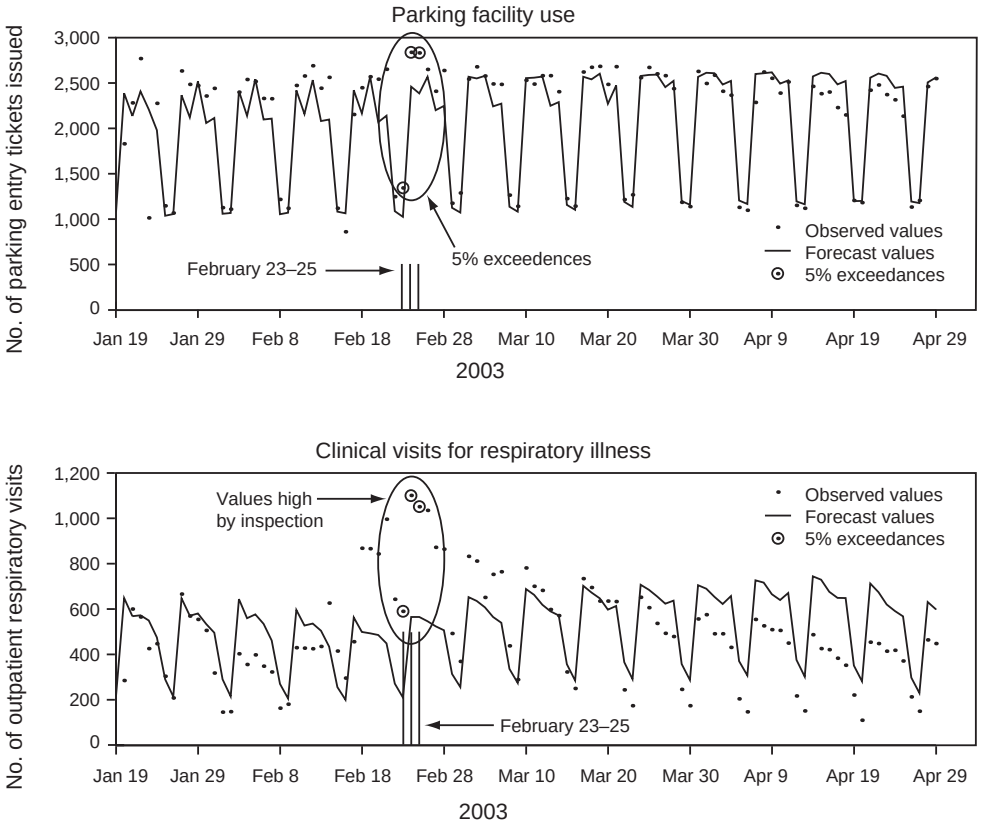
Objective: This study aimed to determine whether a nontraditional source (i.e., parking facility use data) reflected a sudden communitywide surge in health-care facility use associated with widespread news coverage of an unexpected local cluster of respiratory illness-related deaths among children.

Methods: Two data sources were collected and compared for the period in which the cluster occurred: 1) daily parking facility use data from a parking structure serving a medical center complex and 2) regional counts of outpatient respiratory visits to military treatment facilities made by military members and their families. Daily localized forecasts of expected parking use and outpatient visits were generated on the basis of recent historic counts to reduce cyclic influences (e.g., day-of-week effects). Daily variations in parking and clinic volume and differences between actual volume and forecast volume were analyzed for statistical significance.

Results: A statistically significant increase in actual parking facility use compared with expected use was identified, coincident with both the statistically significant increase in actual outpatient respiratory visits compared with forecast visits and with the period of local news reporting on the cluster of deaths (February 23–25, 2003) (Figure). No other variations in the parking or outpatient-visit data during that calendar quarter had similar statistical significance.

Conclusion: Syndromic surveillance efforts can be supported by standard analytic and statistical examination of nonclinical, real-world data.

FIGURE. Observed versus expected parking facility use and observed versus expected outpatient respiratory visits associated with a cluster of childhood respiratory illness deaths — one community, January 19–April 29, 2003



Effects of Sensitivity and Specificity on Signal-to-Noise Ratios for Detection of Influenza-Associated Aberrations

William W. Thompson
National Immunization Program, CDC

Corresponding author: William W. Thompson, CDC, 1600 Clifton Rd., MS E-61, Atlanta, GA 30333. Telephone: 404-639-8256; Fax: 404-639-8834; E-mail: wct2@cdc.gov.

Abstract

Introduction: Influenza-associated outcomes have been used to test and validate alternative aberration-detection methods, yet a limited number of studies have examined the effects of using different outcomes with varying levels of sensitivity and specificity for influenza.

Objectives: Influenza aberration-detection models developed by CDC were applied to daily death outcomes by using city-level mortality data.

Methods: Influenza surveillance data were obtained from the World Health Organization, and city-level mortality data were obtained from CDC's National Center for Health Statistics. Deaths were categorized by *International Classification of Diseases, Ninth Revision* (ICD-9) and *Tenth Revision* (ICD-10) codes. Age-specific log-linear regression models were used to identify influenza-associated aberrations in death outcomes.

Results: For pneumonia and influenza deaths, the models accounted for 49% and 83% of the variance for persons aged <65 years and persons aged ≥65 years, respectively. Influenza accounted for 4.4% of the variance among persons aged <65 and 8.2% of the variance among persons aged ≥65. Seasonal variation accounted for the greatest percentage of explained variance for pneumonia and influenza deaths; day-of-week, holiday, and post-holiday variables accounted for <1% of the explained variance. For respiratory and circulatory deaths, the models accounted for 89% of the variance in outcome both for persons aged <65 and persons aged ≥65. Influenza accounted for 1.2% of the variance among persons aged < 65 and 6% of the variance among persons aged ≥65. Seasonal variation accounted for substantially less of the explained variance in death outcome for persons aged <65; time trends accounted for substantially more of the variation when compared with models applied to persons aged ≥65.

Conclusions: Substantial differences were identified in the signal-to-noise ratios by influenza-associated death outcomes. Certain confounders (e.g., age, time, and season) are key factors when identifying influenza-associated aberrations.

Empirical Evaluation of Space-Time Models for Surveillance of Disease Maps

Carmen L. Vidal Rodeiro, A. Lawson
University of South Carolina, Columbia, South Carolina

Corresponding author: Carmen L. Vidal Rodeiro, Department of Epidemiology and Biostatistics, Norman J. Arnold School of Public Health, 800 Sumter St., Columbia, SC 29208. Telephone: 803-777-1232; Fax: 803-777-2524; E-mail: cvidal@sc.edu.

Abstract

Introduction: Interest in statistical methods for public health surveillance has increased in recent years.

Objectives: Different space-time models for counts of disease were compared to assess their ability to detect changes in risk patterns across space and time.

Methods: Space-time models for estimating disease risk should be able to describe the overall space-time behavior of a disease and should also be sensitive to changes in its spatio-temporal structure. For this study, the observed count of disease cases in a region was assumed to be a Poisson variable. Logarithms of relative risk parameters were assumed to follow normal distributions with mean that incorporated potential risk factors and variance matrix that incorporated the possibility of spatial dependence (e.g., correlation induced by unmeasured variables). Space-time models in different scenarios representing possible changes in risk patterns over space and time were fitted.

Results: As a goodness of fit measure, the deviance information criterion was used. It demonstrated statistically significant increases in the years in which changes in risk were generated. Analysis of the p-value surface, residuals, and surveillance residuals (difference between observed data for 1 year and data expected under a model when fitted for previous years) proved that an unusual event happened in the counties and years with changes; therefore, those data were not representative of what was expected under the model. Where no changes in risk were generated, the p-values indicated that the model produced an optimal fit.

Conclusions: Although existing methods can be used for disease surveillance, additional methods that are more sensitive to the sequential nature of the surveillance task are needed.

Toxic Exposure Surveillance System

William A.Watson,[1] T. Litovitz,[1] C. Rubin,[2] E. Kilbourne,[3] M. Belson,[2] M. Patel,[2] J. Schier,[2] A. Funk[2]
[1]*American Association of Poison Control Centers, Washington, D.C.;* [2]*Division of Environmental Hazards and Health Effects, National Center for Environmental Health, CDC;* [3]*Agency for Toxic Substances and Disease Registry, National Center for Environmental Health, CDC*

Corresponding author: William A. Watson, American Association of Poison Control Centers, 3201 New Mexico Ave., N.W., Suite 330, Washington, D.C. 20016. Telephone: 202-363-7630; Fax: 202-362-3240; E-mail: watson@poison.org.

Abstract

Introduction: The Toxic Exposure Surveillance System (TESS) is a national, real-time surveillance database that includes all human exposures reported to participating U.S. poison control centers since 1985. More than36 million human poison exposures have occurred since December 2002. The database is continuously updated, with approximately 6,500 new human exposure cases added daily.

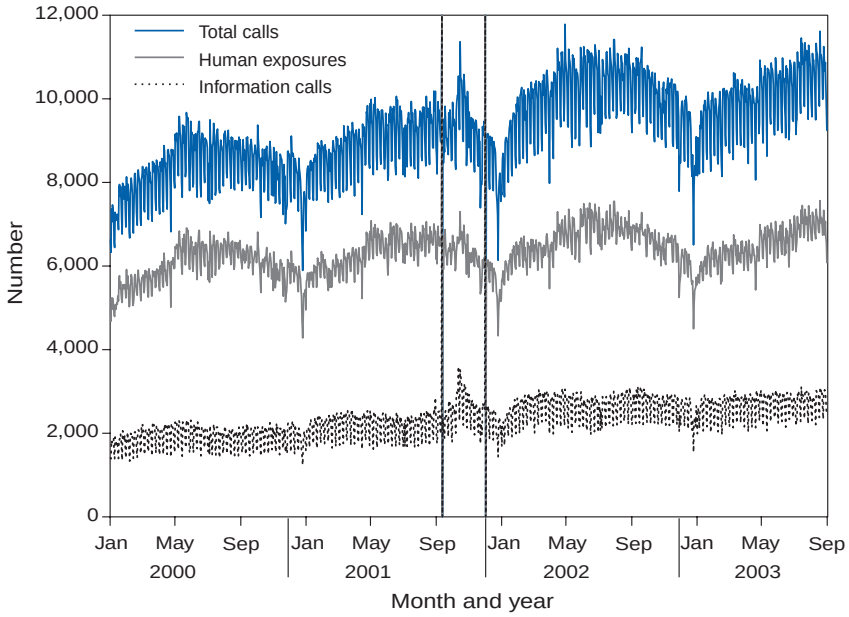
Objectives: This paper describes TESS and the current toxicosurveillance methods being applied to TESS for earliest possible identification of potential instances of chemical terrorism and other events of potential public health importance that require additional investigation.

Methods: Poisoning cases are managed and entered into TESS by poison-information specialists at each U.S. poison control center. The specialists collect data as part of triage and case management and code these data according to standardized definitions. Approximately 44% of cases receive follow-up, allowing for determination of the clinical course and outcome of the exposure. TESS searches for aberrations in hourly case volume for each poison control center and for daily frequency of clinical effects. Multiple surveillance case definitions and queries for presence of specific substances are used to identify possible sentinel cases for review. Query results are interpreted by clinical toxicologists, and individual cases producing signals are reviewed for clinical and surveillance significance. Reporting poison control centers are contacted for additional information as needed.

Results: Daily total case counts demonstrate the effect of the anthrax-related events of October–November 2001 on poison center case volume, with an increase in both information and human exposure cases (Figure). An increase in calls to the New York City Poison Control Center regarding food poisoning and what to do with spoiled food after the power blackout of August 2003 demonstrated the potential for TESS data to identify local changes in specific call types that might not be recognized in national data. TESS surveillance detected the malicious contamination of coffee with arsenic in Maine. Local and federal chemical and biologic terrorism-preparedness exercises have also generated surveillance signals.

Conclusions: TESS adds a real-time set of toxicity data to national, regional, and local surveillance that can rapidly identify local and national events of public health importance, including intentional chemical attacks.

FIGURE. Time series of information calls, human exposure reports, and total cases received by U.S. poison control centers, as reported to the Toxic Exposure Surveillance System — United States, January 1, 2000–September 1, 2003



Note: The increase in volume during October and November 2002 reflects the increase in calls to poison centers regarding anthrax-related events.

What's Strange About Recent Events, Version 3.0 — Accounting for a Changing Baseline

Weng-Keen Wong,[1] A. Moore,[1] G. Cooper,[2] M. Wagner[2]
[1]*Carnegie Mellon University, Pittsburgh, Pennsylvania;* [2]*University of Pittsburgh, Pittsburgh, Pennsylvania*

Corresponding author: Weng-Keen Wong, University of Pittsburgh, 200 Lothrop St., 8084 Forbes Tower, Pittsburgh, PA 15213. Telephone: 412-246-5824; Fax: 412-246-5985; E-mail: wwong@cbmi.pitt.edu.

Abstract

Introduction: This paper extends the algorithm outlined in an earlier version of this paper by detecting anomalous patterns in health-care data while accounting for temporal trends in the data (e.g., fluctuations caused by day-of-week effects and seasonal variations in temperature and weather).

Objectives: What's Strange About Recent Events (WSARE) 2.0 compared the distribution of recent data against a baseline distribution obtained from raw historic data. However, this baseline is affected by different fluctuations in the data (e.g., day-of-week effects and seasonal variations). Creating the baseline distribution without taking such trends into account can lead to unacceptably high false-positive counts and slow detection times.

Methods: This paper replaces the baseline method of WSARE 2.0 with a Bayesian network, which produces the baseline distribution by taking the joint probability distribution of the data and conditioning on attributes that are responsible for the trends.

Results: WSARE 3.0 is evaluated on a simulator that contains different temporal trends. Annotated results on real emergency department data are also included.

Conclusions: WSARE 3.0 is able to detect outbreaks in simulated data with almost the earliest possible detection time while keeping a low false-positive count.

Practical Evaluation of Electronic Disease Surveillance Systems for Local Public Health

Cynthia B. Yund,[1] M. Ottaway,[1] D. Brannen[2]
[1]*Hamilton County General Health District, Cincinnati, Ohio;* [2]*Greene County Combined Health District, Xenia, Ohio*

Corresponding author: Cynthia Yund, 250 Howard Taft Rd., 2nd Fl., Cincinnati, OH 45219. Telephone: 513-946-7809; Fax: 513-946-7890; E-mail: cynthia.yund@hamilton-co.org.

Abstract

Introduction: Practical use of electronic disease surveillance systems is in the public health domain. Academia and the private sector have provided multiple surveillance options. The onus lies with public health to determine which system best supports the current infrastructure. Traditionally, public health has not had the capacity for such systems.

Objectives: A process was established to evaluate multiple electronic surveillance products for a population of >1.6 million persons. The geographic area encompasses 16 local health jurisdictions within eight southwestern Ohio counties consisting of urban, suburban, and rural populations.

Methods: Seven viable surveillance systems were identified through an Internet search. Members researched selected systems according to published criteria for evaluation of electronic disease surveillance systems, including vendor, validation, flexibility, expandability, operation, timeliness, reliability, notification, usability, security, compatibility, and supportability. Systems were rated by group consensus as acceptable (1) or unacceptable (0) on each of the criteria. A total score was assigned. Scores were adjusted (+1, 0, or –1) according to feasibility of local implementation on the basis of need for physical and human resources.

Results: Total adjusted scores ranged from 2 to 12 (Table). The Real-Time Outbreak and Disease Surveillance system was identified as superior and most feasible for local implementation.

Conclusions: Use of a group process to research the feasibility of using syndromic surveillance in local health jurisdictions increased the group's knowledge about the benefits of early warning indicators and facilitated discussion on viable systems.

TABLE. Rating of syndromic surveillance systems by a group of local health department representatives

Syndromic surveillance system	Criteria score*	Adjustment	Total score
A	10	0	10
B	9	-1	8
C	5	0	5
D	9	+1	10
E	9	-1	8
F	2	0	2
G	11	+1	12

* Systems were rated by group consensus as acceptable (1) or unacceptable (0) on each of the following criteria: vendor, validation, flexibility, expandability, operation, timeliness, reliability, notification, usability, security, compatibility, and supportability.

2004 Syndromic Surveillance Conference

Boston, Massachusetts, November 3–4, 2004

Abstract submission deadline is September 15, 2004
Early bird registration ends October 1, 2004

Tufts Health Care Institute

ALFRED P. SLOAN FOUNDATION 1934

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

Conference agenda and abstract submission at www.syndromic.org

The *Morbidity and Mortality Weekly Report (MMWR)* Series is prepared by the Centers for Disease Control and Prevention (CDC) and is available free of charge in electronic format and on a paid subscription basis for paper copy. To receive an electronic copy each week, send an e-mail message to *listserv@listserv.cdc.gov*. The body content should read *SUBscribe mmwr-toc*. Electronic copy also is available from CDC's World-Wide Web server at *http://www.cdc.gov/mmwr* or from CDC's file transfer protocol server at *ftp://ftp.cdc.gov/pub/publications/mmwr*. To subscribe for paper copy, contact Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402; telephone 202-512-1800.

Data in the weekly *MMWR* are provisional, based on weekly reports to CDC by state health departments. The reporting week concludes at close of business on Friday; compiled data on a national basis are officially released to the public on the following Friday. Address inquiries about the *MMWR* Series, including material to be considered for publication, to Editor, *MMWR* Series, Mailstop C-08, CDC, 1600 Clifton Rd., N.E., Atlanta, GA 30333; telephone 888-232-3228.

☆U.S. Government Printing Office: 2004-633-140/00048 Region IV