MMWR™

Morbidity and Mortality Weekly Report

Supplement August 26, 2005 / Vol. 54

Syndromic Surveillance

Reports from a National Conference, 2004

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, 2004. MMWR 2005: 54(Suppl).

Specific: [Author(s)]. [Title of particular article]. In: Syndromic Surveillance: Reports from a National Conference, 2004. MMWR 2005; 54(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.

The material in this supplement originated in the National Center for Public Health Informatics, John W. Loonski, MD, (Acting) Director.

CONTENTS

CONTENTS (*Continued*)

Preface

Public health surveillance continues to broaden in scope and intensity. Public health professionals responsible for conducting such surveillance must keep pace with evolving methodologies, models, business rules, policies, roles, and procedures. The third annual Syndromic Surveillance Conference was held in Boston, Massachusetts, during November 3–4, 2004. The conference was attended by 440 persons representing the public health, academic, and private-sector communities from 10 countries and provided a forum for scientific discourse and interaction regarding multiple aspects of public health surveillance. The conference was sponsored by the Alfred P. Sloan Foundation, CDC, Tufts Health Care Institute, and the U.S. Department of Homeland Security and organized by a Scientific Program Planning Committee; members of the committee are listed at http://www.syndromic.org/syndromicconference/2004/course_book/TAB_2.pdf.

During the conference, 134 presentations were given, including 18 at plenary sessions, 60 oral presentations, and 56 poster presentations. The entire list of presentations is available at http://www.syndromic.org/syndromicconference/2004/course_book/TAB_1.pdf.

After the conference, an editorial committee was formed, consisting of members of the planning committee. The board conducted a peer-review process to select abstracts and manuscripts for publication. A total of 36 abstracts and 45 manuscripts were submitted. Each submission was evaluated, scored according to preset criteria by at least two reviewers, and discussed by members of the committee. Time and resource limitations precluded inclusion of all submissions for publication. The manuscripts and abstracts contained in this supplement represent a sampling of the relevant topics and perspectives for this complex subject area. The manuscripts are categorized into five content areas: 1) overview, policy, and systems; 2) data sources; 3) analytic methods; 4) simulation and other evaluation approaches; and 5) practice and experience.

During the conference, a session was held to discuss the possible formation of a professional society to advance the field of disease surveillance. This nonprofit entity will be incorporated to advance the science of surveillance. Identified proposed functions include serving as the institutional home of the annual conference, maintaining and expanding a website, and coordinating work groups to advance specific scientific projects. Interest in this new society reflects the importance of this field and the requirement for communication in operational surveillance. A more formalized social basis for focus in the field and adaptive business rules might permit 1) increased integration among the needed scientific cultures and disciplines; 2) maturation of approaches to surveillance methods, technology, standards, and evaluation; 3) increased interaction and more productive partnerships between the respective public health level roles; and 4) outreach to integrate perspectives and operational realities for public health that include not only infectious disease and biologic terrorism preparedness and response but also counterterrorism and national security concerns. Finally, such a society might provide a forum for consolidation of a much-needed professional peer group and network for surveillance system operators and data monitors. More discussion of this topic will take place at the 2005 Syndromic Surveillance Conference, which will be held in Seattle, Washington, during September 13–15, 2005.

The program committee, editorial committee, and the editorial staff of *MMWR* all deserve recognition for their work in organizing the conference and preparing these proceedings. Special thanks are given to Haobo Ma, MD, MS, BioSense Program, National Center for Public Health Informatics, CDC, who coordinated the preparation of these reports.

— *Henry R. Rolka, Chair, Editorial Committee*
Chief Scientist/Statistician
National Center for Public Health Informatics, CDC

Peer Reviewers

The following persons served as peer reviewers for these proceedings: Deepak Agarwal, PhD, AT&T Labs — Research; John R. Allegra, MD, PhD, Morristown Memorial Hospital Residency in Emergency Medicine, Emergency Medical Associates Research Foundation; Nelson Arboleda, MD, National Center for Infectious Diseases, CDC; Lawrence E. Barker, PhD, National Center for Injury Prevention and Control, CDC; Andrew L. Baughman, PhD, National Immunization Program, CDC; Marco Bonetti, PhD, Universita' Bocconi and Harvard University; Claire V. Broome, MD, Coordinating Center for Health Information and Service, CDC; John S. Brownstein, PhD, Harvard Medical School; David L. Buckeridge, MD, PhD, Stanford University; James W. Buehler, MD, Rollins School of Public Health, Emory University; Howard S. Burkom, PhD, Johns Hopkins University Applied Physics Laboratory; Man-Huei Chang, MPH, National Center for Public Health Informatics, CDC; Wendy W. Chapman, PhD, University of Pittsburgh; Dennis G. Cochrane, MD, Morristown Memorial Hospital Residency in Emergency Medicine, Emergency Medical Associates Research Foundation; Duncan L. Cooper, MSc, MRes, West Midlands Health Protection Agency, UK; Urania G. Dafni, Department of Public Health, School of Nursing, University of Athens; Andrew G. Dean, MD, Columbia University International Family AIDS Program; Timothy J. Doyle, MPH, National Center for Public Health Informatics, CDC; Luiz Duczmal, PhD, Univ. Federal de Minas Gerais, Brazil; Vanja Dukic, PhD, University of Chicago; Laura M. Forsberg, Harvard School of Public Health; Kathleen F. Gensheimer, MD, Maine Department of Health and Human Services; Joe F. Gonzalez, Jr., MS, National Center for Health Statistics, CDC; Colin R. Goodall, PhD, AT&T Labs — Research; Ivan J. Gotham, PhD, New York State Health Department; Nancy W. Grady, PhD, Scientific Application International Coorporated; Timothy A. Green, PhD, National Center for HIV, STD, and TB Prevention, CDC; Samuel L. Groseclose, DVM, National Center for HIV, STD, and TB Prevention, CDC; Julia E. Gunn, MPH, Boston Public Health Commission; Richard Heffernan, MPH, New York City Department of Health and Mental Hygiene; Kelly J. Henning, MD, New York City Department of Health and Mental Hygiene; Lan Huang, PhD, National Institutes of Health; Kathy J. Hurt-Mullen, MPH, Montgomery County Department of Health and Human Services; James N. Hyde, MA, Tufts University School of Medicine; Denise Koo, MD, Office of Workforce and Career Development, CDC; Martin Kulldorff, PhD, Harvard Medical School and Harvard Pilgrim Health Care; Andrew B. Lawson, PhD, University of South California; William B. Lober, MD, University of Washington; Joseph S. Lombardo, MS, Johns Hopkins University Applied Physics Laboratory; Haobo Ma, MD, National Center for Public Health Informatics, CDC; J.A. Magnuson, PhD, Oregon Health Services; Kenneth D. Mandl, MD, Harvard Medical School; Nicola M. Marsden-Haug, MPH, Walter Reed Army Institute of Research; Sarah Michalak, Los Alamos National Laboratory Statistical Sciences Group; Ramal Moonesinghe, PhD, National Center for Health Marketing, CDC; Farzad Mostashari, MD, New York City Department of Health and Mental Hygiene; David J. Muscatello, MPH, New South Wales Department of Health, Sydney, Australia; Elena N. Naumova, PhD, Tufts University School of Medicine; Karen L. Olson, PhD, Children's Hospital Boston and Harvard Medical School; Julie A. Pavlin, MD, Walter Reed Army Institute of Research; Rick Picard, PhD, Los Alamos National Laboratory; Daniel A. Pollock, MD, National Center for Infectious Diseases, CDC; Julie A. Rawlings, MPH, Texas Department of State Health Services; Janise Richards, PhD, Office of Workforce and Career Development, CDC; Robert T. Rolfs, MD, Utah Department of Health and CDC; Henry R. Rolka, MS, National Center for Public Health Informatics, CDC; Kyumin Shim, PhD, National Center for Public Health Informatics, CDC; David W. Siegrist, MA, MSM, Potomac Institute for Policy Studies; Daniel M. Sosin, MD, Coordinating Office of Terrorism Preparedness and Emergency Response, CDC; Sotirios Tsiodras, MD, PhD, University of Athens Medical School, Hellenic Center for Infectious Diseases Control; Michael A. Stoto, PhD, RAND; Donna F. Stroup, PhD, Coordinating Center for Health Promotion, CDC; Jerome I. Tokars, MD, National Center for Public Health Informatics, CDC; Fu-Chiang Tsui, PhD, University of Pittsburgh; Sylvan Wallenstein, PhD, Mount Sinai School of Medicine; Garrick L. Wallstrom, PhD, RODS Laboratory, University of Pittsburgh; Don Weiss, MD, New York City Department of Health and Mental Hygiene; Christopher W. Woods, MD, Duke University and Durham VAMC; and Byung-Kwang Yoo, MD, PhD, National Center on Birth Defects and Developmental Disabilities, CDC.

Overview, Policy, and Systems

Federal Role in Early Detection Preparedness Systems

Claire V. Broome
Coordinating Center for Health Information and Service, CDC

Corresponding author: Claire V. Broome, CDC, 1600 Clifton Rd., NE, MS D-68, Atlanta, GA 30333. Telephone: 404-639-7860; Fax: 404-639-7770; E-mail: cvb1@cdc.gov.

Disclosure of relationship: The contributor of this report has disclosed that she has no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

This report discusses CDC's role in the early detection of threats to public health, including linkage of detection to other preparedness functions, and provides an update on recent progress in improving these capabilities. CDC's role has been fivefold: 1) identifying and adopting information system standards; 2) providing funding; 3) defining critical surveillance functionalities; 4) accelerating Internet-based surveillance systems and electronic reporting of laboratory results; and 5) implementing BioSense, a secure, Internet-accessible, national early detection system that includes syndromic surveillance. Ongoing iterative efforts and consultation are needed to ensure future progress.

Introduction

The U.S. public health system is responsible for the prompt detection and investigation of, and response to, threats to the population's health, whether caused by a known organism, a previously uncharacterized disease, or a covert or overt terrorist event. Accomplishing these functions requires a coordinated effort of local, state, and federal public health entities, working with multiple partners, including clinical care delivery, public policy, first responders, the public, and law enforcement. Information systems to support these complex activities by partners in multiple organizations and areas must be interoperable, incorporating information system standards to facilitate sustainable, real-time delivery of important data and to make alerts and information available to the public health partners that verify, investigate, and respond to outbreaks.

Preparedness both requires and benefits from the active collaboration of federal, state, and local public health partners. Data for the early detection functions that are part of these interoperable information systems are obtained 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 data, work and school absenteeism data, veterinary health data, and information from environmental sensing devices). This report discusses CDC's role in early detection, including linkage of detection to other preparedness functions, and provides an update on recent progress.

CDC's Role in Early Detection Systems

CDC has played a key role in identifying and adopting national standards for data and system architecture to achieve the needed interoperable systems. The Public Health Information Network (PHIN) defines standards for technology, data, vocabulary, and information security to enable the consistent exchange of health, disease tracking, and response data among public health partners, protect the security of these data, and 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 (*1*).

CDC also participates actively in key national standards development organizations to ensure that public health needs are considered as these national standards evolve. In 2004, CDC worked with Health Level 7, the leading international standards development organization that addresses clinical and administrative standards for health, to provide input on public health functions that should be considered as part of the recently balloted draft standard for trial use for electronic health records (EHRs). CDC also continues to participate in national consortia (e.g., Connecting for Health) to accelerate use of EHRs by providers (*2*). The more rapidly providers adopt standards-based, interoperable EHRs, the more readily a broader range of clinical data for early detection can be available electronically from more providers in both ambulatory and hospital settings.

Provision of funding for preparedness systems and personnel is also a key federal role. Since Fiscal Year 2002, CDC and the Health Resources and Services Administration (HRSA) have provided $1.1–1.4 billion dollars annually in funding to state and local health departments to enhance preparedness activities. The CDC and HRSA funding guidance to state health departments requires states to use PHIN standards when they

invest preparedness funding in developing or modifying information systems (*3*).

The funding guidance also identifies specific functions that states must support, including enhanced surveillance and early detection capacity. Enhanced surveillance requires strengthening personnel and systems needed for surveillance, as well as strategic decisions about what approach to syndromic surveillance is appropriate for the specific local or state circumstances. This report discusses CDC contributions to this core state and local responsibility, first in the context of disease-specific surveillance and then in the context of systems that include surveillance of syndromes and prediagnostic data sources.

To enhance disease-specific surveillance, CDC is working with partners to implement the surveillance component of PHIN, the National Electronic Disease Surveillance System (NEDSS) (*4*). Disease-specific surveillance is relevant to early detection in multiple ways: 1) it might complement telephone notifications of unusual events; 2) it might detect outbreaks of which public health authorities were otherwise unaware; and 3) it can provide complementary information, and potentially infrastructure, to support syndromic surveillance activities. States can either use NEDSS standards to develop or modify state surveillance systems or implement the NEDSS Base System, a highly configurable software program developed by CDC and its partners to enhance surveillance at the state and local level (*4*). NEDSS incorporates multiple approaches to improve the timeliness and completeness of disease-specific surveillance systems that enable clinicians, laboratories, and local health department investigators to use the Internet to enter data into a database at the health department. This approach makes information on a reported case available at the state or local health department without the delays of data entry or mailing a form. As of April 2005, a total of 27 states* were using surveillance systems that utilize Internet-based data entry.

NEDSS standards also support automatic electronic laboratory result (ELR) reporting from a clinical diagnostic laboratory information system to state and local health departments for positive test results (i.e., those that identify a notifiable disease or condition). Communicable disease surveillance focuses on laboratory reports because a high proportion of notifiable diseases can be identified on the basis of laboratory test results. ELR has been well documented to increase the number of cases reported to public health two- to threefold, as well as to dramatically increase the timeliness of reports (*5*).

As of March 2005, a total of 26 states† were receiving ELR reports for certain conditions (e.g., hepatitis and meningitis) included in communicable disease surveillance.

NEDSS also supports electronic transmission of surveillance data about cases from states to CDC. Although all states currently perform this function through either the National Electronic Telecommunications System for Surveillance (NETSS) or NEDSS, the NEDSS notification format enhances surveillance by providing more complete information about each reported case and sending notifications as soon as the state approves the notification, rather than in a weekly batch.

CDC has developed BioSense, a secure early detection system accessible through the Internet that includes surveillance of syndromes. BioSense enables local and state health departments to view syndromic data relevant to their areas from multiple sentinel national data sources (e.g., Veterans Administration hospitals and military treatment facilities). This application has been described in detail elsewhere (*6*). BioSense accomplishes two key federal roles. First, it provides a single location for a state or local health department to monitor, instead of trying to identify signals from multiple Internet sites that use different data sources and different approaches to analysis and alerting. This approach also combines analysis of syndromic data with specimens collected in the same area through BioWatch, which places environmental air samplers in key locations (*7*). Second, BioSense provides a means of coordinating data requests and changes in analytic approaches with national data sources, which should be more efficient than attempting bilateral approaches between these sources and the 50 states or the thousands of local health departments. Realizing this efficiency requires a process for ongoing collaboration among all participants. BioSense also can function as a platform for evaluating utility of data sources, syndrome categorization, and analytic algorithms, in addition to providing an opportunity for collaborations to address key research questions.

CDC's involvement in syndromic surveillance at state and local health departments is a complex and evolving area. As states and local health departments consider whether to develop or enhance syndromic surveillance capacity within their health departments, what has been learned to date regarding the usefulness (or lack thereof) of various data sources should be incorporated into their plan. Assessing what level of population coverage for that area is already provided in national platforms (e.g., BioSense) is also appropriate. Pragmatic considerations include the availability of information

* Alaska, Arizona, Colorado, Delaware, Florida, Georgia, Idaho, Illinois, Kansas, Kentucky, Louisiana, Michigan, Nebraska, Nevada, New Jersey, New York, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, Vermont, Virginia, and West Virginia.

† Alaska, Arizona, Colorado, Delaware, Florida, Georgia, Hawaii, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Massachusetts, Michigan, Minnesota, Nebraska, New Hampshire, New Jersey, New York, North Dakota, Oklahoma, Oregon, Pennsylvania, Texas, and Wisconsin.

technology resources needed to add and maintain additional data sources and the public health resources needed to investigate inevitable false alarms. Because it includes information from all states and all cities in which BioWatch environmental collectors are in use, BioSense can provide syndromic surveillance information for local health departments that do not have that capacity. The public health resources needed to investigate inevitable false alarms are also a concern with BioSense. However, local jurisdictions can work with CDC to determine levels of sensitivity and specificity that are appropriate for their areas.

CDC is planning to implement provision of data from regional and national data sources to the relevant jurisdiction. The availability of these data, whether from BioSense or from state and local systems, will provide opportunities for integrating surveillance systems; this integration will make increasingly moot the distinction between disease-specific systems and early detection systems that include syndromic or prediagnostic data.

Using PHIN standards for BioSense and for local and state information systems will facilitate interoperability across systems. CDC also is providing technical assistance and specific software tools for use by state, local, and federal partners to accomplish needed cross-cutting functions. For example, the PHIN Messaging System is software for standards-based, secure bi-directional transport of messages between institutions (*8*). Interoperability also requires adherence to detailed national data standards, including defined controlled vocabularies. CDC is facilitating this process with the PHIN Vocabulary Access and Distribution System, which includes Internet-accessible standard reference tables (*9*). Implementing interoperable systems also requires a detailed level of specifications beyond the agreed upon, high-level standard. CDC has worked collaboratively with partners to provide that level of detail (e.g., through message implementation guides [*10*] that specify message format, including data content).

Conclusion

Substantial progress has occurred in the early detection area from the partnership among local, state, and federal public health entities. In 2005, BioSense received a substantial increase in funding, which should accelerate its ability to provide a wider range of data sources; partners are participating in a working group to help define the highest priority data sources. Accelerated progress depends on making preparedness systems interoperable and avoiding an isolated focus on early detection alone. This imperative underscores the complexity of the task, requiring concrete progress at the state and local level in implementing a PHIN-compatible surveillance, alerting, and response infrastructure, as well as accelerated progress by CDC in providing the needed specifications and tools. All partners are aware of the urgency, but also the difficulty, of accomplishing these public health objectives; continued consultation and iterative efforts are needed to ensure further progress. Finally, with the recognition that systems will continue to evolve, plans for evaluation and subsequent modification should be incorporated into the ongoing work.

Acknowledgments

The progress on PHIN, NEDSS, and BioSense described in this manuscript reflects the collaborative work of participants from local, state, and federal organizations who have provided vision, thoughtfulness, commitment, and unstinting hard work. This manuscript also benefited from thoughtful review by and comments from Henry Rolka, National Center for Public Health Informatics, and Robert Pinner, National Center for Infectious Diseases, CDC.

References

1. CDC. Public Health Information Network: functions and specifications, Version 1.2. Atlanta, GA: US Department of Health and Human Services, CDC; 2002. Available at http://www.cdc.gov/phin/architecture/index.htm.
2. The Markle Foundation. Connecting for health 2005. New York, NY: The Markle Foundation; 2005. Available at http://www.connectingforhealth.org.
3. CDC. Continuation guidance for cooperative agreement on public health preparedness and response for bioterrorism—budget year four. Appendix 4: IT specifications and functions, December 18, 2002. Atlanta, GA: US Department of Health and Human Services, CDC; 2002.
4. CDC. Public Health Information Network: software and solutions. Atlanta, GA: US Department of Health and Human Services, CDC; 2003. Available at http://www.cdc.gov/phin/software-solutions/index.html.
5. Effler P, Ching-Lee M, Bogard A, Ieong M, Nekomoto T, Jernigan D. Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods. JAMA 1999;282:1845–50.
6. Bradley C, Rolka H, Loonsk JL. BioSense: implementation of a national early event detection system. In: Syndromic surveillance: reports from a national conference, 2004. MMWR 2005;54(Suppl):11–9.
7. US Department of Homeland Security. BioWatch: early detection, early response. Washington, DC: US Department of Homeland Security; 2005. Available at http://www.milnet.com/wh/DoHS/BioWatchFactSheetFINAL.pdf.
8. CDC. Public Health Information Network messaging system. Atlanta, GA: US Department of Health and Human Services, CDC; 2004. Available at http://www.cdc.gov/phin/software-solutions/phinms/index.html.
9. CDC. Public Health Information Network vocabulary standards and specifications. Atlanta, GA: US Department of Health and Human Services, CDC; 2005. Available at http://www.cdc.gov/phin/vocabulary/index.htm.
10. CDC. Public Health Information Network message implementation guides. Atlanta, GA: US Department of Health and Human Services, CDC; 2003. Available at http://www.cdc.gov/phin/architecture/implementation_guides.

BioSense: Implementation of a National Early Event Detection and Situational Awareness System

Colleen A. Bradley,[1,2] H. Rolka,[1] D. Walker,[1] J. Loonsk[1]
[1]National Center for Public Health Informatics, CDC;
[2]Science Applications International Corporation, Atlanta, Georgia

Corresponding author: Colleen A. Bradley, CDC, 1600 Clifton Rd., NE, MS E-06, Atlanta, GA 30333. Telephone: 404-498-6312; Fax: 404-498-6145; Email: Cbradley1@cdc.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

BioSense is a CDC initiative to support enhanced early detection, quantification, and localization of possible biologic terrorism attacks and other events of public health concern on a national level. The goals of the BioSense initiative are to advance early detection by providing the standards, infrastructure, and data acquisition for near real-time reporting, analytic evaluation and implementation, and early event detection support for state and local public health officials. BioSense collects and analyzes Department of Defense and Department of Veterans Affairs ambulatory clinical diagnoses and procedures and Laboratory Corporation of America laboratory-test orders. The application summarizes and presents analytical results and data visualizations by source, day, and syndrome for each ZIP code, state, and metropolitan area through maps, graphs, and tables. An initial proof of a concept evaluation project was conducted before the system was made available to state and local users in April 2004. User recruitment involved identifying and training BioSense administrators and users from state and local health departments. User support has been an essential component of the implementation and enhancement process. CDC initiated the BioIntelligence Center (BIC) in June 2004 to conduct internal monitoring of BioSense national data daily. BIC staff have supported state and local system monitoring, conducted data anomaly inquiries, and communicated with state and local public health officials. Substantial investments will be made in providing regional, state, and local data for early event detection and situational awareness, test beds for data and algorithm evaluation, detection algorithm development, and data management technologies, while maintaining the focus on state and local public health needs.

The BioSense Initiative

BioSense is a CDC initiative to support enhanced early detection and situational awareness for possible biologic terrorism attacks and other events of public health concern on a national level. It is the primary early event detection component of CDC's Public Health Information Network (*1*). BioSense Initiative goals include the advancement of analytics for pre-diagnostic and diagnostic data; collaboration with state, local, and regional systems to provide data in near-real time; increased sharing of approaches and technology among federal, state, and local levels of public health; and the promotion of national standards and specifications to ensure integration with other public health systems (*2*).

The BioSense software application and the BioIntelligence Center (BIC) are two key components of CDC's BioSense Initiative. The BioSense application is an Internet-based software system for collecting, analyzing, and visualizing data reported to BioSense. Since June 2004, BIC has conducted monitoring and investigation of BioSense national data daily and supports state and local system monitoring and data anomaly investigations. Although the BioSense Initiative involves broader activities in the public health context, this report is primarily focuses on surveillance use of the BioSense application.

The BioSense Application

Overview

The purpose of the BioSense application is to provide early event detection and situational awareness critical for biologic terrorism surveillance and routine public health event management. BioSense uses near-real time reporting of health data, performing analysis and data visualization techniques on diagnostic and pre-diagnostic electronic data sources and providing the results to state and local public health departments for use in detecting and characterizing events of potential public health importance. BioSense summarizes and presents analytical results and data visualizations by source, day, and syndrome for each state and metropolitan area (MRA) through

maps, graphs, and tables. States and MRA jurisdictions are defined by a set of ZIP codes.

Data Sources

BioSense has implemented three national data sources: *Department of Defense (DoD) Military Treatment Facilities,* Department of *Veterans Affairs (VA)* treatment facilities, and Laboratory Corporation of American (LabCorp®) test orders. Approximately 700 DoD and 1,100 VA medical facilities report data to BioSense (*3,4*). LabCorp operates a nationwide network of 31 primary testing locations and more than 1,100 patient service centers (*5*). Data are received and analyzed daily and historical data are available; DoD data have been collected since May 2003; VA, December 2003; and LabCorp, June 2004. Since October 2004, the average number of daily records received for DoD has been 98,000; VA, 151,800; and LabCorp, 137,600. Both DoD and VA provide ambulatory-care data in the form of *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) diagnosis codes and current procedural terminology (CPT®) medical procedure codes. LabCorp provides test orders and ICD-9-CM codes associated with the reason for the orders. All data sources include additional information with each record (e.g., patient age, sex, ZIP code of residence, and facility identifier and ZIP code). Experts from different agencies participated in mapping each data source to 11 syndrome categories: botulism-like, fever, gastrointestinal, hemorrhagic illness, localized cutaneous lesion, lymphadenitis, neurologic, rash, respiratory, severe illness and death, and specific infection (*6,7*).

Statistical Analysis

The calculations for analytics that appear in BioSense are pre-generated daily. Small area regression and testing (SMART) is an adaptation of a generalized linear mixed modeling (GLMM) technique (*8*). The SMART model takes into account multiple comparisons and includes parameters for ZIP code, day of the week, holiday, and day after a holiday. The model also uses sine and cosine terms for seasonal cyclic variation. Parameters are calculated weekly for each date-source-syndrome-ZIP code combination. Predicted values are generated by the model, and the observed data counts for each combination are compared with these predicted values daily under the assumption of a Poisson distribution of events. Significant differences between the observed and predicted values are indicated in the application.

The second analytical technique is an adaptation of a cumulative sum (CUSUM) approach that is used in stand-alone drop-in surveillance (*9*). It is used as a short-term surveillance technique to indicate recent data changes through the comparison of moving averages. Days that have variation higher than two standard deviations from the moving averages are indicated in the application. Because of high variability within the data, individual CUSUM values are calculated for each date-source-syndrome combination at the state or MRA level, rather than for individual ZIP codes.

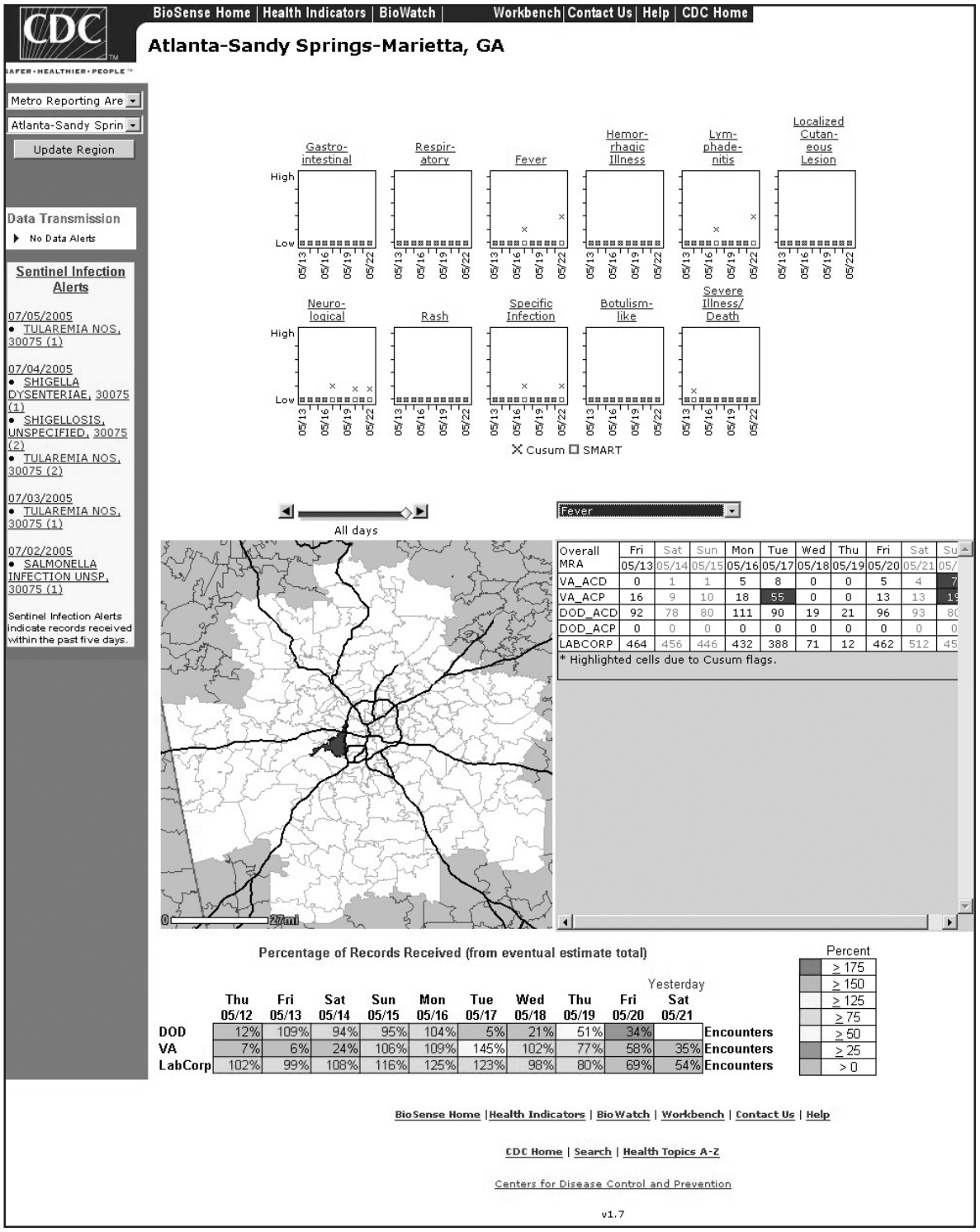
BioSense Home Page

The BioSense application home page provides analytical results to users approved to access data for their jurisdiction (Figure 1). These results indicate data anomalies that might require further investigation. On the left side of the home page, options are listed for changing the region being viewed, data transmission notes regarding the national data sources, and the Sentinel Infection Alerts section, which displays records of ICD-9 codes received that CDC has designated as potential biologic terrorism agents (*10*). In the center of the home page, the syndrome "punch cards" provide the jurisdictional analytical results for each syndrome across all data sources. For this display, elevated SMART or CUSUM scores for a state or MRA are indicated in the punch cards. There are elevated CUSUM scores for several syndromes (fever, lymphadenitis, neurological, and specific infection) (Figure 1). If a particular punch card is selected, detailed analytical information for that syndrome is presented in graphs, maps, and tables. Beneath the punch cards, the jurisdictional map portrays spatial results for analytics, and the time shift feature allows the user to determine temporal patterns. The table to the right of the map provides data source and ZIP code counts, with counts highlighted if associated with an elevated CUSUM or SMART score. The percentage of records received (table at the bottom of the BioSense home page) displays data receipt status, by data source, as a percentage equal to the number of records received out of the number expected to be received. The expected number is based on a historical day of week average. Low percentages indicate that full data delivery might not have occurred, whereas percentages substantially ≥100% indicate that 1) data might have been duplicated in the transmission process, 2) a recent increase in the number of facilities reporting data occurred, or 3) a genuine increase occurred in the number of clinic visits or laboratory tests ordered.

Health Indicator Pages

The Health Indicator pages provide access to the data visualization components for the purpose of evaluating data patterns across sources, geographically, and temporally. The syndrome consolidated line graphs present data for each source on one graph so the user can determine patterns across sources

FIGURE 1. BioSense home page*



* Demonstration data.

for each syndrome (Figure 2). Selecting a syndrome consolidated graph allows the user to view data for that syndrome in graphic, spatial, and tabular presentations. The line graph presentation permits users to compare jurisdictional and national data patterns for each data source (although state and local users do not have access to detailed national data) (Figure 3). The patterns in the VA ambulatory-care diagnosis count data for Atlanta closely mirror the national data (Figure 3). The map presentation illustrates data by ZIP code for each source (Figure 4). The user can view detailed information regarding individual ambulatory-care visits and laboratory-test orders within the tabular presentation (Figure 5). Different data selection and display options provide flexibility when viewing the visualizations, including the option to evaluate data patterns at the ZIP code level, view up to 1 year of data, and examine data for certain age and sex combinations.

BioSense Application Implementation

Initial CDC Evaluation

The BioSense application initial design, data acquisition, and system development involved several challenges. These challenges included the nontraditional nature of the surveillance system, the use of novel data sources, and the lack of pre-defined user requirements. Therefore, an initial "proof of concept" evaluation project was conducted in November 2003 before the system was made available to state and local users. This evaluation involved incorporating known embedded signals into BioSense data to determine if these data aberrations could be distinguished in the application. Such signals were identifiable in the application based on the results of the proof of concept evaluation. Evaluators provided feedback, which resulted in modifications to the initial system design that made it more useful.

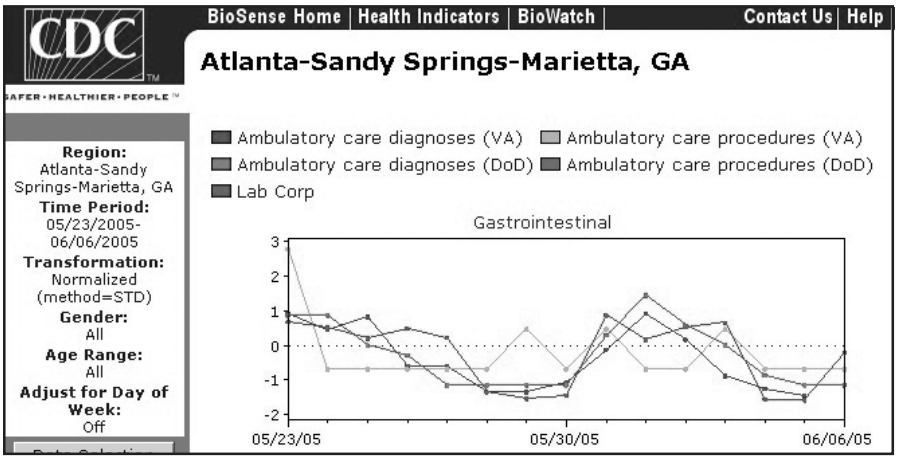
Implementation for State and Local Public Health Officials

Before BioSense was ready to release to state and local public health users in April 2004, state and local public health leaders identified appropriate BioSense administrators within each state and MRA. These administrators were program officials responsible for granting access to the application and appropriate state, metropolitan area, and ZIP code level data and were the points of contact for data aberrations observed in BioSense. State level users were granted access to view BioSense data for their state as well as any metropolitan areas within their state, whereas users at the local level were granted access only to MRA level data. Local users whose jurisdiction did not fall within a BioSense MRA were granted access to state or ZIP code level data, if the state administrator granted approval. BioSense administrators were identified for 49 states and approximately 30 major MRAs, and during April–December 2004, approximately 300 users were approved to access BioSense through the Secure Data Network.

As a critical information system, BioSense leverages security services and protections for key CDC information systems. The services and protections provided by security mechanisms (e.g., Secure Data Network [SDN]) are inclusive of user identity management and authentication as well as authorization controls to ensure appropriate access to BioSense data. In addition, the security measures used by BioSense facilitate compliance with different federal laws and regulations, including: Privacy Act of 1974, Health Insurance Portability and Accountability Act of 1996 (HIPAA), Federal Information Security Management Act of 2002 (FISMA), E-Government Act of 2002, and Homeland Security Presidential Directive (HSPD) -12.

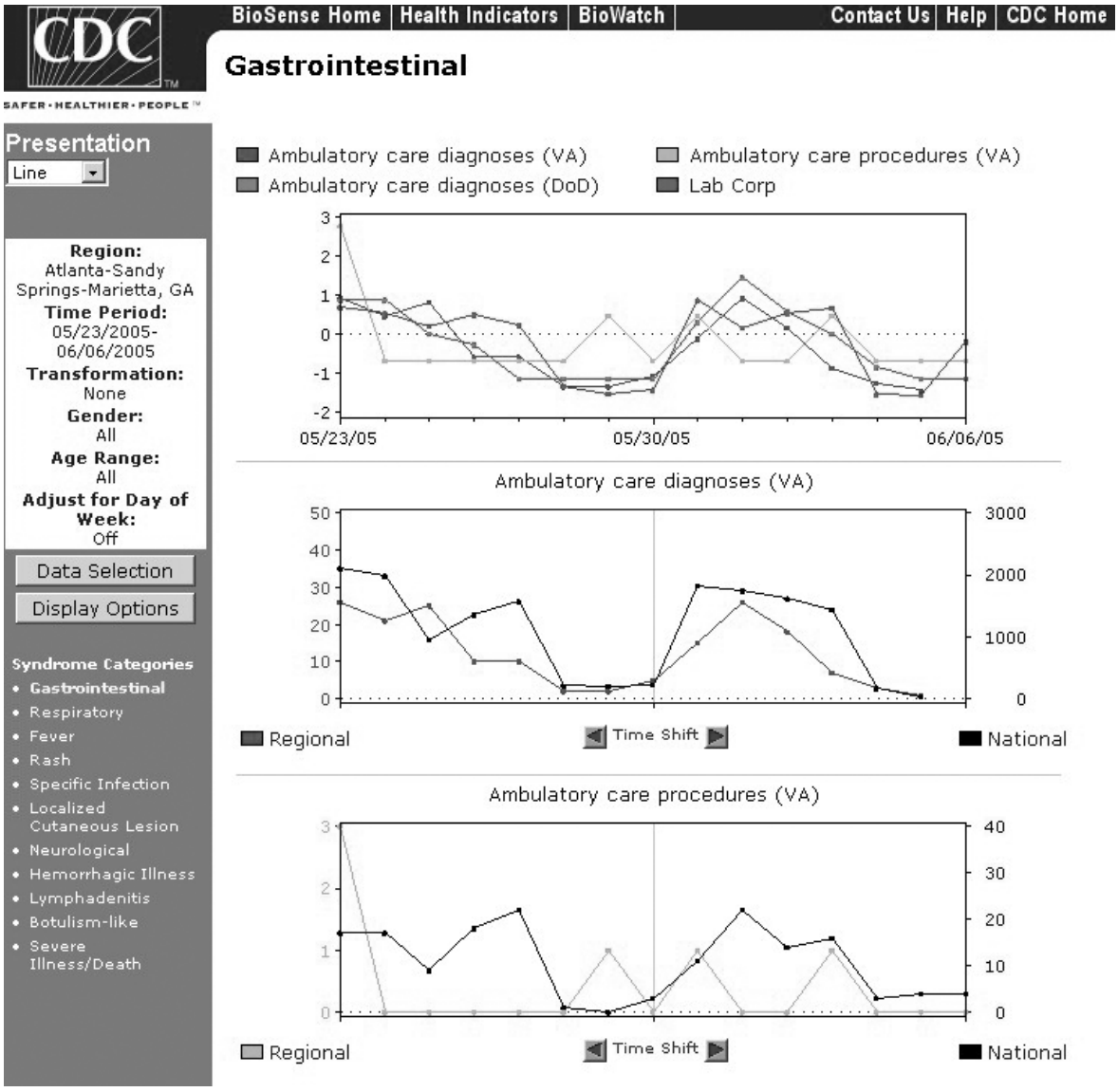
In-depth training regarding the numerous system functionalities, analytics, and data visualizations was initiated. Training sessions were conducted on a bimonthly basis and an interactive telephone bridge and web-based presentation (webinars) was used. The webinars provided a general overview of the BioSense system and its functionalities, the tools users needed to begin monitoring data, a forum for addressing questions, and information regarding

FIGURE 2. Health indicators gastrointestinal syndrome consolidated line graph*



* Demonstration data. Data transformed using standard deviation.

FIGURE 3. Health indicators gastrointestinal syndrome line graph presentation



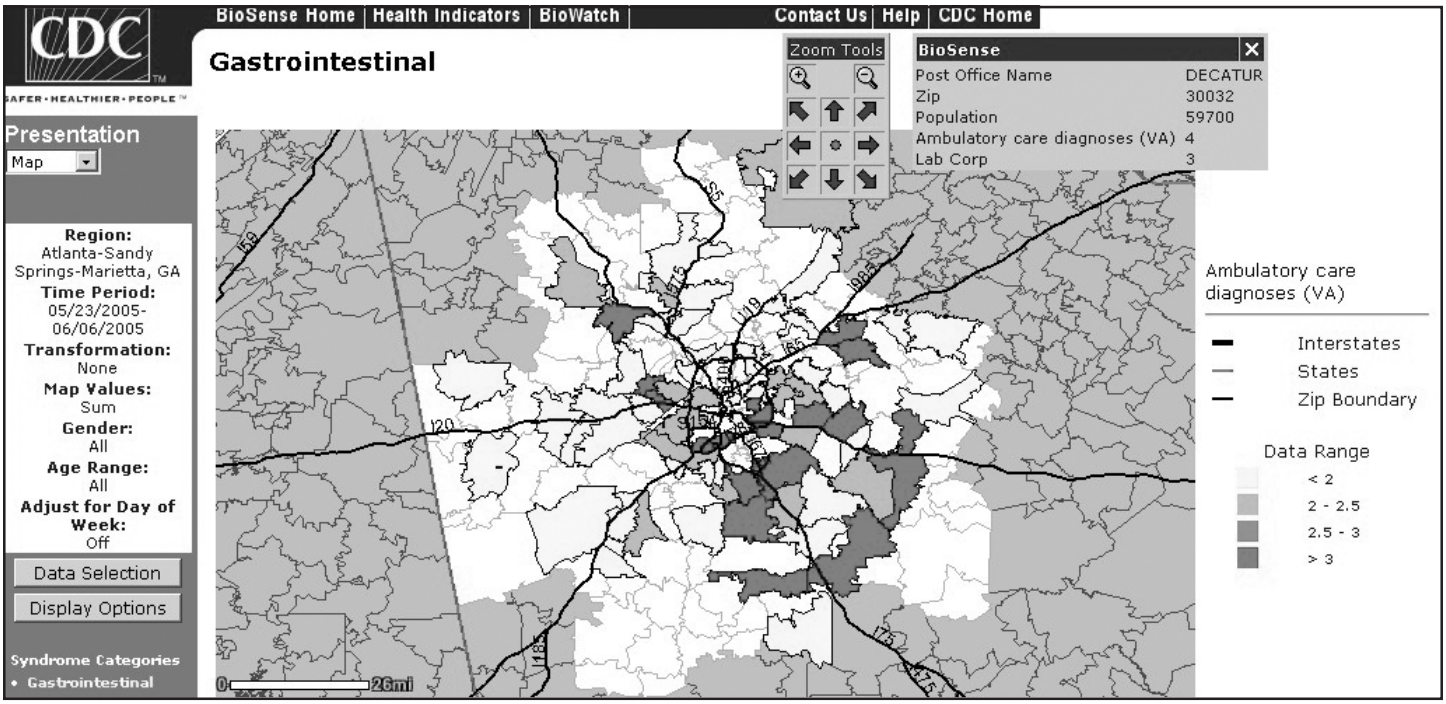
new system functionalities and data sources. During June–December 2004, approximately 250 BioSense users participated in one of 14 training sessions.

In addition to training, user support played two major roles: 1) the provision of requested information and assistance to state and local public health officials and 2) the opportunity to obtain user feedback for incorporation into system development and enhancement. Through the BioSense help email address and the technical help desk phone line, hundreds of questions were routed to the appropriate BioSense team member for response. In an example of a state BioSense administrator request, historical influenza season respiratory syndrome data was provided for establishing baseline patterns and tracking influenza season patterns for 2004–2005. In response to local user requests, customized BioSense MRA jurisdictions have been created. During the G-8 summit in June 2004, the Brunswick, Georgia, Savannah, Georgia, and Coastal Carolina, South Carolina MRA regions were created to improve surveillance during that important convention. Other requested MRA regions that were created included Pierce County, Washington, and Research Triangle, North Carolina.

BioIntelligence Center

CDC initiated BIC in June 2004 to support state and local early event detection capabilities. The BIC functions were to

FIGURE 4. Health indicators gastrointestinal syndrome map presentation of jurisdictional Department of Veterans Affairs ambulatory-care diagnosis counts*



* Demonstration data.

FIGURE 5. Health indicators gastrointestinal syndrome tabular presentation of jurisdictional Department of Veterans Affairs ambulatory-care diagnoses*

Gastrointestinal

Ambulatory care diagnoses (VA) - Gastrointestinal - 06/02/2005

zip=

Rows with the value [ ] in column [Analysis Zip] exclude include clear Help

Analysis Zip	Analysis Zip Key *	Patient Zip	Patient State	Health Facility Visited	Health Facility Home	Patient ID	Age	Gender	ICD-9	ICD-9 Description	Visit ID	Message ID
30338	P	30338	GA	508	508	5080000000197665	77	M	789.00	ABDOMINAL PAIN,UNSPECIF.	K0C60-ATG	da0448f0-b731-4027-8001-00065BF4A933
30315	P	30315	GA	508	619	5080000000067480	51	F	789.00	ABDOMINAL PAIN,UNSPECIF.	K0H8J-ATG	2a9df2c0-b6f6-4027-8001-00065BF676EE
30252	P	30252	GA	508GA	508	5080000000192701	50	F	787.91	DIARRHEA	K0B8R-ATG	5cbaebb0-b731-4027-8001-00065BF4A933
30324	P	30324	GA	508	508	5080000000113276	59	M	789.04	ABDOMINAL PAIN,LEFT LOWER	K07R6-ATG	40243e50-b6f7-4027-8001-00065BF676EE
30236	P	30236	GA	508	508	5080000000055638	67	M	787.02	NAUSEA ALONE	K0D45-ATG	8c0485b0-b732-4027-8001-00065BF4A933
30106	P	30106	GA	508	508	5080000000189744	69	M	787.2	DYSPHAGIA	K0JH4-ATG	bf039a50-b6f6-4027-8001-00065BF676EE
30122	P	30122	GA	508GF	508	5080000000156643	39	M	789.00	ABDOMINAL PAIN,UNSPECIF.	K0CVR-ATG	00dac260-b732-4027-8001-00065BF4A933
30032	P	30032	GA	508	508	5080000000165847	58	M	787.1	HEARTBURN/PYROSIS	K0B84-ATG	4c163580-b731-4027-8001-00065BF4A933
30236	P	30236	GA	508GA	508	5080000000155390	67	M	789.00	ABDOMINAL PAIN,UNSPECIF.	K0BNP-ATG	4e1e1370-b731-4027-8001-00065BF4A933
30680	P	30680	GA	508GH	508	5080000000234737	72	M	787.91	DIARRHEA	K08PB-ATG	cb28af70-b730-4027-8001-00065BF4A933
30033	P	30033	GA	508	626	5080000000155211	83	M	535.50	GASTRITIS/GASTRODUOD. W/O	K0B6V-ATG	342de8a0-b731-4027-8001-00065BF4A933
30045	P	30045	GA	508	612	5080000000218632	48	F	787.02	NAUSEA ALONE	K09W5-ATG	4d6daf30-b731-4027-8001-00065BF4A933
30310	P	30310	GA	508	508	5080000000106545	73	M	789.00	ABDOMINAL PAIN,UNSPECIF.	K0F67-ATG	a658f090-b732-4027-8001-00065BF4A933
30012	P	30012	GA	508	508	5080000000243058	59	M	558.9	GASTROENTERITIS/COLITIS N	K08JV-ATG	79e37aa0-b730-4027-8001-00065BF4A933
30224	P	30224	GA	508	508	5080000000270375	64	M	558.9	GASTROENTERITIS/COLITIS N	K08M1-ATG	4c2bb950-b731-4027-8001-00065BF4A933
30236	P	30236	GA	508	508	5080000000011722	61	M	787.2	DYSPHAGIA	K0NH3-ATG	d61153c0-b6f8-4027-8001-00065BF676EE
30047	P	30047	GA	508	508	5080000000128542	40	M	789.00	ABDOMINAL PAIN,UNSPECIF.	K0BDR-ATG	ddb46f90-b6f8-4027-8001-00065BF676EE
30187	P	30187	GA	508	508	5080000000262659	72	M	787.02	NAUSEA ALONE	K0B46-ATG	804c9a60-b731-4027-8001-00065BF4A933

* Demonstration data.

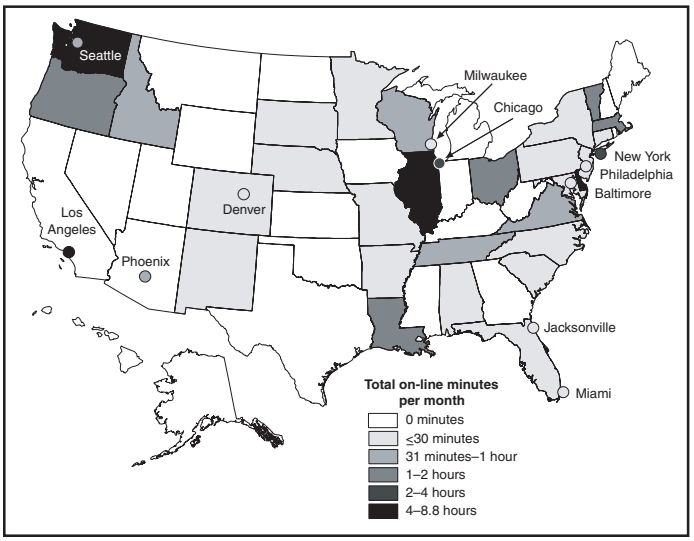
conduct daily monitoring and investigation of BioSense national data, support state and local system monitoring and data anomaly investigations, engage in communication with state and local public health officials in all relevant data anomaly investigations, and develop standard operating procedures for data evaluation. Daily monitoring included investigating, analyzing, and tracking data aberrations (*11*). BIC monitors also played an active role in system troubleshooting and in generating ideas for system enhancements. Examples included improved labeling of data visualizations, increased capability to navigate between pages, and descriptions provided of ICD-9-CM codes associated with individual patient visits.

BIC has actively tracked system use since its inception (Figure 6). Because BioSense monitoring by state and local public health officials varied, the BIC activities ensured that all state and MRA data were being evaluated regularly.

BIC has provided state and local public health officials with surveillance support for major events (e.g., political conventions and presidential debates) and day-to-day evaluation of data to assist users to better understand monitoring methods and data aberrations. Communication with state and local BioSense users regarding data aberrations of potential concern has been a mutual learning process as CDC and state and local users work to characterize the data in BioSense.

User-requested BIC reports provided a summary of jurisdictional data activity within BioSense. Monitors examined the available data and determined if particular situations might require further monitoring or investigation. An example of such a situation included increased syndrome activity for a particular ZIP code or set of adjacent ZIP codes over subsequent days, with epidemiologic patterns related to the disease diagnoses or patient age and sex (*11*). Reports regarding diagnoses of diseases presented in the Sentinel Infection Alerts were communicated to state and local public health officials upon request. Because BIC monitors reviewed data for a set of jurisdictions daily, they were familiar with data activity for that particular area of the country. State and local officials were able to contact BIC monitors and receive responses to questions, feedback, and requests for assistance.

FIGURE 6. BioSense monitoring time, by state and local public health officials, December 2004

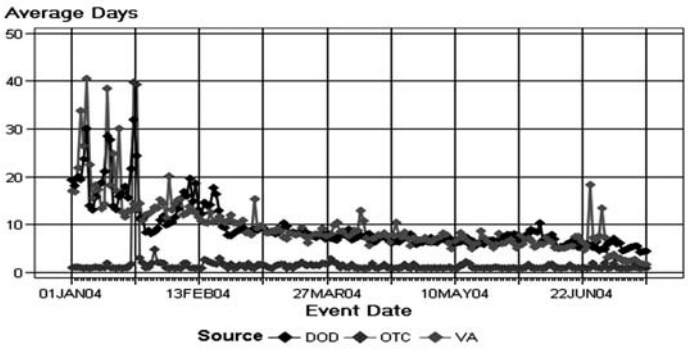


Data-Related Concerns

BIC has encountered multiple data-related concerns since it was established. The data include duplicate or updated records. Appropriate de-duplication methods were used to remove these records that had been previously transmitted. The data were opportunistic and noisy. For the empirical process for analysis, no sampling design existed; therefore, the potential for confounding between reporting volume and event intensity existed.

Data lag between the time of the patient encounter and the time when the related data were accessible in the application was a key issue and varied among data sources and reporting facilities within the same data source. Incomplete data were available for generation of the analytical results and for evaluating data patterns, although data lag has improved. During January–June 2004, the average number of days between the patient clinical encounter and the availability of this information in the BioSense application decreased and stabilized (Figure 7). The percentage of records received table was presented to illustrate that unusual analytical results might occur because of issues related to data transmission and lag rather than because of true health activity (Figure 1). Data lag had implications for inaccuracies in producing analytical results,

FIGURE 7. Average days from patient visit date to date available in the BioSense application, by patient visit date, January–June 2004



and methods to adjust for the potential inaccuracies in the algorithmic outputs will be evaluated.

Data coverage was also an issue. Certain data sources were limited in their representation of the health status of the general population. VA patients were primarily older, whereas DoD included both service personnel and their dependents; both VA and DoD data represented subsets of the population. The data were national; however, geographic coverage varied by data source. The analysis, visualization, and reporting (AVR) ZIP code for the DoD data was the medical facility ZIP code (rather than ZIP code of patient residence) because DoD personnel might be stationed a substantial distance from their home residence. Therefore, the DoD data provided excellent coverage, but only for the ZIP codes where medical facilities were located (Figure 8). LabCorp data were analyzed and presented by the patient residence ZIP code when available. When this ZIP code was not available, either the medical facility or laboratory testing facility ZIP codes were used. LabCorp data coverage was more complete in the eastern United States and in metropolitan areas (Figure 9). The VA analysis, visualization, and reporting ZIP code was the patient residence ZIP code, so coverage was broader than for DoD. A higher number of patient visits usually occurred in the eastern United States and in areas of high population density (Figure 10).

Conclusion and Future Directions

In using BioSense, one lesson learned was that public health officials tended to prefer varied analytic approaches and data sources. BioSense implementation provided insight into the usefulness of analytical techniques in guiding users to potentially important health activity. As a result, frequent training and data aberration monitoring support were required. Detailed information regarding the data loading performance, analytics, data sources, syndrome mappings, application functionalities, and data selection and display options were provided. Communicating with state and local system users as well as continuing training and user support were essential as the system was developed. Characterization of data aberra-

FIGURE 8. Average number of weekly patient visits — Department of Defense ambulatory-care diagnoses and procedure data, January–November 2004

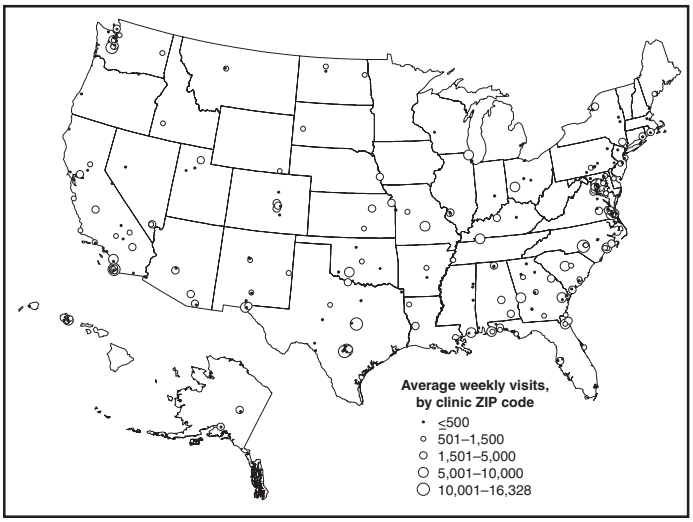
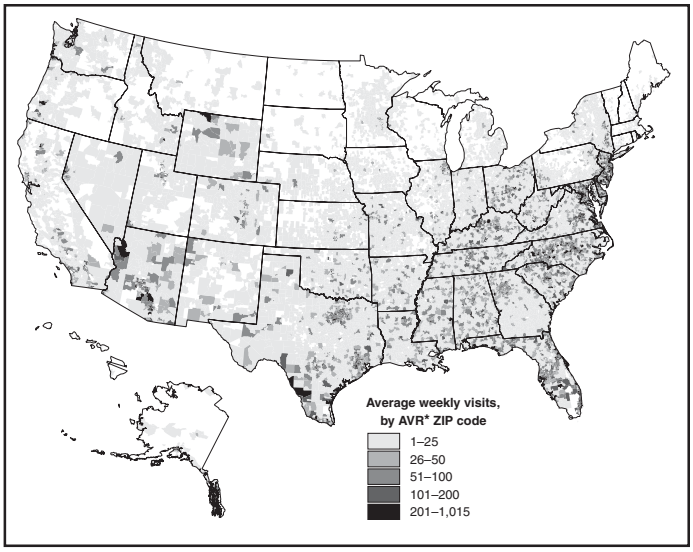
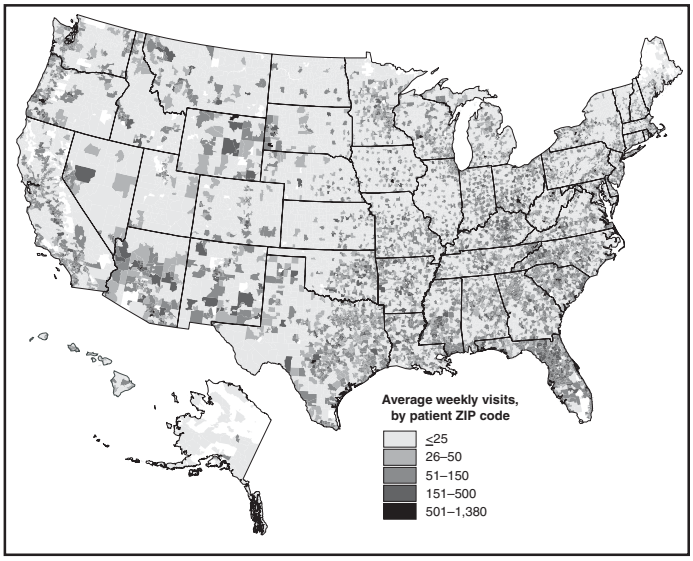


FIGURE 9. Average number of weekly laboratory-test orders — Laboratory Corporation of America laboratory-test order data, January–November 2004



* Analysis, visualization, and reporting.

FIGURE 10. Average number of weekly patient visits — Department of Veterans Affairs ambulatory-care diagnoses and procedure data, January–November 2004



tions has involved collaboration among members of BIC and state and local public health officials and has been a mutual learning experience.

As implementation of the BioSense Initiative continues, substantial investments will be made in 1) regional, state, and local data sources for early event detection and situational awareness, 2) test beds for data source and algorithm evaluation, 3) algorithm development and advancement, and 4) data management technologies. Key considerations in providing data for early event detection and situational awareness will include adherence to Public Health Information Network Preparedness requirements and standards as well as use in local, state, and national public health systems (*12*). The use of test beds for data and algorithm evaluation will allow for rigorous evaluation of the use of investigational data sources, detection algorithms, and approaches. Algorithm development and advancement will include an emphasis on implementation in a standard technical environment and the importance of multiple data source integration for increased sensitivity and specificity. Implementation of a spatio-temporal scanning statistic is being planned (*13*). User-defined analytical thresholds and syndrome categories will also be explored. Data management approaches will include probabilistic, population-based profiles of health events or outbreaks, and text parsing. Natural language processing, data mapping capabilities for chief complaints and other data sources, and pre-analysis data processing and smoothing will also be explored.

The focus on state and local public health needs will be maintained. The framework is being developed for a working group that will foster relationships and communication among local, state, and federal public health officials, and facilitate information exchange regarding data sources, analytics, monitoring practices, and other aspects of early event detection and situational awareness. The creation of additional user-requested customized MRAs and the incorporation of state and local level data will enable state and local users to better monitor the health status of their jurisdictions. BIC will contribute toward developing standard operating procedures for early event detection and anomaly investigation, notification of specific events of concern, and collaboration with state and local public health officials in monitoring BioSense data. The BioSense Initiative will continue to better prepare the public health community for biologic terrorism–related and other surveillance challenges in the 21st century.

Acknowledgments

The authors thank Roseanne English-Bullard and Paul McMurray for their contributions of data analyses and maps addressing data-related issues; Leslie Sokolow, for comments and the map illustrating state and local public health use of BioSense; Shirley Willson, for technical assistance; Nancy Grady, for contributing information regarding data; and Toby Slusher and David King, for information regarding CDC systems security.

References

1. CDC. PHIN preparedness (Draft for discussion). Atlanta, GA: US Department of Health and Human Services, CDC; 2004. Available at http://www.cdc.gov/phin/eed.pdf.
2. Loonsk JW. BioSense—a national initiative for early detection and quantification of public health emergencies. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004; 53(Suppl):53–5.
3. Department of Defense. Department of Defense DMIS ID resource page. Silver Spring: MD: Department of Defense; 2005. Available at http://www.dmisid.com/cgi-dmis/download.
4. Department of Veterans Affairs. Department of Veterans Affairs facilities locator and directory. Department of Veterans Affairs: Cincinnati, OH; 2004. Available at http://www1.va.gov/directory/guide/home.asp?isFlash=1.
5. Laboratory Corporation of America. Burlington, NC: Laboratory Corporation of America; 2005. Available at http://www.labcorp.com/about_labcorp/index.html.
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. Ma H, Rolka H, Mandl KD, et al. Lab order data implementation in BioSense early outbreak detection system. In: Syndromic surveillance: reports from a national conference, 2004. MMWR 2005;54(Suppl):27–30.
8. 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.
9. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response early aberration reporting system (EARS). J Urban Health 2003;80(Suppl):89i–96i.
10. CDC. Bioterrorism agents/diseases, by category. Atlanta, GA: US Department of Health and Human Services, CDC; 2004. Available at http://www.bt.cdc.gov/agent/agentlist-category.asp.
11. Sokolow L, Grady N, Walker D, et al. Deciphering data anomalies in BioSense. In: Syndromic surveillance: reports from a national conference, 2004. MMWR 2005;54(Suppl):133–9.
12. CDC. Public Health Information Network (PHIN). Atlanta, GA: US Department of Health and Human Services, CDC; 2004. Available at http://www.cdc.gov/phin.
13. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. J R Statistical Soc 2001;164:61–72.

Guidelines for Constructing a Statewide Hospital Syndromic Surveillance Network

Zygmunt F. Dembek, K. Carley, 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.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *The process to initiate a comprehensive and inexpensive statewide hospital emergency department-based syndromic surveillance system (HEDSS) in Connecticut can serve as a template for others.*

Objectives: *With limited financial resources, the Connecticut Department of Public Health (CDPH) determined the requirements necessary to establish and routinely conduct hospital emergency department (HED)-based syndromic surveillance.*

Methods: *A statewide survey assessed ability and willingness of Connecticut hospitals to participate in HED syndromic surveillance. The New York City HED-based system protocol and analysis programs, available without financial charge, were modified for use in Connecticut. This system is based on hospitals sending daily standardized files of chief complaint data through encrypted e-mail or an FTP protocol to CDPH with subsequent categorization into syndromes using a SAS program. Anticipating regional surveillance needs during the Republican National Convention in New York City (RNC), CDPH initiated HEDSS in August 2004.*

Results: *Most Connecticut HEDs were willing and able to participate on a voluntary basis. Beginning in July 2004, hospital recruitment began. By the time of the RNC, 11 of 32 Connecticut hospitals participated in HEDSS. Since then, an additional six HEDs have joined.*

Conclusion: *Establishing a voluntary statewide HEDSS was possible using an existing, readily available protocol with minimum financial resources and consensus from a statewide workgroup over a several-month time period.*

Introduction

Many public health jurisdictions have implemented emergency department (ED) chief-complaint-based syndromic surveillance systems as part of their effort at public health preparedness (*1*). In Connecticut, a hospital admissions syndromic surveillance system (HASS) had been implemented (*2*) but not an outpatient emergency department system. After observing the experience of others to determine the utility and appropriate level of investment, during 2004, the decision was made to develop a prospective hospital ED chief complaint-based syndromic surveillance system. The goal was to have readily available data to monitor and prospectively detect unusual disease activity at times of heightened public health alert (e.g., during the Republican National Convention [RNC] in New York City), and to investigate and monitor possible outbreaks, including influenza, as detected through other surveillance initiatives. The criteria for the system included voluntary participation, potential for automated transmission from participating EDs (no sustained labor involved), financial sustainability (if possible, no ongoing software maintenance or licensing costs), simplicity of operation (no sustained outside consultant help needed), and minimum daily personnel costs (5 day per week operation under most circumstances).

Methods

Consensus-Building and Readiness Assessment

A sequential consensus-building process was used to determine interest and readiness of hospital EDs to participate and to select a system with the desired characteristics. To determine interest and readiness, the Connecticut Department of Public Health (CDPH) met bimonthly for 8 months (November 2003–June 2004) with representatives from the state's hospitals to assess ED data collection status and determine the best method of instituting statewide hospital ED-based syndromic surveillance. For biologic terrorism planning

purposes, Connecticut is divided into two hospital regions, from which representatives at the central hospitals in each region (i.e., Centers of Excellence) regularly meet with CDPH to conduct centralized planning for all hospitals. The workgroup consisted of administrative, medical, and information technology experts from the Centers of Excellence and representatives of CDPH and the Connecticut Hospital Association. An interest and technologic readiness survey was developed and a survey of all hospital EDs conducted during February 2004.

During workgroup meetings, it was decided to use an existing ED-based syndromic surveillance system rather than to design one *de novo*. Speakers from four existing systems were invited to present their systems to the workgroup. They included the New York City (NYC) syndromic surveillance system (*3*), the Harvard University National Bioterrorism Syndromic Surveillance Demonstration Program system (*4*), the Department of Defense system (ESSENCE) (*5*) and the University of Pittsburgh system (RODS) (*6*). Subsequently, a group decision was made as to which system to implement in Connecticut.

HEDSS Implementation and Operation

In July 2004, anticipating regional surveillance needs during the RNC, CDPH resurveyed the state's hospitals to determine their ability to participate in a voluntary ED syndromic surveillance effort by the end of August in time for the RNC. Initially, hospitals that began participation before the end of August 2004 were also promised $2,000 to cover relevant IT expenses to develop the ED data file for export and to write the electronic program for its creation and transmission on a daily basis. Later, the compensation period was extended to the end of September 2004. HEDSS was then initiated during late August 2004.

The HEDSS system is based on the New York City ED syndromic surveillance system protocol (*3*). To implement the system, a standardized file format was requested from participating hospitals for daily data transmission of ED visit data (Table). The file contains the hospital of origin; the time and date that each patient was seen in the ED; the age, sex, and home ZIP code of the patients; and chief-complaint information. Each day, participating hospitals send a file containing this information for each patient ED visit for the preceding 24-hour period. All participants use a secure method of data transmission including encrypted password-protected e-mail (two hospitals), daily e-mail notification allowing CDPH access to the hospital's password-protected secure FTP website (one hospital), and posting the file directly to the Connecticut secure FTP website (14 hospitals). Participating hospitals send CDPH unique daily data files and do not access each other's files.

The files are downloaded daily from the state website to an epidemiologist's desktop for merging and analysis. FileZilla shareware, available free from SourceForge.net (http://filezilla.sourceforge.net) is used as a viewer for this purpose. FileZilla is an FTP/SFTP client for Windows and can act as a site manager with the potential to perform multiple simultaneous transfers, Secure Sockets Layer (SSL) and Kerberos Generic Security Services (GSS) authentication/encryption with speed and a simple interface.

Chief-complaint data are transformed into syndromes for analysis using SAS code components developed by the New York City Department of Health and Mental Hygiene (NYCDOHMH). These were available free from NYCDOHMH and were modified for use by available staff. Daily analysis (using SAS macro codes developed by NYCDOHMH) generates tables for syndrome totals; syndrome by hospital; chief-complaint text; a line list; and syndromes indicative of anthrax, botulism, plague, and smallpox.

Results

The initial readiness survey indicated that 97% of the state's 32 acute-care hospitals collected ED chief complaint data, and 90% entered this data into an electronic database, most (97%) within 24 hours. In comparison, 53% entered a patient's discharge diagnosis into a database within 72 hours. A total of 77% of all hospitals could produce electronic extracts of

TABLE. Standardized file format for participation in the hospital emergency department syndromic surveillance (HEDSS) network — Connecticut

Field name	Description	Format	Mask	Length	Start	End
Hospital name	Unique label	Character	xxxx	15	1	15
Date	Arrival date	Character	mm/dd/yyyy	10	16	25
Time	Arrival time	Character	hh:mm	5	26	30
Age	Age in years	Character	999	3	31	33
Sex	'M' 'F' or 'U'	Character		1	34	34
ZIP	Pt home ZIP code	Character	99999	5	35	39
Chief complaint	Free-text Do not truncate	Character		200	40	239

ED chief-complaint data, 86% were willing to share deidentified data, and 76% would participate in an ED syndromic surveillance system. Ultimately, the system that was selected for adaptation in Connecticut was the NYC system. It was chosen for several reasons. First, using it would enable Connecticut to be fully compatible with New York in terms of definition of syndromes and exchange of information. Connecticut is part of the NYC metropolitan area with at least 100,000 commuters to the city daily. In addition, the cost for implementation and maintenance was very low. The system is simple for hospitals to participate in, the SAS program used for analysis was free and readily able to be adapted to Connecticut needs, and NYCDOHMH staff were available to share their experiences with implementation. All implementation work was done with available staff at CDPH.

Despite short notice during the summer, by the time of the RNC, 11 of 32 Connecticut hospitals were participating in HEDSS. A daily surveillance summary was sent to NYCDOHMH and in turn, a NYC metro-area syndromic surveillance report, including status reports from Connecticut, New York, New Jersey, NYC, and New Jersey counties, was provided daily by NYCDOHMH to all participating NYC metropolitan health agencies. Currently, 17 hospitals participate in the HEDSS.

Historically, if nonresidents have a reportable illness and are identified in a Connecticut hospital, this is reported to CDPH and in turn, the case is reported to the neighboring jurisdiction. If a syndromic disease alert occurred in hospitals within adjoining jurisdictions, communication of this information would be made to the pertinent health authorities. CDPH does not report Connecticut-specific patient data to out-of-state health authorities, nor do out-of-state health authorities report their residents' patient-identifiable data to CDPH.

The full utility of the HEDSS is being evaluated. The methodology used to evaluate the HEDSS efficacy includes:

- Comparisons of historic disease syndrome events with HASS. Those syndromes mutually reported (e.g., respiratory illness, fever, and rash illness) from daily HEDSS analysis are compared with biweekly HASS analysis. HEDSS analysis is completed within 48–72 hours after the event occurrence versus weekly for HASS. Therefore, HEDSS presents more recent data trends and, if accurate, widespread disease occurrences should later be reflected in HASS and through community-reported disease.
- Comparisons with other area-wide hospital syndromic disease events. For example, during the RNC, daily syndromic disease trends and significant events were shared by the respective health agencies between Connecticut, NYC, New York state and neighboring New York counties, and New Jersey. Large disease events (e.g., an influenza outbreak) should be reflected in syndromic surveillance reporting of shared illness within neighboring health jurisdictions.
- The state's total burden of ED visits and disease syndromes, to include those hospitals not participating in HEDSS, is unknown. However, the 17 hospitals in HEDDS are located in seven of the eight Connecticut counties and comprise approximately 65% of the number of staffed hospital beds statewide. Six of these hospitals are located in the two counties (Litchfield and Fairfield) adjoining New York and comprise approximately 70% of the staffed beds in those counties. As more hospitals participate in HEDSS, data will reflect whether the State's ED visits and disease syndromes are reflected in similar proportions.
- Disease trend analysis will continue to be conducted for counties or other geographic areas by examining ZIP code information (spatial trends), and for ED visit time (hourly) or successive days (temporal trends).

Discussion

An important factor for success of HEDSS was the commitment for statewide hospital participation. The CDPH-Hospital workgroup helped establish the interest and requirements for syndromic surveillance in Connecticut by mutual agreement through regular meetings over an 8-month period.

Although HEDSS is simple conceptually, generation of compatible datasets from different entities for merging is difficult. The most difficult facet of obtaining successful data transmission by each participating hospital to CDPH was ensuring that comprehensive data specifications were met for individual hospital ED datasets. This effort was successful once the responsible hospital IT staff or hospital IT contractor fully understood HEDSS data requirements and agreed to participate. It was often necessary for the hospital IT staff/contractor to identify a responsible person within their organization to perform the initial data formatting for the hospital's chief-complaint data to ensure that the data were automatically sent daily to CDPH and to have an established contact person should lapses in formatting or data transmission occur.

The full utility of HEDDS is still to be determined. The program is being assessed to monitor the onset and level of activity of influenza, and results are being compared with those obtained from the HASS. No direct patient identification information is collected by HEDDS (i.e., patient name, social security number, or hospital identification number). However, follow-up on individual case reports can be done with a phone call to the hospital ED to obtain case-specific data.

The time of the patient visit to the ED and the patient's age and sex make it easy for the hospital to identify the person. HEDSS might be useful to identify individual cases of disease syndromes of interest during investigation of possible outbreaks.

Every report of adult onset chickenpox is followed by a CDPH case investigation, including cases identified by HEDSS. Pediatric cases of chicken pox also are examined to determine if they have been reported to CDPH through other reporting mechanisms. Given the 48–72 hour time lag from case presentation in the ED to CDPH analysis, complete case-specific information is obtained within 49–96 hours after a patient was seen at the hospital ED.

Conclusion

Establishing a voluntary statewide hospital ED syndromic surveillance system is possible using an existing readily available protocol with minimum financial resources over a several-month time period. Connecticut has a statewide hospital ED surveillance network that includes approximately 50% of all acute care hospitals and 65% of the state's hospital capacity. The success in implementing HEDSS is due, in part, to the consensus building that occurred between CDPH and the state's hospitals. HEDSS can be established by any state, large city, or county that wants to have hospital ED-based syndromic surveillance with a minimum investment and IT support.

Acknowledgments

The authors wish to acknowledge the contributions of the Centers of Excellence ED syndromic surveillance workgroup including Mark Tepping of Bridgeport Hospital, Christopher Cannon, Elaine Chapman, Elaine Forte, and Cindy Pugliese of Yale New Haven Hospital and Sandra Brown, Richard Kubica, Vivian Lane, Patricia Montanaro and Dr. Alan J. Smally of Hartford Hospital.

References

1. Sosin DM. Syndromic surveillance: the case for skillful investment. Biosurety and Bioterrorism: Biodefense Strategy, Practice and Science 2003;1:247–53.
2. Dembek ZF, Carley K, Siniscalchi A, Hadler J. Hospital admissions syndromic surveillance—Connecticut, September 2001–November 2003. MMWR 2004;53:50–2.
3. Heffernan R, Mostashari F, Das D, Karpati A, Kuldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerging Infectious Diseases 2004;10:858–64.
4. Yih WK, Caldwell B, Harmon R, et al. National bioterrorism syndromic surveillance demonstration program. MMWR 2004;53:43–6.
5. 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: 32–42.
6. 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;53:28–31.

Data Sources

Implementation of Laboratory Order Data in BioSense Early Event Detection and Situation Awareness System

Haobo Ma,[1] H. Rolka,[1] K. Mandl,[2] D. Buckeridge,[3,4] A. Fleischauer,[1] J. Pavlin[5]

[1]*CDC, Atlanta, Georgia;* [2]*Children's Hospital Boston, Boston, Massachusetts and Harvard Medical School, Boston, Massachusetts;* [3]*Palo Alto Veterans Health Care, Palo Alto, California;* [4]*Stanford University, Stanford, California;* [5]*Walter Reed Army Institute of Research, Silver Spring, Maryland*

Corresponding author: Haobo Ma, CDC, 1600 Clifton Rd., NE, MS E-06, Atlanta, GA 30333, Telephone: 404-498-6283; Fax: 404-498-3225; E-mail: HMa@cdc.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Laboratory test orders constitute an early outbreak data source. CDC receives laboratory order data in HL7 format from the Laboratory Corporation of America (LabCorp) and plans to use the data in the BioSense Early Event Detection and Situation Awareness System.*

Methods: *These LabCorp data contain information on tests ordered and include the type of test ordered and the* International Classification of Diseases, Ninth Revision, Clinical Modification *(ICD-9-CM)-coded reasons for the order. A consensus panel was formed to group test orders on the basis of expert opinion into eight standard syndrome categories to provide an additional data source for early outbreak detection. A laboratory order taxonomy was developed and used in the mapping consolidation phase. The five main classes of this taxonomy are miscellaneous functional tests, fluid screening tests, system-specific tests, tests for specific infections (by primary manifestation), and tests for specific noninfectious diseases.*

Results: *Summary of numbers of laboratory order codes in each syndrome category are fever (53), respiratory (53), gastrointestinal (27), neurological (35), rash (37), lymphadenitis (20), localized cutaneous lesion (11), and specific infection (63).*

Conclusion: *With the daily use of laboratory order data in BioSense, the actual distribution of laboratory order codes in syndrome groups can be evaluated, allowing modification of the mapping.*

Introduction

Laboratory test orders constitute an early outbreak data source because test ordering occurs earlier in patients' health service than laboratory test results or final diagnoses. Three basis exists for using laboratory order data for early outbreak detection. First, laboratory orders reflect the physician's assessment of the patient's condition and intent to confirm or differentiate diagnoses. Second, when laboratory order data feeds directly from national laboratories are used, it can provide good population coverage. Finally, laboratory orders in large hospitals or national laboratories are transmitted in the standard Health Level 7 (HL7) format electronically, which provides timely data collection and helps ensure data quality. Use of HL7 is increasingly important when constructing automated early outbreak detection systems to handle large data volume. In addition, standard laboratory terminology such as the Logical Observation Identifiers, Names, and Codes (LOINC) has been well adopted for laboratory orders in commercial labs (*1*). Using LOINC codes in laboratory order data improves interoperability and scalability for the detection system.

CDC's BioSense program receives laboratory order data from LabCorp, which operates a nationwide network of 31 primary testing locations, and approximately 1,100 patient service centers. It tests more than 340,000 specimens daily

Methods

Laboratory Order Data

CDC receives a daily transmission of HL7 laboratory orders from LabCorp. Data files are sent to CDC from LabCorp using the Public Health Information Network Messaging System (PHINMS) with digital certificates and encryption. Immediately upon receipt, HL7 messages are mapped to XML format and stored in a data warehouse.

The laboratory order data contain patient age, sex, ZIP code, timestamp information and LabCorp propriety codes and corresponding LOINC codes for each laboratory test ordered. Each laboratory message might have more than one laboratory order placed in one patient encounter. Each laboratory order has one or more subcomponents. For example, a hepa-

titis B laboratory order will include various antibodies and antigens as its subcomponents. In addition, *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) codes are used as reasons for laboratory orders. CDC BioSense does not receive laboratory order results associated with each order.

All data received are anonymous and cannot be traced back to the patient. Use of patient ZIP codes and year of birth are not sufficient for patient identification in the large population areas surveyed.

Consensus Panel

When the laboratory order data are identified, laboratory order codes are mapped into syndrome groups. The objective for the mapping was to provide useful information for early event detection.

A consensus panel of 19 persons was organized with expertise in surveillance, infectious diseases, and medical informatics to perform the mapping. Participants included representatives from the Council of State and Territorial Epidemiologists, the Department of Homeland Security, Harvard Medical School, the Johns Hopkins University Applied Physics Laboratory, Stanford Medical Informatics, the Walter Reed Army Institute of Research, the New York City Department of Hygiene and Mental Health, and CDC. The expertise of domain experts was used to ensure the validity and reliability of the mapping result.

Mapping Procedures

The 11 syndrome groups used in BioSense have been defined by a multiagency working group. The syndromes included fever, respiratory, gastrointestinal, lymphadenitis, specific infection, localized cutaneous lesion, rash, neurologic, botulism-like, hemorrhagic illness, and severe illness or death potentially caused by infectious disease. Detailed syndrome definitions are available on the CDC Emergency Preparedness and Response website (http://www.bt.cdc.gov/surveillance/syndromedef). For the laboratory test mapping, the working group determined that laboratory order codes for botulism-like illness, hemorrhagic illnesses, and severe illness or death do not exist; LabCorp data feeds and these syndromes were not included in the mapping.

Domain experts determined the mapping, which was performed in stages. At each stage, a subset of the panel worked on the mapping problem, and the rest of the panel reviewed the results and arrived at a consensus. Four group members volunteered for the first round of mapping. They mapped each laboratory order to one or more syndromes using a simple mapping form. For example, a laboratory order with the name of influenza A and B antibodies, quantitative can be mapped to respiratory syndrome group. After receiving the four sets of results, three panel members consolidated the results. During these processes, other panel members provided input and advice.

The syndrome mapping of each laboratory order code was combined and represented as numbers. The number in combined results reflected the number of experts' selections in that syndrome group for a specific laboratory order code.

To reduce variations in the mapping results, the following constraints were enforced on the mapping:

- Mapping results are consistent within the same LabCorp laboratory order panel or profile group.
- Mapping results are consistent with LabCorp's online documentation of its use and specimen.
- Laboratory order codes classified in the same disease or pathogen group in a laboratory order code taxonomy are mapped in the same way.

Laboratory Order Code Taxonomy

A taxonomy of laboratory order codes was created. The five top-level classes of this taxonomy are miscellaneous functional tests (e.g., Coombs test and HLA screening), fluid screening tests (e.g., antibody identification and cultures), system-specific tests (e.g., tests for respiratory or gastrointestinal systems), tests for specific infections by primary manifestation (e.g., Cytomegalovirus or Lyme disease), and tests for specific noninfectious diseases (e.g., Lupus test).

This taxonomy provides a mechanism for classifying laboratory order codes systematically and for identifying agreement in mappings. Grouping laboratory orders for the same disease or pathogen enabled easy determination of the convergence and variations in the mapping results.

Implementation

After grouping laboratory order codes and identifying variations, additional constraints were placed on the mapping to facilitate laboratory order surveillance in a production system.

- Two or more experts agreed on syndrome mapping. The consensus panel reviewed and approved the results.
- Mapping was consistent in three ways as specified previously.
- Specific considerations included that the fever syndrome category could be selected only if three experts agreed because fever syndrome is more general than the other syndrome groups; the specific infection syndrome was not selected if a laboratory order code could be mapped to other syndromes.

Data Management

CDC BioSense receives a large amount of laboratory order data from LabCorp facilities. Many HL7 messages are updates for the same laboratory order. Although all received messages are maintained in the data warehouse, the most recent message of the same laboratory order is flagged for use in the BioSense web interface and for data analysis.

Results

The consensus panel's first meeting was in May 2004. In June, the simple mapping table was created in ACCESS, and four experts volunteered to conduct the first round of mapping. Four mapping results were received and consolidated in July. The final result was reported to the working group in September 2004.

Of 309 laboratory order codes, 246 were mapped into eight syndrome categories (Table 1). Some laboratory order codes were mapped to more than one syndrome group. For example, laboratory orders for tuberculosis were mapped as both fever and respiratory syndrome groups. Gastrointestinal syndrome group contains 27 laboratory order codes (Table 2). Respiratory syndrome group contains 53 LabCorp laboratory order codes (Table 3).

In addition, of 309 laboratory order codes, 63 were regarded by the working group as nonapplicable in the syndrome grouping. For example, bleeding and coagulation laboratory orders, nonspecific orders (e.g., complete blood counts), and orders for specific purposes (HLA typing) were determined nonapplicable.

Laboratory order data have been used on the CDC BioSense surveillance production system since November 2004. Both internal CDC BioSense monitors and state and local BioSense monitors can review the data daily for abnormal patterns. The data management team in CDC BioSense has created deduplicated data files in SAS format to facilitate data monitors' queries of records.

TABLE 1. Summary of mapping results

Syndrome group	Order code count*	Frequency (%)†
Fever	53	19.2
Respiratory	53	14.0
Gastrointestinal	27	12.5
Neurological	35	4.1
Rash	37	11.7
Lymphadenitis	20	3.0
Localized cutaneous lesion	11	3.7
Specific infection	63	31.9

* Some laboratory order codes mapped to more than one syndrome groups.
† Frequencies were calculated based on data from 2004.

TABLE 2. Laboratory order in gastrointestinal syndrome group

Laboratory order code	Laboratory order name
180885	*Helicobacter pylori,* Culture
163683	*H. pylori,* IgM, IgG, IgA Ab
162289	*H. pylori,* IgG, Abs
180764	*H. pylori,* Stool Antigen
163170	*H. pylori,* IgA
163204	*H. pylori,* IgM Ab
086181	*Clostridium difficile,* Toxin A
180448	*C. difficile,* Toxin B/Cytotoxin
008045	*C. difficile,* Culture
180141	*Campylobacter,* Culture
180356	Enterohemorrhagic *E. coli,* Culture
187013	Adenovirus (40/41)/Rotavirus
185041	Adenovirus (40/41), Direct EIA
138307	Norovirus, RT-PCR
006866	Rotavirus Detection by EIA
008755	*Cryptosporidium* Smear, Stool
006874	Amebiasis Antibodies
183145	*Cyclospora* Smear, Stool
182204	*Giardia lamblia*, Direct, EIA
188110	*Giardia*, EIA; Ova/Parasites
008144	Stool Culture
182410	Stool Culture, *Yersinia* Only
182311	Stool Culture, *Vibrio* Only
008656	White Blood Cells (WBC), Stool
008607	Occult Blood, Stool
008623	Ova/Parasites Exam, Routine
016766	Fecal Reducing Substances

Discussion

Laboratory order data might provide information that can facilitate early event detection. To reduce the number of categories under surveillance and enable the integration of laboratory orders with other surveillance data sources in BioSense, it is helpful to map laboratory order codes to syndromes before analysis.

To fully cover the received laboratory order codes, mapping was initiated with LabCorp's local codes. LabCorp has mapped most of its proprietary codes to LOINC codes. Results can be converted to the LOINC-based mapping in the future. In the mapping process, domain experts mapped LabCorp codes directly to 11 syndromes. A laboratory order taxonomy was applied in the consolidation phase, which helped reduce inconsistencies. Taxonomy also provides finer granularity as disease classifications for laboratory order codes. Disease-based classification will provide flexibility in surveillance.

This mapping process has several limitations. First, only four domain experts volunteered in the first round mapping. A larger group of experts could help the mapping results converge and reduce inconsistency. Second, some pathogens infect various organ systems. The precise mapping for their laboratory order codes could only be achieved when the specimen collected was also considered. For example, *Haemophilus influenzae* could be a cause of meningitis as well as pneumo-

TABLE 3. Laboratory order in respiratory syndrome group

Laboratory order code	Laboratory order name
182675	AFB Culture/Smear, Broth, Suscep
182402	AFB Culture and Smear, Broth
008466	Organism ID, Mycobacteria
550087	*Mycobacterium tuberculosis* Detection, PCR
188540	*M. tuberculosis*, PCR/Culture
086876	*Mycoplasma pneumoniae,* Culture
138420	*M. pneumoniae,* PCR
163758	*M. pneumoniae,* IgG/IgM Abs
163741	*M. pneumoniae,* IgG Ab
163212	*M. pneumoniae,* IgM Ab
096065	Adenovirus Group Ab, Qn
138164	Adenovirus Detection by PCR
185033	Virus, Adenovirus by DFA
086173	*Bordetella pertussis* Smear, DFA
164384	*B. pertussis,* IgA Ab, Quant
164541	*B. pertussis,* IgG/M/A Ab, Quant
180224	*B. pertussis,* Nasophar Culture
138677	*Bordetella*, Para&Pertussis PCR
161745	*B. pertussis,* IgG Ab, Quant
163030	*B. pertussis,* IgG/IgM Ab, Quant
161752	*B. pertussis,* IgM Ab, Quant
008169	Beta-Hemolytic Strep, A Only
018788	*Streptococcus pneumoniae,* Ag
018804	Beta Strep (Group B) Antigen
008532	*Chlamydia psittaci,* Culture
138263	*C. pneumoniae,* PCR
018762	*Haemophilus influenzae* B Ag
138271	*H. influenzae* B IgG
182295	Influenza A Only by Direct EIA
186023	Viral Culture, Rapid, Influenza
096487	Influenza A/B Ab, Quant
186064	Influenza A & B, Immunoassay
096214	Parainfluenza Virus Antibody
086868	*Legionella* species, Culture
182246	*L. pneumophila,* Ur Ag
085506	*L. pneumophila* by DFA
188227	*L. pneumophila*/Culture
096131	RSV Ab, Quant
185017	Virus, RSV by DFA
014548	RSV by EIA
008342	Upper Respiratory Culture
180810	Lower Respiratory Culture
186015	Viral Culture, Rapid, Respirator
164608	*Brucella abortus,* IgG, EIA
164624	*B. abortus,* IgM, EIA
163709	Diphtheria Antitoxoid Ab
016774	Q Fever Antibodies, IgG
058503	Respiratory Infection Prof A
057877	Respiratory Infection Prof B
058669	Respiratory Infection Prof D
091454	Fungal Antibodies, Quant
164319	*Histoplasma,* Abs, Qn, DID
006742	Tularemia Agglutinins

nia. If the specimen was collected from cerebrospinal fluid, this laboratory order could be mapped to the neurologic syndrome; for a sputum sample, it would be mapped to the respiratory syndrome.

LabCorp messages include both laboratory order codes and ordering reasons in ICD-9-CM codes. The correlation between laboratory orders and their order reasons should be explored. Because ICD-9-CM code mapping has been defined, the correlation might help improve laboratory order mapping.

With daily monitoring of laboratory order data in BioSense, the actual use of laboratory order codes in syndrome groups should be evaluated, allowing modification of mapping. Certain issues in the use of laboratory order data require further study, such as avoiding redundancy when a series of laboratory orders, which can be grouped into the same syndrome group, was ordered on the same patient by more than one physician and defining an event when a patient received laboratory orders during a continuous period.

Conclusion

To implement the laboratory order data, laboratory order codes were mapped to syndrome groups by domain experts *a priori*. In addition, a laboratory order taxonomy was created to facilitate grouping laboratory order codes and identifying inconsistencies. Finally the data management team created deduplicated data for monitors' use.

Acknowledgments

Members of the working group included John Abellera, Council of State and Territorial Epidemiologists; Martha Davis, U.S. Department of Homeland Security; Kenneth Mandl, Harvard Medical School; Steven Babin, Joseph Lombardo, Stephen Jacob, Johns Hopkins University; David Buckeridge, Stanford Medical Informatics; Timothy Endy, Shilpa Hakre, Cara Olsen, Julie Pavlin, Walter Reed Army Institute of Research; Farzad Mostashari, New York City Department of Health and Mental Hygiene; Aaron Fleischauer, Samuel Groseclose, Mary Hamilton, Henry Rolka, Richard Hopkins, Daniel Jernigan, Haobo Ma, CDC. The authors also appreciate LabCorp for their data provision and their consultation availability. The authors also thank Roseanne English and Paul McMurray in the CDC's BioSense BioIntelligence Center for their data management support.

References

1. McDonald CJ, Huff SM, Suico JG, et al. LOINC: a universal standard for identifying laboratory observations: a 5-year update. Clin Chem 2003;49:624–33.

Use of Medicaid Prescription Data for Syndromic Surveillance — New York

Jian-Hua Chen, K. Schmit, H. Chang, E. Herlihy, J. Miller, P. Smith
New York State Department of Health, Albany, New York

Corresponding author: Jian-Hua Chen, New York State Department of Health, Rm. 503, Corning Tower, Empire State Plaza, Albany, NY 12237. Telephone: 518-474-4394; Fax: 518-473-2301; E-mail: jhc04@health.state.ny.us.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *This study explored the utility and value of Medicaid prescription data for statewide syndromic surveillance.*

Methods: *Daily Medicaid claims forms are transmitted to the New York State Syndromic Surveillance Project as summary counts by ZIP Code, age category, sex, and 18 medication groups. The CUSUM statistic is used to analyze the data daily at the county level with a 7–10-day moving average baseline. The system was evaluated following an outbreak of pertussis in a small institutional setting in a rural county by comparing the county's CUSUM signals for prescriptions for macrolide antibiotics with the onset of the outbreak.*

Results: *A case of pertussis was suspected on July 21, 2004, and was reported to the New York State Department of Health on July 22. Treatment and prophylaxis were initiated on July 22. CUSUM analysis flagged a county-wide increase in macrolide antibiotics on the day of treatment/prophylaxis for the first case and contacts in the outbreak. The following week, approximately 300 contacts received prophylaxis (not all of whom were Medicaid clients), resulting in CUSUM signals during the week (July 28 and 29).*

Conclusion: *Medicaid prescription data are routinely collected and readily available for syndromic surveillance. This data source has shown potential value as an indicator of disease activity, as in this case study, in an area where a high concentration of Medicaid recipients reside. However, for the surveillance system to be considered a useful early warning system, additional study is required to determine the best methods for selecting from the automatically generated CUSUM signals those that might be important for public health.*

Introduction

The New York State (NYS) Medicaid Program provides health-care coverage for 34% of the population in New York City and 4%–20% of the populations in the 57 counties in the rest of the state. Enrollees in the Medicaid Program receive approximately 30% of prescriptions filled in NYS. Medicaid activity is concentrated in New York City. However, except for the most sparsely populated areas of upstate NYS, Medicaid enrollees are widely distributed throughout urban, suburban, and rural communities in the rest of the state.

Pharmaceutical data have long been used to monitor the public health (*1*). Recent examples include over-the-counter (OTC) drug sales at pharmacies and groceries (*2–4*) and prescriptions for specific populations, such as members of the military (*5*). This case study demonstrates the use of Medicaid records of pharmaceutical activity for syndromic surveillance.

Methods

The New York State Department of Health (NYSDOH) uses an electronic system (eMedNY) for Medicaid management, with a subsystem for reimbursement of prescriptions and OTC medications. The Syndromic Surveillance Project office receives a daily dataset consisting of the number of prescriptions filled in 18 drug categories (Table 1), aggregated by the ZIP code of residence, age group, and sex of the patient. Every day since March 2003, approximately 20,000 records have been transmitted to the Syndromic Surveillance Project office. Approximately 95% of records are for prescriptions filled during the preceding 24–48 hours.

The CUSUM statistic, a measure of how much higher a current observation is than a reference baseline, compares each day's volume with a short-term historic baseline (7–10-day moving average as used in the CDC model) (*6*). Every day, CUSUM-flagged observations are examined by state staff for indications that they might represent an acute disease out-

TABLE 1. Medicaid prescription drug categories included in New York State's syndromic surveillance system

Analgesics (narcotic)
Analgesics (nonnarcotic)
Antacids
Antiasthmatic medications
Antibiotics
Cephalosporins, first- and second-generation
Cephalosporins, third- and fourth-generation
Fluoroquinolones
Macrolide antibiotics
Penicillin G and ampicillins
Penicillinase-resistant, extended spectrum, and penicillin combinations
Tetracyclines
Antidiarrheal medications
Antihistamines
Cough, cold, and allergy medications
Electrolyte mixtures
Herpes agents
Influenza agents
Systemic and topical nasal products

break. Staff consider the recent and long-term trends of the daily observations in assessing each flagged observation by comparison with previous highest levels or weekly, monthly, or seasonal cycles in prescription activity.

This case study compares county-level CUSUM signals for prescriptions for macrolide antibiotics with timing of events in a pertussis outbreak in a rural county in July 2004. In addition, for purposes of comparison, the number of counties in upstate NYS with and without CUSUM signals for high levels of macrolide antibiotic prescriptions was compared with the number of each county's pertussis cases reported by local health departments to NYSDOH's Confidential Case Reports database.

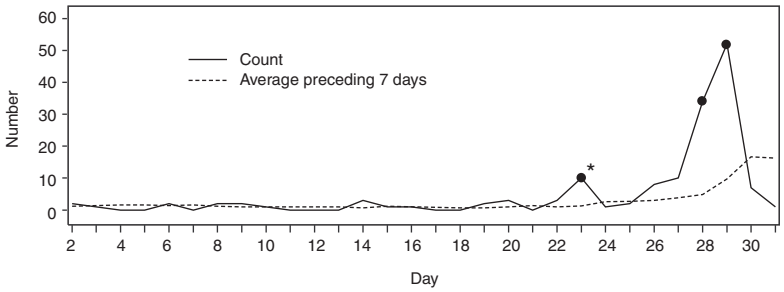
Results

A case of pertussis was reported on July 22, 2004, in a rural county of upstate NYS (county A). The next day, the patient was treated, contacts were administered prophylaxis with macrolide antibiotics, and the Vaccine Preventable Disease Surveillance Officer at NYSDOH was notified. Several persons administered the antibiotics were residents of a group home for mentally retarded adults with Medicaid health-care coverage. These prescriptions generated a macrolide antibiotic CUSUM signal in the Medicaid data for July 23 (n = 10). The following week, additional contacts were administered prophylaxis, and signals were generated on July 28 (n = 34) and 29 (n = 52) (Figure 1 and Table 2). Overall, 13 persons in county A had a polymerase chain reaction that was positive for *Bordetella pertussis*. During July 23–30, according to the county health department, approximately 300 persons received prophylactic treatment with macrolides; the Medicaid surveillance system detected 125 prescriptions (the patients receiving treatment plus approximately one third of all persons receiving prophylaxis treated).

The count of 10 prescriptions on July 23 was high for the short term but not unusually high for the preceding year (Figure 2). Staff did not identify the count of July 23 as unusual for routine follow-up. The CUSUM signals of July 28 and 29 were immediately brought to the attention of epidemiology staff, who confirmed that a number of cases had been identified and that an investigation was ongoing.

During July 2004, a total of 90 CUSUM signals for macrolide antibiotics were generated in 37 of the 57 counties of upstate NYS. Of the 90 signals, 54 (60%) were generated by counts higher than a very small (≤3) daily baseline average. During the same month, outbreaks were identified and reported through normal surveillance in 10 counties. The Medicaid prescription data generated CUSUM signals for five of these 10 counties in July 2004, but those counties that did have signals had a higher disease rate for the month than the counties without signals (mean: 8.9 versus 3.5 cases per 100,000 population; $p<0.0001$), even though the percentage of total population enrolled in Medicaid was lower in the counties with signals (mean: 11.7 versus 15.3; $p<0.0001$). For

FIGURE 1. Daily prescription counts — county A, New York, July 2004

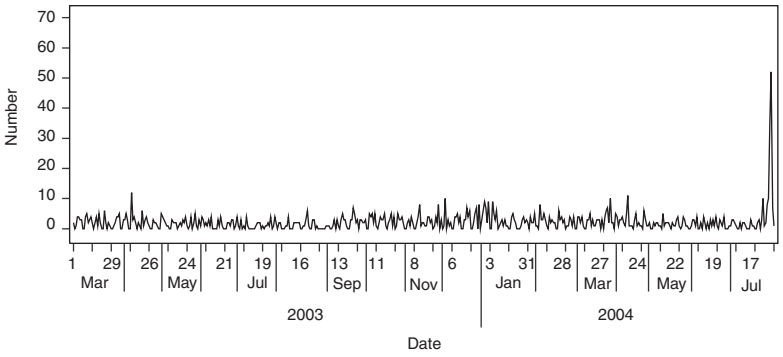


* Count generated a CUSUM signal.

TABLE 2. Sequence of events in pertussis outbreak — county A, New York, July 2004

Date	Event
July 21	Pertussis (whooping cough) suspected in client of adult group home.
July 22	Case is confirmed, and county and state departments of health are notified.
July 23	Case is treated and contacts receive prophylaxis with macrolide antibiotics. Count of Medicaid prescriptions generated a CUSUM flag.
July 26–30	Prophylaxis administered to wider group of contacts in group home and health-care workers.
July 28–29	CUSUM flags generated and recognized as unusually high counts.

FIGURE 2. Daily prescription counts — county A, New York, March 2003–July 2004



detecting county-level pertussis outbreaks during July 2004, the Medicaid prescription data for macrolide antibiotics had a sensitivity of 5/10 = 0.5 and a specificity of 15/47 = 0.32.

Discussion

NYS experience in using Medicaid pharmaceutical data contributes to the experience with early detection surveillance systems. In this case study, the surveillance system identified an outbreak that had already been reported to NYSDOH; therefore, its warning was not early. The value of this example is in demonstrating that a subset of the CUSUM signals can be linked to disease outbreaks (for pertussis in July 2004, sensitivity = 50%). If the same number of Medicaid recipients in county A had received macrolides without a specific cause already having been identified, notification of the increase by NYSDOH to the county local health department (LHD) should have initiated investigations of the cause. The next challenge is to discover the best methods for selecting from the automatically generated signals those that are most likely to be early or real-time indicators of disease activity. This sort of alert could be useful for early detection of disease outbreaks (natural or intentional) not yet recognized by public health agencies. For example, NYSDOH notified LHD of an increase in Medicaid prescriptions in the category "Influenza Agents" in another rural NYS county in early spring 2005. Investigation by LHD resulted in early notification of a small outbreak.

Another benefit of Medicaid pharmaceutical data is the availability of the ZIP code of the patient's residence. When county-wide increases in prescriptions are noted, the distribution by ZIP code of recent prescription claims might help in identifying an outbreak. However, in the pertussis outbreak in county A, the residences of persons receiving prophylaxis were spread out over a wide area, and no clustering was found.

A limitation of the Medicaid prescription CUSUM analyses is the generation of a large number of false-positive signals for disease outbreaks (in July 2004, 32 signals for increased prescriptions for macrolide antibiotics were not linked to any known outbreaks). Furthermore, drugs for less specific uses might generate positive signals without being linked to increased incidence of a particular disease. Users of the CUSUM analyses need to determine the best methods for selecting from the CUSUM signals those that might be important for public health.

Evaluations of medication surveillance document a wide range of conclusions on the usefulness of the data, depending on such factors as study population characteristics, type of medication, and disease outcome. For example, a year-long comparison of prescriptions for antidepressants and anxiolytics with matched outpatient records from a military health-care system serving 4.5 million persons reported a sensitivity of 0.76 and a specificity of 0.94 for the prescription data's performance as an indicator of diagnosed disorders (*5*); these measures are much higher than those reported here (sensitivity = 0.5; specificity = 0.32). However, a study of OTC medications among a large population (in six urban areas in three states) reported a high correlation between OTC electrolyte sales and hospitalization of children aged ≤5 years with various respiratory and diarrheal diseases (*4*). In comparison, use of sales volumes of OTC cough, cold, and antidiarrheal medications in spatial analyses in New York City has produced an average of two spatial signals per month for each syndrome, but none has led to early detection of a localized outbreak (*3*).

Among the possible alternatives to relying only on the daily CUSUM analyses, the following are being evaluated: aggregating county-level counts by week for comparison with other, more traditional surveillance methods (e.g., weekly Medicaid prescriptions for influenza agents versus weekly reports from sentinel physicians of percent of patients seen with influenza-like illness); use of the spatiotemporal scan statistic at the ZIP code level to determine if clusters of elevated Medicaid prescriptions are linked to disease occurrence (e.g., Medicaid prescriptions for electrolyte mixtures indicate an increase in incidence of reported shigellosis cases in children); and correlation of statewide Medicaid prescriptions with the timing of seasonal increases in infectious disease (e.g., the increase in Medicaid prescriptions for influenza agents and the increase in laboratory-confirmed influenza across the state). Development of any analytic methods that identify prescription increases linked to actual disease activity on the basis of these studies will have to be evaluated for their usefulness to public

health surveillance. Examples of useful early or real-time detection of disease include indications of an upsurge in influenza-like illness in counties without reporting sentinel physicians; detection of a clustering of increased prescriptions in ZIP codes in adjacent counties when an increase in disease is not noticeable at the county level; and confirmation of the absence of unusual prescription activity when the possibility of a biologic terrorism threat is being investigated.

Conclusion

Medicaid prescription data are routinely collected and can be readily available for syndromic surveillance. This data source has potential value as an indicator of disease activity. Certain features of this outbreak were conducive to its detection through Medicaid prescription data. First, the majority of cases and a substantial proportion of contacts receiving prophylaxis were enrolled in Medicaid. Second, the number of potential contacts was quite large, given the institutional setting of the outbreak, resulting in many prophylactic treatments that generated a CUSUM signal. Finally, pertussis treatment and prophylaxis are usually done with macrolide antibiotics, although this class of drug also has other uses.

References

1. Welliver RC, Cheny JD, Boyer KM, et al. Sales of nonprescription cold remedies: a unique method of influenza surveillance. Pediatr Res 1979;13:1015–7.
2. Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proc NAS 2002;99:5237–40.
3. CDC. New York City syndromic surveillance systems. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:25–7.
4. 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.
5. Pavlin JA, Murdock P, Elbert E, et al. Conducting population behavioral health surveillance by using automated diagnostic and pharmacy data systems. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:166–72.
6. Hutwagner L, Thompson W, Seeman M, et al. The bioterrorism preparedness and response early aberration reporting system (EARS). J Urban Hlth 2003;80:89–96.

Poison Control Center–Based Syndromic Surveillance for Foodborne Illness

Mary P. Derby,[1] J. McNally,[1] J. Ranger-Moore,[1] L. Hulette,[2] R. Villar,[1] T. Hysong,[1] E. MacNeill,[2] M. Lebowitz,[1] J. Burgess[1]
[1]*The University of Arizona, Tucson, Arizona;* [2]*Pima County Health Department, Tucson, Arizona*

Corresponding author: Mary P. Derby, Environmental and Community Health, Mel and Enid Zuckerman College of Public Health, The University of Arizona, 1435 North Fremont, Tucson, AZ 85719. Telephone: 520-620-6936; Fax: 520-882-5014; E-mail: mpderby@email.arizona.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Objectives: *This retrospective study evaluated the usefulness of a poison control center (PCC) data collection system in Tucson, Arizona for early detection of foodborne disease outbreaks.*

Methods: *A search of a PCC database identified callers with gastrointestinal symptoms attributable to suspected foodborne illnesses whose calls were received during January 1–March 31, 2000. For each foodborne illness–related call, PCC coding was compared with a predefined diarrhea/gastroenteritis syndrome. PCC calls also were evaluated by using ZIP code, age, sex, and date of symptom onset to determine if call classifications matched any laboratory-confirmed cases reported to a county health department.*

Results: *An independent review generally agreed with the PCC's classification of calls. When calls and cases were compared, only one potential match was identified.*

Conclusion: *Although confirmatory diagnostic information was not available, PCC calls were not duplicative of cases evaluated by the county health department, which suggests that they represent two independent data sets. PCC data might provide a useful addition to surveillance data reported to public health agencies for the early detection of foodborne disease outbreaks. These results will now be used to develop collaborative prospective surveillance systems.*

Introduction

Increased concern for public safety and the risk of emerging infectious diseases has prompted interest in the development and implementation of syndromic surveillance systems. The aim of these systems is to enhance traditional public health surveillance systems by collecting and analyzing real-time (i.e., instantaneous) or near real-time data on symptoms from nontraditional sources (e.g., pharmaceutical sales or school absentee records) before laboratory confirmation of an illness is received (*1–3*). By focusing on trends or unexpected patterns in data, syndromic surveillance systems might assist public health authorities in detecting outbreaks more rapidly than can be performed by using traditional public health surveillance systems (*4,5*). Public health departments are using syndromic surveillance systems to monitor diarrheal disease or influenza-like symptoms (through pharmaceutical sales data) (*6,7*) and influenza-like illness (through emergency medical services dispatch records) (*8*) and to detect outbreaks of illnesses that might be infectious or related to a biologic terror attack (through data on emergency department [ED] and urgent care visits) (*9,10*).

Public health departments are responsible for detecting, managing, and preventing foodborne disease outbreaks (FBDOs). Foodborne diseases account annually for an estimated 76 million illnesses, 325,000 hospitalizations, and 5,000 deaths in the United States (*11,12*). Traditional public health FBDO surveillance systems collect data when ill persons seek medical care. However, only an estimated 8% of persons with gastroenteritis symptoms seek medical care, and only a minority of those who do seek care undergo diagnostic testing necessary to determine an etiology for their illness (*13,14*). In addition, substantial delays can occur while awaiting laboratory results. Because PCCs operate 24 hours a day, collect real-time symptom data from callers who seek treatment advice, and provide referral assistance before callers seek medical care, forming partnerships with PCCs might enhance public health surveillance systems. A recent Institute of Medicine report underscored the need to integrate PCC networks with the public health system to improve surveillance systems (*15*).

Investigation of foodborne illnesses in Pima County (2000 population: 843,746), Arizona, is primarily the responsibility of the Pima County Health Department (PCHD), which operates a traditional public health FBDO surveillance

system. All suspected and laboratory-confirmed cases of foodborne illnesses reported to PCHD by health-care providers and all restaurant complaint calls received by PCHD are investigated. The Arizona Poison and Drug Information Center (APDIC) provides case-specific treatment and referral information to callers who complain of foodborne illness. APDIC evaluates complaints received independently of PCHD. The goals of this study were to evaluate the ability of APDIC's data collection system to provide early detection of FBDOs and to recommend ways to improve FBDO surveillance efforts by integrating the two surveillance systems.

Methods

PCC Data Collection

APDIC is one of two PCCs operating in Arizona, serving approximately 2 million persons statewide (excluding residents of Maricopa County, who call the Banner PCC). Staffed by trained pharmacists and certified by the American Association of Poison Control Centers, APDIC receives 180–250 calls daily. In 2000, of 68,433 calls received, approximately 31,000 involved exposures, of which 783 (2.5%) involved suspected foodborne illness or an illness related to food ingestion. An additional 38,433 information (nonexposure) calls were received, of which 202 (0.5%) were related to questions about foodborne illness or food products. When a call is received, a PCC specialist records the caller's suspected exposure (e.g., food poisoning) and provides consultation on the basis of the caller's suspected exposure, age, underlying medical history, and acuity and duration of reported symptoms. The PCC specialist also enters call data, including an assessment (i.e., exposure reason), into an electronic database (Toxicall,® Computer Automation Systems, Inc., Aurora, Colorado). For all exposure-related calls, the PCC specialist enters the date and time the call was received; the caller's sex, age, and ZIP code; type of call (exposure or information); acuity of symptoms; reason for exposure (as assessed by the PCC specialist); exposure site, time, and route; clinical effects, therapies, and medical outcome; and any multiple contacts. Because a caller might report more than one ill person within a household, the PCC specialist also assigns a case-record number and completes a medical history and disposition for each contact identified. To identify multiple contacts, the PCC specialist cross-references all case-record numbers in each caller's and contact's case record. If the call is anticipated to require PCC follow-up, the caller's name and telephone number are also recorded. The PCC specialist may elect to record additional call information in the narrative section of the database. To ensure quality control, every tenth record received daily is reviewed for accuracy of data entry.

Call Confirmation and Comparison to Syndrome Definition

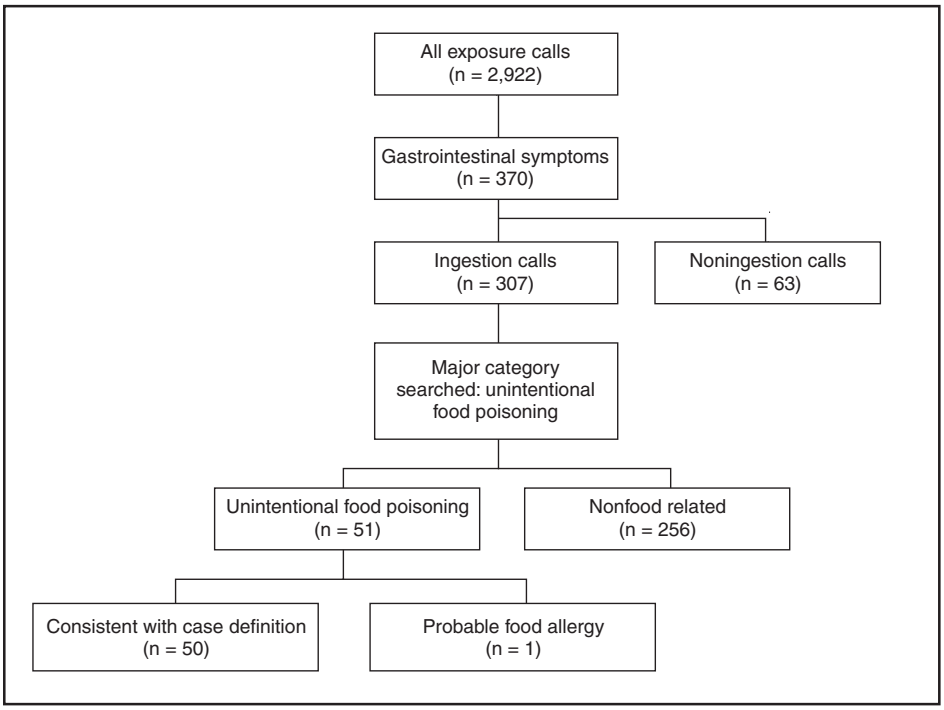
All calls to APDIC concerning human exposure received during January 1–March 31, 2000, were reviewed retrospectively. Three factors were considered in constructing the syndrome definition: 1) APDIC lacks a uniform process for coding suspected foodborne illnesses; 2) the PCC specialist codes calls according to how callers report exposure history (i.e., chief complaint); and 3) multiple syndromic definitions exist for diarrhea/gastroenteritis. The syndrome definition used was modified from that used by the Arizona Department of Health Services for drop-in syndromic surveillance conducted during the 2001 World Series in Arizona (i.e., vomiting, abdominal pain, or any other gastrointestinal [GI] distress) (*16*). Foodborne illness was defined broadly as a syndrome or associated symptoms recognized by a caller who reported ingestion of suspected contaminated food and any of the following GI complaints: abdominal pain, nausea, vomiting, or diarrhea. The selected syndrome definition used in this study included symptoms associated with, but not limited to, foodborne illnesses. Callers with GI symptoms associated with an underlying medical condition (e.g., irritable bowel syndrome, reaction to medication, or chemical exposure not related to ingestion of contaminated food) were excluded.

A nurse reviewed call data records from APDIC's electronic database by following a protocol that identified all possible calls with suspected foodborne illness, reviewed each identified caller's record, and abstracted preselected variables from each record. Within the protocol used, the search process (Figure) involved a review of all Pima County calls by a nurse to determine if coding of calls by PCC specialists conformed to the syndrome definition. Standardized variables were collected and entered into a database for analysis. Information collected included each caller's age, sex, and ZIP code; symptoms; presumed exposure history; exposure site, time, and route; multiple contact(s); referral recommendation(s); and APDIC coding of the caller's suspected diagnosis. Sensitivity, specificity, and positive predictive value (PPV) were calculated on the basis of APDIC coding of callers' suspected foodborne illness compared with the nurse reviewer's assessment of APDIC call data, using the syndrome definition.

Call Confirmation and Comparison to Laboratory-Confirmed Cases

PCHD provided data on cases of laboratory-confirmed illnesses during January 1–March 31, 2000. Variables collected on laboratory-confirmed illnesses included the caller's sex, age, and ZIP code; the date a laboratory confirmed the results; the

FIGURE. Search process used to abstract foodborne illness complaint calls and outcomes from the Arizona Poison and Drug Information Center database — Pima County, Arizona, January 1–March 31, 2000



date of onset of symptoms; and the date PCHD was notified of the confirmed report. To determine whether APDIC foodborne illness complaint calls also were reported to and evaluated by PCHD, an attempt was made to match calls to PCHD laboratory-confirmed cases by using date of onset of symptoms, ZIP code, age, and sex. Onset of symptoms was considered a critical variable because it most likely represented the date a person would have called APDIC. However, for 49% of PCHD cases, the onset of symptoms date was unavailable. The unavailability rate varied (range: 25%–62%) for each microorganism identified through laboratory confirmation (Table 1). For those laboratory-confirmed PCHD cases for which the date of the patient's onset of symptoms was unavailable, an onset date was estimated for each type of foodborne illness (i.e., campylobacter, giardiasis, hepatitis A, shigellosis, and salmonellosis) by using the average interval from date of onset of symptoms to the laboratory-confirmed date for cases for which the patient's symptom onset date was known. The average interval, ±3 days, was used to look for matches to APDIC calls by ZIP code, age, and sex. Differences in age, sex, and incubation period were evaluated by using chi-square analyses. Data were analyzed by using SPSS (Chicago, Illinois) and Stata 8.0 (College Station, Texas).

The study analysis focused on Pima County callers to APDIC and laboratory-confirmed PCHD cases. This study was conducted in accordance with the Health Insurance Portability and Accountability Act, which requires that a minimum of identifiable data be used to ensure patient privacy, and data that would identify callers and cases were not collected. This study underwent human subjects review and was approved by The University of Arizona Institutional Review Board.

Results

APDIC Call Description

During January 1–March 31, 2000, APDIC received 6,768 calls reporting human exposure; 2,922 (43.2%) were from Pima County residents. The database search process and outcome were limited to Pima County callers (Figure). Of 370 calls reporting human exposure with recorded GI symptoms,

TABLE 1. Description of Pima County Health Department laboratory-confirmed foodborne illnesses — Pima County, Arizona, January 1–March 31, 2000

Laboratory-confirmed foodborne illness	No. of confirmed cases	No. of confirmed cases with symptom onset date	Average interval between symptom onset date to laboratory	Incubation period*
Campylobacteriosis	21	12	6	2–5 days
Giardiasis	4	3	12	3–25 days
Hepatitis A	22	11	6	15–50 days
Shigellosis	9	5	8	12–96 hrs
Salmonellosis	21	8	5	12–36 hrs
Total	**77**	**39**		

* **SOURCE:** Heymann D, ed. Control of communicable diseases manual. 18th ed. Washington, DC: American Public Health Association; 2004.

307 (83%) were ingestion related; 51 (14%) calls were recorded as unintentional food poisoning, of which 50 were consistent with the syndrome definition. Twenty-nine (58%) persons whose reported symptoms were consistent with the syndrome definition called APDIC ≤24 hours of suspected reported exposure, and seven (14%) persons called >24 hours after suspected reported exposure. For 14 (28%) callers, exposure history was undetermined. Sixteen (32%) callers reported an exposure outside the home, and 19 (38%) involved more than one person. Only two (4%) callers were referred to a health-care facility or ED for immediate medical evaluation. Adults accounted for the highest number of symptomatic calls, with 18 (36%) known to be aged ≥25–64 years. Six (12%) calls were received from adults reporting illness among persons aged 15–24 years, six (12%) reporting illness among children aged 5–14 years, and three (6%) reporting illness among children aged <5 years (Table 2).

Sensitivity, Specificity, and PPV

Sensitivity, specificity, and PPV were calculated for APDIC's coding of each call, and compared with the nurse reviewer's assessment of the call data (Table 3). Of the 256 calls involving GI symptoms and ingestion but coded by APDIC as nonfood-related, eight (3%) likely represented potential cases of foodborne illness. Including these eight calls, sensitivity of the search strategy and APDIC coding for foodborne illness was 86% (50/58). Only one false-positive call was identified, which yielded a specificity of 99.6% (248/249). The symptoms and exposure reported from this caller were more consistent with a probable food allergy. Of 51 calls identified by the search strategy, 50 (98%) were consistent with the syndrome definition, providing a PPV of 98% (50/51).

TABLE 2. Number and percentage of Arizona Poison and Drug Information Center (APDIC) callers whose symptoms were consistent with the syndrome definition for foodborne illness and of persons with Pima County Health Department (PCHD) laboratory-confirmed cases of foodborne illness, by age and sex — Pima County, Arizona, January 1–March 31, 2000

Characteristic	APDIC callers No.	(%)	PCHD laboratory-confirmed cases No.	(%)
Age group (yrs)				
<5	3	(6)	17	(22)
5–14	6	(12)	14	(18)
15–24	6	(12)	10	(13)
25–64	18	(36)	25	(33)
≥65	3	(6)	8	(10)
Other*	14	(28)	3	(4)
Sex				
Female	28	(56)	45	(58)
Male	22	(44)	32	(42)
Total	**50**	**(100)**	**77**	**(100)**

* The exact age of 14 callers (28%) was unknown because of overlapping age categories used in the Toxicall® database, which listed 13 persons as age ≥20 years and one person as age ≥60 years.

TABLE 3. Number of calls coded as unintentional food poisoning* by Arizona Poison and Drug Information Center (APDIC) staff, compared with nurse reviewer's assessment of call data — Pima County, Arizona, January 1–March 31, 2000

APDIC staff coded as unintentional food poisoning	Consistent with syndrome definition		Total
	Yes	No	
Yes	50	1	**51**
No	8†	248	**256**
Total	**58**	**249**	**307**

* With gastrointestinal symptoms and ingestion.
† Five calls were recorded as "unintentional general" (i.e., unintentional exposure) yet were consistent with the syndrome definition; two calls were consistent with the syndrome definition, although symptoms might not have been caused by the postulated beverage; and one call recorded as "unintentional occupational" (unintentional workplace exposure or injury) was consistent with a foodborne illness.

APDIC Calls Compared to PCHD-Confirmed Cases

A total of 77 laboratory-confirmed cases were reported to PCHD. Only one of the APDIC calls was a potential match to a confirmed PCHD case. For this potential match, both the APDIC call and PCHD case had the same date of symptom onset, date of call, ZIP code, and sex; however, the caller's age was different by 1 year. No significant differences in age or sex were noted between callers and cases in the APDIC and PCHD databases. However, the incubation period for the 77 laboratory-confirmed cases (≥2 days for 47 [61%] and <24 hours for 30 [39%] cases) was statistically different ($p<0.001$) from the incubation period for the APDIC calls (<24 hours for 29 [58%] calls).

Discussion

PCCs provide a new source of real-time data that might help improve surveillance for FBDOs, which are characterized by illness attributable to a common food or water source among two or more persons (*12*). During the 3-month study period, APDIC received 51 calls regarding foodborne illness, of which 50 were consistent with the diarrhea/gastroenteritis syndrome definition used. Although the majority of calls represented single occurrences of illness, 19 (38%) calls involved more than one person and thus might have been representative of outbreaks. Only one potential match was identified as a PCHD laboratory-confirmed illness, indicating that the APDIC data collection system is not duplicative of cases

reported to and evaluated by PCHD during the same period. Although clinical validation studies are needed to determine the exact etiology of these illnesses, APDIC might help PCHD identify FBDOs earlier through detection of new cases of foodborne illness.

Although APDIC has multiple coding categories that PCC specialists can use to code suspected cases of foodborne illnesses, the defined search strategy identified calls regarding potential foodborne illness reported to and classified by APDIC with high specificity and reasonable sensitivity. If only callers with GI symptoms and a history of ingestion were considered, a high degree of agreement existed between the syndrome definition and coding by the PCC specialist for those calls classified as unintentional food poisoning. The entire database was searched for GI symptoms and ingestion, and all cases of potential foodborne illness were reviewed, thus ensuring that all cases of misclassification were identified.

Because PCCs report data regarding symptoms, not diagnoses, the syndrome definition used to evaluate calls for foodborne illness was critical. Studies have not compared the accuracy of callers' chief complaints compared with that of PCC specialists' assessment for foodborne illness. However, a recent study that evaluated the accuracy of using ED patients' chief complaints classified ED visits into syndromes and identified agreement for GI syndrome (*17*). Other research on ED visits has determined that *International Classification of Diseases* (ICD) codes are more accurate than a patient's chief complaint, possibly because they are based on a physician's diagnosis (*18*). In constructing the GI syndrome definition, consideration was given to multiple definitions discussed in the literature. The syndrome definition used was consistent with that used by ADHS for drop-in surveillance; however, other definitions have been used (*9*). A limitation of the syndrome definition used is that it included symptoms associated with, but not limited to, foodborne illness. For outbreak detection to be improved, a standard syndrome-based definition should be developed for APDIC to collect and code call data consistent with an established syndrome-based definition.

Conclusion

Comparing the incubation period noted for the majority of APDIC callers (<24 hours) with that for persons whose cases were identified by PCHD suggests that the two data sets are identifying persons from different populations. Preformed toxins are typically associated with gastroenteritis after a short incubation period (usually 2–4 hours) (*19*). Chemical contaminants also produce similar symptoms ≤30 minutes after ingestion (*12*). In contrast, bacterial pathogens, including *Shigella* and *Salmonella*, might be associated with longer incubation periods (range: 12–36 hours for *Salmonella* and 12–96 hours for *Shigella*) (*19*). Another possibility is that callers inaccurately identified their previous meal as the source of their illness. Because the etiologic source of the callers' reported illnesses could not be validated, this matter remains unresolved, underscoring the necessity for external validation of sources.

This study is subject to multiple limitations. Demographic and clinical data are missing from both the APDIC and PCHD databases. Not all suspected exposures could be confirmed through laboratory testing, and the study period was brief. Symptom onset dates were not available for 49% of PCHD laboratory-confirmed cases. In an attempt to compensate, onset dates were calculated on the basis of the average interval from the date of symptom onset to the laboratory confirmation date. This was done separately by type of PCHD case (e.g., campylobacter and hepatitis A). One reason why APDIC calls could not be matched to PCHD cases might be that natural variation was lost because of use of a group mean, underscoring the need for more complete data in future studies. Consistent with other foodborne surveillance studies, the illness reported might have been acquired not through what the caller identified as previously consumed contaminated food but instead through an earlier contaminated food source, contaminated water, person-to-person contact, or direct contact with animals (*20*).

Despite these limitations, this study demonstrates a potential role for PCC foodborne illness call data and public health agencies' foodborne illness surveillance systems. The two systems detect GI syndromes and could potentially overlap if both the PCC and the public health department used a well-designed, specific syndromic screen and maintained data regarding the frequency and severity of foodborne illnesses (i.e., for detection of FBDOs). PCCs provide 24-hour coverage with trained PCC specialists who might enhance health department surveillance, particularly during the evenings and on weekends. However, successful integration of these systems requires external clinical validation of PCC foodborne illness calls.

On the basis of this analysis, prospective surveillance of APDIC's real-time foodborne illness complaint calls will be implemented. Callers to APDIC whose illnesses are consistent with the syndrome definition and who pose the highest transmission risk (e.g., food handlers, child-care providers or attendees, and callers with an increase in severity of symptoms) will, upon informed consent, be referred to PCHD for evaluation of their symptoms. Similarly, an independent nurse reviewer will conduct follow-up regarding consenting callers

whose symptoms are consistent with the syndrome definition but who are not referred to PCHD to determine the extent of their illness and any medical treatment received. This prospective surveillance will allow for a clinical validation of callers' illnesses by determining sensitivity, specificity, and PPV by using a laboratory test as a standard on those calls referred to PCHD.

Additional studies are needed to assess the effectiveness of a prospective, integrated approach to foodborne illness surveillance. Fundamental to success of these efforts is public awareness of what types of illnesses should be reported and where they should be reported, the development of standardized definitions of syndromic illnesses, and consistent coding of calls by PCC specialists. Studies are needed to determine whether a temporal relationship exists between APDIC calls and PCHD cases; if such a relation is demonstrated, prospective monitoring of increases in APDIC calls might be predictive of increases in the number of foodborne illness cases.

PCCs might improve public health syndromic surveillance for foodborne illnesses because they capture early, real-time symptom data from the broadest possible range of sources (i.e., callers with mild to severe illness). Once clinical evaluation has confirmed callers' foodborne illnesses, then computerized PCC databases could be linked and programmed with detection algorithms to alert public health officials of increases in cases or geographic clustering of cases. The PCHD surveillance system is limited in its ability to capture real-time foodborne illness symptom data. APDIC might provide a useful addition to the PCHD's syndromic surveillance system and might assist in early detection of FBDOs.

Acknowledgment

This research was supported in part through a National Institute of Health Predoctoral Training Fellowship, NIH grant number 5F31NR00824-03.

References

1. Mandl KD, Overhage M, 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. 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.
3. Sosin DM. Syndromic surveillance: the case for skillful investment. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 2003;1:247–53.
4. Rand Center for Domestic and International Health Security. Syndromic surveillance: an effective tool for detecting bioterrorism? Santa Monica, CA: Rand Center for Domestic and International Health Security; 2004.
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. Miller JR, Mikol Y. Surveillance for diarrheal disease in New York City. J Urban Health 1999;76:388–90.
7. Das D, Mostashari F, Weiss D, Balter S, Heffernan R. Monitoring over-the-counter pharmacy sales for early outbreak detection—New York City, August 2001–September 2003.In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53 (Suppl):235.
8. Greenko J, Mostashari F, Fine A. Clinical evaluation of the emergency medical services (EMS) ambulance dispatch-based syndromic surveillance system, New York City. J Urban Health 2003;80(Suppl 1):i50–6.
9. Heffernan R, Mostashari F, Das D, et al. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 2004;10:858–64.
10. Mundorff MB, Gesteland P, Haddad M, Rolfs R. Syndromic surveillance using chief complaints from urgent-care facilities during the Salt Lake 2002 Olympic winter games. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):254.
11. Mead PS, Slutsker L, Dietz V, et al. Food-related illness and death in the United States. Emerg Infect Dis 1999;5:607–25.
12. Olsen SJ, MacKinnon LC, Goulding JS, et al. Surveillance for foodborne-disease outbreaks—United States, 1993–1997. MMWR 2000;49(No. SS-1).
13. Khan AS, Swerdlow DL, Juranek DD. Precautions against biological and chemical terrorism directed at food and water supplies. Public Health Rep 2001;116:3–14.
14. Oregon Department of Human Resources. Food practices and diarrheal diseases: *vox populi*. CD Summary 1998; Vol. 47, no. 24:1. Available at http://oregon.gov/DHS/ph/cdsummary/1998/ohd4724.pdf.
15. Institute of Medicine. Forging a poison prevention and control system. Washington, DC: The National Academies Press; 2004.
16. Fleischauer AT, Silk BJ, Schumacher M, et al. The validity of chief complaint and discharge diagnosis in emergency department-based syndromic surveillance. Acad Emerg Med 2004;11:1262–7.
17. 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.
18. 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.
19. Heymann D, ed. Control of communicable diseases manual. 18th ed. Washington, DC: American Public Health Association; 2004.
20. CDC. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food—selected sites, United States, 2003. MMWR 2004;53:338–43.

Monitoring Over-The-Counter Medication Sales for Early Detection of Disease Outbreaks — New York City

Debjani Das, K. Metzger, R. Heffernan, S. Balter, D. Weiss, F. Mostashari
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 St., Box No. CN22A, New York, NY 10013. Telephone: 212-788-4318; Fax: 212-788-5470; E-mail: ddas@health.nyc.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Over-the-counter (OTC) medications are frequently used during the initial phase of illness, and increases in their sales might serve as an early indicator of communitywide disease outbreaks. Since August 2002, the New York City (NYC) Department of Health and Mental Hygiene (DOHMH) has tracked OTC medication sales to enhance detection of natural and intentional infectious disease outbreaks.*

Objectives: *This report describes the surveillance system and presents results from retrospective analyses and a comparison between citywide trends in OTC medication sales and emergency department (ED) visits.*

Methods: *Sales data transmitted daily to DOHMH are categorized into two groups: influenza-like illness (ILI), which includes cough and influenza medications, and gastrointestinal illness (GI), which includes major brand and generic antidiarrheals. Cyclical, linear regression models were used to identify significant ($p<0.05$) increases in the daily ratio of ILI to analgesics sales (analgesics are used as a denominator in the absence of total sales). Daily and weekly average ratios of GI to analgesic sales were analyzed. Citywide trends in OTC ILI and GI medication sales were compared with ED visits for fever/influenza and diarrhea syndromes.*

Results: *Citywide ILI drug sales were highest during annual influenza epidemics and elevated during spring and fall allergy seasons, similar to trends in the ED fever/influenza syndrome. ILI sales did not consistently provide earlier warning than the ED system of communitywide influenza. GI drug sales increased during the fall and peaked during early winter and after the blackout of August 2003. Unlike ED diarrheal visits, GI medication sales did not substantially increase during late winter (February–March).*

Conclusion: *Citywide OTC medication sales can provide indications of communitywide illness, including annual influenza epidemics. Antidiarrheal medication sales were more sensitive to increases in GI caused by norovirus and influenza than illness caused by rotavirus. OTC medication sales can be considered as an adjunct syndromic surveillance system but might not be as sensitive as ED systems.*

Introduction

Over-the-counter (OTC) medications are commonly taken before or instead of seeking medical care (*1–3*). OTC medication sales, therefore, might be an early indicator of communitywide illness. One of the first signs of a large waterborne cryptosporidiosis outbreak in Milwaukee in 1993 was newspaper reports that local pharmacies had sold out of antidiarrheal medications (*4,5*). A retrospective report confirmed that increases in sales of Immodium® (McNeil Consumer and Specialty Pharmaceuticals, Fort Washington, Pennsylvania), Pepto Bismol® (Proctor and Gamble, Cincinnati, Ohio), and Kaopectate® (Pfizer, New York, New York) were the earliest indicators of widespread illness, suggesting that prospective monitoring of OTC medication sales might have resulted in earlier detection of the outbreak (*4*). In England, increases in electronic-point-of-sale pharmacy sales occurred 2 weeks before an increase in emergency department (ED) visits during the winter (*6*), and correlations between OTC antinausea and antidiarrhea medication sales and ED admissions have been reported (*7*).

New York City (NYC) began monitoring OTC medication sales in 1996 as part of its waterborne disease surveillance program. Aggregate, weekly counts of OTC antidiarrheal medications sold were transmitted from one large pharmacy chain to the Department of Environmental Protection and reviewed weekly (*8*). In 2000, the New York City Department of Health and Mental Hygiene (DOHMH) asked pharmaceutical

retailers to establish daily, automatic data transmission to improve timeliness and expand the system to cover medications for influenza-like illness (ILI). This report describes the methods developed during the first 32 months of operation, presents results from a retrospective analysis of data from that period, and compares citywide trends in OTC medication sales with comparable syndromes from the DOHMH ED syndromic surveillance system (*9*).

Materials and Methods

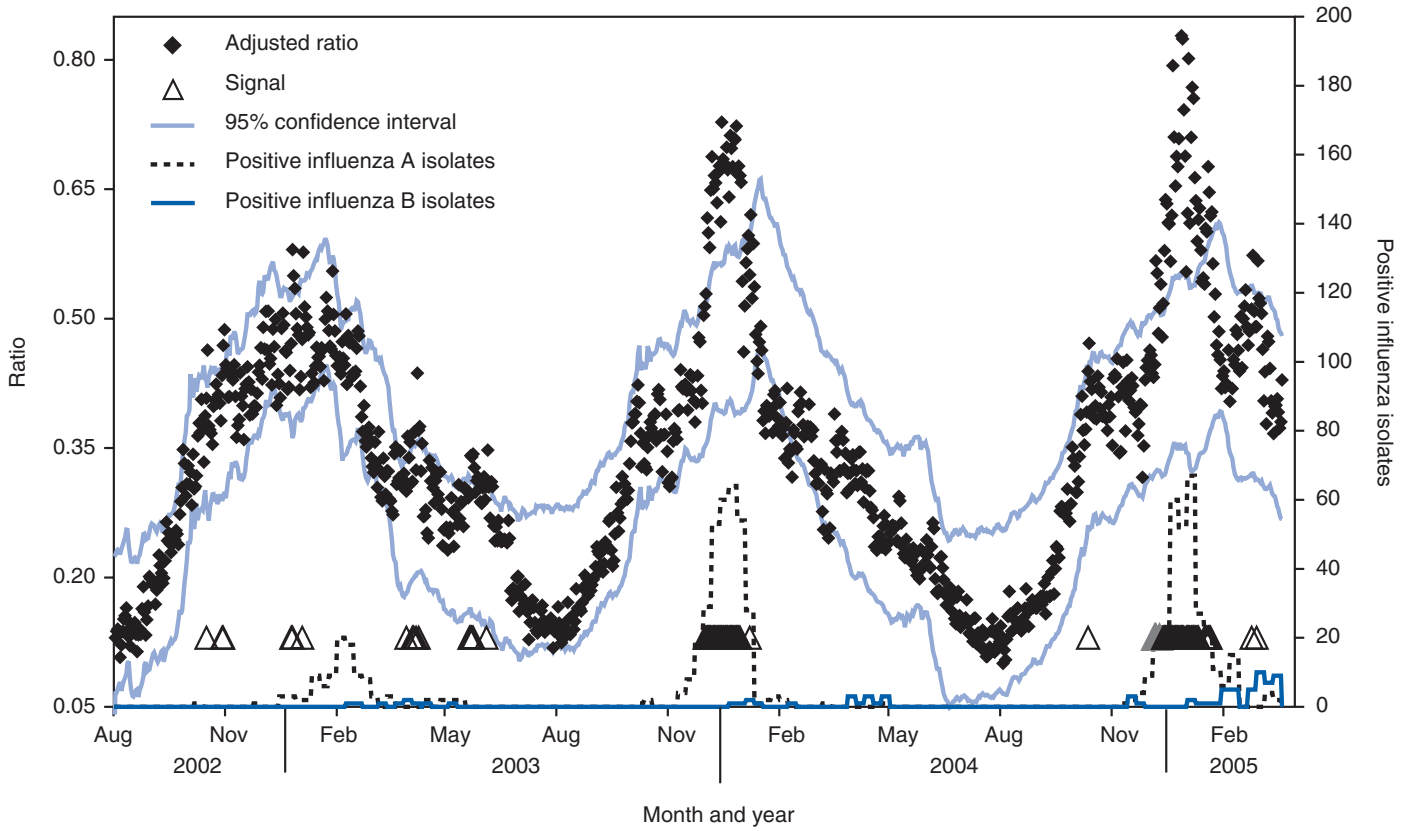
Since August 2002, OTC medication sales data have been transmitted daily by file transfer protocol (FTP) from a central pharmacy database. Each daily file contains information on the previous day's sales of the following medications: cold drugs, analgesics, vapor rubs (for colds), vitamins, stomach aids, and first aid materials. Data elements include the number of units sold, drug name, department and subdepartment, store location, whether or not the item was on promotion, and the number of units in stock (an indication of whether the item was available for sale that day). Data are checked for completeness, duplicate records, and other errors before being appended to an archive for analysis. Baseline data were obtained beginning in August 2001.

Because of the multiple possible names and formulations for each drug and because new items are frequently introduced into the market, a simple and flexible method was sought to categorize drugs into syndrome groups. For the ILI category, the estimated 400 drug names in the cold subdepartment, which accounted for 47% of the total drug sales reported to DOHMH, were examined. To identify the subset of these drug names that was most closely associated with influenza activity, the ratio was calculated of sales during peak influenza season (November 2001–January 2002) to nonpeak season (August 2001–October 2001). These periods were based on the number of positive influenza isolates identified by the World Health Organization (WHO) sentinel laboratories serving NYC (Figure 1). A review of the 50 drugs with the highest ratios suggested that selecting any drug with text strings "flu" or "tussin" in the name would capture most influenza-associated sales. Drugs not included in the ILI category because their sales were not highly correlated with influenza activity included 1) multicold symptom relievers (e.g., Alka Seltzer Plus® [Bayer, Morristown, New Jersey] and Nyquil® [Proctor and Gamble, Cincinnati, Ohio]); 2) decongestants (e.g., Sudafed® [Pfizer, New York, New York]); 3) analgesic brand name cold products (e.g., Tylenol Cold® [McNeil Consumer and Specialty Pharmaceuticals, Fort Washington, Pennsylvania]), and 4) chest rubs (e.g., Vicks Vaporub® [Proctor and Gamble, Cincinnati, Ohio]).

For citywide temporal ILI analysis, the outcome was the daily ratio of ILI drug sales to sales of analgesics. When the ratio was used, rather than just the number of ILI sales, noise was reduced in the data by partially controlling for variation in total OTC medication sales volume as a result of store hours, consumer behavior, and other unmeasured factors. Analgesics were chosen as the denominator, because total sales data were not available. Analgesics sales were relatively stable throughout the year and accounted for approximately 26% of sales reported to DOHMH. The analgesic category included the following brand names and their generic equivalents: Advil® (Wyeth Consumer Healthcare, Madison, New Jersey); Aleve® (Bayer, Morristown, New Jersey); Bayer® (Bayer, Morristown, New Jersey); Bufferin® (Bristol-Myers, New York, New York); Ecotrin® (GlaxoSmithKline, Brentford, Middlesex, UK); Excedrin® (Bristol-Myers, New York, New York); Motrin® (McNeil Consumer and Specialty Pharmaceuticals, Fort Washington, Pennsylvania); and Nuprin® (Bristol-Myers, New York, New York).

In this report, prospective, daily analyses of the citywide ratio of ILI to analgesics sales for each day during August 1, 2002–March 31, 2005, were mimicked. A cyclical linear regression model computed the difference between the expected and the observed ratios. The model was based on the classic Serfling influenza model as applied previously to NYC ambulance dispatch data (*10,11*). For each daily analysis, the most recent 7 days of data were censored to ensure that recent increases or decreases in sales would not affect the model. Terms in the model included sine and cosine to capture seasonal cycles, day-of-week (dummy variables with Sunday as reference), holidays, postholidays, promotional sales (the proportion of total sales that were on promotion), and the weekly number of positive influenza A and B isolates identified by NYC's three WHO influenza reference laboratories. Influenza isolate data were censored for the previous 14 days to ensure that recent increases could be attributed to communitywide influenza outbreaks when needed. Holidays were categorized into major winter holidays (Thanksgiving, Christmas, and New Year's Day) and other national holidays (Martin Luther King Day, President's Day, Memorial Day, Independence Day, and Labor Day). If a holiday was on a weekend, the official government work holiday (Friday or Monday) was coded as a holiday. Days after holidays were coded as postholidays. The model also included a measure of recent increases in allergy medication sales to partially control for increases in ILI drug sales associated with the spring and fall allergy seasons. The measure found to be most useful was a z-score recalculated daily as z = (mean daily sales during previous week) – (mean daily sales during the period 4–10

FIGURE 1. Citywide trends and signals in the adjusted ratio of influenza-like illness to analgesic over-the-counter (OTC) drug sales and positive isolates of influenza A and B — New York City, August 1, 2001–March 31, 2005*



* Temporal signals p<0.01 and 95% confidence intervals from daily linear regression. The OTC ratio is adjusted for day of week, major national winter holidays, and the day after these holidays (Thanksgiving, Christmas, New Years, and Martin Luther King observance day).

weeks before date being analyzed)/(standard deviation of daily sales during the period 4 to 10 weeks before date being analyzed).

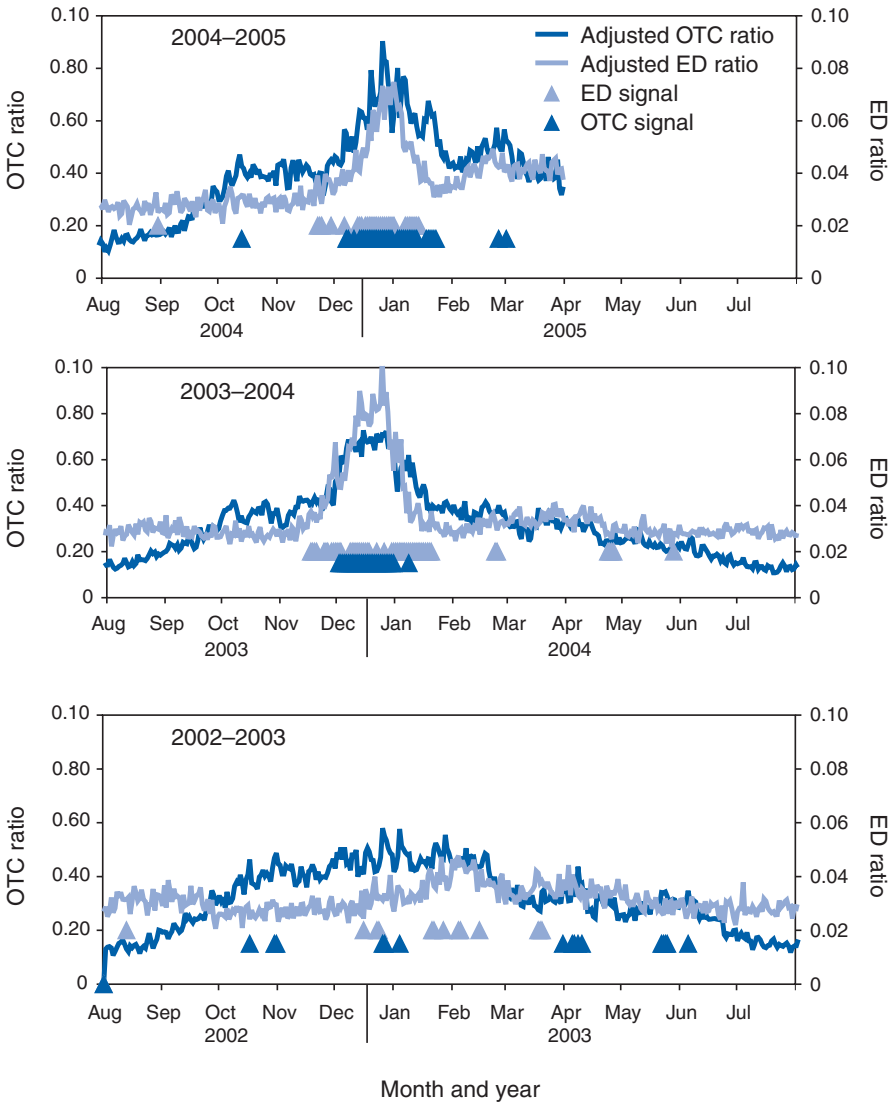
The model-adjusted ratio and confidence limits generated from each daily analysis were plotted (Figure 1). A signal was defined as any observation that exceeded the 95% upper confidence limit of the model. These results were compared with trends in the day-of-week adjusted ratio of ED fever/influenza-like syndrome (Figure 2) (*9*). To characterize the relative timing of influenza-associated trends in the ED and OTC medication time series, the ED time series was limited to influenza season (October–April), and the correlation was examined between 1) the weekly average ratio of OTC ILI to analgesics and 2) the weekly average ratio of fever/influenza-like syndrome visits to other syndrome visits (for various lag periods ranging from -14 days to +14 days).

For the antidiarrheal GI category, common antidiarrheal drugs were identified, including Immodium®, pink bismuth formulas, Kaopectate, Maalox® (Novartis, Basel, Switzerland), and generic antidiarrheals (*12*). Immodium accounted for the highest volume (51%) of sales in this category. Kaopectate was removed from the GI category because its weekly sales indicated low correlation with weekly sales of Immodium during peak diarrhea season (November 2002–February 2003; Spearman correlation coefficient $r^2 = 0.46$). The ratio of OTC GI sales to analgesics sales and 7-day moving average have been plotted (Figure 3) along with the ratio of ED diarrhea syndrome to other ED visits (*9*).

Results

During the reporting period, August 1, 2002–March 31, 2005, the mean daily total sales reported to DOHMH was 34,883 (standard deviation [SD] = 9,475). Mean daily sales for ILI, GI, and analgesics were 2,383 (SD = 1,229); 1,132 (SD = 231); and 6,638 (SD = 1,249), respectively. A total of 99% of sales were reported to DOHMH by the next calendar day.

FIGURE 2. Citywide trends and signals in over-the-counter (OTC) influenza-like illness category and emergency department (ED) fever/influenza category — New York City, August 1, 2002–March 31, 2005*



* The OTC ratio is adjusted for day of week and major national winter holidays and the day after these holidays (Thanksgiving, Christmas, New Years, and Martin Luther King observance day). The emergency department ratio is adjusted for day of week.

Citywide trends in the ratio of ILI to analgesic sales followed a consistent cyclical pattern with highs in the winter and lows in the summer (Figure 1). Within these annual cycles, observed increases were associated with epidemic influenza and the spring and fall allergy seasons. A series of 31 signals over 38 days occurred during the 2003–04 peak influenza season. Similarly, 48 signals occurred over 57 days during the 2004–05 peak influenza season. Ten additional signals occurred during April–June 2003 (Figure 1).

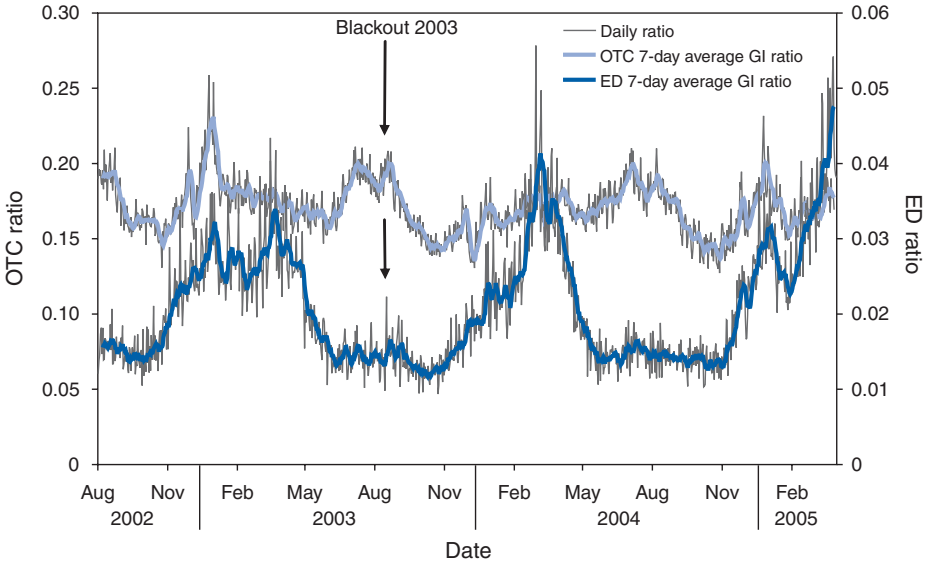
ILI trends were compared with ED visits classified under the fever/influenza-like syndrome (Figure 2). Overall, the weekly ratio of ILI to analgesics sales was highly correlated with the weekly ratio fever/influenza-like syndrome visits to other syndrome visits (r^2 = 0.60; $p<0.001$). Slightly lower correlations were identified by moving the OTC time series forward and backward in time by 7 and 14 days in relation to the ED fever/influenza time series (14 days before, r^2 = 0.55; 7 days before, r^2 = 0.59; 7 days after, r^2 = 0.54; and 14 days after, r^2 = 0.44). The arrival of communitywide influenza for the 2004–05 season was signaled by ED fever/influenza surveillance in late November 2004, whereas the first series of OTC signals began 6 days later. During the 2003–04 peak influenza season, the ILI ratio doubled and the ED fever-influenza ratio tripled. The series of consecutive ED signals associated with this increase began 2–3 weeks earlier than the first ILI signals (Figure 2). No overlapping signals occurred during the mild 2002–03 influenza season (Figure 2).

Trends in the ratio of GI sales to analgesic sales are plotted against the ratio of ED GI visits to other visits (Figure 3). The weekly ratio of GI to analgesics sales was a substantially lower correlation with the weekly ratio ED GI to other visits (r^2 = 0.24; $p<0.005$). Increases in both systems were detected in late autumn (coinciding with documented institutional norovirus outbreaks and suspected communitywide norovirus activity), during peak influenza, and during the blackout in August 2003. Although the most significant increases in ED GI visits occurred annually during late winter (February–March), no corresponding increase in GI sales was observed. Increases during the summer in GI sales were not accompanied by increases in ED GI visits.

Discussion

In the first 32 months of OTC surveillance in NYC, the system served as an adjunct to other indications of citywide illness. Sustained, statistically significant increases in ILI drug

FIGURE 3. Citywide actual and 7-day moving average of over-the-counter (OTC) ratio of gastrointestinal (GI) to analgesic category sales and emergency department (ED) ratio for diarrhea to other visits — New York City, August 1, 2002–March 31, 2005



sales were observed during annual influenza epidemics, and increases in GI drug sales occurred during fall norovirus season and immediately after the August 2003 blackout. No localized disease outbreaks were detected with the system in any syndrome. In addition, although the purchase of OTC medications is hypothesized to occur earlier in the course of illness than visits to a health provider, a consistent pattern was not observed when OTC sales were compared with ED visits.

Multiple factors might contribute toward the challenges of OTC syndromic surveillance. The high rate of background sales unrelated to illness, which might include consumer "stockpiling" of OTC medicines, obscures purchases for acute illness. Differences in store hours and local consumer behavior add to the variance in sales, posing challenges to data modeling and spatial analysis to detect local clustering in OTC sales. Drugs comprise multiple formulations, and new drugs enter the market regularly, making syndrome categorization difficult. Perhaps the most challenging problem of routine OTC surveillance is how to respond to signals. No information is available concerning the person purchasing the medication, and direct investigation is not possible either with individual pharmacies or consumers.

Despite these challenges in multiple syndromic surveillance systems, in the United States, OTC sales data are used. The Electronic Surveillance for the Early Notification of Community-based Epidemics (ESSENCE II) project is based on an electronic medical chart system that captures both medical visits and OTC and prescription sales (*13*). A key advantage of the system is that it links OTC sales, prescription sales, and medical visit information in one system, allowing for patient-level follow-up. Drugs were categorized into syndromes by correlating discharge diagnoses with drugs purchased or prescribed. A limitation of the system is that it is restricted to military personnel and their families. Another system, the National Retail Data Monitor (NRDM) operated by the University of Pittsburgh, receives data daily from >15,000 retail stores nationwide (*14,15*). Drugs are grouped into 18 categories (e.g., antidiarrhea, antifever, and cold relief) and can be viewed through a website interface by public health authorities for aberrations in sales within their region. NRDM has detected both communitywide influenza and at least one localized GI outbreak, which was identified based on increases in electrolyte sales, which NYC does not track (*16*).

Conclusion

OTC syndromic surveillance might be useful as an early indicator of disease outbreaks. To date in NYC, however, the system has served primarily to corroborate large-scale illness trends detected in other syndromic disease outbreak systems. Future possible enhancements include obtaining the total number of transactions, by store by day, for use as a denominator to better control for unmeasured consumer behavior. In addition, to improve coverage, the feasibility of integrating local systems with data received from the NRDM system is being evaluated (*14*). DOHMH has requested to expand the list of drugs to include pediatric formulations, including electrolyte products; the current ILI and GI categories focus on adult formulations. The use of OTC surveillance for disease outbreak detection in NYC should be increased as these improvements are implemented.

References

1. McIsaac WJ, Levine N, Goel V. Visits by adults to family physicians for the common cold. J Fam Pract 1998;47:366–9.
2. Vingilis E, Brown U, Hennen B. Common colds: reported patterns of self-care and health care use. Can Fam Physician 1999;45:2644–6, 2649–52.

3. Metzger KB, Hajat A, Crawford M, Mostashari F. How many illnesses does one emergency department visit represent? Using a population-based telephone survey to estimate the syndromic multiplier. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Supp):106–11.
4. Morris R, Naaumova EN, Griffith JK. Did Milwaukee experience waterborne cryptosporidiosis before the large documented outbreak in 1993? Epidemiology 1998;9:264–70.
5. 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.
6. Davies GR, Finch RG. Sales of over-the-counter remedies as an early warning system for winter bed crises. Clin Microbiol Infect 2003;9: 858–63.
7. Edge VL, Pollari F, Lim G, et al. Syndromic surveillance of gastrointestinal illness using pharmacy over-the-counter sales: a retrospective report of waterborne outbreaks in Saskatchewan and Ontario. Can J Public Health 2004;95:446–50.
8. Miller JR, Mikol Y. Surveillance for diarrheal disease in New York City. J Urban Health 1999;76:388–90.
9. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorf M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 2004;10:858–64.
10. Mostashari F, Fine A, Das D, Adams J. Use of ambulance dispatch data as an early warning system for community-wide influenza-like illness, New York City. J Urban Health 2003;80(Suppl):i43–9.
11. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963;78:494–506.
12. Kelly W, ed. Nursing 2003 drug handbook. Philadelphia, PA: Lippincott Williams & Wilkins; 2003.
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 (Suppl):i32–42.
14. Wagner MM, Robinson JM, Tsui FC, et al. The National Retail Data Monitor for Public Health Surveillance. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):40–2.
15. Wagner MM, Robinson JM, Tsui FC, Espino JU, Hogan WR. Design of a national retail data monitor for public health surveillance. J Am Med Inform Assoc 2003;10:409–18.
16. Realtime Outbreak and Disease Surveillance (RODS) Laboratory at University of Pittsburgh. National Retail Drug Monitor case studies. Pittsburgh, PA: RODS Laboratory at University of Pittsburgh; 2004. Available at http://rods.health.pitt.edu.

Experimental Surveillance Using Data on Sales of Over-the-Counter Medications — Japan, November 2003–April 2004

Yasushi Ohkusa, M. Shigematsu, K. Taniguchi, N. Okabe
Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Tokyo, Japan

Corresponding author: Yasushi Ohkusa, Infectious Disease Surveillance Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-Ku, Tokyo 162-8640, Japan. Telephone: 81-3-5285-1111; Fax: 81-3-5285-1129; E-mail: ohkusa@nih.go.jp.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Objectives: *This report describes a study to explore the possibility of using data on sales of over-the-counter (OTC) medications as part of a routine syndromic surveillance system aimed at early detection of infections of public health concern. A retrospective evaluation was conducted of sales of OTC medications used to treat the common cold. This report discusses the correlation of these data to influenza activity in Japan during the 2003–04 influenza season and evaluates the potential of using such data to predict influenza epidemics.*

Methods: *Data from approximately 1,100 pharmacies throughout Japan collected during November 2003–April 2004 were analyzed. OTC sales data were compared with influenza incidence data (one weekly and two daily data sets) to determine correlations and predictability. Adjusted R-square was used as an index of goodness-of-fit in the estimation. Data reflecting daily influenza activity were obtained from the National Surveillance of Daily Influenza Outpatients and the Mailing List–Based Influenza Epidemic Database. National sentinel surveillance data for influenza from approximately 5,000 sites nationwide also were analyzed.*

Results: *Although a correlation was demonstrated between sales of OTC medications used to treat the common cold and concurrent influenza activity, analysis of sales data alone was not sufficient to determine influenza activity in advance even when sales promotion effects were excluded from the analysis.*

Conclusion: *Because visiting a health-care provider costs more than purchasing OTC medications, the hypothesis was formed that an ill person will purchase OTC medications first and visit a physician only if the condition does not resolve or worsens. The results of this study do not provide any clear evidence to support this hypothesis. For this reason, OTC sales do not appear to be a good candidate for a national real-time detection system for influenza epidemics in Japan.*

Introduction

In 2000, the first syndromic surveillance prototype in Japan was initiated by the Japanese Ministry of Health, Labour, and Welfare (MHLW) in the Kyushu area during the G-8 summit meeting to assist in the early detection of an act of biologic terrorism or an unusual cluster of tropical diseases imported by travelers from tropical areas (*1*). This limited-scale surveillance involved 17 medical institutes in two prefectures for <1 month. Data for the surveillance system were reported through facsimile transmissions for five syndromic categories (i.e., respiratory, gastrointestinal, neurological, cutaneous-mucous membrane-bleeding, and nonspecific). The second (and the first nationwide) syndromic surveillance system was implemented during May 20–July 14, 2002, in connection with the Japan-Korea 2002 World Cup soccer games. The Internet-based surveillance, which was conducted by MHLW and the Infectious Disease Surveillance Center of the National Institute of Infectious Diseases (NIID), grouped hospitalized patients by symptoms into the same five syndromic categories used in 2000. Both ad hoc syndromic surveillance systems operated during high-profile events and were conducted successfully, and their data were matched with those diseases with the same clinical features that were collected later by routine national surveillance. For example, this second ad hoc syndromic surveillance detected a cluster of viral meningitis and a regional outbreak of measles successfully, thereby illustrating the potential of these data in assisting with early detection of disease. However, further improvements are required to detect pandemic influenza or a possible biologic terrorist attack in time to minimize its consequences.

The goal of the early detection syndromic surveillance system is to conduct routine (not ad hoc) surveillance that complements existing surveillance systems and to detect increases in the number of patients before they report to hospitals with severe conditions. Data concerning sales of

over-the-counter (OTC) medications, emergency department (ED) visits, ambulance calls, and other factors were assessed as tentative candidates for early detection of disease outbreaks (*2,3*). Because no routine syndromic surveillance for respiratory syndrome had been conducted previously in Japan, the effectiveness of OTC surveillance in early detection was compared with multiple influenza surveillance systems that were already in place. This report presents interim findings from the OTC sales surveillance.

Methods

Data Source

Commercially available data collecting reported daily sales of OTC medications in all forms (e.g., tablets, powder, granules, and syrup) used to treat the common cold from 1,100 pharmacies throughout Japan were obtained. So-called combination or general common-cold medications were chosen for examination because use of such medications has long been accepted in Japanese society as the first and most common treatment for influenza-like illness (ILI). These medications usually consist of a combination of antipyretic analgesics (e.g., acetaminophen or ibuprofen), antitussives (e.g., dihydrocodeine phosphate or noscapine), expectorants (e.g., bromohexine hydrochloride, guaifenesin, or potassium guaiacolsulfonate), exogenous enzyme (e.g., lysozyme chloride), bronchodilator (e.g., dl-methylephedrine hydrochloride), antihistaminics (e.g., carbinoxamine maleate or mequitazine), vitamins (e.g., vitamin B1, B2, or vitamin C), and others (e.g., herbal medicines or caffeine). The category also includes combined herbal medicines that are licensed for common cold treatment.

Data were collected by a private marketing company from randomly chosen pharmacies covering approximately 2.0% of the 50,000 pharmacies in Japan. The influenza season was defined as November–April. Sales data collected during November 2003–April 2004 were subjected to retrospective analysis to examine the suitability of OTC sales surveillance for early detection of unexpected rare events. OTC sales data were compared with reliable sentinel surveillance data for influenza collected during November 2003–April 2004 by the National Epidemiological Surveillance of Infectious Diseases (NESID) and with data on influenza activity collected daily by two other surveillance systems from clinics, hospitals, and health-care providers. In Japan, sentinel reporting of clinical cases of ILI is mandatory, with or without laboratory tests or confirmation. Data (e.g., the number of influenza outpatients, by age and age group) are collected weekly from 5,000 sentinel surveillance sites (including 3,000 pediatricians and 2,000 internal medicine clinics or departments) nationwide covering one tenth of all clinics and hospitals in Japan for all influenza-related visits. Two daily influenza activity information sources are 1) reported numbers of cases of ILI reported by the National Surveillance of Daily Influenza Outpatients (Daily Case Reporting [DCR]), which collects data from 10% of selected sentinel medical institutions and 2) voluntary reporting by clinicians to the Mailing List–Based Influenza Epidemic Database (MLflu). DCR is operated by NIID and began operating in January 2004 for the 2003–04 influenza season; it collects data regarding the number of outpatients who received a diagnosis of ILI either clinically or by diagnostic test from 500 sentinel sites in clinics and hospitals. Date of onset is not included in the reported data, which makes this surveillance vulnerable to the-day-of-the-week effect (i.e., few patient visits reported during the weekend and more on the following Monday). MLflu is operated by volunteer pediatricians and began operating in December 2003 for the 2003–04 influenza season; it collects data from approximately 350 pediatricians regarding outpatients who have received a diagnosis of influenza by rapid test. Cases reported through MLflu are more likely to reflect actual influenza activity. Date of onset is reported, so the surveillance system is free from the-day-of-the-week effect. However, because reporting is voluntary, the number and representativeness of participants varies during the influenza season.

Analysis

A model was created to estimate influenza activity from the OTC sales information during a 6-month period, as follows:

$$\log(\text{influenza activity in period } t) = \alpha + \beta \log(\text{OTC sales in period } t\text{-}j) + \varepsilon$$

OTC sales data were then adjusted for the-day-of-the-week effect and compared with three other different influenza activity surveillance systems (sentinel surveillance, DCR, and MLflu) to examine the number of lead-days by OTC sales. The adjusting procedure consisted of two steps, as follows:

$$\text{Adjusted OTC sales in period } 1 = \text{Replaced OTC sales in period } 1$$

$$\text{Adjusted OTC sales in period } t = \frac{2k}{k(k+1)}\ \text{Replaced OTC sales in period } t + \sum_{j=1}^{k-1} \frac{2(k-j)}{k(k+1)}\ \text{Adjusted OTC sales in period } t\text{-}j$$

$$\text{for } t > 1 \text{ and } t \leq 6$$

$$\text{Adjusted OTC sales in period } t = 7/28 \text{ Replaced OTC sales in period } t + \sum_{j=1}^{6} \frac{(7-j)}{28} \text{ Adjusted OTC sales in period } t\text{-}j \quad \text{for } t > 6.$$

The data set was adjusted by replacing data for weekends, holidays, and the day before and after weekends or holidays with data for the nearest preceding nonholiday weekday. Then the replaced data were smoothed to the past by taking a moving average from the current period to 1 week previous, giving a relatively heavier weight to the nearer days, and gradually reducing the weight for the far past. Although this adjusting procedure did not require future data, the adjustment result might be affected (pulled) from the data used for the replacement and smoothing procedure.

Comparative analysis of OTC sales with one weekly and two daily data sets recording influenza incidence was performed to determine correlations and predictability. Adjusted R-square was used as an index of goodness-of-fit in the estimation.

Results

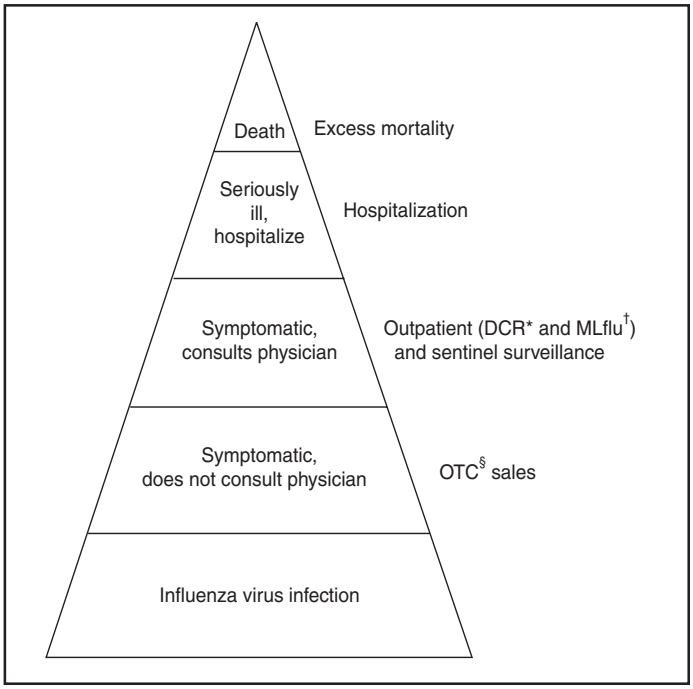
Because national surveillance data do not capture the number of persons who consult a health-care provider for general respiratory symptoms, data regarding consultations for influenza symptoms were used as a substitute to assess lead time of OTC information. Influenza surveillance in Japan was designed to report all potential influenza patients from at least one system for robust detection of influenza activity other than hospitalization (Figure 1). The case definition of influenza used for both outpatient sentinel surveillance and DCR was based on clinical symptoms, which resulted in reporting of patients with ILI.

For this analysis, the hypothesis used was that the majority of persons who were infected by influenza virus and who experienced mild symptoms would choose to self-treat with OTC medications and that those persons whose condition subsequently became more serious would then consult a physician later. Data of sales of OTC medications used to treat the common cold, readily provided as commercial databases, were assumed to reflect the population of preclinical visits by persons with ILI. Data on outpatient visits were represented by sentinel surveillance, DCR, and MLflu. An increase in OTC sales of medications used to treat the common cold was assumed to indicate an initial increase of ILI, and the lead time of the sales to the influenza activity was expected to be observed.

OTC sales per pharmacy were tracked, and the time trend of sales per pharmacy, which was adjusted for the-day-of-the-week effect and then smoothed, was given as a line (Figure 2). Multiple peaks of different size were observed during the 5-month surveillance period, with the consistent underlining trend being that sales were higher in winter and decreased toward spring. Peaks observed were in early and mid-December, early February, and late March. The third peak observed occurred during late January–early February and corresponded with the peak of ILI sentinel reporting generally recorded during influenza seasons; a subsequent period of decline toward spring was also matched. However, the pattern of the early influenza season was fairly discrete between the two data sets (Figure 3).

Adjusted OTC sales data also were compared with adjusted influenza data from DCR to identify a similar pattern during the height of the influenza season (Figure 4). DCR for clinically confirmed ILI is case-based and includes the patient's age and age group, date of visit, performance of rapid test, and result of a rapid diagnostic test as a single thread of information. Because data are reported by clinics and hospitals, numbers were low on Saturdays and Sundays and high on Mondays; consequently, numbers were adjusted for the-day-of-the-week effect. As with sentinel surveillance, DCR also indicated a different pattern early in the influenza season, and the peak coincided with the third peak of OTC sales. Characteristically, no rise in DCR was observed to match the last peak of OTC sales during late March.

FIGURE 1. Relationship of influenza status and influenza-related surveillance



* Daily case reporting of the National Survey of Daily Influenza Outpatients.
† Mailing List–Based Influenza Epidemic Database.
§ Over-the-counter medications.

FIGURE 2. Time-trend of adjusted over-the-counter (OTC) sales per pharmacy, by date — Japan, November 2003–April 2004

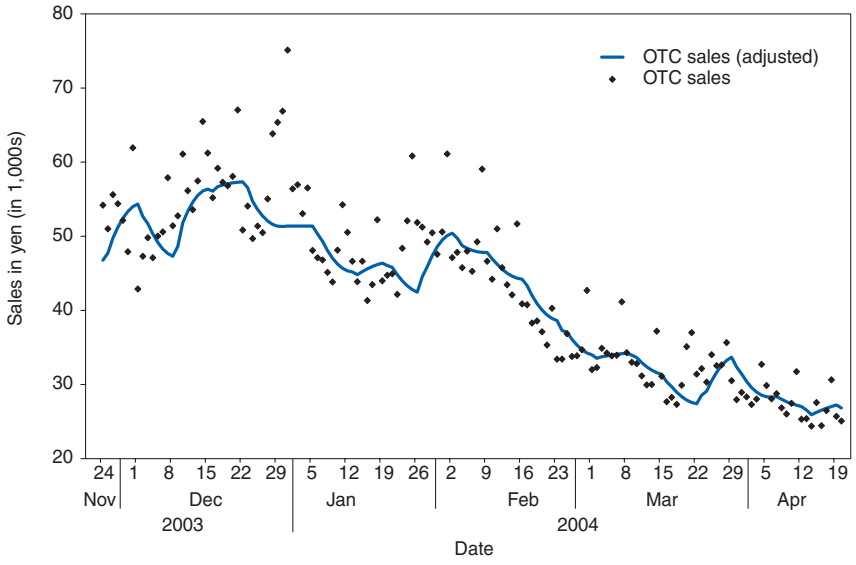
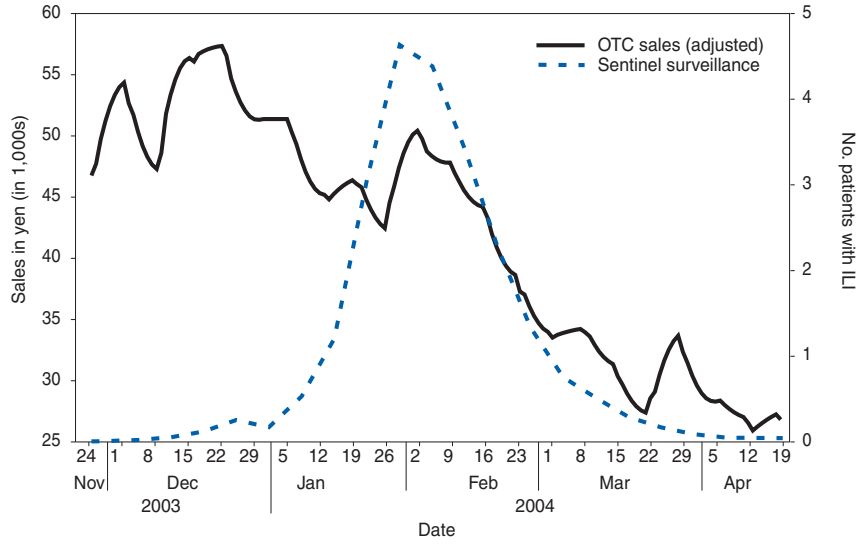


FIGURE 3. Comparison of over-the-counter (OTC) sales per pharmacy (adjusted) with number of patients with influenza-like illness (ILI) reported per sentinel point by national sentinel surveillance, by date — Japan, November 2003–April 2004



MLflu data were reported voluntarily by physicians interested in influenza preparedness. Information collected through the case-based reporting system included the patient's age, date of illness onset, date of visit, type of rapid diagnostic test used, type of influenza virus (A or B) diagnosed, and name of antivirals or other common cold medications prescribed. The date of onset was available for MLflu, which made it free from the-day-of-the-week effect. Additionally, this system was able to provide the number of laboratory-confirmed cases of influenza (i.e., those diagnosed by rapid diagnosis tests). A limitation of this system was that the number of participants varied during the season (low at the beginning and the end of the season). Interest of the clinicians participating in MLflu was high when ILI was rapidly increasing but decreased after the peak period ended (Figure 5). The effect of this variance in the reporting rate should be considered when interpreting the results. As with the other two influenza surveillance systems, MLflu indicated a different pattern from the OTC medicine surveillance at the beginning of the influenza season (Figure 5). However, for the third peak, the rise in sales of OTC medications did not coincide with the peak of MLflu reporting. Instead, the peak observed by MLflu preceded sales by 1–2 weeks (Figure 5). No matched peak was observed for the one during March.

OTC sales data were compared with other influenza activity surveillance data to determine lead time (i.e., the number of days that OTC sales elevation preceded an increase in the number of influenza patients) (Figure 6). Fitness among DCR declined as lead time became longer. The highest adjusted R-square was obtained when OTC data led by 1 day. Conversely, fitness among sentinel surveillance or MLflu rose when lead time was longer. In the case of sentinel surveillance or MLflu, OTC sales appeared to lag behind influenza activity. A peak in OTC sales observed at the end of 2003 was suspected to reflect influenza activity.

Discussion

Syndromic surveillance in Japan has been conducted on an ad hoc basis during high-profile events (*1*). A short-term, labor-intensive analysis system was used that was expensive and resource-intensive to run on a daily basis. To date, several routine influenza surveillance systems have been implemented in Japan. However, each system by itself is unable to provide sufficient information to prepare for the potential emergence of pandemic influenza or related diseases. None of three currently existing influenza surveillance systems might be able to detect the early stage of a pandemic because all systems detect patients only at the point of consultation. In addition, each surveillance system has certain limitations. For example, the national sentinel surveillance provides reliable mandatory reporting but captures only the number

of patients who visit sentinel clinics and hospitals without collecting sufficient qualitative information. These data are reported weekly, with a 1-week delay during which data are compiled. DCR captures additional qualitative information but reports include only the date of visit. MLflu reports the number of patients who receive a diagnosis for influenza with rapid testing. In each surveillance system, timeliness, accuracy, and representativeness have been traded off for other advantages. The rationale for using readily available data of OTC sales for monitoring is to establish routine early detection surveillance for pandemics and other unexpected events to complement those surveillance systems.

The lead time for OTC sales was compared with influenza surveillance to evaluate the timeliness of sales data for detecting seasonal influenza epidemics. An estimated 72% of Americans with cough, cold, influenza, or sore throat often purchase OTC medications early in the course of their illnesses (*4*). Increases in OTC sales were expected to precede an increase in patient visits to hospitals, assuming that consumer behavior in Japan is similar to that in other developed countries (i.e., persons purchase OTC medications when they first feel ill and then visit clinics or EDs if their illness becomes more serious). Although OTC sales correlated well with contemporary influenza activity (*2,3*), a clear lead time was lacking, and analysis of OTC sales data indicated no evidence of advance detection of influenza activity. Additionally, difficulties were encountered in interpreting sales increases in late December from influenza surveillance alone. The increase appeared to reflect preparation for a long holiday season accelerated by year-end discount promotions but not an actual increase in influenza activity. However, further analysis excluding this sales promotion effect was also not able to determine any influenza activity in advance.

These results indicate that sales data on OTC medications used to treat common colds have a low potential for predicting increased influenza activity in Japan. Multiple factors might account for this outcome. Because the analysis was performed only on a national level, the study did not take into account regional varia-

FIGURE 4. Comparison of over-the-counter (OTC) sales per pharmacy (adjusted) with number of patients with influenza-like illness (ILI) per hospital or clinic recorded through daily case reporting (DCR) of the National Survey of Daily Influenza Outpatients (adjusted), by date — Japan, January–April 2004

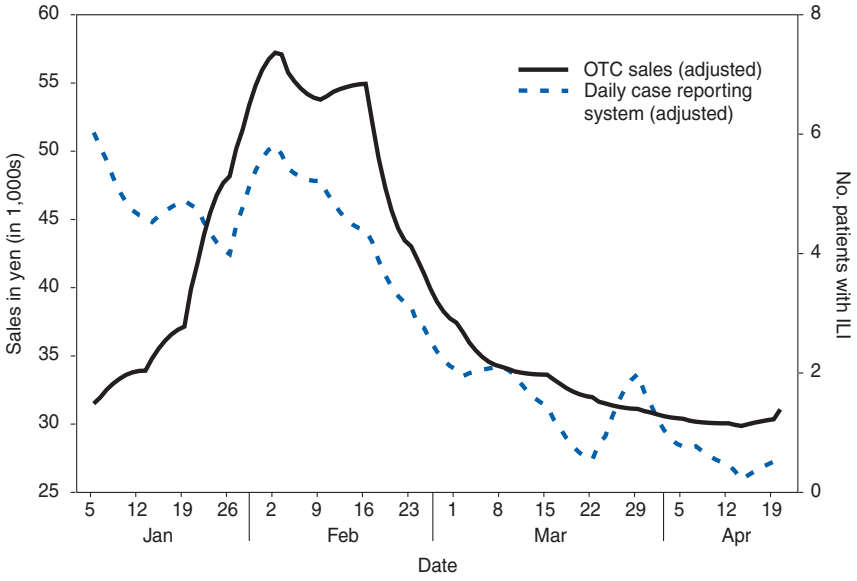
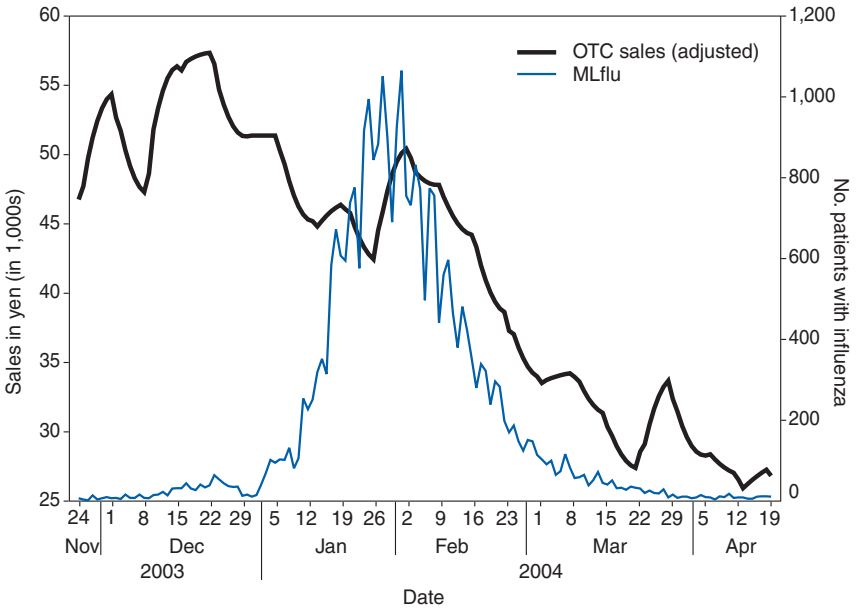
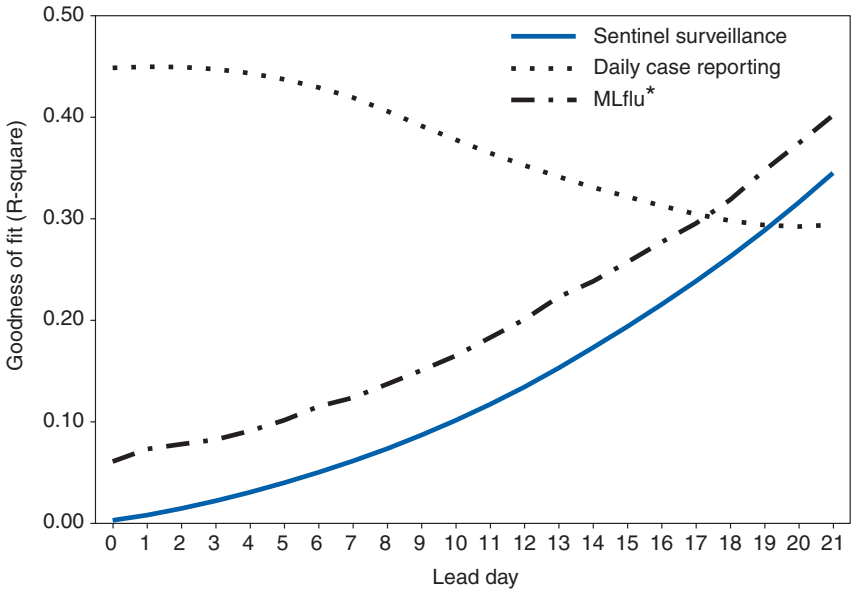


FIGURE 5. Comparison of over-the-counter (OTC) sales per pharmacy (adjusted) with number of patients with influenza reported through the Mailing List–Based Influenza Epidemic Database (MLflu), by date — Japan, November 2003–April 2004*



* MLflu reporting system was activated in November 2003 and officially launched in December 2003 for the 2003–04 influenza season.

FIGURE 6. Goodness of fit (adjusted R-square) between over-the-counter (OTC) sales and other influenza activity surveillances at different OTC lead times for 2003–04 influenza season — Japan, November 2003–April 2004



* Mailing List–Based Influenza Epidemic Database.

tions in influenza activity and variations in when the influenza season began. Variation of lead time of OTC sales to the actual disease incidence by locality has been suggested previously (*5*); therefore, to assess the real situation, smaller geographic areas must be analyzed. The next step to confirm correlations will be to break down the analysis at the prefecture level for 47 prefectures, with and without the effects of sales promotions. However, commuters cross prefecture borders frequently every day, and spatial correspondences or noncorrespondences of OTC sales and physician visits might remain biased in certain instances as a result of inexact geographic data.

The choice of OTC medications selected for this study might have contributed to the outcome. The study was limited to medications used to treat the common cold, which were already grouped in the commercialized sales reporting database. However, in certain cases of early stages of influenza, persons might purchase more symptom-oriented medications (e.g., antipyretic analgesic, antitussive, and antihistaminic medications). To include the entire sales rise attributable to ILI in the analysis, medications in those categories should be examined to formulate a suitable product group to use as precursor for detecting increased ILI as soon as data become available (*5*).

As the copayment proportion of payment for medical care by consumers continues to rise, a gradual move toward self-medication is under way in Japan. Consequently, the potential value of using OTC medication sales data as an indicator of disease outbreaks should continue to rise. However, Japanese consumers are still relatively reluctant to take an active role in decision making regarding their own health care. In addition, the majority of Japanese have easy access to medical care, and the national health insurance system provides a high degree of coverage. As a result, persons who are ill are more likely to visit a clinic at an early stage of illness. The introduction of antiviral agents (e.g., oseltamivir) that require a physician's prescription also has promoted medical assistance–seeking behavior during the influenza season. All of these factors combined might have influenced the study results.

Conclusion

The results presented in this report are tentative. Thorough data cleaning and additional analysis are required before a final decision is made concerning the use of OTC medication sales data as part of a national real-time syndromic surveillance system. Further studies are planned, including a geographic breakdown analysis, analysis with exclusion and inclusion of sales promotion effects (other than the year-end discount promotion), choice of methods for statistical analysis, and analysis taking into account bargain sales and associated promotion types and trial surveillance concerning respiratory symptoms in a limited area.

Acknowledgments

Naruo Saito and physicians participating in MLflu surveillance provided data for the 2003–04 influenza season. Shizuka Matsuoka provided technical support for a poster presentation of this research. The research was funded by a grant from the Ministry of Health, Labour, and Welfare, Japan.

References

1. Matsui T, Takahashi H, Ohyama T, et al. An evaluation of syndromic surveillance for the G-8 Summit in Miyazaki and Fukuoka, 2000. Kansenshogaku Zasshi 2002;76:161–6.
2. Ii M, Ohkusa Y. Estimating price sensitivity of the demand for medical services for minor ailment econometric estimates using information on illness and symptoms. Jpn Econ Rev 2002;55:154–66.
3. Ii M, Ohkusa Y. Should the coinsurance rate be increased in the case of the common cold? An analysis based on an original survey. J Jpn Int Econ 2002;13:353–71.
4. Labrie J. Self-care in the new millennium: American attitudes towards maintaining personal health. Washington, DC: Consumer Healthcare Products Association; 2001.
5. Magruder SF, Lewis SH, Najmi A, Florio E. Progress in understanding and using over-the-counter pharmaceuticals for syndromic surveillance. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):117–22.

Analytic Methods

Public Health Monitoring Tools for Multiple Data Streams

Howard S. Burkom,[1] S. Murphy,[1] J. Coberly,[1] K. Hurt-Mullen[2]
[1]*Johns Hopkins Applied Physics Laboratory, Laurel, Maryland;*
[2]*Montgomery County Department of Health and Human Services, Silver Spring, Maryland*

Corresponding author: Howard S. Burkom, Johns Hopkins Applied Physics Laboratory, 11100 Johns Hopkins Rd., MS 8-220, Laurel, MD 20723. Telephone: 240-228-4361; Fax: 240-228-5950; E-mail: Howard.Burkom@jhuapl.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *In concert with increased concerns regarding both biologic terrorism and new natural infectious disease threats (e.g., severe acute respiratory syndrome [SARS] and West Nile virus), as a result of advances in medical informatics, various data sources are available to epidemiologists for routine, prospective monitoring of public health. The synthesis of this evidence requires tools to find anomalies within various data stream combinations while maintaining manageable false alarm rates.*

Objectives: *The objectives of this report are to establish statistical hypotheses to define the compound multivariate problem of surveillance systems, present statistical methods for testing these hypotheses, and examine results of applying these methods to simulated and actual data.*

Methods: *Canonical problems of parallel monitoring and consensus monitoring are considered in this report. Modified Bonferroni methods are examined for parallel monitoring. Both multiple univariate and multivariate methods are applied for consensus monitoring. A multivariate adaptation of Monte Carlo trials, using the injection of epidemic-curve–like signals in the multiple data streams of interest, is presented for evaluation of the various tests.*

Results: *The Monte Carlo test results demonstrate that the multiple univariate combination methods of Fisher and Edgington provide the most robust detection performance across the scenarios tested. As the number of data streams increases, methods based on Hotelling's T^2 offer added sensitivity for certain signal scenarios. This potential advantage is clearer when strong correlation exists among the data streams.*

Conclusion: *Parallel and consensus monitoring tools must be blended to enable a surveillance system with distributed sensitivity and controlled alert rates. Whether a multiple univariate or multivariate approach should be used for consensus monitoring depends on the number and distribution of useful data sources and also on their covariance structure and stationarity. Strong, consistent correlation among numerous sources warrants the examination of multivariate control charts.*

Introduction

In concert with increased concerns regarding both biologic terrorism and new natural infectious disease threats (e.g., severe acute respiratory syndrome [SARS] and West Nile virus), as a result of advances in medical informatics, data sources are available to epidemiologists for routine, prospective monitoring of public health. Persons who daily monitor these data sources must synthesize recent, disparate evidence to make decisions about possible public health concerns. To synthesize evidence, tools must be used that can find anomalies in single data streams and in various stream combinations while maintaining manageable false alarm rates. In this report, a framework is proposed for this synthesis; basic epidemiologic hypotheses for routine testing are presented and standard algorithms from other fields are adapted for testing them. Implementation principles are offered, but derived multivariate strategies depend on the available data streams, on their individual and covariate behavior, and on the goals and resources of the monitoring organization.

Objectives

Two prototype health monitoring problems are considered. The first problem, the parallel monitoring problem, is the monitoring of time series representing different physical locations (e.g., counties or treatment facilities) that are possibly stratified by other covariates (e.g., syndrome type or age group). The issue is how to maintain sensitivity while limiting the number of alerts that arise from testing the multiple resulting time series. Empirical thresholds can be chosen to achieve

expected background alert rates of *m* per week or per month, where *m* depends on the investigation capacity and tolerance of the monitoring system. The second problem, the consensus monitoring problem, is the testing of a single hypothesis by using multiple sources of evidence. For example, suppose that the null hypothesis is that no current outbreak of gastrointestinal disease exists in the monitored population. Usually, evidentiary time series available to test this hypothesis are syndromic counts of emergency department visits, outpatient clinic office appointments, and sales of over-the-counter remedies. The question is how to make decisions based on all available data sources (i.e., whether to combine results of tests applied to the individual time series or to apply a multivariate algorithm to the collection of visits, appointments, and sales) and how to implement the chosen strategy most effectively.

The purview and resources of a group performing surveillance will determine the importance of these two monitoring problems. This report discusses statistical hypothesis tests specific to these problems, presents an approach for evaluating and comparing these tests, and applies this approach to gain preliminary insights. Algorithms used to implement some of these tests are either implemented or being reviewed in the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) biosurveillance systems (*1*).

Methods

Multiple methods that are presented combine the results of separate hypothesis tests. The computed p values provide test results amenable to these methods in the interval [0, 1]. Whereas a typical null hypothesis is that the current value(s) belong to some assumed probability distribution, the epidemiologic null hypothesis is that disease incidence is at expected levels. Even for tests of clinical diagnosis data, the rejection of the statistical hypothesis might not imply rejection of the epidemiologic one. In examples in this report, nominal thresholds of p = 0.05 and p = 0.01 are used, but in practical circumstances with unknown time series distributions from various data sources, the probabilistic interpretation is frequently unclear, and empirical threshold choices might be necessary for robust detection performance. Historical data sets or simulations can be used to make such threshold choices. Statistical properties of the methods (i.e., parallel monitoring methods, consensus methods, multiple univariate methods, and multivariate methods) depend on constant data variance (and on constant covariance for the multivariate methods); therefore, estimates of variance and other baseline parameters should be updated regularly to adapt to nonstationary behavior. The term "alert" is used to denote a threshold crossing, whereas false "alert" will refer to a threshold crossing unrelated to an increase in incidence of the disease type of concern.

Parallel Monitoring Methods

Multiple testing can lead to uncontrolled alert rates as the number of data streams increases. For example, suppose that a hypothesis test is conducted on a time series of daily diagnoses of influenza-like illness. In a one-sided test, this test results in a statistical value in some distribution that yields a probability *p* that the current count is as large as observed. For a desired Type I error probability of α, the probability is then $(1-\alpha)$ that an alert will not occur in the distribution assumed for data with no underlying aberrations. Therefore, for the parallel monitoring problem of interest, if such tests are applied to n independent data streams, the probability that no background alerts occur is $(1-\alpha)^n$, which increases quickly for practical error rates a (exceeding 0.5 for $\alpha = 0.05$ and $n > 13$).

A common method of controlling this multiple testing problem is to replace the probability threshold "a" with the Bonferroni bound α / N, where N is the number of monitored data streams (*2*). The resulting criterion is sufficient but usually not necessary for ensuring an overall Type I error rate of at most a, and it frequently results in an increased loss of sensitivity. Several published modifications (*3–6*) of the Bonferroni procedure maintain the error rate of "a" with less stringent rejection criteria. Let $P_{(1)},\ldots, P_{(N)}$ be the p-values sorted in ascending order. In a stagewise rejective multiple test (*3*), combined null hypothesis was used if for any j, j = 1...N, i.e.:

$$P_{(j)} < j \cdot a / C \cdot N, \text{ where } C = \sum 1/j.$$

This criterion provides an overall error rate of "a" for C = 1 if the tests are independent (*4*), and this relaxed criterion has been demonstrated to maintain this error rate for multiple common multivariate data sets with positive correlation (*6*). These improvements were applied when it was demonstrated (*7*) that they control the false discovery rate (FDR), or expected ratio of false alerts to the total alert count. For example, FDR methods have been used to monitor CUSUM results of hospital data streams from multiple districts in the National Health Service of the United Kingdom (*8*).

For a large number of data streams and a well-defined signal, Bonferroni modifications might substantially improve detection performance (*9*). However, for a limited number of data streams (e.g., 24–36 counties or treatment facilities) these criteria yield alert rates within approximately 0.1 of the rates from the Bonferroni bound, unless the data are highly correlated.

The manner in which the number of alerts increase with the number of simultaneous tests is presented (Table 1). The

TABLE 1. Counts and ratios* with nominal alerts comparing simple minimum, Bonferroni, and Simes parallel monitoring approaches†

	Minimum		Simes		Bonferroni	
No. streams/ Technique	No. alerts	Rate	No. alerts	Rate	No. alerts	Rate
Poisson data streams: correlation coefficient = 0.0						
1	43	0.046	43	0.046	43	0.046
2	89	0.095	48	0.051	48	0.051
3	135	0.145	47	0.050	47	0.050
4	169	0.181	49	0.052	49	0.052
5	212	0.227	48	0.051	48	0.051
6	249	0.267	48	0.051	48	0.051
7	279	0.299	49	0.052	49	0.052
8	319	0.342	52	0.056	51	0.055
9	362	0.388	49	0.052	49	0.052
10	397	0.425	50	0.054	47	0.050
11	430	0.460	56	0.060	54	0.058
12	461	0.494	56	0.060	54	0.058
Poisson data streams: correlation coefficient = 0.5						
1	45	0.048	45	0.048	45	0.048
2	89	0.095	49	0.052	48	0.051
3	122	0.131	50	0.054	47	0.050
4	141	0.151	51	0.055	45	0.048
5	161	0.172	48	0.051	39	0.042
6	179	0.192	45	0.048	36	0.039
7	201	0.215	45	0.048	41	0.044
8	221	0.237	45	0.048	42	0.045
9	234	0.251	49	0.052	44	0.047
10	243	0.260	48	0.051	43	0.046
11	254	0.272	48	0.051	43	0.046
12	269	0.288	46	0.049	41	0.044
Syndromic data streams: mean correlation coefficient = 0.037§						
1	61.25	0.066	61.25	0.066	61.25	0.066
2	117.95	0.126	75.15	0.080	74.50	0.080
3	170.44	0.182	84.30	0.090	83.28	0.089
4	219.00	0.234	92.03	0.099	90.54	0.097
5	263.95	0.283	96.83	0.104	95.05	0.102
6	305.56	0.327	102.41	0.110	100.09	0.107
7	344.10	0.368	104.26	0.112	101.71	0.109
8	379.81	0.407	110.52	0.118	108.48	0.116
9	412.94	0.442	112.84	0.121	111.75	0.120
10	443.71	0.475	117.74	0.126	117.20	0.125
11	472.33	0.506	120.50	0.129	119.75	0.128
12	499.00	0.534	129.00	0.138	127.00	0.136

* Total = 934 days.
† In computations, three sets of 12 time series were used: uncorrelated Poisson data, Poisson data with pairwise correlation coefficients of 0.5, and large-county syndromic counts that included outbreak events among several of the series.
§ Alert counts averaged over all combinations for each row.

columns are alert rates obtained by combining p values from a control chart method applied to a data set containing 934 days of daily counts. The first two data sets represent independent and correlated background noise. The first was a set of 12 independent Poisson-distributed time series with a mean value of 50. The second data set consisted of 12 series that were similar except that each pairwise correlation coefficient was approximately 0.5. The third data set contained 12 series of syndromic counts from a large (median count: >10 per day) county that included outbreak events among several of the series. Data use agreements prevent disclosure of the actual values, but individual time series were Poisson-like, with variance/mean ratios (excluding one 6-week influenza season) >1.5 in only two of the series and with a mild day-of-week effect present in only one. The mean pairwise correlation coefficient was 0.037, whereas the median was 0.023; therefore, they can be viewed as independent. An exponentially weighted moving average (EWMA) control chart appropriately scaled to return values (*10*) was applied to all three sets of time series. Such charts have been applied widely for hospital surveillance (*11–13*). The first column (Table 1) indicates the number of series combined in each row, increasing from a single series in the top row to all 12 at the bottom. The three columns for each data set indicate the daily alert rates as the fractions of days on which the combined values decline below a = 0.05, computed by three methods. If the p values for the day in ascending order are $P_{(1)} \ldots P_{(N)}$, then

Minimum method: alert if $P_{(1)} < \alpha$,

Simes method: alert if Mininum $(N \cdot P_{(j)} / j) < \alpha$, $j = 1 \ldots N$,

Bonferroni method: alert if $P_{(1)} / N < \alpha$.

The multiple-testing problem is illustrated for each data set by the rapid increase in the alert rate using the minimum p value. For the two simulated data sets, the Simes and Bonferroni alert rates are comparable to formerly published values (*6*). The columns computed from the actual syndromic counts indicate realistic alert rates in the presence of scattered signals. The combination methods control the increase in alert rates with the number of data streams, and for these nearly independent streams, the Simes alert rates are only slightly (Table 1) above the Bonferroni rates. This difference increases as data streams are added, as their correlation increases, and as the alerting threshold "a" is increased. These factors should be considered in the choice of a parallel monitoring method intended to control alert rates.

Regarding the Simes method, only p values below the nominal threshold "a" affect the result. No consensus effect exists (see Consensus Monitoring Methods); the method applied to 10 p values of 0.06 returns 0.06. The Simes criterion (*3–5*) does not specify which of the data streams is anomalous; a procedure (*4*) is to reject the null hypothesis for all streams with p values below the largest one that satisfies the Simes method inequality. More conservative closed-form criteria (*5*) have been developed that indicate which component hypotheses to reject, and the designers of large, complex systems with hundreds of simultaneous data streams should consider these criteria.

Consensus Monitoring Methods

The consensus problem is the combination of clinical and nonclinical evidence to gain sensitivity in disease monitoring. The scope of this report is restricted to prospective monitoring with daily or more frequent syndromic data so that hypothesis testing can apply. The data streams can be combined at more than one level. Multiple univariate and multivariate strategies are considered.

Multiple Univariate Methods

The multiple univariate methods resemble those of the Parallel Monitoring section except that the p values are combined to produce a single p value $p^* = f(p_1,\ldots,p_n)$ with the consensus property that multiple near-critical values can produce a critical one. Multiple such functions are possible; two methods adapted from use in independent, sequential clinical trials are considered. The first method is Fisher's rule (*14*), a function of the product of the p values. The statistic is

$$F = 2\sum_j \ln\left(p_j\right)$$

For independent tests, values of this quantity form a χ^2 distribution with 2n degrees of freedom. As a multiplicative method, it is more sensitive to a few small p values than to a broader number of moderate values. The recommendation is to use the Fisher's Rule if the objective is to extract a single decision on whether to avoid the overall null hypothesis and avoid considering the individual $p_{j,}$.

The second statistic is Edgington's method (*15*), an additive method that calculates the resultant p value as

$$p_E = \frac{S^n}{n!} - \binom{n}{1}\frac{(S-1)^n}{n!} + \binom{n}{2}\frac{(S-2)^n}{n!} - \binom{n}{3}\frac{(S-3)^n}{n!} + \Lambda$$

where S is the sum of the n p values. The summation continues until (S-j) is no longer positive. This additive method is more sensitive to multiple, near-critical values. For approximately 24–36 data streams, this formula cannot be computed accurately. In such cases, the expression

$$(\text{mean}(p) - 0.5) \,/\, (0.2887 \,/\, \sqrt{n})$$

gives a z-score with a Gaussian probability that is a close approximation to this formula (*16*).

If the data streams are independent, Edgington's method gives fewer alerts than Fisher's method at nominal thresholds but is more sensitive to data correlation. Edgington's method is recommended if the number of data streams is modest (e.g., <12 data streams) and the user wants a sensitive consensus indicator in addition to the individual test results. This need has been expressed by epidemiologist users of the ESSENCE biosurveillance systems and is common among large system users who require some summarization but are skeptical of bottom-line results that hide the contributions of individual evidence sources.

An example (Table 2) of the performance of these methods applied to the 2.5 years of syndromic data (Table 1) is presented. Again, the multiple-testing problem is evident, and both the Fisher and Edgington methods control the alert rate growth with the number of data streams. Each entry in the row representing *r* data streams is the mean of the alert rates computed for all combinations of the 12 streams taken *r* at a time. Mean alert counts indicate the number of the combination alerts that are picked up by the individual tests. Edgington's method gives smaller alert rates because of the independence of the data streams, and the majority of the alerts found with Fisher's method are also individual stream alerts. However, because it is an additive method, small single p values do not necessarily cause alerts in Edgington's method; therefore, if the system is not also monitoring single streams, the use of Fisher's method or both methods is recommended.

Multivariate Methods

The motivation to use fully multivariate methods is that they can detect distributed faint outbreak evidence that might be lost in the individual hypothesis tests, and strong correlation among the data sources might also be exploited. Efforts have been confined to multivariate statistical process control (MSPC) charts based on Hotelling's T^2 as applied in surveillance efforts in related fields (*17*). The T^2 statistic can be written as

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

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.

TABLE 2. Comparison of multiple univariate approaches based on Fisher and Edgington combination rules*

No. streams used	No. alerts from any stream	Edgington combination alerts	Edgington excluding single alerts	Fisher combination alerts	Fisher excluding single alerts
1	61.3	61.3	0.0	61.3	0.0
2	118.0	49.7	18.1	69.9	4.2
3	170.4	46.7	16.7	74.8	2.6
4	219.0	46.7	14.2	79.1	2.4
5	264.0	46.8	12.6	83.5	2.0
6	305.6	47.0	11.4	87.2	1.8
7	344.1	47.4	10.3	90.1	1.7
8	379.8	48.0	9.5	93.0	1.5
9	412.9	48.4	8.7	95.9	1.3
10	443.7	49.7	8.2	98.7	1.3
11	472.3	51.6	7.9	100.5	1.1
12	499.0	58.0	10.0	104.0	1.0

* Methods were applied to the 2.5 years of syndromic data. Entries are counts and ratios of days with nominal alerts averaged over all stream combinations for a fixed number of time series in each row.

Whereas T^2 can be viewed as a multidimensional z-score, this method has been generalized to obtain other multivariate control charts. A multivariate EWMA chart (MEWMA) has demonstrated improved run length characteristics (*18*) with health surveillance data (*19*). In MEWMA, the data vector is replaced by the exponentially weighted moving average:

$$Z_j = R\,X + (1\text{-}R)\,Z_{j-1}$$

where R is a diagonal matrix of smoothing coefficients, and the covariance matrix is a scalar multiple of the data covariance matrix S in the usual application where equal smoothing coefficients are used (*18*). This method is demonstrated in the Results section. Analogous multivariate CUSUMs have also been applied to surveillance data, with the Crosier's multivariate cumulative sum (MCUSUM) method (*20*) applied to syndromic data from multiple hospitals (*21*) and Pignatiello's MCUSUM (*22*) applied to yearly, spatially distributed counts of breast cancer incidence (*23*). Whereas the attraction of these multivariate methods is their signal sensitivity, they are also sensitive to noise background changes. Hotelling's T^2 has been described (*24*) as "particularly bad at distinguishing location shifts from scale shifts." Combined univariate methods are directional in that they might be quick to detect shifts in just a few data sources but less sensitive to shifts in more general directions (*23*). These methods are omnidirectional, a property that can be useful in detecting an earlier signal, but can also cause false alerts if a change in the covariance matrix occurs that is irrelevant to any outbreak signal of interest.

Results

Evaluation Methodology

In this section, simulation is used to test the detection performance of some of the consensus monitoring methods discussed previously. In this testing, direct choice of the minimum p value was compared with the Fisher and Edgington multiple univariate methods and with Hotelling's T^2 and Lowry's MEWMA among the multivariate methods.

For background data, eight time series simulating 700 days of syndromic data counts were formed by random draws from a Poisson distribution with a mean of 100. The individual and consensus alerting methods were applied to the unaltered background data to find threshold p values corresponding to three alert rates considered practical for public health monitoring: one alert every 2 weeks, every 4 weeks, and every 6 weeks.

For the signal to be detected, injected cases attributable to a presumed outbreak were added to the background data. These data epicurves were stochastically drawn from an ideal incubation period distribution to test the ability of each method to detect outbreak-like signals (*10*). This procedure differs from the standard method of adding a fixed quantity to the process mean to find the average run length of a control chart (*1*).

The incubation period distribution (*25*) was used to estimate the idealized curve for the expected number of new symptomatic cases on each outbreak day. The lognormal parameters were chosen to give a median incubation period of 3.5 days, consistent with the symptomatology of known weaponized diseases (*26*) and a temporal case dispersion consistent with previously observed outbreaks (*25*).

The stochastic epicurves were drawn from the resulting lognormal distribution. To challenge the algorithms, the number of attributable cases on the peak day of the outbreak was set at one standard deviation of the background data. The total outbreak size N was this peak value divided by the maximum of the lognormal probability density function. Individual incubation periods were then chosen with a set of N random lognormal draws and rounded to the nearest day. The number of cases to add for each day after onset was then the number of draws that were rounded to that day. The evaluation process was to add one of these stochastic epidemic curves to the background time series at a randomly chosen start day beyond an 8-week start-up period for the alerting method and then to run the alerting methods on the time series to determine whether the thresholds were exceeded for each practical alert level. Algorithm performance can be precisely measured in this process because the start and duration of each simulated outbreak are known.

This process was repeated for 100 trials; for each alert level, algorithm sensitivity was measured as the ratio of the trials for which the algorithm output exceeded threshold for that level during the outbreak injection interval.

Tables of sensitivity calculations are indicated for sets of 2, 4, 6, and 8 background data streams in which simultaneous stochastic signals were added to all streams in a set (Table 3). The table represents two background data sets: the three left-hand columns are results for a set with independent data streams, and the three right-hand columns are for a set computed so that each pairwise correlation coefficient among the eight streams was 0.5. Key features among the method comparisons exist.

- The multiple univariate methods indicate uniformly better sensitivity than the multivariate methods among the independent data streams for this transient signal type. For the correlated data, the multivariate methods have better sensitivity for six and eight data streams, an advantage that should continue to increase with the number of streams.

TABLE 3. Sensitivity measured as ratio of number of events detected out of 100 simulated runs, comparing five methods for consensus monitoring*

No. streams/ Sensitivity†/ Technique	Signal with Peak 1σ injected in each stream					
	Correlation: ρ = 0.0			Correlation: ρ = 0.5		
	Background alert interval (days)			Background alert interval (days)		
	14	28	42	14	28	42
2						
Minimum	0.79	0.64	0.55	0.87	0.68	0.55
Fisher	0.83	0.72	0.57	0.83	0.72	0.67
Edgington	0.81	0.67	0.54	0.84	0.70	0.69
Hotel	0.58	0.40	0.28	0.70	0.41	0.30
Lowry	0.68	0.49	0.39	0.68	0.52	0.46
4						
Minimum	0.71	0.53	0.41	0.83	0.70	0.59
Fisher	0.84	0.68	0.63	0.87	0.73	0.65
Edgington	0.90	0.71	0.61	0.86	0.70	0.46
Hotel	0.51	0.27	0.18	0.69	0.60	0.51
Lowry	0.47	0.31	0.20	0.70	0.57	0.47
6						
Minimum	0.89	0.69	0.55	0.86	0.69	0.62
Fisher	0.99	0.96	0.91	0.85	0.69	0.46
Edgington	0.96	0.94	0.94	0.90	0.46	0.31
Hotel	0.57	0.46	0.38	0.81	0.66	0.59
Lowry	0.60	0.49	0.37	0.81	0.70	0.55
8						
Minimum	0.80	0.60	0.40	0.91	0.70	0.60
Fisher	0.92	0.88	0.85	0.93	0.56	0.37
Edgington	0.98	0.92	0.87	0.68	0.34	0.23
Hotel	0.61	0.45	0.44	0.74	0.55	0.48
Lowry	0.60	0.48	0.39	0.74	0.57	0.46

* Background data were random Poisson time series with means of 100 and with fixed pairwise correlation coefficients. Stochastic point-source epicurves, with a peak value of 1 background standard deviations, were injected into all data streams for each run.
† Sensitivity as a function of background alert interval.

TABLE 4. Sensitivity measured as ratio of number of events detected out of 100 simulated runs, comparing five methods for consensus monitoring*

No. streams/ Sensitivity†/ Technique	Signal with Peak 2σ injected in each stream					
	Correlation: ρ = 0.0			Correlation: ρ = 0.5		
	Background alert interval (days)			Background alert interval (days)		
	14	28	42	14	28	42
2						
Minimum	0.93	0.85	0.74	0.99	0.90	0.81
Fisher	0.92	0.82	0.70	0.89	0.82	0.77
Edgington	0.69	0.49	0.44	0.75	0.50	0.40
Hotel	0.40	0.26	0.15	0.42	0.25	0.21
Lowry	0.35	0.24	0.16	0.30	0.21	0.16
4						
Minimum	0.74	0.61	0.51	0.92	0.80	0.72
Fisher	0.77	0.50	0.43	0.72	0.62	0.50
Edgington	0.61	0.39	0.30	0.57	0.27	0.18
Hotel	0.26	0.17	0.10	0.37	0.21	0.13
Lowry	0.21	0.13	0.10	0.31	0.13	0.07
6						
Minimum	0.75	0.57	0.49	0.90	0.78	0.71
Fisher	0.70	0.55	0.41	0.60	0.41	0.27
Edgington	0.59	0.46	0.38	0.45	0.23	0.15
Hotel	0.37	0.22	0.17	0.43	0.20	0.18
Lowry	0.34	0.18	0.14	0.30	0.15	0.13
8						
Minimum	0.69	0.49	0.32	0.88	0.73	0.63
Fisher	0.69	0.45	0.39	0.59	0.33	0.22
Edgington	0.65	0.40	0.25	0.42	0.21	0.14
Hotel	0.44	0.29	0.24	0.37	0.25	0.18
Lowry	0.34	0.19	0.15	0.30	0.18	0.11

* Background data were random Poisson time series with means of 100 and with fixed pairwise correlation coefficients. Stochastic point-source epicurves, with a peak value of 2 background standard deviations, were injected into one data stream for each run.
† Sensitivity as a function of background alert interval.

- For the independent data streams, the controlled alert rates of the Fisher and Edgington methods yield sensitivity advantages over the simple minimum method, increasing with the number of data streams.
- Whereas the Edgington method has effective sensitivity for independent data, the correlation degrades its performance critically at the stricter alert levels. The Fisher method indicates the same effects to a lesser degree.

These observations suggest that the choice of multiple univariate or multivariate methods should depend on the number of data streams monitored and on their correlation characteristics. For a large enough collection of highly correlated data streams, the alert rate for even the Fisher method would suffer, but multivariate methods might retain sensitivity.

The same comparisons are demonstrated (Table 4) but with the signal injected into only one stream (i.e., the consensus factor in these methods offers no advantage). For the single signal results in this table, the peak value was increased to twice the standard deviation of the background data. The multivariate methods fared poorly in these runs, even in the correlated data. For the consistently correlated data, the sustained sensitivity of the simple minimum suggests two principles for system designers. First, attempts to remove cross-correlation with modeling of known features (seasonal, day-of-week effects) should be tried in the univariate algorithms so that the combination methods will be more robust. Second, the parallel monitoring methods should be employed to remain aware of individual algorithm outputs.

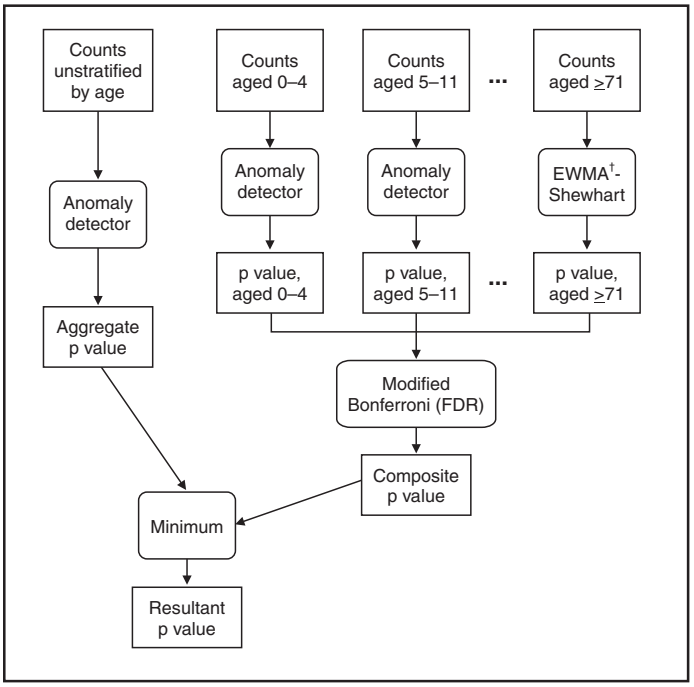
Conclusion

This report presents and examines statistical tools for systematizing the prospective monitoring of public health by using various spatially distributed time series data. The modeled multiple-stream scenarios illustrate the need to blend the parallel and consensus monitoring tools to achieve a system with distributed sensitivity and controlled alert rates.

Parallel monitoring methods in which modified Bonferroni criteria are used are important to retain detection power at controlled alert rates. A flowchart (Figure) is presented, envisioned as part of a distributed, syndromic monitoring system diagram, for a method to combine agestratified and unstratified monitoring methods for increased sensitivity to localized outbreaks, which have proven difficult (*27*) for large systems to detect.

The choice of a multiple univariate or multivariate approach for consensus monitoring depends on the number and distribution of useful data sources and also on their covariance structure and stationarity. Strong, consistent correlation among multiple sources warrants the examination of multivariate control charts. However, whereas explicitly multivariate methods offer the possibility of increased sensitivity, careful attention must be given to data interrelationships when using them. These charts must be proven sufficiently robust to customary variation in the correlation among data streams to ensure that the signals are not overwhelmed by multivariate noise.

FIGURE. Schematic illustrating parallel monitoring approach for increased, distributed sensitivity*



*Age in years.
†Exponentially weighted moving average. EWMA-Shewhart describes a composite control chart already in use for hospital surveillance.

References

1. Lombardo J, Burkom H, Pavlin J. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II), Framework for Evaluating Syndromic Surveillance Systems. In: Syndromic surveillance: report from a national conference, 2003. MMWR 2004;53(Suppl):159–65.
2. Ryan TP. Statistical methods for quality improvement. New York, NY: John Wiley & Sons; 1989:64.
3. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni Test. Biometrika 1988;75:383–6.
4. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika 1986;73:751–4.
5. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800–2.
6. Sarkar SK, Chang C-K. The Simes method for multiple hypothesis testing with positively dependent test statistics. J Amer Stat Assoc 1997;92:1601–8.
7. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc B 1995;57:289–300.
8. Marshall C, Best N, Bottle A, Aylin P. Statistical issues in prospective monitoring of health outcomes across multiple units. J Royal Stat Soc A 2004;167(Pt 3):541–59.
9. Miller CJ, Genovese C, Nichol RC, et al. Controlling the false discovery rate in astrophysical data analysis. Astronomical Journal 2001;122:3492.
10. Burkom H, Hutwagner L, Rodriguez R. Using point-source epidemic curves to evaluate alerting algorithms for biosurveillance. In: Proceedings of the 2004 American Statistical Association, Statistics in Government Section [CD-ROM]; 2004. Toronto, Canada: American Statistical Association; January 2005.
11. Morton AP, Whitby M, McLaws M-L, et al. The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. J Qual Clin Prac 2001;21:112–7.
12. Benneyan JC. Statistical quality control methods in infection control and hospital epidemiology: introduction and basic theory. Infect Control Hosp Epidiol 19;4(Pt 1):194–214.
13. Benneyan JC. Statistical quality control methods in infection control and hospital epidemiology: chart use, statistical properties, and research issues. Infect Control Hosp Epidemiol 19;4(Pt 2):265–83.
14. Bauer P, Kohne K. Evaluation of experiments with adaptive interim analyses. Biometrics 1994;50:1029–41.
15. Edgington ES. An additive method for combining probability values from independent experiments. J Psychol 1972;80:351–63.
16. Edgington ES. A normal curve method for combining probability values from independent experiments. J Psychol 1972;82:85–9.
17. Ye N, Cheng Q, Emran S, Vilbert S. Hotelling's T^2 multivariate profiling for anomaly detection. In: Proceedings of the 2000 IEEE Workshop on Information Assurance and Security; 2000. West Point, NY: IEEE; June 2002.
18. Lowry CA, Woodall WH. A multivariate exponentially weighted moving average control chart. Technometrics 1992;34:46–53.
19. Hong B, Hardin M. A report of the properties of the multivariate forecast-based processing scheme. In: Proceedings of the Joint Statistical Meetings; 2004. Toronto, Canada: American Statistical Association; August 2004.
20. Crosier RB. Multivariate generalizations of Cumulative Sum Quality-Control Schemes. Technometrics 1988;30:291–303.
21. Stoto M. Multivariate methods for aberration detection: a simulation report using the District of Columbia's syndromic surveillance data. In: Proceedings of the 2004 National Syndromic Surveillance Conference [Oral Presentation]. Boston, MA; 2004.
22. Pignatiello JJ, Runger GC. Comparisons of multivariate CUSUM charts. J Qual Technol 1990;22:173–86.

23. Rogerson PA, Yamada I. Monitoring change in spatial patterns of disease: comparing univariate and multivariate Cumulative Sum Approaches. Stat Med 2004;23:2195–214.
24. Hawkins D. Multivariate quality control based on regression-adjusted variables. Technometrics 1991;33:61–75.
25. Sartwell PE. The distribution of incubation periods of infectious disease. Amer J Hyg 1950;51:310–8. Am J Epidemiol 1995(Reprint); 141:301–18.
26. US Army Medical Research Institute of Infectious Diseases. Medical Management of Biological Casualties. 4th ed. Ft. Detrick, MD: US Army Medical Research Institute of Infectious Diseases; September 2000.
27. Balter S. Three years of emergency department gastrointestinal syndromic surveillance in NYC: what have we found? In: Proceedings of the National Syndromic Surveillance Conference; 2000 [Oral Presentation]. Boston, MA; 2004.

Use of Multiple Data Streams to Conduct Bayesian Biologic Surveillance

Weng-Keen Wong,[1] G. Cooper,[1] D. Dash,[2] J. Levander,[1] J. Dowling,[1] W. Hogan,[1] M. Wagner[1]
[1]*RODS Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania;*
[2]*Intel Research, Santa Clara, California*

Corresponding author: Greg Cooper, Center for Biomedical Informatics, University of Pittsburgh, 8084 Forbes Tower, 200 Lothrop Street, Pittsburgh, PA 15213. Telephone: 412-647-7113; Fax: 412-647-7190; E-mail: gfc@cbmi.pitt.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Emergency department (ED) records and over-the-counter (OTC) sales data are two of the most commonly used sources of data for syndromic surveillance. The majority of detection algorithms monitor these data sources separately and either do not combine them or combine them in an ad hoc fashion. This report outlines a new causal model that combines the two data sources coherently to perform outbreak detection.*

Objectives: *This report describes the extension of the Population-wide Anomaly Detection and Assessment (PANDA) Bayesian biologic surveillance algorithm to combine information from multiple data streams. It also outlines the assumptions and techniques used to make this approach scalable for real-time surveillance of a large population.*

Methods: *A causal Bayesian network model used previously was extended to incorporate evidence from daily OTC sales data. At the level of individual persons, the actions that result in the purchase of OTC products and in admission to an ED were modeled.*

Results: *Preliminary results indicate that this model has a tractable running time consisting of 209 seconds for initialization and approximately 4 seconds for every hour's worth of ED data, as measured on a Pentium-4 three-Gigahertz machine with two Gigabytes of RAM.*

Conclusion: *Preliminary results for surveillance using a new Bayesian algorithm that models the interaction between ED and OTC data are positive regarding the run time of the algorithm.*

Introduction

Syndromic surveillance systems routinely monitor data concerning sales of over-the-counter (OTC) medications and records of chief complaints of persons reporting to hospital emergency departments (EDs) (*1,2*). If a disease outbreak occurs in a region, its effects are often expected to be seen in both data sources (*3,4*). Although ED and OTC data sources contain the signal of an outbreak, detection algorithms generally monitor each data type separately, which limits the surveillance system's detection capabilities.

Persons with initial symptoms of disease are assumed often to treat themselves before seeking medical care (*5–7*). Consequently, an outbreak signal is expected to appear first in OTC medication sales data and then later in ED data (*3*). Although an early signal is expected to appear in OTC data, this signal typically will be weak. Furthermore, OTC data are often reported as a univariate time series in which regional sales volume for a particular category of product (e.g., sales of cough medications) is recorded daily, and case-level data about sales transactions are not available. If such data were to be available, a multivariate detection algorithm could be applied, and the additional information about each transaction might be exploited to improve detection capability. In contrast, ED chief complaint data do contain case-level information (e.g., admission date and time, age, sex, home ZIP code, and chief complaint) about each patient admitted. These data can be used to improve a detection algorithm's capability by identifying known spatial, temporal, demographic, and symptomatic patterns of the disease in the data. Nevertheless, an outbreak signal typically is expected to appear later in ED data than in OTC data.

Developing a detection algorithm that integrates the two data sources would combine the advantages of both data types and might help monitors determine that an outbreak is occurring. The key difficulty with this data-fusion approach is in measuring the relationship between data sources when an outbreak occurs. Correlations between OTC and ED data

during an outbreak cannot be estimated from data because because no training data exist that capture the effects of a large-scale epidemic on these data sources during the same period. The majority of existing data-fusion approaches treat data sources as independent (*8–10*). However, despite the absence of training data, a certain amount of background knowledge does exist about the plausible relationship between OTC and ED data for particular diseases that can be used to model the actions of persons that result in possible OTC medication purchases and ED admissions.

This report extends the Population-wide Anomaly Detection and Assessment (PANDA) algorithm described previously (*11*). That algorithm used a causal Bayesian network to model a population of persons. The original PANDA algorithm was designed to monitor only ED chief complaint data. This report enhances PANDA to simultaneously monitor two data sources of different granularity: aggregated regional counts for OTC sales and multivariate ED records for individual patients. Although the Bayesian network can be used to model the effects of any noncontagious disease outbreak (i.e., those not involving person-to-person transmission) in a geographic area (*11*), this report focuses on monitoring to detect an outdoor, aerosolized release of an anthrax-like agent within a countywide region.

Methods

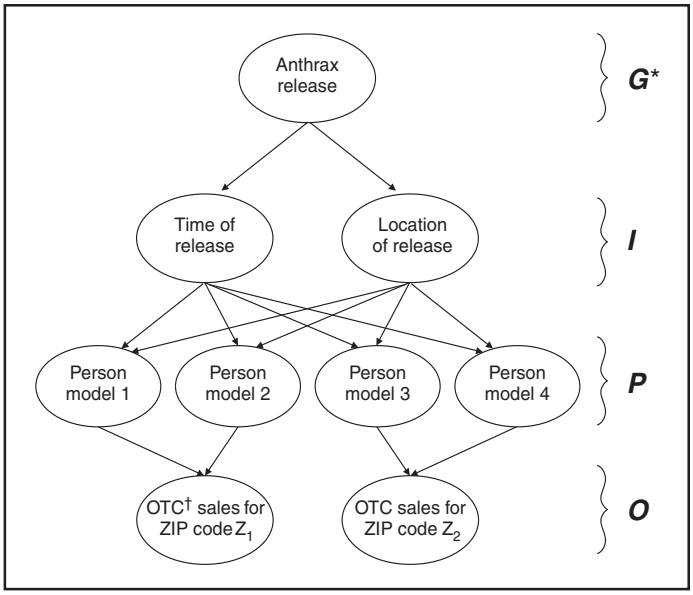
The key aspect of the PANDA algorithm is the explicit modeling of each person in the population as a subnetwork of the overall causal Bayesian network. In this report, these persons are referred to as "person models"; however, models could be generalized to entities that provide information about disease outbreaks (e.g., biosensors or livestock). The advantage of modeling each person in a population is that it allows the algorithm to have substantial representational power and flexibility. Having a subnetwork for each person enables users to represent different types of background knowledge coherently in the model. For instance, to model an aerosolized anthrax release, designers can build into the model a temporal assumption about the incubation time of anthrax and a spatial assumption that the release will take the shape of a downwind plume (*12,13*). In addition to the power in representing prior knowledge, modeling an entire population allows the model to combine spatial, temporal, demographic, and symptomatic evidence to derive a posterior probability of a disease outbreak. With respect to representational flexibility, individual modeling allows new types of knowledge and evidence to be readily incorporated into the model. For example, radiology reports (which are especially useful indicators of an anthrax attack) can easily be added to the model as a new evidential variable. Finally, the majority of the background knowledge of the characteristics of respiratory anthrax disease is at an individual rather than a population level.

Generic Model

In this model, the causal Bayesian network is used to detect an outbreak caused by an aerosolized release of an anthrax-like biologic agent (Figure 1). This network is an instantiation of a generic model for infectious but noncontagious diseases. The generic model can be partitioned into four sets of nodes, as follows:

- **Global nodes** (***G***). These represent global features common to all persons. Included in this set is a target node (*T*) indicating the occurrence of a disease outbreak. Monitoring the state of the target node (e.g., anthrax release) permits users to derive an updated posterior probability that a disease outbreak is occurring. The larger set could include other variables (e.g., the national terror alert level or information about major local sports events or political conventions).
- **Interface nodes** (***I***). For the sake of simplicity, only time and location of release are included in the model outlined in this report. As the model is refined, other nodes (e.g., amount of release, type of anthrax powder, and meteorologic information) will be added to the interface layer.

FIGURE 1. The anthrax model used by the Population-wide Anomaly Detection and Assessment (PANDA) algorithm



* ***G*** = global nodes (i.e., features common to all persons), ***I*** = interface nodes (i.e., time and location of release), ***P*** = person nodes (i.e., evidence noted for each person), and ***O*** = populationwide evidence nodes (i.e., evidence aggregated for a particular set of persons).
† Over-the-counter medications.

- **Person models** ($\boldsymbol{P}$) = $\{P_1, P_2, \ldots, P_n\}$. For each person, evidence observed on an individual basis will be entered.
- **Populationwide evidence nodes** ($\boldsymbol{O}$) = $\{O_1, O_2, \ldots, O_m\}$. These represent evidence aggregated for a particular set of persons (e.g., those living in a particular geographic region). In the model outlined in this report, the set $\boldsymbol{O}$ consists of aggregate OTC sales of a particular type (e.g., cough medication sales) over a particular ZIP code.

The generic model makes the following three assumptions that are intended to facilitate the calculation of the probability of the target node T given the evidence:

- **Assumption 1:** $\forall i, P_i \perp\!\!\!\perp P_{-i} \mid \boldsymbol{I}$ and any arc between a node I in $\boldsymbol{I}$ and a node X in some person model P_i is oriented from I to X. The symbol $\perp\!\!\!\perp$ means "is independent of" and the notation P_{-i} means all person models except P_i. The fact that the model does not condition on the populationwide evidence in $\boldsymbol{O}$ might appear counterintuitive, but the evidence in $\boldsymbol{O}$ is not used in calculating the contribution of the evidence in $\boldsymbol{P}$ to the posterior probability.
- **Assumption 2:** $\boldsymbol{G} \perp\!\!\!\perp \boldsymbol{P} \mid \boldsymbol{I}$ and any arc between a node G in $\boldsymbol{G}$ and a node I in $\boldsymbol{I}$ is oriented from G to I.
- **Assumption 3:** The person models P_i contain arcs that are oriented toward the populationwide evidence nodes in $\boldsymbol{O}$.

Thus, from Assumptions 1 to 3, arcs are not allowed directly between the person models. For noncontagious diseases that might cause outbreaks, these assumptions are reasonable when $\boldsymbol{I}$ contains all the factors that influence the status of an outbreak disease in persons in the population. For example, in the case of a biologic terrorist–released agent, such information includes the time and location of release of the agent. A key characteristic of nodes in $\boldsymbol{I}$ is that they have arcs to the nodes in one or more person models, and they induce the conditional independence relationships described in Assumptions 1 and 2. In contrast, for contagious diseases (i.e., those involving person-to-person transmission), arcs are needed between person models because persons can infect each other. Once these three assumptions no longer hold, inference becomes much more computationally expensive, and the current optimizations that allow PANDA to run efficiently do not hold.

Anthrax Model

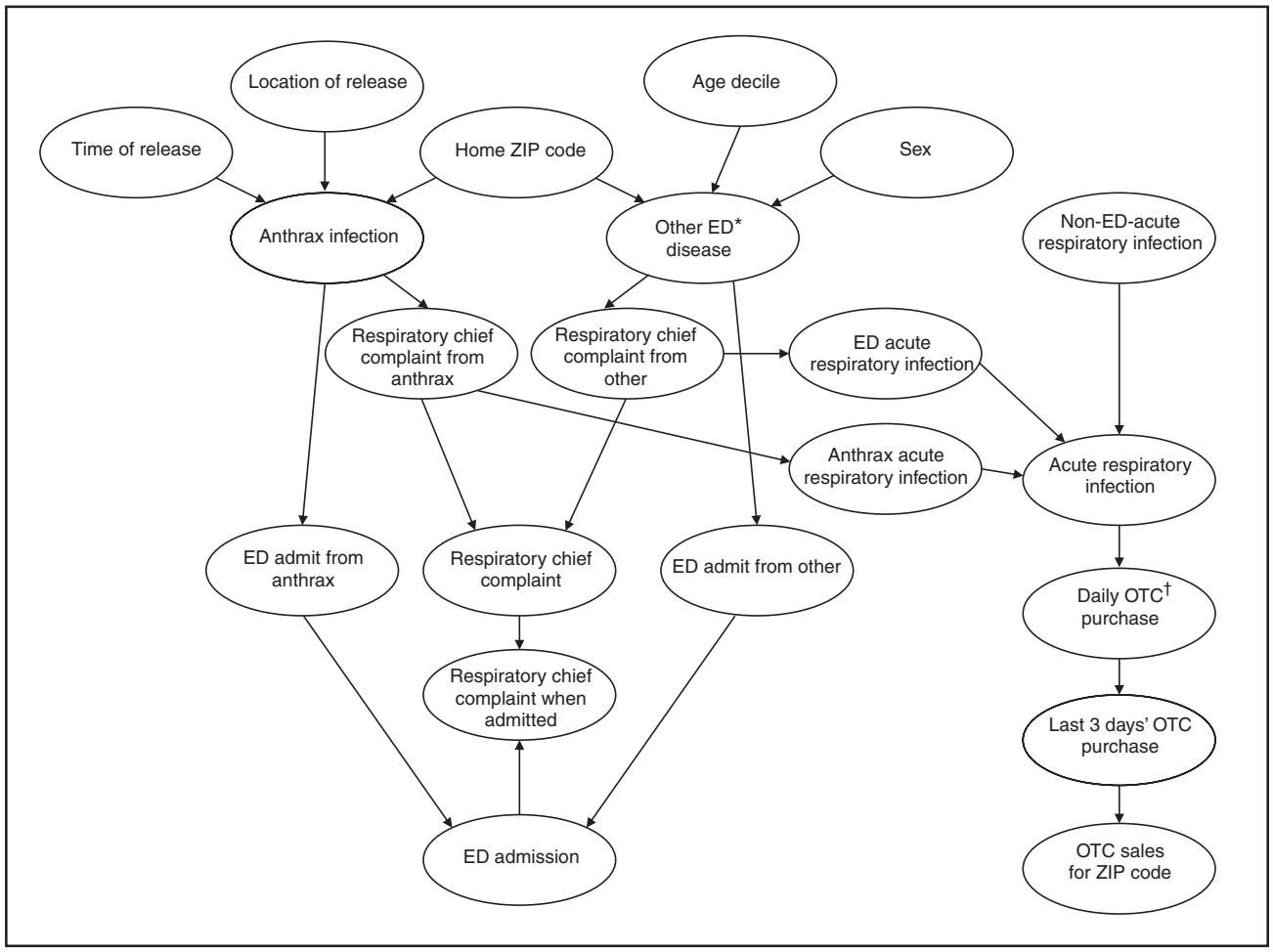
In this prototype model, the simplifying assumption is made that persons living in a particular ZIP code purchase OTC medications only within that ZIP code. Consequently, the OTC purchases in each ZIP code are independent of each other. Because this assumption will be violated in the event of a large-scale biologic terrorist attack, this issue will be addressed subsequently; however, this assumption was used for the initial prototype. One straightforward way to avoid this assumption is to model the populationwide evidence $\boldsymbol{O}$ as OTC sales for an entire region (Allegheny County, in this model); however, doing so would lose spatial information that might be helpful in detecting an outbreak. The OTC Sales for ZIP code nodes are integer-valued nodes representing the aggregate number of units of OTC medications sold throughout the specified ZIP code. These nodes are considered to be observed nodes because they are instantiated with values from the OTC data.

The structure of the person model (Figure 2) was created on the basis of expert judgment. Certain nodes in this model are temporal and modeled for 3 days; that duration was selected for use in the prototype as the shortest period of time meaningful for modeling a disease outbreak. For modeling an anthrax outbreak, this period will be extended to 2 weeks. Evidence nodes whose values are observed in the ED data include home ZIP code, age decile, sex, respiratory chief complaint when admitted, and ED admission.

The parameters of certain nodes in this model were estimated from a training set consisting of 1 year's worth of HIPAA compliant ED patient data from certain hospitals in western Pennsylvania during 2000 or from a training set of OTC data from 2004 (*2*). The parameters of other variables were obtained from U.S. Census data about the region. Respective probabilities (whether prior or conditional) for the rest were derived as a logical function of their parents or assessed subjectively on the basis of the published literature and general knowledge about infectious diseases. Because of space restrictions, this report describes only six nodes that differ from the original PANDA model (*1*):

- **Anthrax acute respiratory infection node.** This node indicates whether a person has an acute respiratory infection (ARI) attributable to anthrax. This node models the presence of ARI during a 3-day period, similar to the anthrax infection node.
- **ED acute respiratory infection node.** This node indicates whether a person who reports to an ED has an acute respiratory infection (ARI) attributable to an ED disease other than anthrax. This is a 3-day temporal node.
- **Non-ED acute respiratory infection node.** This node indicates whether a person who does not report to an ED has an acute respiratory infection (ARI) attributable to a disease other than anthrax. This is a 3-day temporal node.
- **Acute respiratory infection.** This node indicates whether a patient has ARI. This node is a "logical or" of three Boolean nodes: anthrax acute respiratory infection, ED

FIGURE 2. The person model used by the Population-wide Anomaly Detection and Assessment (PANDA) algorithm



* Emergency department.
† Over the counter. OTC sales for ZIP code is part of the populationwide evidence layer. Although the OTC sales for ZIP code is only shown with input from a single person model, it actually has arcs from all person models in a given ZIP code (see Figure 1).

acute respiratory infection, and non-ED acute respiratory infection.

- **Daily OTC purchase.** This node captures the probability of purchasing a respiratory-related OTC medication (e.g., a cough medication) today, yesterday, the day before yesterday, or never. The conditional probability of this variable was derived by using OTC sales of respiratory-related medications in the countywide region being modeled.
- **Previous 3 Days' OTC purchase.** This Boolean node describes whether a respiratory-related OTC medication was purchased during the previous 3 days by the person being modeled (*2*).

Inference

The goal of PANDA is to monitor the state of the target node T, which captures the probability of a disease outbreak occurring. PANDA calculates the posterior probability of T as new ED and OTC data arrive. Let $\boldsymbol{o}$ be the set of populationwide evidence (i.e., OTC sales volume for each ZIP code in the countywide region). Similarly, let $\boldsymbol{e}$ be the collective set of evidence from individual persons (i.e., case information from persons who recently visited EDs in the region). From ED data, demographic data from the most recent U.S. Census can be used to infer information about persons who have not been recently admitted to an ED. The sets $\boldsymbol{e}$ and $\boldsymbol{o}$ can be expressed as follows:

$$\boldsymbol{e} = \{X = e : X \in P_i, P_i \in \boldsymbol{P}\} \text{ and } \boldsymbol{o} = \{X = o : X \in O_j, O_j \in \boldsymbol{O}\}.$$

The goal of the algorithm is to calculate the probability of a disease outbreak given the OTC and the ED data. Mathematically, this objective is expressed as:

Equation 1. $P(T \mid \boldsymbol{o}, \boldsymbol{e}) = k \cdot P(\boldsymbol{o}, \boldsymbol{e} \mid T) \cdot P(T)$,

where the proportionality constant is

$$k = 1 \Big/ \sum_T P(\boldsymbol{o}, \boldsymbol{e} \mid T) \cdot P(T).$$

The term $P(T)$ can be calculated by using Bayesian network (BN) inference on just the portion of the model that includes $\boldsymbol{G}$. Performing BN inference over just the nodes in $\boldsymbol{G}$ is much preferable to inference over all the nodes in X, because the number of nodes in X is approximately 10^7 in the current model. Because the set $\boldsymbol{I}$ renders the nodes in $\boldsymbol{P}$ (including $\boldsymbol{e}$) independent from the nodes in $\boldsymbol{G}$, the term $P(\boldsymbol{o}, \boldsymbol{e} \mid T)$ can be derived as follows:

Equation 2. $P(\boldsymbol{o}, \boldsymbol{e} \mid T) = \sum_i P(\boldsymbol{o}, \boldsymbol{e} \mid \boldsymbol{I} = \boldsymbol{i}) \cdot P(\boldsymbol{I} = \boldsymbol{i} \mid T)$

The above summation can be very demanding computationally, because $\boldsymbol{e}$ usually contains many nodes. The term $P(\boldsymbol{o}, \boldsymbol{e} \mid \boldsymbol{I} = \boldsymbol{i})$ can be factored as follows:

$$P(\boldsymbol{o}, \boldsymbol{e} \mid \boldsymbol{I}) = P(\boldsymbol{o} \mid \boldsymbol{e}, \boldsymbol{I}) \cdot P(\boldsymbol{e} \mid \boldsymbol{I})$$

The term $P(\boldsymbol{o} \mid \boldsymbol{e}, \boldsymbol{I})$ can be considered to be the conditional contribution of the OTC evidence to the posterior probability $P(T \mid \boldsymbol{o}, \boldsymbol{e})$, whereas the term $P(\boldsymbol{e} \mid \boldsymbol{I})$ can be considered as the conditional contribution of the ED evidence.

Incorporating ED Evidence

The term $P(\boldsymbol{e} \mid \boldsymbol{I})$ is calculated efficiently by using equivalence classes and incremental updating (*11*). Space can be saved and inference time reduced by using equivalence classes to group persons who are indistinguishable on the basis of their evidence. Persons in the same equivalence class have the same values for the home ZIP code, age decile, sex, respiratory chief complaint when admitted, and ED admission nodes. Incremental updating dramatically reduces inference time by avoiding the necessity to calculate $P(\boldsymbol{e} \mid \boldsymbol{I})$ for the entire population every time new ED data arrive.

Incorporating OTC Evidence

For OTC evidence to be incorporated into the posterior probability, the probability $P(\boldsymbol{o}, \boldsymbol{e} \mid \boldsymbol{I})$ must be computed. If, for the purpose of this initial prototype, the simplifying assumption is made that persons living in a specific ZIP code purchase OTC medications only within their home ZIP code, then OTC purchases for each ZIP code are independent of each other, conditioned on the nodes in $\boldsymbol{I}$. OTC purchases within a given equivalence class are also assumed to be capable of being modeled with a binomial distribution, and the distribution of OTC purchases within a given ZIP code is assumed to be capable of being modeled as the sum of independent binomial distributions of the equivalence classes within that ZIP code. Let $\boldsymbol{Z}$ be the set of all ZIP codes in the region under surveillance, and let O_{Z_k} be the variable representing the OTC cough-medication sales volume for ZIP code Z_k. Furthermore, let the observed OTC cough-medication sales volume during the previous 3 days for ZIP code Z_k be o_{Z_k}. The ZIP-code independence assumption allows the probability $P(\boldsymbol{o} \mid \boldsymbol{e}, \boldsymbol{I})$ to be factored as follows:

Equation 3. $$P(\boldsymbol{o} \mid \boldsymbol{e}, \boldsymbol{I}) = \prod_{Z_k \in \mathbf{Z}} P(O_{Z_k} = o_{Z_k} \mid \boldsymbol{e}, \boldsymbol{I})$$

To model the probability $P(O_{Z_k} = o_{Z_k} \mid \boldsymbol{e}, \boldsymbol{I})$, which corresponds to the probability of the OTC cough-medication data for ZIP code Z_k, the contribution of the equivalence classes that belong to this ZIP code need to be determined. Let $\boldsymbol{\Omega}_{Z_k}$ be the set of equivalence classes that have home ZIP codes equal to Z_k. For the sake of clarity, assume only one person model exists (Figure 2) that is common to all persons in the population; in general, as many person models can exist as is useful to represent different types of persons. An equivalence class Q_j is defined by a tuple $\boldsymbol{e}_j$, which is a (possibly incomplete) set of values for the evidence nodes in the person model. An example of such a tuple is {home ZIP code = 15213, age decile = 2, sex = male, respiratory chief complaint when admitted = true, ED admission = today}.

The OTC sales volume for each equivalence class Q_j is modeled as a binomial distribution with parameters n_j and p_j. The parameter n_j is simply the number of persons currently within the equivalence class. The second parameter p_j represents the probability of an individual in the equivalence class making an OTC cough-medication purchase within the previous 3 days. This probability is calculated by conditioning on the evidence $\boldsymbol{e}_j$ that defines the equivalence class and computing the probability that the last 3 days' OTC purchase equals *true* by performing Bayesian network inference on the person model.

Distributions for each equivalence class in $\boldsymbol{\Omega}_{Z_k}$ are combined by using a normal approximation to the binomial distribution (*14*) to represent the OTC sales distribution for each equivalence class. The normal approximation is needed because no efficient way exists to derive the distribution over the sum of binomial variates directly. In contrast, deriving the distribution over the sum of normal variates is straightforward. With this approximation, a binomial distribution with parameters n_j and p_j can be converted into a normal distribution with mean $n_j p_j$ and variance $n_j p_j (1-p_j)$. The distribution for the entire ZIP code is therefore represented as a

normal distribution $N_{Z_k}(\mu_{Z_k}, \sigma^2_{Z_k})$ with mean μ_{Z_k} and variance $\sigma^2_{Z_k}$ that are as follows:

$$\mu_{Z_k} = \sum_{Q_j \in \Omega_{Z_k}} n_j p_j \text{ and } \sigma^2_{Z_k} = \sum_{Q_j \in \Omega_{Z_k}} n_j p_j (1-p_j)$$

Finally, to derive the probability of observing the OTC sales for each ZIP code, the following is computed:

$$P(O_{Z_k} = o_{Z_k} \mid \boldsymbol{e}, \boldsymbol{I}) = \int_{O_{Z_k}-0.5}^{O_{Z_k}+0.5} N_{Z_k}(x; \mu_{Z_k}, \sigma^2_{Z_k}) dx.$$

Results and Discussion

The average running times for the ED model described previously (*1*) and for the ED+OTC model described in this report were compared, with both models operating on a Pentium-4 three-Gigahertz machine with two Gigabytes of RAM. The initialization time for the ED+OTC model (209.7 seconds) is nearly four times that of the ED model (45.3 seconds). However, when actually processing the 3-day window of data, the models take approximately the same amount of time (3.0 seconds for the ED model and 3.9 seconds for the ED+OTC model). For example, suppose PANDA is run in real-time starting at t = 72 hours. The algorithm first performs an initialization phase and then processes the data from t = 0 to t = 71 hours. When data accumulate for the next hour, PANDA moves its window of cases forward by 1 hour and analyzes the data from t = 1 to t = 72 hours. For each subsequent 72-hour window, the running time of PANDA is approximately 4 seconds. Even with the OTC extension, the new PANDA model is capable of processing all current data well before the next hour's worth of data arrives. These timing results indicate that the method is practical for real-time biosurveillance. The false-positive rate and detection time of this approach will be evaluated by using data created by injecting simulated anthrax cases into existing ED and OTC data streams.

Related Work

The algorithms used in syndromic surveillance have been described previously (*15*). Two approaches have been suggested (*10*) for using a spatial scan statistic (*16*) to combine multiple data sources in performing syndromic surveillance. The first method treats the multiple sources as covariates. The spatial scan statistic is calculated by using the sum of the observed counts from the data sources and the sum of the expected counts. One of the main problems of this approach is that a data source with a large count might mask data sources with smaller counts. An alternative approach is to calculate the log likelihood ratio for each data source and sum these ratios to form the scan statistic, similar to an approach taken previously (*11*). Combining multiple univariate statistical process control methods by using a consensus method technique (*9,17,18*) has also been suggested. However, the consensus method assumes independence among the data sources. To capture the correlation between data streams, multivariate methods (e.g., Hotelling's T^2 [*19*], MCUSUM [*20*] and MEWMA [*21*]) have been used on multiple univariate signals (*17*). In these multivariate methods, the covariance matrix for the data streams is typically estimated from a baseline period. If the covariance matrix changes substantially during an outbreak, then this estimate will not capture the actual relationship between the data streams during an outbreak.

Conclusion

This report has introduced a data-fusion approach to biosurveillance that is based on modeling the effects of an outbreak disease (excluding diseases associated with person-to-person transmission) on individual persons in the population. The causal Bayesian network (*1*) was extended to incorporate evidence from both ED and OTC data by modeling the actions of individuals in terms of purchasing OTC products and visiting the ED. This data-fusion model can process a 3-day window of ED and OTC data in approximately 4 seconds, making it a feasible algorithm for real-time surveillance. In future work, the model will be extended to cover a 2-week period and relax the ZIP code independence assumption. A thorough and high-fidelity evaluation of the detection algorithm will be performed that will involve injecting simulated anthrax cases into actual ED and OTC data streams.

Acknowledgments

This research was supported by grants from the National Science Foundation (IIS-0325581), the Department of Homeland Security (F30602-01-2-0550), and the Pennsylvania Department of Health (ME-01-737).

References

1. Espino J, Wagner M, Szczepaniak C, et al. Removing a barrier to computer-based outbreak and disease surveillance: The RODS Open Source Project. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):32–9.
2. Wagner MM, Tsui F-C, Espino J, et al. National retail data monitor for public health surveillance. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):40–2.
3. 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.

4. Wagner M, Pavlin J, Brillman JC, Stetson D, Magruder S, Campbell M. Synthesis of research on the value of unconventional data for early detection of disease outbreaks. Washington, DC: Defense Advanced Research Projects Agency, BioALIRT Program; 2004.
5. Labrie J. Self-care in the new millennium: American attitudes towards maintaining personal health. Washington, DC: Consumer Healthcare Products Association; 2001.
6. McIsaac WJ, Levine N, Goel V. Visits by adults to family physicians for the common cold. J Fam Pract 1998;47:366–9.
7. Zeng X, Wagner MM. Modeling the effects of epidemics on routinely collected data. J Am Med Inform Assoc 2002;9(6 Suppl 1):S17–22.
8. Burkom H, Elbert Y, Feldman A, Lin J. Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):67–73.
9. Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. J Urban Health 2003;80(2 Suppl 1):i57–65.
10. Kulldorff M, Yih K, Kleinman K, Platt R, Mostashari F, Duczmal L. The space-time permutation scan statistic for multiple data streams. In: Proceedings of the National Syndromic Surveillance Conference [CD-ROM]. Boston, MA: Fleetwood Multimedia, Inc; 2004.
11. Cooper GF, Dash DH, Levander JD, Wong WK, Hogan WR, Wagner MM. Bayesian biosurveillance of disease outbreaks. In: Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence; July 8–11, 2004, Banff, Alberta, Canada. Arlington, VA: Association for Uncertainty in Artificial Intelligence Press; 2004:94–103. Available at http://www.cbmi.pitt.edu/panda/publications.html.
12. Hogan WR, Cooper GF, Wagner MM. A Bayesian anthrax aerosol release detector. RODS Technical Report 2004.
13. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proceedings of the National Academy of Sciences USA 2003; 100:4346–51.
14. DeGroot MH. Probability and statistics. 2nd ed. Reading, MA: Addison-Wesley; 1989.
15. Wong W-K. Data mining for early disease outbreak detection [Dissertation]. Pittsburgh, PA: Carnegie Mellon University; 2004.
16. Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods 1997;26:1481–96.
17. Burkom H, Coberly J, Murphy S, Elbert Y, Hurt-Mullen K. Public health monitoring tools for multiple data streams. In: Syndromic surveillance: reports from a national conference, 2004. MMWR 2005; 54(Suppl):55–62.
18. Edgington ES. A normal curve method for combining probability values from independent experiments. J Psychol 1972;82:85–9.
19. Hotelling HH. Multivariate Quality Control. In: Eisenhart C, Hastay MW, Wallis WA, eds. Techniques of statistical analysis. New York, NY: McGraw-Hill; 1947:111–84.
20. Crosier RB. Multivariate generalizations of cumulative sum quality-control schemes. Technometrics 1988;30:291–303.
21. Lowry CA, Woodall WH. A multivariate exponentially weighted moving average control chart. Technometrics 1992;34:46–53.

Space-Time Clusters with Flexible Shapes

Vijay S. Iyengar
IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, New York

Corresponding author: Vijay S. Iyengar, IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598. Telephone: 914-945-3407; Fax: 914-945-2141; Email: vsi@us.ibm.com.

Disclosure of relationship: The contributor of this report has disclosed that he has no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Detection of space-time clusters plays an important role in epidemiology and public health. Different approaches for detecting space-time clusters have been proposed and implemented. Many of these approaches are based on the spatial scan statistic formulation. One key aspect of these cluster detection methods is the choice of cluster shape.*

Objectives: *In this report, the effect of using flexible shapes for clusters is explored by discussing the issues that need to be considered and evaluated.*

Methods: *The first issue is the flexibility of the shape and its ability to model the disease cluster being studied. Another subtle and related factor is that with a more flexible shape, clusters can appear more often by chance, which will be reflected in the p value obtained through Monte Carlo hypothesis testing. Choosing more complex cluster shapes can affect the computational requirements and also constrain the cluster detection approaches that could be applied.*

Results: *The New Mexico brain cancer data set is used to illustrate the tradeoffs. The analysis of these data should not be construed as a comprehensive investigation from the public health perspective. The data set is used to illustrate and compare clusters with two different shapes, cylinder and square pyramid. The results indicate the insights that can be gained from these shapes, individually and collectively.*

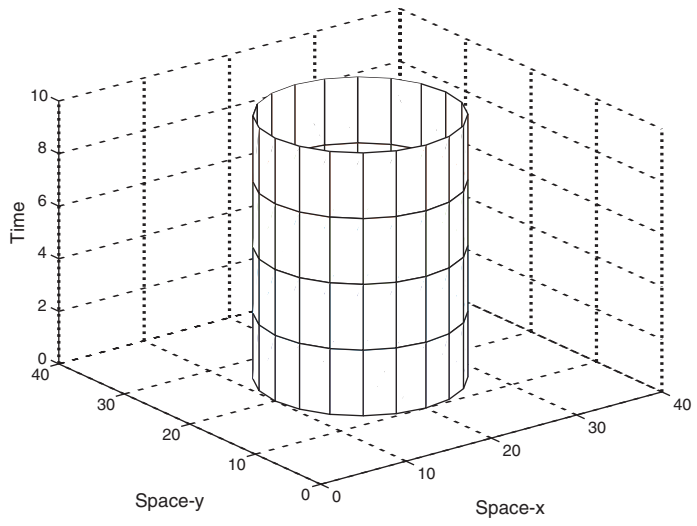
Conclusion: *The domain expert should choose the cluster shape, being aware of the disease being modeled and the analysis goals. For example, a flexible shape like the square pyramid can model either growth or shrinkage and movement of the disease and might provide insights on its origin. In addition, performing the analyses with more than one shape can lead to increased insights regarding the disease cluster.*

Introduction

The spatial scan statistic (*1–3*) has been applied to both retrospective and prospective applications (*4*) in epidemiology and public health. A family of analysis methods has been developed for different models of the underlying disease cluster (e.g., Bernoulli model and Poisson model). Examples of the Poisson model are used to illustrate the concepts presented. For the Poisson model, events are allowed to be generated by an inhomogeneous Poisson process (e.g., the number of disease events in a region over a time interval can be expected to be proportional to the corresponding population, assuming that no other factors are relevant).

These models have been implemented in SaTScan™, a software used to detect space-time clusters (*5*). SaTScan detects space-time clusters by using cylindrical windows (Figure 1) with a circular geographic base and the height of the cylinder corresponding to a certain interval in time. Geographic locations are specified discretely (e.g., centers of counties) to

FIGURE 1. A three dimensional view of a cluster with a cylindrical shape



SaTScan. Input data to SaTScan include the number of cases and population information at these discrete locations at different times. SaTScan evaluates a set of cylindrical windows by considering all those spatially centered at any point in a user-specified grid and exhaustively varying the cylinder's radius and time duration. The evaluation computes the likelihood ratio of the alternative hypothesis (an elevated event rate within the cylindrical window) and the null hypothesis (the rate is the same inside and outside the window). For the Poisson model, this likelihood function (*1*) is proportional to

$$LR = \begin{cases} \left(\frac{c}{n}\right)^{c} \times \left(\frac{C-c}{C-n}\right)^{C-c}, & case\ I \\ 0, & otherwise \end{cases}$$

where C is the total number of cases over the entire space and time, c is the number of cases within the window, and n is the expected number of cases within the window under the null hypothesis. Case I refers to the condition when the window has more cases than expected under the null hypothesis, and LR is zero when this condition is not true. The cylindrical window with the highest value of the likelihood function is the resulting cluster R. The multiple hypotheses testing problem is overcome in SaTScan by using Monte Carlo methods by generating synthetic datasets for the entire space-time region in which the event counts are independently generated, conforming to the Poisson model for each location and time. Each of these synthetic datasets is analyzed to determine its most dominant cluster and its likelihood function value. The likelihood that the cluster R could have occurred by chance under the null hypothesis (p value) can be determined by using these Monte Carlo experiments.

The use of cylindrical space-time windows for the clusters examined can limit the fit to the disease being analyzed. For example, the cylindrical shape cannot model growth or shrinkage of a disease cluster over time nor can it model movement over time. The square pyramid shape was proposed as an approach to overcome these limitations (*6*). This cluster shape is illustrated (Figures 2 and 3) in three dimensional and two dimensional views. The three dimensional view (Figure 2) is a cluster growing with time. The axis of the pyramid does not need to be orthogonal to the two spatial axes, allowing the cluster to model movement of the disease. This attribute is clear from the two dimensional view (Figure 3), where the squares represent the geographic extent at discrete times in the cluster time interval. The use of this flexible shape results in substantially increased computational requirements. The computational issue can be addressed by using a randomized search heuristic for the strongest cluster instead of the grid-based pseudoexhaustive approach used in SaTScan (*6*). This report explores the issues related to the choice of the shape used for space-time clusters.

FIGURE 2. A three dimensional cluster with a square pyramid shape

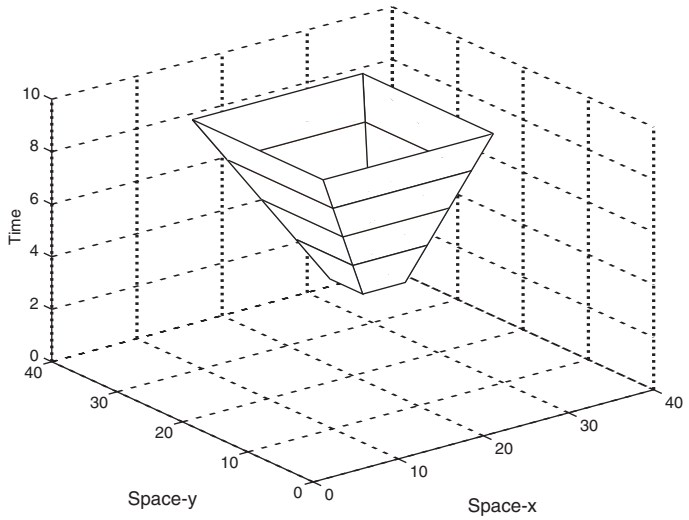
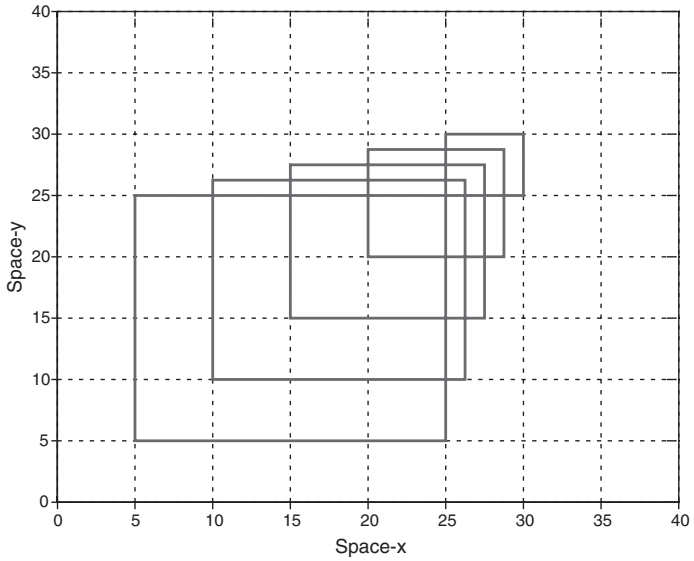


FIGURE 3. A two dimensional view of a cluster with a square pyramid shape



Methods

Choice of Cluster Shape

The first criterion to consider when choosing the cluster shape is the fit to the disease cluster being modeled. All of the available information about the disease cluster can be used to

determine which characteristics are important to model. For example, if modeling the growth of the disease cluster over time is important, using a shape that can represent this behavior is preferable. Choosing a shape arbitrarily that allows more flexibility than is needed also has shortcomings.

As discussed previously, the goal is to detect strong clusters that are also significant when compared with those that can occur by chance under the null hypothesis. An arbitrarily complex shape will increase the chances that the detected cluster is not significant, because the chance of finding strong clusters in the synthetic data of the Monte Carlo experiments also increases.

The fit of the shape to the disease cluster also needs to be balanced with the computational need for the shape being considered. This second criterion, namely the computational need, has to be considered in conjunction with the search algorithm used. A search algorithm can be exhaustive by considering all possible clusters of the chosen shape. For example, in the retrospective analysis of the New Mexico brain cancer data (*7*), cylindrical space-time clusters were used. The disease occurrences and population counts were provided for each of the 32 counties included in the data. In an exhaustive analysis of cylindrical time-space clusters, all possible circular cross-sections would have to be considered, with each circular cross-section represented uniquely by the subset of counties included. Using a regular grid for the centers of the circular cross-sections and then exhaustively considering all possible radii might not be exhaustive, depending on the positions of the county centers and the choice of the grid. For cylindrical clusters, choosing a grid fine enough to be spatially exhaustive for a given data set might be practical. However, this approach might not be computationally effective for all data sets, and efficient algorithms that guarantee exhaustive exploration by cylindrical clusters need to be developed. Exhaustive methods might not be practical for more complex shapes. For example, the computational need is significantly higher for the square pyramid shape (*6*). One practical solution to detect square pyramid clusters is to use heuristic search, based on randomized algorithms (*6*). However, the p value computed by any method that is not guaranteed to be exhaustive needs to be validated (*6*).

The use of these criteria is illustrated by evaluating the applications of two different cluster shapes to the New Mexico brain cancer data (*7*). This data set was analyzed by using cylindrical clusters with cross-sections restricted to having one of the 32 county locations as its center (*2*). To extend the analysis by using a more complex shape that can model either growth (or shrinkage) and movement over time, the consideration would be restricted to convex three dimensional shapes because allowing nonconvex shapes is too general for the modeling goal. The three dimensional convex hull would be the least restrictive convex shape, but it is still too general for the goal at hand, which is to model either growth or shrinkage, but not both. Truncated pyramids are adequate to model growth (or shrinkage) over time. The pyramid can model movement if its axis is not restricted to be orthogonal to the spatial plane. The degrees of freedom can be limited by choosing a regular polygon for the pyramid cross-section. Whereas the square cross-section is used as an example in this report, similar analysis can be performed with other regular polygons for the cross-section. The flexibility could have increased by allowing irregular polygons for the cross-sections. However, an attempt to use an irregular polygon (rectangle) for the cross-section of the pyramid was not successful. Obtaining effective convergence behavior for the randomized algorithm was substantially more difficult with this extra degree of freedom. As mentioned previously, the p values would also be expected to worsen if the cross-sections were not restricted to regular polygons. The truncated cone was also considered to be another cluster shape candidate. A regular polygon was chosen over the circle for the cross-section, because the computations with planes in the case of the polygon were simpler when they involved linear equations.

In the next section, the New Mexico brain cancer data (*7*) are used to compare the results of the analyses by using two shapes, the cylindrical and square pyramid clusters. The intent of the analysis is not to conduct a public health investigation but to simply use this data set to illustrate the effect of cluster shape.

Results

The data set (*7*) contains occurrences of 1,175 cases of brain cancer in 32 counties in New Mexico during 1973–1991. Occurrences are aggregated at the temporal granularity of a year. Population information is provided for each year. Three covariates are provided: age group, sex, and ethnicity. First, only the first two covariates will be considered. The third covariate, ethnicity, will be added, and the effect of this addition will be discussed. The inclusion and exclusion of covariates are not based on domain knowledge of their effect on cancer but are designed to merely illustrate the effect of the cluster shape in two different situations.

Considering Covariates: Age Group and Sex

The Poisson formulation for the spatial scan statistic provides adjusting for covariates by using indirect standardization (*1*). The two covariates, age group and sex, are adjusted

for, and it is assumed that both covariates are relevant to the disease being analyzed. And the analysis is intended to find clusters that cannot be explained by these two covariates.

First, results for the cylindrical clusters detected by using the SaTScan system will be presented (*3*). The results in the first column (Table 1) are generated by using the default mode in SaTScan without an explicit grid. The strongest cluster detected in this mode had a log likelihood ratio of 11.07 and a p value of 0.013; it includes 16 counties over a 5-year period, 1985–1989. This mode misses analyzing multiple potential cylindrical clusters and can be illustrated by using an explicit fine grid of size 1 Cartesian coordinate to perform the analysis. A fine grid is used to better approximate an exhaustive analysis for cylindrical clusters. The characteristics of the strongest cluster detected in this mode are included in the second column (Table 1). The cluster detected in this mode is stronger, with a log likelihood ratio of 13.70, and has a smaller spatial extent that included only 12 counties over the same 5-year period (1985–1989). The significant differences in the clusters detected by these two modes demonstrate the effect of approaching exhaustive analysis with the fine grid. Only the cylindrical clusters detected by using the fine grid will be considered in the rest of this report.

Second, the results for the square pyramid cluster are presented (*6*). The characteristics of the strongest cluster detected by this heuristic search are presented in the last column (Table 1). The number of cases included in this square pyramid cluster (284) is larger than the number in the cylindrical cluster (265), which was detected by using a fine grid; it also extends over a longer period (1982–1989). The p value of 0.038 computed by using 999 Monte Carlo replications is higher than the 0.004 value for the cylindrical cluster (Table 1), but the cluster is significant using the threshold of 0.05.

The cylindrical and square pyramid clusters can be compared by using the three dimensional and two dimensional views (Figures 4 and 5). In the two dimensional view (Figure 5), the county locations are indicated by an asterisk. This view also illustrates the square cross-sections of the pyramid for each of the 8 years. For the 5 years (1985–1989) common to both clusters, the square cross-sections of the pyramid are indicated with solid lines. The first 3 years (1982–1984) of the square pyramid cluster are not included in the cylindrical cluster and are marked with dashed lines. The square pyramid cluster originates with six counties during the beginning of 1982 but expands to include 15 counties during the end in 1989. In contrast, the cylindrical cluster with a spatial extent is marked by the circle covering 12 counties for the 5-year

TABLE 1. Results for brain cancer data, by age group and sex as covariates — New Mexico, 1973–1991

Characteristic	Cylindrical cluster (No grid)	Cylindrical cluster (Fine grid)	Square pyramid cluster
Log likelihood ratio	11.07	13.70	16.918
No. of cases	317 (249.09 expected)	265 (195.33 expected)	284 (204.92 expected)
Overall relative risk	1.273	1.357	1.386
p-value	0.013	0.004	0.038
Centroid coordinates	(60,67)	(89,81)	NA*
Cross-section radius	68.96	50.25	NA
Years	1985–1989	1985–1989	1982–1989

* Not applicable.

FIGURE 4. A three dimensional view of a comparison of the cylindrical and square pyramid clusters

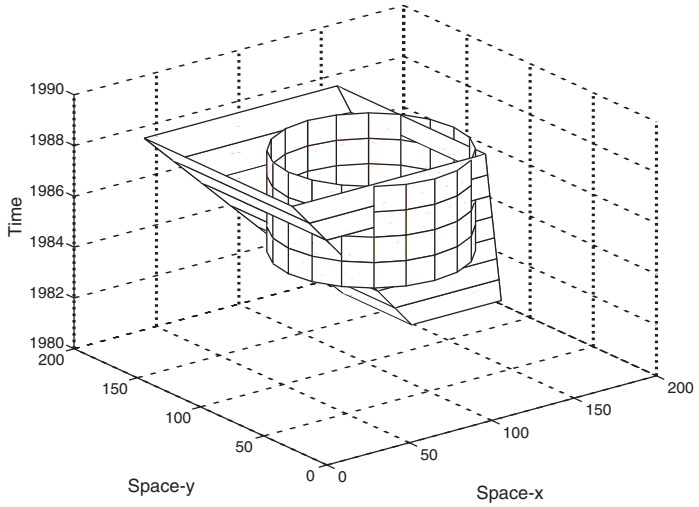
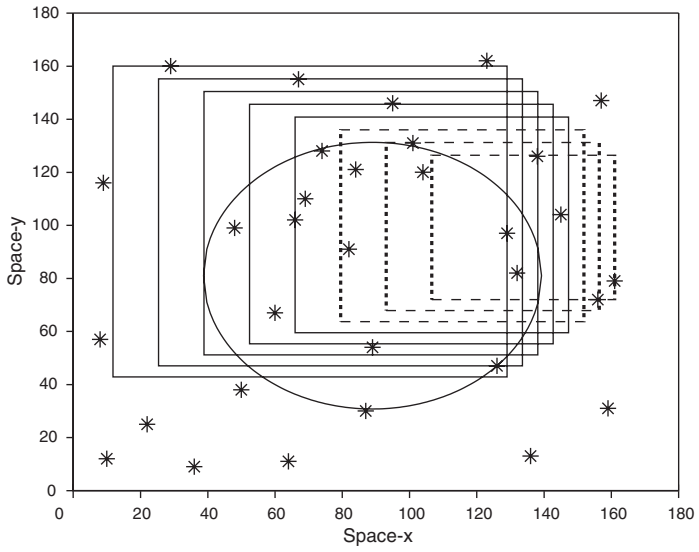


FIGURE 5. A two dimensional view of a comparison of the cylindrical and square pyramid clusters



period, 1985–1989. The square pyramid cluster also indicates movement over time in addition to the growth as certain counties at the right of the two dimensional view get dropped in the latter years.

Together, the two dimensional and three dimensional views provide visualization of the cylindrical and square pyramid clusters, indicating key aspects (e.g., overlap). The visualization suggests that the detected cylindrical cluster could be viewed as a reasonable approximation of the detected square pyramid, given the constraints of its shape. Therefore, the analyses with these two shapes can be viewed as supporting each other.

Adding Covariate: Ethnicity

In this section, the covariate ethnicity is included in this analysis. This covariate can take one of three values: white, black, or other. The spatial distribution of the covariate in 1987 is illustrated (Figure 6). The bar chart (Figure 6) indicates the population fractions for the ethnicity values, black and other, for each of the 32 counties. The figure illustrates wide variation of the ethnicities over the counties. Analyzing this distribution for each year included in the cluster also indicates shifts in the distribution over time. Therefore, factoring out this covariate can be expected to affect the cluster detected.

The results for both cluster shapes are presented (Table 2). The strongest cylindrical cluster is the same as in the two covariate cases in the section, Considering Covariates: Age Group and Sex. The log likelihood ratio of the strongest cylindrical cluster is lower (12.86) and the p value higher (0.01), with this additional covariate factored out. The analysis with cylindrical clusters differs in multiple aspects from the comprehensive public health perspective described in the literature (*2*). The temporal trend is not factored out, and the explicit fine grid in the analysis is also used. These factors affect the results and prevent a direct comparison. However, the set-up to a direct comparison between the two cluster shapes will be used. The square pyramid cluster detected in the previous subsection is not the strongest cluster any more when this additional covariate is factored out. The log likelihood ratio for the square pyramid cluster decreases to 16.05. Moreover, even the strongest square pyramid cluster detected with a log likelihood ratio of 16.208 is not significant with a p value estimate of 0.054 (using the earlier threshold of 0.05). Further investigation is needed to determine if the lack of a significant square pyramid cluster in this case is a result of the increased flexibility of the shape or a result of slow convergence of the randomized heuristic.

FIGURE 6. Distributions of the ethnicity covariate, 1987

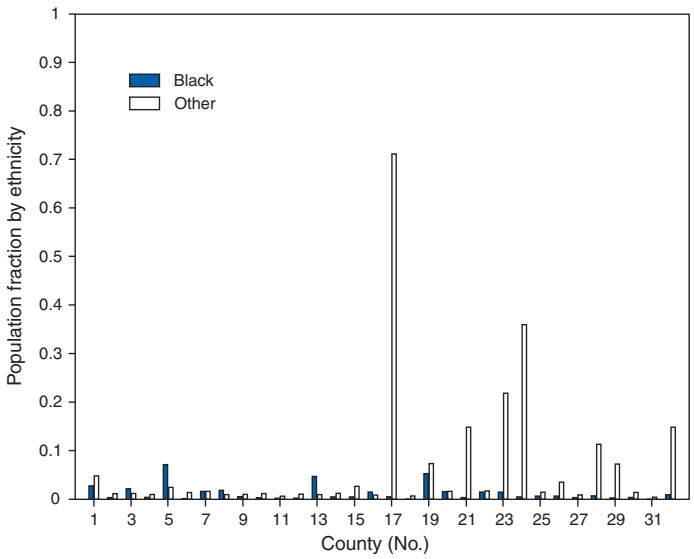


TABLE 2. Results for brain cancer data, by age group, sex, and ethnicity as covariates — New Mexico, 1973–1991

Characteristic	Cylindrical cluster (Fine grid)	Square pyramid cluster
Log likelihood ratio	12.86	16.208
No. of cases	265 (197.34 expected)	420 (330.14 expected)
Overall relative risk	1.343	1.272
p-value	0.010	0.054
Centroid coordinates	(89,81)	NA*
Cross-section radius	50.25	NA
Years	1985–1989	1983–1991

* Not applicable.

Computational Issues and Usage

The implementation of the prototype took approximately 40 hours on an IBM Intellistation M-Pro computer with an Intel P4 processor running at 2.2 Ghz to perform each square pyramid analysis discussed previously. This high computational cost of the heuristic approach can limit its usage in a surveillance application where the analysis has to be performed frequently. However, when the computational cost is not an issue, clusters with complex shapes can be used in a surveillance application by using the methodology described in the literature (*3*).

Conclusion

The purpose of considering the different sets of covariates in this report was to illustrate and compare the behavior of cluster detection methods with different underlying shapes.

The actual set of covariates that needs to be adjusted for in any data set should be determined by the domain expert performing the analysis. The domain expert should also choose the cluster shape, keeping in mind the disease being modeled and the analysis goals. For example, a flexible shape like the square pyramid can model either growth (or shrinkage) and movement of the disease cluster and might provide certain insights into its origin. However, computational considerations might limit the analysis to use heuristic approaches that can only estimate the strongest cluster and, more importantly, its p value. Performing the analyses with more than one shape can lead to greater insights about the disease cluster. Moreover, more confidence is gained in these insights when the results of the analyses with different shapes support each other as illustrated in the previous example.

Acknowledgments

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 author and do not necessarily reflect the views of AFRL or DARPA.

References

1. Kulldorff M. A spatial scan statistic. Communications in statistics: Theory and Methods 1997;26:1481–96.
2. Kulldorff M, Athas W, 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.
3. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. J Royal Statistical Society 2001;A164:61–72.
4. 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.
5. Kulldorff M. Information Management Services Inc. SaTScan™ version 3.1: software for the spatial and space-time scan statistics. Technical Report, 2002. Available at http://www.satscan.org.
6. Iyengar VS. On detecting space-time clusters. In: Proceedings of the 10th Association for Computing Machinery (ACM) International Conference on Knowledge Discovery and Data Mining; August 22–25, 2004; Seattle, WA. New York, NY: ACM Press; 2004:587–92.
7. National Cancer Institute. Brain cancer in New Mexico. Bethesda, MD: Division of Cancer Prevention, Biometry Research Group. Technical Report Data set (1973–1991). Available at http://www3.cancer.gov/prevention/bb/brain_nm.html.

INFERNO: A System for Early Outbreak Detection and Signature Forecasting

Elena N. Naumova,[1] E. O'Neil,[1] I. MacNeill[2]
[1]*Tufts University School of Medicine, Boston, Massachusetts;* [2]*University of Western Ontario, London, Canada*

Corresponding author: Elena N. Naumova, Department of Public Health and Family Medicine, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111. Telephone: 617-636-2462; Fax: 617-636-4017; E-mail: elena.naumova@tufts.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Objective: *Public health surveillance systems that monitor daily disease incidence provide valuable information about threats to public health and enable public health authorities to detect enteric outbreaks rapidly. This report describes the INtegrated Forecasts and EaRly eNteric Outbreak (INFERNO) detection system of algorithms for outbreak detection and forecasting.*

Methods: *INFERNO incorporates existing knowledge of infectious disease epidemiology into adaptive forecasts and uses the concept of an outbreak signature as a composite of disease epidemic curves.*

Results: *Four main components comprise the system: 1) training, 2) warning and flagging, 3) signature forecasting, and 4) evaluation. The unifying goal of the system is to gain insight into the nature of temporal variations in the incidence of infection. Daily collected records are smoothed initially by using a loess-type smoother. Upon receipt of new data, the smoothing is updated; estimates are made of the first two derivatives of the smoothed curve, which are used for near-term forecasting. Recent data and near-term forecasts are used to compute a five-level, color-coded warning index to quantify the level of concern. Warning algorithms are designed to balance false detection of an epidemic (Type I errors) with failure to correctly detect an epidemic (Type II errors). If the warning index signals a sufficiently high probability of an epidemic, the fitting of a gamma-based signature curve to the actual data produces a forecast of the possible size of the outbreak.*

Conclusion: *Although the system is under development, its potential has been demonstrated through successful use of emergency department records associated with a substantial waterborne outbreak of cryptosporidiosis that occurred in Milwaukee, Wisconsin, in 1993. Prospects for further development, including adjustment for seasonality and reporting delays, are also outlined.*

Introduction

Daily disease monitoring through public health surveillance systems provides valuable information about threats to public health. Substantial outbreaks can be caused by emerging new pathogens (e.g., Severe Acute Respiratory Syndrome and West Nile virus) and evolving well-known ones (e.g., cryptosporidiosis). Modern surveillance systems require efficient statistical tools for early detection of rapid changes in disease incidence and forecasting the extent of an outbreak, and more rigorous methodology is needed (*1,2*). Such tools should accommodate vital features of surveillance data that relate to the nature of diseases, their etiology and epidemiology, and characteristic properties of the data. For example, seasonal patterns of diseases should be considered in outbreak detection algorithms, potential reporting delays should be taken into account in estimating the size of the infected population, and appropriate adjustments should be made in outbreak-detection and forecasting algorithms.

This report describes an innovative approach for outbreak detection and forecasting outbreaks, the INtegrated Forecasts and EaRly eNteric Outbreak (INFERNO) detection system. INFERNO is a system of adaptive algorithms for early outbreak detection and forecasting the extent of detected infectious disease outbreaks (*3*). The system uses the concept of an outbreak signature (i.e., a composite of elementary distributions of incubation times associated with exposure and population characteristics) and an adaptive forecasting approach. This report discusses the INFERNO system by using as examples retrospective evaluation of a daily time series of physician-diagnosed cases of nonspecific gastroenteritis that occurred in Milwaukee, Wisconsin, in 1993, in association with a well-documented waterborne outbreak of cryptosporidiosis (*4*).

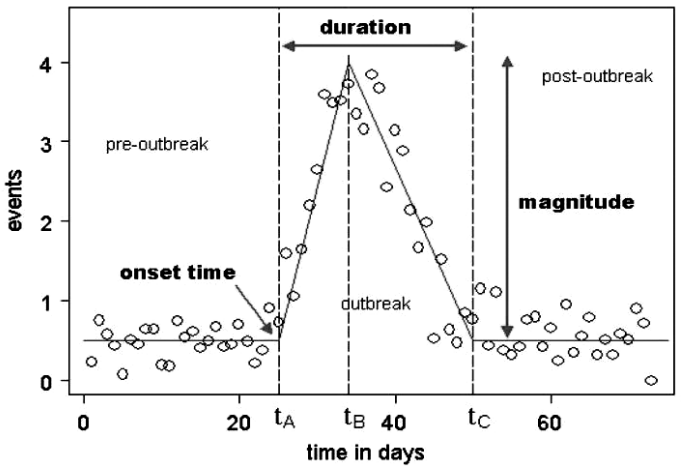
Methods

Infectious disease is initiated by the introduction of a pathogen into a susceptible population. For the majority of exposed persons, a certain incubation period exists between the time a pathogen is acquired and the time of its clearance. In certain cases, exposure results in clinical manifestation. The period between exposure and onset of clinical signs and symptoms is referred to as the incubation period. Duration of incubation time depends on host immune reactivity. Typically, a case of infection is recorded as disease symptoms are developed and confirmed tests are performed. From a modeling point of view, reported cases of infections are the realization of a random process that can include both observable and unobservable parts. The observable part of a process depends on the size of the population in which exposure results in a symptomatic event and on detection and recording of such events.

For a substantial class of waterborne and foodborne enteric infections, symptoms typically are mild and self-limiting; the observable part might be substantially smaller than the part that is unobservable and will vary over time. For example, epidemiologic studies of cryptosporidiosis incidence demonstrate that although approximately 70% of a population might exhibit markers of recent infection, typically only 2%–5% have reported symptoms (*5*). Waterborne and foodborne enteric infections often manifest by alternating periods of low and high incidence. Daily cases of infection might represent two distinct modes: 1) an endemic mode, in which incidence is low and observed disease incidence normal; and 2) an epidemic mode, in which increased incidence of waterborne or foodborne enteric infections is caused primarily by an increase of either a fraction of the susceptible subpopulation or the dose of the exposure.

Infectious disease events occur in the form of a series of dependent observations. A sequence of daily cases of infection reflects a temporal composition of recorded events. Any outbreak comprises time of onset, magnitude, and duration. A time series of daily cases can be illustrated schematically as a simple single point-source exposure outbreak (Figure 1). In a single point-source exposure outbreak, all subjects are assumed to have been exposed to the same dose at the same time. An outbreak starts at time t_A, when the mean of an infectious process begins to change. An outbreak reaches its maximum at time t_B, l-days after a spike in exposure, where l is the mean latent period. The mean of an infectious process declines to a preoutbreak level at time t_C, k-days after a spike in exposure, where k is a maximum incubation period. A dark blue line is used to reflect the temporal pattern of mean disease incidence with respect to 1) time of onset, t_A, 2) duration, $t_C - t_A$, and 3) peak time of disease incidence (Figure 1).

FIGURE 1. Time series of a simple single-point source exposure outbreak, by time of onset, duration, and magnitude — Milwaukee, Wisconsin, 1993



The actual data from which this temporal pattern is observed are the realization of an infectious disease process, often referred to as an epidemic curve.

The temporal pattern of mean disease incidence closely approximates the distribution of incubation times in the symptomatic population only if 1) every person with a recorded case is exposed at the same time and to the same dose, 2) the reported time is the latent time, and 3) the latent period is proportional to the incubation period. For outbreaks caused by sources of infection other then person-to-person transmission (i.e., intentional or environmental outbreaks), this statement might be valid, but to expect these three conditions to be satisfied in real-life settings seems impractical (*4,6*). Nevertheless, if a temporal pattern of mean disease incidence is assumed to be a composite of elementary distributions of incubation times associated with exposure and population characteristics, the knowledge of pathogen-specific incubation times can be used to detect an outbreak and forecast its magnitude and duration. Such composites form a unique signature. For monitoring systems, a signature has two potential uses. First, when a signature is applied to streamline data in real time, it produces a long-term forecast. Second, a signature allows retrospective identification and quantification of similar temporal patterns from historic data.

Results

INFERNO is a system of adaptive algorithms for outbreak detection and forecasting that is based on the concept of outbreak signature forecasting. This section discusses the four components of the system (training, warning and flagging, signature forecasting, and evaluation) by using a retrospective daily time series of cases of nonspecific gastroenteritis associ-

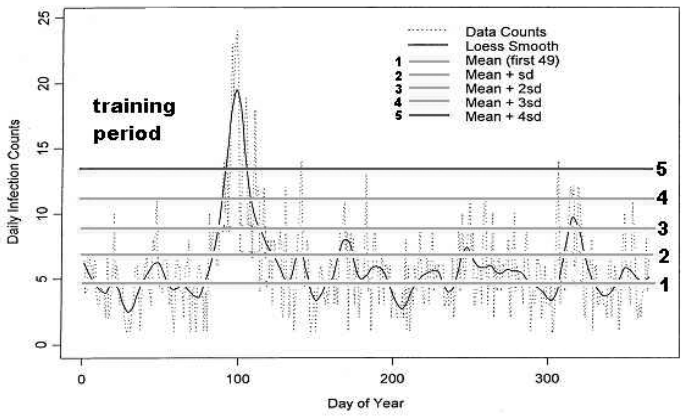
ated with a 1993 waterborne outbreak of cryptosporidiosis that occurred in Milwaukee, Wisconsin.

Training

The purpose of this component is to gain insight into the nature of endemic temporal variations in the incidence of infection. As in any other outbreak detection system or algorithm, this component includes examination of retrospectively collected data and provides estimates for baseline parameters (*7*). In this system, daily cases of infection represent two modes of the process: endemic and epidemic. Although this distinction is somewhat artificial, it is useful for modeling purposes. In this analysis, training is limited to a 7-week period during which the infection is assumed to be in endemic mode, which is supported by visual inspection and posthoc analysis using INFERNO. This set of data is used to determine the average rate of gastroenteritis events and the extent of their variability. Using the standard deviation of the training set as a multiplier, four levels of departure from the mean are estimated (Figure 2). The training component currently contains the algorithms to assess potential temporal fluctuations associated with seasonal fluctuations. The training for the epidemic mode is also important for refining the signatures to detect different patterns and to develop a library of observed epidemic curves.

A loess smooth is superimposed on all available daily counts of nonspecific gastroenteritis (Figure 2). The issue of selecting a proper smoothing technique deserves special attention. The span of the smoother must be long enough that random fluctuations in the background infection rates are not chased, yet short enough to react to the onset of an outbreak. To achieve such balance for a year of data, the selected span ratio of the nonparametric smoother is equal to 25/365. This span ratio implies a window size of 25 days, an arbitrary but reasonable choice to cover the duration of an incubation period for cryptosporidiosis, which might be up to 21 days. The data indicate an outbreak associated with cryptosporidiosis infection during the period starting at 75 days and ending at 130 days. As an indication of the effectiveness of the smoothing in capturing the essentials of the daily counts, cumulative data sums are compared with the smoothed function (*3*).

FIGURE 2. Time series of daily cases of nonspecific gastroenteritis, with superimposed loess-smoothed curve and four levels of departure from the mean* — Milwaukee, Wisconsin, 1993



* 7-week training period.

Warning and Flagging

The purpose of this component is to quantify the level of concern in the streamlined (i.e., frequently updated) data to switch from endemic to epidemic mode in forecasting procedure. Daily collected reports are smoothed by using a loess-type smoother. Smoothing is updated daily, and estimates are made of the first two derivatives of the smooth curve indicating the gradient of change in mean incidence. These loess- and derivative-based estimates are used to build a near-term forecast (Figure 3). Three marks indicate the prediction for the next day during 3 consecutive days, if the records are only available up to day 80, 81, 82. Next, a long-term forecast is built by extending the near-term forecast to almost 1 month in advance as if the records are available only up to a day 80, 81, 82 (Figure 3). To produce the forecasts, a trajectory is created of the mean between two points: the time of the last observation (point A) and 1 month in advance (point B) (*3*). The trajectory is estimated by using the first two derivatives of the smoothed function at point A and the values of a truncated version of the Taylor's series expansion of the function for the point B. The forecasts are quite different from the actual observations, indicating that a switch should be made from the endemic to the epidemic mode that is essential for modeling.

The recent data and the near-term forecasts are used to quantify a switch point. The warning algorithm uses four severity indexes: 1) actual counts at the given day, 2 the mean estimate 1 week before the given day, 3) the estimate of the mean at the given day, and 4) the maximum of the forecast for the next 7. The sum of four severity indexes for a given day forms a basis for flagging. A given day is flagged by using five color-coded categories, and depending on the flag value, either the attenuated forecasts or the signature forecasts are performed. A time series of cases is transformed to a time series of flag values, which in any given day contains past, present, and predicted future information (Figure 4). During the outbreak, the daily flag values clearly exceed the red code, indicating the period in which signature forecasting is used.

FIGURE 3. Near term (panel A) and extended (panel B) forecast shown as if records are available for only up to a day 80, 81, 82 — Milwaukee, Wisconsin, 1993

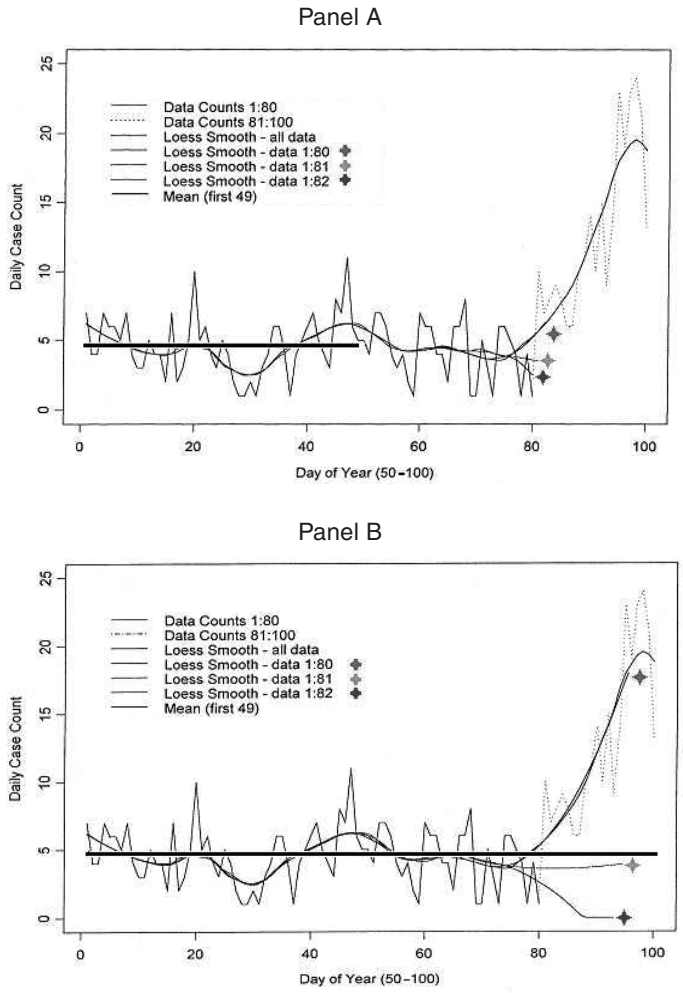
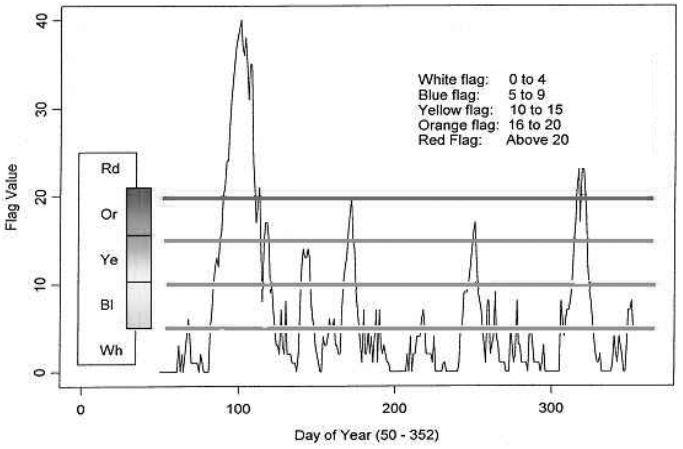


FIGURE 4. Time series of cases transformed to a time series of flag values by using five color-coded categories — Milwaukee, Wisconsin, 1993



Signature Forecasting

If the flag value signals a sufficiently high probability of an epidemic mode, then a forecast of the possible size of the outbreak is made. Currently, a model for the mean value function implemented in the system is:

Equation 1. $\mu(t) = f(t) + \sum_i \delta_{(t1i,t2i)}(t)\, S_i(t - t_{1i}),\ t \geq 0,$

where $f(t)$ is a function representing the mean of the endemic counts of cases at time t, S(.) is a function characterizing the shape of an outbreak and is referred to as the base signature function for disease outbreaks, $Si(.)$ is $S(.)$ amplified to accord with the extent of the ith outbreak, t_{1i} is the date of the onset of the ith outbreak, t_{2i} is the date of termination or resolution of the ith outbreak, and $\delta_{(t1i,t2i)}$ is a function that assumes the value 1 during the period of the ith outbreak and otherwise is zero (*3*).

Examples of the background, or endemic, function, $f(t)$, are:

Equation 2. $f(t) = \mu, t \geq 0$

for the case of constant background infection rates, and

Equation 3. $f(t) = \mu + a \sin(\varpi t) + b \cos(\varpi t),\ t \geq 0$

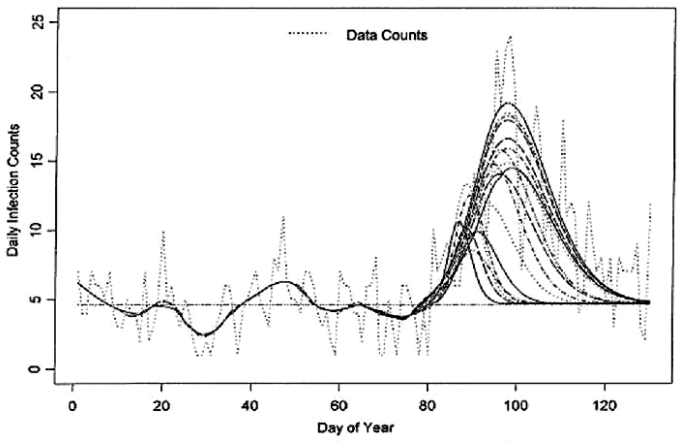
for the case of seasonal fluctuations in the background infection rates. As a possible model for the random process generating the time series of daily counts for a particular infection, a time-dependent Poisson process with intensity function $\mu(t)$ defined by Equation 1 is chosen.

To build a signature forecast, a family of gamma distributions is chosen. This nonnegative, right-skewed distribution is used to approximate a distribution of incubation times and can be easily generated by using statistical software. Depending on the level of concern, an appropriate curve is selected from a library of distributions and serves as a signature base (Figure 5). A signature curve also can be simulated as a set of incubation time distributions according to an intensity function given by Equation 1, which on the average gives the signature shape. For streamlined data, the fitting is updated daily. The fitted signature curve is the long-term forecast.

Evaluation

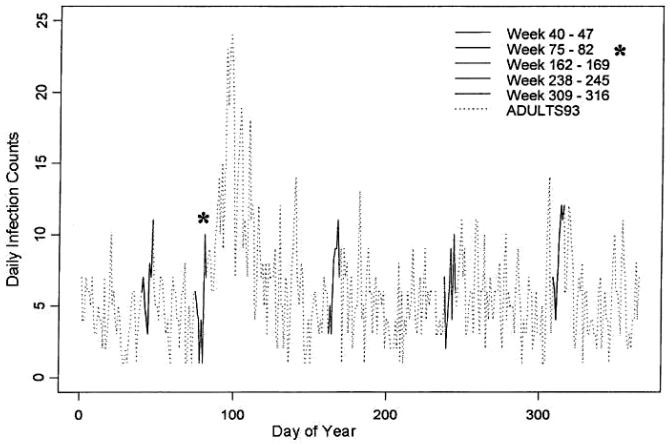
The purpose of this component is to quantify uncertainty associated with the predicted size of the outbreak. The algorithms for computing the warning index have to be designed to balance false prediction of an epidemic (Type I errors) with failure to correctly predict an epidemic (Type II errors). Initiation of the signature forecast and selection of the signature base are determined by the daily flag value. In the example, five separate week-long sets of data selected from the raw time

FIGURE 5. Time series of daily cases of nonspecific gastroenteritis, with superimposed loess-smoothed curve and signature fits for days 83–101 — Milwaukee, Wisconsin, 1993



series that appear to have a similar pattern to the week that launched the outbreak of cryptosporidiosis during day 75–130 are highlighted (Figure 6). If the same signature forecasting procedure were applied to each of the four other sequences, a Type I error would be committed in each case. To avoid potential error, the warning index has to be corrected. A data-driven inflator was developed to balance Type I and Type II errors. The price for the use of this data-driven inflator is a significant delay in making a definitive forecast of the extent of an outbreak when flag values are yellow or blue. However, this delay would be considerably less when flag values are orange or red.

FIGURE 6. Time series of gastroenteritis cases with five highlighted sets of data selected from the raw time series that appear to have a similar pattern to the week that launched an outbreak of cryptosporidiosis — Milwaukee, Wisconsin, 1993



Conclusion

Library of Signature Curves

The effectiveness of the forecast depends upon the extent to which the signature curve captures the shape of outbreaks of the infection under consideration. In the system under discussion, epidemiologic observations for incubation periods of cryptosporidiosis are considered. The lag between the time of exposure and the time of emergency department visit for gastroenteritis symptoms is assumed to be approximately 8 days (*8*). However, this period might vary among sensitive subpopulations; for example, it might decrease among children or the elderly (*9*). Different strains of a pathogen might have different incubation times. Acquired immunity to a pathogen might shorten the incubation period and reduce the number of symptomatic cases. Theoretically, mean incubation time might be inversely proportional to the inoculum dose at low doses. A library of signature base curves for infections with various epidemic properties should be developed. A period between exposure and an outcome of interest (e.g., disease onset, emergency department visit, or hospitalization time) can be thought of as a random variable. Although a lognormal distribution, a classic model for incubation time of infectious disease, has been shown to be robust to many biological factors; multiple distributional forms (e.g., Gamma, Weibull, inverse Gaussian, exponential, and Poisson) for a continuous or discrete random variable of time to event have been applied for investigating waterborne cryptosporidiosis (*10–13*).

Provisional Data and Reporting Delay

In the proposed system, retrospectively collected data are used, but the reporting mechanism is assumed to make data available the day after an event occurs. Certain delays in reporting are inevitable. Time-consuming testing or report submission procedures might cause systematic delays in the streamline surveillance data. The INFERNO framework allows users to correct for a systematic reporting delay. For example, the warning index that uses an actual record on a given day and three estimates (weekly mean for a week before a given day, a predicted mean for a given day, and a maximum of forecasts for next few days) could be adjusted for a lagged systematic delay. A small simulation study was conducted to initiate an analysis of systematically delayed reporting. On any given day, i, a report was assumed to contain a number of time-distributed cases,

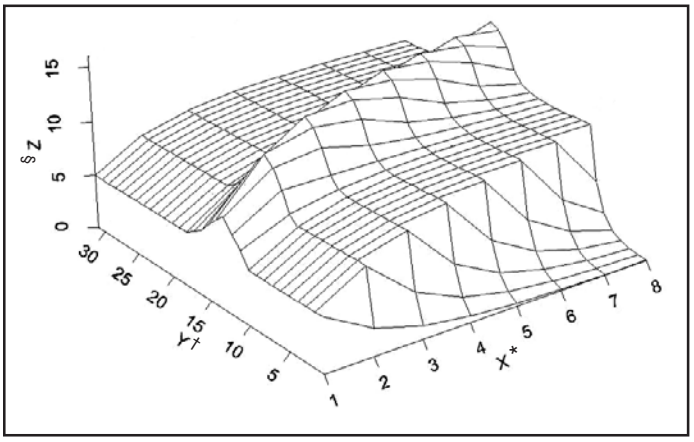
$Yi = \{y_{0i}, y_{1i}, \ldots, y_{Ki}\}$, so the number consists of y_0-cases occurring on a given i-day, y_1-cases occurring 1 day before the i-day, y_2-cases occurred 2 days before, etc. Thus, $Zi = \Sigma\, \alpha_{i-k}\, Yik$. For demonstration purposes, an exponential weighting

$\alpha_{i-k} = 0.5^{(k+1)}$, where $k = 0 - K$, and $K = 8$ be the maximum length of delay in days, was used. A 31-day long-time series was simulated, with 10 cases each day for an endemic background level (the first 10 and the last 14 days) and 7 days of outbreak by using Poisson-distributed multipliers of 1.25, 1.45, 1.68, 1.55, 1.38, 1.2, and 1.1. The results of the simulation were presented by using the three-dimensional temporal exposure-response surface (TERS) (*5,9*), in which the x-axis reflects the delay in reporting in days, the y-axis imitates the simulated time series obtained as provisional data, and the z-axis reflects restored daily counts with an 8-day lag-distributed scheme (Figure 7). The TERS plot depicts the speed in reporting recovery. On day 1, the simulated provisional daily counts would consist of 10/2 = 5 cases for endemic level. On day 2, counts would consist of five cases for a second day and 10/4 = 2.5 for the previous day, the number of cases for day 1 would be equal to 7.5. By day 8, completed reporting is 99% for an endemic level and 96% for an outbreak. This simple simulation study suggests an approach for a detailed analysis of provisional data that might be helpful in assessing delays, developing adjusting schemes, and quantifying potential duration for a training period in the surveillance systems. Better understanding of reporting barriers and better adjusting for delays in outbreak detection and forecasting are needed.

Seasonality Adjustment

Although the algorithms for warning and forecasting of an outbreak size in this system do not rely on extensive historical recording, a long training period is needed for a proper seasonality adjustment. The Milwaukee outbreak occurred in the spring, an unusual time for cryptosporidiosis, which typically exhibits a seasonal increase during late summer–early fall (*14*). The detection of an outbreak close to or during a disease seasonal peak is a difficult task. The forecasting might be sensitive to a degree of seasonal adjustment. Currently, the system offers certain algorithms to assess potential temporal fluctuations associated with seasonal increases using parametric and non-parametric approaches. Additional studies are needed to investigate the effect of seasonality adjustment on forecasting. Integration of additional information obtained from other components of the surveillance (e.g., drug sales, water quality parameters, meteorological information) into the forecasting algorithms might provide better results. Forecasting might also be improved by considering additional factors that might influence incidence and reporting of waterborne infections (e.g., environmental factors [*15*], boil-water orders, media effects, television and radio announcements, day-of-the-week effects, school vacations, and holidays).

FIGURE 7. Three-dimensional surface plot reflecting a simulated reporting delay — Milwaukee, Wisconsin, 1993



* Reflects delay in reporting in days.
† Imitates simulated time series obtained as provisional data.
§ Reflects restored daily counts with lag-distributed scheme.

Terminology Refinement

The concept of what constitutes an outbreak should be refined. In public health literature, the term is used inconsistently and often imprecisely. For example, in certain situations, an outbreak can mean an increase in incidence over the endemic level (i.e., the term is used to refer to the onset of some observed change). In another meaning, the term can signify a degree of magnitude over a period of time that generates a public health concern. In a third context, risk communication, the term can serve as a synonym for epidemic, often with the intention of reducing the public's level of fear. The more is learned about the process of infectious diseases, exposure, and manifestation, the clearer the concept of outbreak will become, allowing it to be used with more rigor. Both the efficiency and the accuracy of outbreak forecasting using the mathematical modeling will depend on clarity and rigor in the use of key terms. With more attention paid to the precision of the common language shared by mathematicians and epidemiologists, forecasting will enable public health authorities not merely to record experiences but also to influence the future (*16*).

Acknowledgment

Support for this research was provided by a grant from the National Institutes of Health (R01 434315).

References

1. Evans S, Kleiman K, Pagano M. Bioterrorism and biosurveillance. American Statistician 2002;7:41–3.
2. Stoto MA, Schonlau M, Mariano LT. Syndromic surveillance: is it worth the effort? Chance 2004;17:19–24.

3. Naumova EN, MacNeill IB. Signature forecasting and early outbreak detection system. Environmetrics 2005;16; In press.
4. Morris RD, Naumova EN, Griffiths JK. Did Milwaukee experience waterborne cryptosporidiosis before the large documented outbreak in 1993? Epidemiology 1998;3:264–70.
5. Egorov A, Frost F, Muller T, Naumova E, Tereschenko A, Ford T. Serological evidence of *Cryptosporidium* infections in a Russian city and evaluation of risk factors for infections. Ann Epidemiol 2004;14:129–36.
6. Ashbolt RH, Coleman DJ, Misrachi A, Conti JM, Kirk MD. An outbreak of cryptosporidiosis associated with an animal nursery at a regional fair. Communicable Diseases Intelligence 2003;27:1–8.
7. Stern L, Lightfoot D. Automated outbreak detection: a quantitative retrospective analysis. Epidemiol Infect 1999;122:103–10.
8. Hunter PR. Waterborne diseases. London, England: John Wiley & Sons; 1997.
9. Naumova EN, Egorov AI, Morris RD, Griffiths JK. The elderly and waterborne *Cryptosporidium* infection: gastroenteritis hospitalizations before and during the 1993 Milwaukee outbreak. Emerg Infect Dis 2003;9:418–25.
10. Brookhart MA, Hubbard AE, van der Laan MJ, Colford JM Jr, Eisenberg JN. Statistical estimation of parameters in a disease transmission model: analysis of a *Cryptosporidium* outbreak. Stat Med 2002;21:3627–38.
11. Gupta M, Haas CN. The Milwaukee *Cryptosporidium* outbreak: assessment of incubation time and daily attack rate. Journal of Water and Health 2004;2:59–62.
12. Naumova EN, Terrin N, Morris RD. A quantitative framework for characterizing waterborne outbreaks. In: Proceedings of the Joint Statistical Meetings. Baltimore, MD: American Statistical Association; 1999:117–22.
13. Naumova EN. Modeling a signature of a single-exposure waterborne outbreak signature with the secondary spread. In: Proceedings of the Joint Statistical Meetings. Indianapolis, IN: American Statistical Association; 2000:76–81.
14. Naumova EN, Chen JT, Griffiths JK, Matyas BT, Estes-Smargiassi SA, Morris RD. Use of passive surveillance data to study temporal and spatial variation in the incidence of giardiasis and cryptosporidiosis. Public Health Rep 2000;115:436–47.
15. Naumova EN, Christodouleas J, Hunter PR, Sued Q. Temporal and spatial variability in cryptosporidiosis recorded by the surveillance system in North West England in 1990–1999. Journal of Water and Health 2005; In press.
16. Hempel C, Oppenheim P. Studies in the logic of explanation. In: Madden EH, ed. The structure of scientific thought. Cambridge, MA: Riverside Press; 1960:22.

High-Fidelity Injection Detectability Experiments: a Tool for Evaluating Syndromic Surveillance Systems

Garrick L. Wallstrom, M. Wagner, W. Hogan
RODS Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania

Corresponding author: Garrick L. Wallstrom, RODS Laboratory, University of Pittsburgh, Suite 550, Cellomics Building, 100 Technology Drive, Pittsburgh, PA 15219. Telephone: 412-383-8141; Fax: 412-383-8135; E-mail: garrick@cbmi.pitt.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *When public health surveillance systems are evaluated, CDC recommends that the expected sensitivity, specificity, and timeliness of surveillance systems be characterized for outbreaks of different sizes, etiologies, and geographic or demographic scopes. High-Fidelity Injection Detectability Experiments (HiFIDE) is a tool that health departments can use to compute these metrics for detection algorithms and surveillance data that they are using in their surveillance system.*

Objective: *The objective of this study is to develop a tool that allows health departments to estimate the expected sensitivity, specificity, and timeliness of outbreak detection.*

Methods: *HiFIDE extends existing semisynthetic injection methods by replacing geometrically shaped injects with injects derived from surveillance data collected during real outbreaks. These injects maintain the known relation between outbreak size and effect on surveillance data, which allows inferences to be made regarding the smallest outbreak that can be expected to be detectable.*

Results: *An example illustrates the use of HiFIDE to analyze detectability of a waterborne* Cryptosporidium *outbreak in Washington, DC.*

Conclusion: *HiFIDE enables public health departments to perform system validations recommended by CDC. HiFIDE can be obtained for no charge for noncommercial use (http://www.hifide.org).*

Introduction

When public health surveillance systems are evaluated, CDC recommends that the expected sensitivity, specificity, and timeliness of surveillance systems be characterized for outbreaks of different sizes, etiologies, and geographic or demographic scopes (*1*). An important approach for computing these metrics is the simulation of outbreaks.

Researchers have developed injection methods in which artificial spikes, perturbations of the surveillance data, are injected into a time series of real surveillance data from nonoutbreak periods (*2–4*). This method is called semisynthetic, because artificial data are injected into real data. After a spike is injected, any outbreak detection algorithm can be run on the injected time series to determine whether the spike can be detected, on what date, and with what false alarm rate. To understand how the detection algorithm would work on average, the injection is then repeated systematically with the inject date moving forward one time unit per repetition (Figure 1). From the results of this procedure, the parameters of sensitivity, false alarm rate, and timeliness for spike detection can be computed. To explore detection algorithm performance at different false alarm rates, the entire procedure is then repeated, varying the detection algorithm alarm threshold. The results are typically displayed graphically by using activity monitor operating characteristic (AMOC) curves (Figure 2) (*5*).

The primary limitation of the semisynthetic technique is that it only determines the smallest spike that could be detected, leaving unanswered the key detectability question, "What is the smallest outbreak that can be detected?" For example, the first published semisynthetic analysis injected spikes that increased linearly in height over time into daily sales data of cough products (*2*). The results indicated that if an outbreak increased sales by a factor of 1.36, it would be detected. The limitation is what size outbreak increases sales by a factor of 1.36 is unknown. A second limitation of semisynthetic analyses conducted is the use of geometrically shaped injects, which might be poor estimates of actual temporal outbreak contours and which do not account for variations in the data as a result of individual behavior.

An alternative injection technique addresses these limitations by forming injections with a shape derived from surveil-

FIGURE 1. An example of a semisynthetic process for creating outbreak data

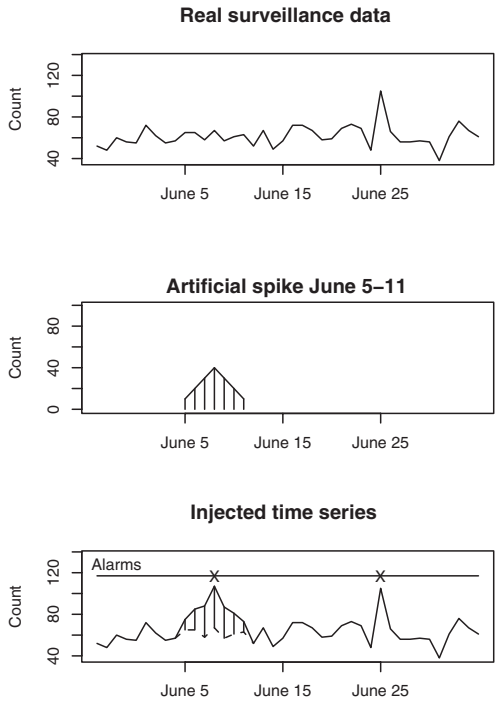
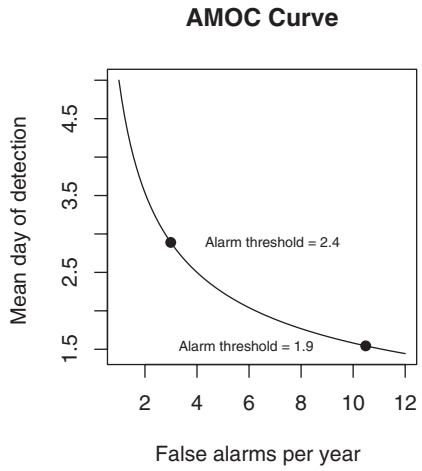


FIGURE 2. A hypothetical example of an AMOC* curve



* Activity monitor operating characteristic.

lance data collected during an actual outbreak (*6*). This inject is called a high-fidelity inject. Moreover, with this technique, the height of the inject is scaled in a method that preserves the known relationship between the magnitude of the real outbreak and the strength of the signal in the surveillance data from the real outbreak. Because of this property, a detectability analysis can be used to determine the smallest outbreak that can be detected. The scaling adjusts for differences in population size and in data completeness, which is the proportion of the data that is available in a jurisdiction. This technique allows health departments to ask whether an outbreak that occurred in some other region would have been detected in their own region, provided that the outbreak region was collecting the same type of surveillance data as the region performing the detectability analysis. The technique can be applied to the majority of surveillance data, including over-the-counter (OTC) sales data and emergency department registrations. However, the two regions might differ in population size, population density, and completeness of surveillance data.

High-Fidelity Injection Detectability Experiments (HiFIDE) is a software tool that uses high-fidelity injects to analyze detectability of surveillance systems. The mechanics of a HiFIDE analysis are similar to those of a semisynthetic analysis, although the software primarily automates this process. A user selects from a user interface the type of surveillance data and outbreak. The user can then easily create a substantial number of injects by using different values of outbreak size (defined as the proportion of the population that would be affected by the outbreak) and surveillance data completeness. HiFIDE combines each inject with real surveillance data to form a time series. HiFIDE then runs a set of detection algorithms on each injected time series, varying the alarm threshold, and summarizes the detectability results in AMOC curves and plots of sensitivity versus timeliness.

This report illustrates how HiFIDE can be used to investigate the detectability of a water-borne *Cryptosporidium* outbreak in the Washington, DC, metropolitan area by assessing data from sales of OTC diarrheal remedies. HiFIDE is used to address the following questions:

- What detection algorithm would be expected to earliest detect a *Cryptosporidium* outbreak in Washington, DC?
- What is the smallest *Cryptosporidium* outbreak we can expect to detect, given the available surveillance data in the city?
- How early can we expect to detect *Cryptosporidium* outbreaks of different sizes?
- How many false alarms per year would we have to tolerate to improve detection?
- How much earlier would detection occur if more pharmacies were successfully recruited to increase the completeness of sales data?

Methods

HiFIDE is a software application that runs under the Microsoft Windows® operating systems. It uses R as a computational backend to enable the rapid inclusion of sophisticated detection algorithms (*7*). However, no knowledge of R is required, because the user interacts with only the graphical user interface of HiFIDE.

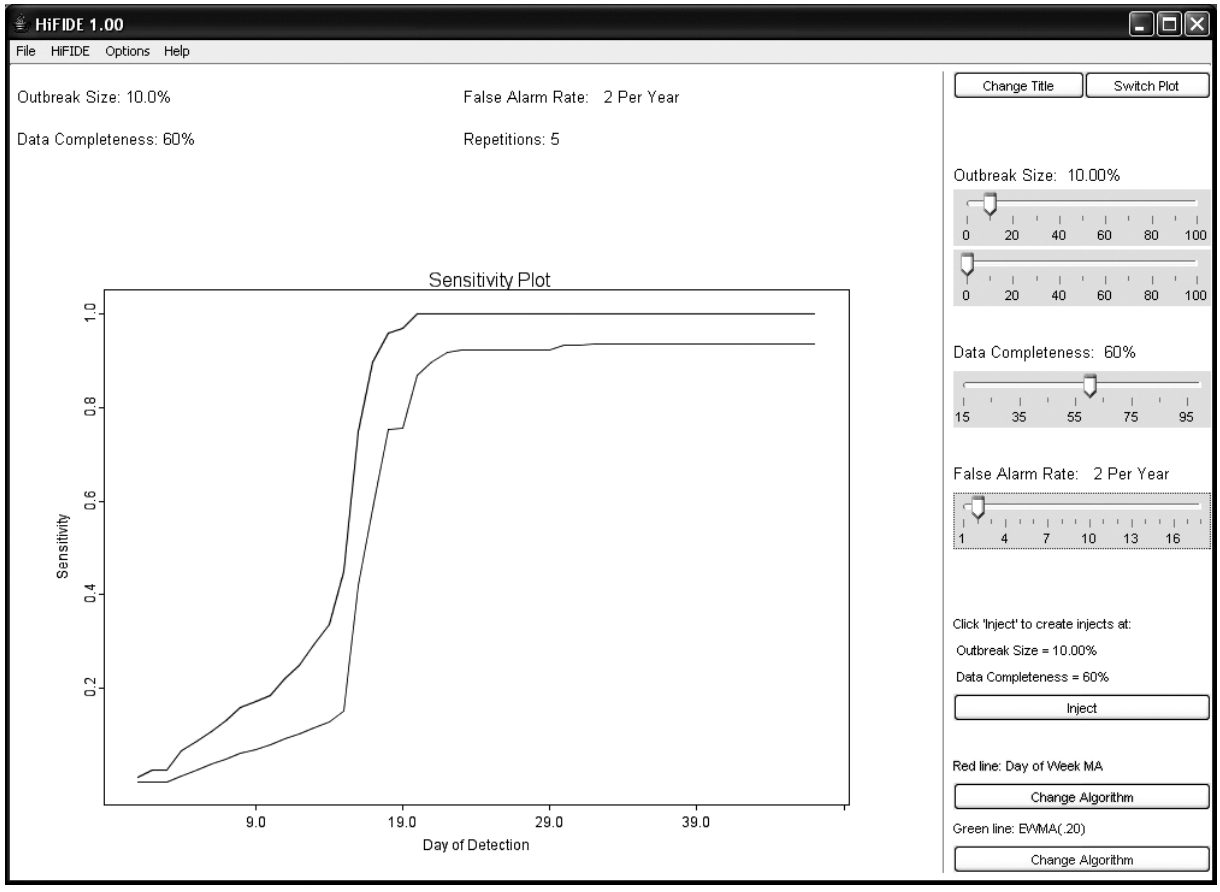
Initiating an Analysis

A user initiates a HiFIDE analysis by selecting an outbreak and surveillance data type from the HiFIDE library of outbreaks and data types. The user then selects a file with sampled surveillance data from their jurisdiction. The data must be organized as a daily time series of counts. Missing data are not allowed, and outbreak-dependent minimum data requirements are checked by HiFIDE to ensure that the time series is long enough to complete the analysis. The user also has the option of providing the data completeness for the sampled surveillance data. If the user provides data completeness, then HiFIDE permits investigation of the effect of varying the data completeness of the surveillance data. For example, the user could investigate the effect that recruiting additional retailers would have on detection timeliness.

Creating Injects

The user creates injects in collections called inject repetitions. An inject repetition consists of one inject for each feasible day in the sampled surveillance time series. Two horizontal sliders on the right of the screen (Figure 3) can be manipulated by the user to adjust the outbreak size and data completeness; the data completeness slider is only active if the data completeness for the sampled surveillance data was provided when the analysis was initiated. A third slider controls the

FIGURE 3. AMOC* curves in HiFIDE† for two detection algorithms§



* Activity monitor operating characteristic.
† High-Fidelity Injection Detectability Experiments.
§ The sliders on the right of the display control the outbreak size and data completeness.

false alarm rate and does not affect the creation of injects. When the user selects "Inject," one inject repetition is added for the selected values of outbreak size and data completeness.

Summarizing Detectability

HiFIDE presents the results of the analysis graphically by using AMOC plots (Figure 3), which depict the relationship between day of detection (computed relative to a reference date) and the false alarm rate, and plots of sensitivity versus day of detection (Figure 4). The user can alternate between these plots by selecting the "Switch Plot" button in the upper right-hand corner. The three sliders control the values of outbreak size, data completeness, and false alarm rate (the false alarm rate slider is only active for the sensitivity versus day of detection plot). HiFIDE reads the values of the sliders and selects values for outbreak size and data completeness that are closest to the slider-specified values for which at least one inject repetition exists. The values used by HiFIDE for the plot are displayed above the plot.

The user can select up to two algorithms to display at a time from the HiFIDE library of algorithms. The results for each of the algorithms are pre-computed when the injects are created, therefore enabling rapid switching between detection algorithms.

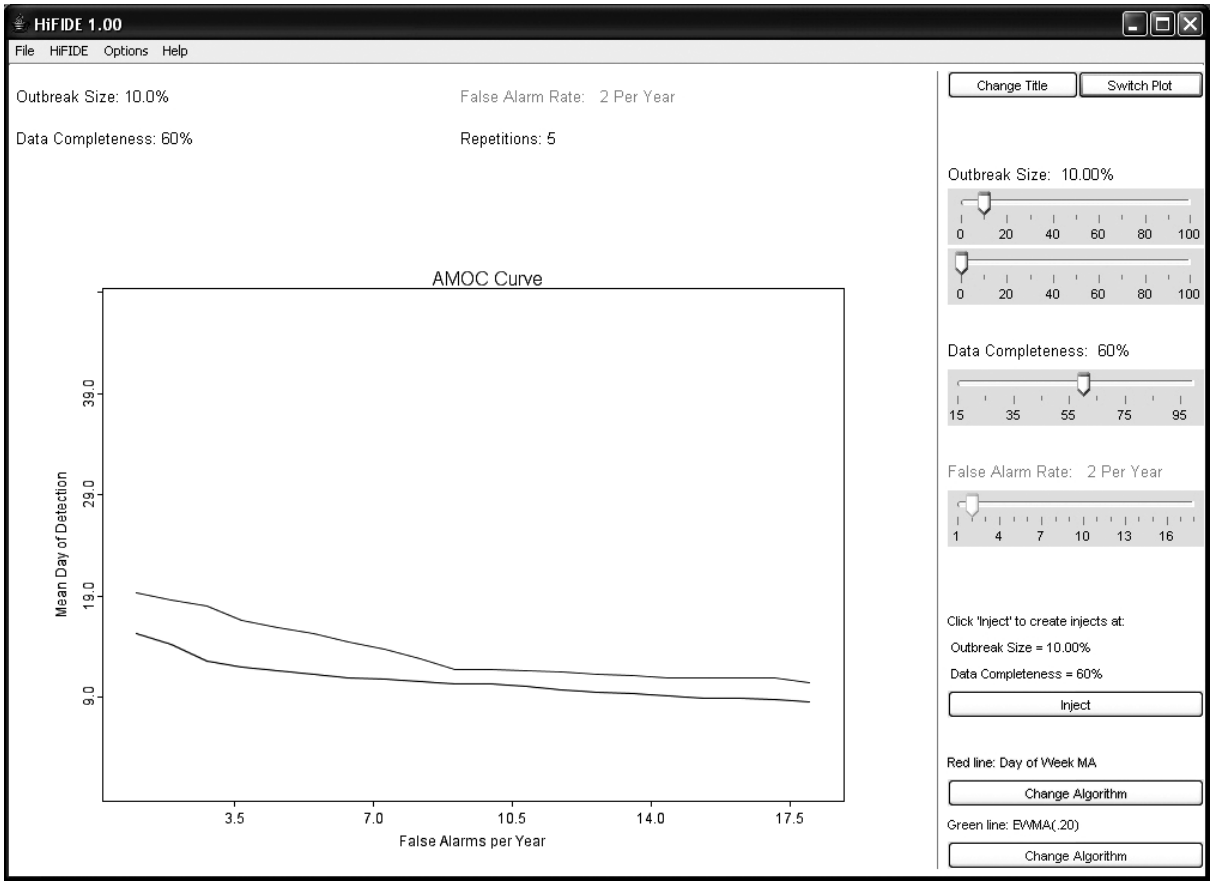
Results

This report illustrates how HiFIDE can be used to investigate the detectability of a water-borne *Cryptosporidium* outbreak in the Washington, DC, metropolitan area by assessing data from sales of OTC diarrheal remedies.

Jurisdictions

The focus of the detectability analysis was the Washington, DC, metropolitan area, which has a resident population of approximately 550,000. Daily sales of antidiarrheal products during August 9–December 20, 2003, were obtained from

FIGURE 4. Sensitivity plots in HiFIDE* for two detection algorithms†



* High-Fidelity Injection Detectability Experiments.
† The sliders on the right control the outbreak size, data completeness, and false alarm rate of the algorithms.

the National Retail Data Monitor (*8*). These sales represented approximately 89% of all such sales.

The analysis is based on a waterborne *Cryptosporidium* outbreak in North Battleford, Saskatchewan. The outbreak began on March 20, 2001, when the solids contact unit at a surface water treatment plant malfunctioned (*9*). Public health officials issued a precautionary drinking water advisory 5 weeks later on April 25 after laboratory-confirmed cases of cryptosporidiosis were identified. The date of the drinking water advisory is a reference date for measuring timeliness. The outbreak affected approximately 36% of the 18,000 residents. The number of weekly retail sales of diarrheal remedies was available (*9*) from a single pharmacy.

Detection Algorithms

Multiple simple detection algorithms were used for this analysis. The algorithms are a day-specific moving average (DSMA), autoregressive integrated moving average (ARIMA [1,0,1]) time series model (*3*), exponentially weighted moving average (EWMA) with weights of 0.05 and 0.2, and CUSUM with a baseline computed via EWMA with weights of 0.05 and 0.2 (*10*). The ARIMA parameters and EWMA weights were set to commonly used values for simplicity rather than by any model selection criteria.

Number of Repetitions

Multiple values of outbreak size and data completeness were used in the analysis. Five outbreak sizes were used with a range of 0.1%–10% in addition to three values of data completeness — 50%, 100%, and 89% — the data completeness for the surveillance data. Five inject repetitions were created for each pair of outbreak size (five sizes) and value of data completeness (three values). The entire computation took approximately 4.5 hours on a personal computer with a 2.4GHz Pentium 4 CPU.

Findings

What detection algorithm would be expected to detect a Cryptosporidium *outbreak in Washington, DC, the earliest?*

At a false alarm rate of four per year, the DSMA algorithm detects outbreaks of all sizes that are considered the earliest (Table 1). The AMOC curves can be inspected to determine how the false alarm rate affects relative algorithm performance. This report illustrates AMOC curves and a sensitivity versus timeliness plot for DSMA and ARIMA for an outbreak of size 1% (Figures 5 and 6); ARIMA detects the outbreak earlier than DSMA at smaller false alarm rates.

What is the smallest Cryptosporidium *outbreak that health departments can expect to be detected, given the available surveillance data in the city?*

The values (Table 2) for DSMA were interpolated to determine the smallest outbreak that is detected 2 weeks before public health response (using the North Battleford experience as a benchmark) at least 75% of the time. The smallest outbreak that is detectable is approximately 0.78% at a false alarm rate of four per year and 3.10% at a false alarm rate of two per year.

How early can health departments expect Cryptosporidium *outbreaks of different sizes to be detected in Washington, DC?*

When using DSMA, health departments can expect to detect an outbreak that affects 10% of the population 26 days before public health response, whereas an outbreak size of 1% is expected to be detected 21 days before public health response. The timeliness of detection for other algorithms and outbreak sizes are illustrated in this report (Table 1).

How many false alarms per year have to be tolerated to improve detection in Washington, DC?

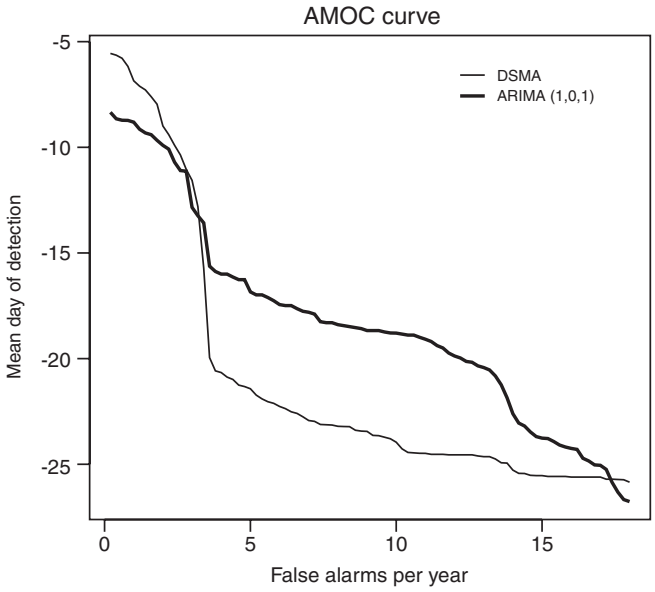
This report illustrates the trade-off between the false alarm rate and timeliness for an outbreak size of 1% for the DSMA and ARIMA algorithms (Figure 5). For example, increasing the false alarm rate from two to four per year improves timeliness by 11.67 days for DSMA and 6.09 days for ARIMA.

TABLE 1. Mean day of detection relative to public health response for a false alarm rate of 4 per year, by algorithm and outbreak size

Outbreak size	DSMA*	ARIMA† (1,0,1)	EWMA§ 0.05	EWMA 0.20	CUSUM¶-EWMA 0.05	CUSUM-EWMA 0.20
0.10%	-16.19	-15.4	-13.4	-6.11	-13.48	-11.43
0.50%	-17.81	-15.44	-13.45	-6.32	-13.52	-11.78
1.00%	-20.66	-16	-13.57	-6.33	-16.19	-12.45
5.00%	-24.76	-23.14	-22.89	-17.52	-23.41	-22.66
10.00%	-26.22	-24.74	-24.67	-22.2	-24.67	-24.18

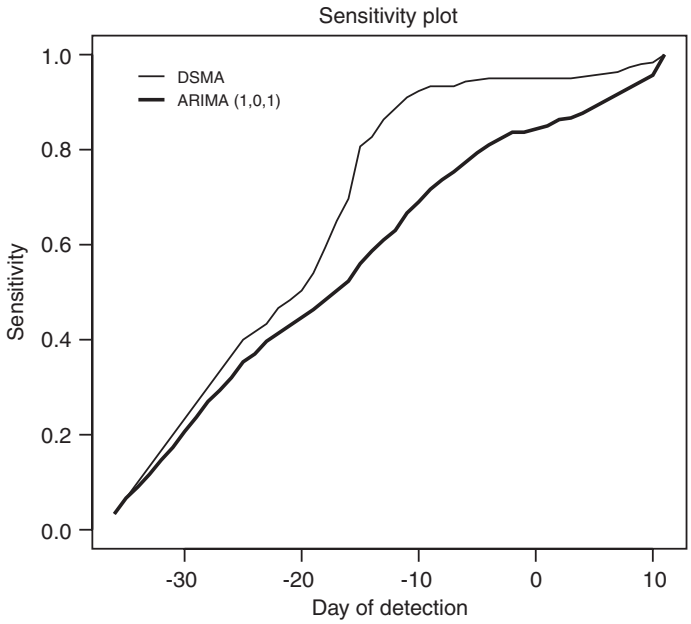
* Day-specific moving average.
† Autoregressive integrated moving average.
§ Exponentially weighted moving average.
¶ Cumulative sum.

FIGURE 5. AMOC* curves for the DSMA† and ARIMA§ (1,0,1) algorithms for the detectability analysis of *Cryptosporidium* — Washington, DC, 2003



* Activity monitor operating characteristic.
† Day-specific moving average.
§ Autoregressive integrated moving average.

FIGURE 6. Sensitivity plots of the DSMA* and ARIMA† (1,0,1) algorithms for the detectability analysis of *Cryptosporidium* — Washington, DC, 2003



* Day-specific moving average.
† Autoregressive integrated moving average.

TABLE 2. Probability (sensitivity) of detection at least 2 weeks before public health response for a false alarm rate of 4 per year, by algorithm and outbreak size

Outbreak size	DSMA*	ARIMA† (1,0,1)	EWMA§ 0.05	EWMA 0.20	CUSUM-EWMA 0.05	CUSUM-EWMA 0.20
0.10%	0.58	0.56	0.5	0.37	0.55	0.55
0.50%	0.65	0.56	0.52	0.38	0.55	0.55
1.00%	0.83	0.59	0.55	0.4	0.6	0.57
5.00%	1	1	1	0.86	1	1
10.00%	1	1	1	1	1	1

* Day-specific moving average.
† Autoregressive integrated moving average.
§ Exponentially weighted moving average.

How much earlier would detection be if more pharmacies were successfully recruited to increase the completeness of sales data?

The current completeness of data is approximately 89%, and limited improvement would be expected in timeliness if coverage was increased to 100%. The timeliness of detection of a 1% outbreak when using the DSMA algorithm at a false alarm rate of four per year only improves from -20.66 days to -20.94 days if the remaining 11% is obtained. The effect that timeliness has on losing pharmacies can also be investigated. If data completeness decreases to 50%, the timeliness increases to -19.74 days.

A limitation of this example is that the North Battleford surveillance data used to estimate the shape of the injects come from only a single pharmacy. However, weekly sales data from a second pharmacy also exhibited a similar temporal effect to the outbreak (*11*).

These results are from a system that aggregates counts for the entire jurisdiction. A full detectability analysis would exploit knowledge of the water distribution system, and substantial improvements in performance might be possible.

Discussion

The primary contribution of HiFIDE is that it enables public health departments to conduct detectability analyses for their jurisdictions. Public health departments can estimate the sizes of outbreaks that are expected to be detectable and the timeliness of their detection by using currently available surveil-

lance data. Decisions regarding whether to allocate resources to improve data completeness by recruiting more retailers or connecting additional hospitals to the surveillance system can also be explored. HiFIDE also enables researchers in the field of biosurveillance to evaluate and compare detection algorithms.

HiFIDE supports detectability analyses for C*ryptosporidium* and influenza when sales of diarrheal remedies and emergency department registrations with constitutional chief complaints are used, respectively. Additional outbreaks and data sources are expected to be added.

Conclusion

In the HiFIDE analysis, the substantial practical information that HiFIDE can provide users of surveillance systems is illustrated. In particular, HiFIDE provides the sensitivity, specificity, and timeliness (metrics recommended by CDC [*1*]) for outbreaks of different sizes, etiologies, and scopes. The HiFIDE tool is available at no charge for noncommercial use (http://www.hifide.org).

Acknowledgments

This report was supported by the Agency for Healthcare Research and Quality (290-00-0009), the Pennsylvania Department of Health (ME-01-737), the Department of Homeland Security (F30602-01-2-0550), and the National Library of Medicine (LM008278-01).

References

1. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V, CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks. MMWR 2004;53(No. RR-5):1–11.
2. 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.
3. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Medical Informatics and Decision Making 2003;3:2.
4. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. PNAS 2003;100:1961–5.
5. Fawcett T, Provost F. Activity monitoring: noticing interesting changes in behavior. In: Proceedings of the 5th Association for Computing Machinery (ACM) SIGKDD International Conference on Knowledge Discovery and Data Mining; 1997. San Diego, CA: Chaudhuri and Madigan; 1997.
6. Wallstrom GL, Wagner MM, Hogan WR. High-Fidelity Injection Detectability Experiments. RODS Technical Report; 2005.
7. R Development Core Team. R: a language and environment for statistical computing. Vienna Austria: R Foundation for Statistical Computing; 2004. Available at http://www.R-project.org.
8. 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.
9. Stirling R, Aramini J, Ellis A, et al. Waterborne cryptosporidiosis outbreak, North Battleford, Saskatchewan, spring 2001. Can Commun Dis Rep 2001;27:185–92.
10. Stoto MA, Schonlau M, Mariano LT. Syndromic surveillance: is it worth the effort? Chance 2004.
11. Stirling R, Aramini J, Ellis A, et al. Waterborne cryptosporidiosis outbreak, North Battleford, Saskatchewan, spring 2001. Health Canada 2001. Available at http://www.health.gov.sk.ca/mc_dp_health_can_epi_report_NB.pdf.

Linked Analysis for Definition of Nurse Advice Line Syndrome Groups, and Comparison to Encounters

Steven F. Magruder,[1] J. Henry,[2] M. Snyder[2]
[1]*Johns Hopkins University, Laurel, Maryland;* [2]*Kaiser Permanente of the Mid-Atlantic States, Rockville, Maryland*

Corresponding Author: Steven F. Magruder, Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Rd., Laurel, MD 20723-6099. Telephone: 443-778-6537; Fax: 443-778-6885; E-mail: steve.magruder@jhuapl.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Nurse advice call centers are a potentially important source of data for syndromic surveillance purposes. For this reason, researchers at Johns Hopkins University Applied Physics Laboratory and Kaiser Permanente of the Mid-Atlantic States (KPMAS) have been collaborating to develop methods to use this data within the ESSENCE II Syndromic Surveillance System in the National Capital Region.*

Objective: *The objective of this report is to present a general method for finding syndrome groups in data sources that can be linked to physician encounters and to determine effective advice call syndrome groups for use with KPMAS advice data.*

Methods: *Advice calls are linked to physician encounters and stratified by patient age. They are placed in groups according to a maximum positive predictive value criterion. The groups are evaluated by correlating the resulting syndrome time series against physician encounter data.*

Results: *Potentially useful advice syndrome groups are found for respiratory, lower gastrointestinal (GI), and total GI syndromes for each age stratum.*

Conclusion: *The time series of the advice data for respiratory, lower GI, and upper GI syndromes accurately predict the physician encounter time series for the corresponding syndromes for each age stratum.*

Introduction

Advice lines are facilities such as nurse call centers that receive telephone calls from persons requiring information, triage, or immediate assistance. Government agencies, healthcare systems (especially HMOs), or private contractors can operate these facilities. Advice lines gather information about a caller's complaints that enables them to allocate appropriate clinical resources to that patient with appropriate urgency. This practice results in an electronic record carrying syndromic information.

Nurse advice data for syndromic surveillance purposes have several potential advantages. The syndromic information is immediately captured in an electronic format. The information has been interpreted by a medical professional. The advice call event might occur either before or instead of a physician encounter (*1*). Nurse advice call volume does not decline as sharply on weekends and holidays as does physician-encounter volume. For these reasons, researchers at Johns Hopkins University Applied Physics Laboratory and Kaiser Permanente of the Mid-Atlantic States (KPMAS) have been collaborating to develop methods to use this data within the ESSENCE II Syndromic Surveillance System in the National Capital Region (*2*). The potential use of this type of data for surveillance has also received attention in the United Kingdom (*3*).

At KPMAS, 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. Nurses operate the advice hotline, administering protocol-driven, medically appropriate advice and scheduling acute-care office visits when necessary. By this process, each call is assigned one of 527 possible advice guidelines (e.g., diarrhea, adult, pinworms, or contact exposure) indicating syndromic information. This report addresses how best to aggregate this information to characterize trends in broad syndrome groups, particularly those defined by CDC relating to critical biologic terrorism–associated agents (*4*). A previous paper on the use of advice guideline groupings for syndromic surveillance evaluated the performance of groupings chosen on the basis of names of the advice protocols and their presumed usage (*1*). For this study, a method was used for finding empirical groupings.

Methods

The KPMAS information system captures the date and time a patient was seen for an outpatient office visit and the date and time the patient called to schedule that appointment. A third record contains the date and time the patient contacted the nurse advice hotline. For this study, a nurse advice hotline call and an outpatient office visit are defined as linked if all of the following criteria were met: 1) the patient identification number assigned to the nurse advice hotline call is the same as the patient identification number assigned to the office visit; 2) the calendar date of the nurse advice hotline call is the same as the calendar date when the patient called for the appointment; and 3) the time of the nurse advice hotline call was equal to, or earlier, than the time the patient called for an appointment.

Using the KPMAS information system, a database was created that contained all nurse advice calls made during calendar year 2002 that could be linked to corresponding physician encounters and *International Classification of Diseases, Ninth Revision* (ICD-9)-coded physician diagnoses. Approximately 38% of all advice calls were linked in this way for approximately 570,500 linked call-encounter pairs. These physician diagnoses provide a standard for classifying a case into a particular syndrome group. Visits were grouped according to the CDC diagnosis-based biologic terrorism syndrome groups (including categories 1 and 3 only) (*4*) (Table 1). Approximately 26% of the linked call-encounter pairs were classified into one of these CDC groups.

For each advice guideline, the fraction of the calls linked to diagnoses in each syndrome group and the fraction linked to diagnoses outside of all the groups (i.e., falls in the nonsyndromic group) were determined. These fractions represent the historic positive predictive value (PPV) of each advice guideline for predicting each syndrome type. Advice syndrome groups were then formed by including in a given group only those guidelines whose PPV is greater for that group than for any other group (including the nonsyndromic group). If PPV was equal for two syndrome groups, the same guideline was assigned to both groups.

TABLE 1. Percentage of syndrome groups linked to nurse advice calls and median lag of visit relative to advice call, 2002

Syndrome group	% Linked to nurse advice calls	Median lag of visit relative to advice call (hrs)
Botulism-like	58	19
Fever	66	4
Gastrointestinal-lower	75	4
Gastrointestinal-upper	68	5
Hemorrhagic illness	56	14
Localized cutaneous lesion	61	6
Lymphadenitis	51	13.5
Neurologic	57	16.5
Rash	51	7
Respiratory	59	6
Severe illness or death potentially caused by infectious disease	73	4

The advice syndrome groups formed by this procedure are then evaluated by comparing time series for advice calls (daily counts in each syndrome group) against physician diagnoses time series in each syndrome group, using a different period than the one used to define the groups. The period used for this analysis was January 2003–November 2004. Rather than estimating marginal probabilities that a call will fall into a given syndrome group, time series were compared because the identification of disease anomalies typically depends on observations of temporal trends and because this allowed inclusion of both linked and unlinked advice calls and physician visits in the evaluation.

For each syndrome group, r^2 (i.e., the fraction of the variance in the physician diagnosis daily counts explained by the nurse advice daily counts) was calculated. This comparison of same-day counts was conducted because of the short observed median time lags (Table 1). Nevertheless, the actual correspondence between calls and encounters might be higher than the same-day comparisons suggest. Except for day of week, holiday, and snow corrections described below, r^2 is calculated directly from the time series of daily counts. No corrections are made for serial autocorrelation or for seasonality; all time scales are assumed to be equally valid for evaluating the correspondence between these two data sources.

For some comparisons, the data by patient age (which is available both in the nurse advice records and in the physician encounter records) was stratified. The age classes included infant (aged 0–1 years), pediatric (aged 2–17 years), and adult. For some comparisons, data were stratified by day of week (i.e., weekends and holidays, Mondays and days after holidays, and other weekdays). When day of week was not stratified, each daily data count was divided by the mean value for its day-of-week class and then the data was smoothed with a 7-day moving average before calculating r^2. This step was necessary to correct for differing day-of-week patterns in the advice and encounter time series. The 7-day average was used to improve the statistical strength of the data when stratifying by age (i.e., daily counts for a given age and syndrome group were sometimes too small to provide meaningful results). The use of 7-day averages also mitigates any loss of correlation caused by small lags between the time of an advice call and the time of a physician encounter.

Before any other processing steps, 8 days were culled from the data set because of large snowfalls, which apparently reduced the numbers of physician encounters without noticeably affecting the numbers of advice calls.

Use of Positive Predictive Value

The goal in defining syndrome groups is to provide a good estimate for any given day of the number of advice calls that will eventually result in physician diagnoses falling in a given syndrome group. Denoting the number of advice calls using guideline i as a_i, the number of those that would be diagnosed in the syndrome group (given a physician examination) as s_i, and the sum of all calls that would be diagnosed into the syndrome group as S, the object is to find coefficients, f_i that minimize the mean squared error, E given by

$$E = Mean[(S - \sum_i f_i a_i)^2]. \quad (1)$$

In this case, the mean is taken over an ensemble of days on which advice call and physician encounter counts are measured. Ignoring correlations between numbers of advice calls following different guidelines, the condition (obtained by differentiating Equation 1 with respect to f_i) for minimizing the squared error can be written as

$$\bar{a}_i(\bar{S} - \sum_j f_j \bar{a}_j) + Cov(a_i, s_i) - Var(a_i) f_i = 0, \text{ for all } I, \quad (2)$$

where overbars denote mean value. If it is further assumed that the number of advice calls that would not be diagnosed in the syndrome group are uncorrelated with those that would, and also that fluctuations both in a and in s follow a Poisson distribution (so that variances equal means), the equation would be

$$\bar{a}_i(\bar{S} - \sum_j f_j \bar{a}_j) + \bar{s}_i - \bar{a}_i f_i = 0, \quad (3)$$

which is solved by

$$f_i = \bar{s}_i / \bar{a}_i. \quad (4)$$

In words, f_i is the positive predictive value of guideline i for predicting the syndrome in question.

Advice syndrome groups could be formed as weighted sums, as implied by Equation 1. However, assigning each advice guideline to a single group adds clarity to the interpretation of the data. When unique assignments are desired, each guideline should be assigned to the group for which it would be most heavily weighted, the one for which it has the greatest PPV.

Results

Using the method described, empirical nurse advice syndrome groups were identified for fever, lower gastrointestinal (GI), upper GI, total GI, rash, and respiratory syndromes. All advice guidelines in the upper and lower GI groups also fall in the total GI group. When estimating whether a guideline belongs in upper or lower GI, the total GI group was not considered as an alternative diagnosis. No advice guidelines had maximum PPV for any of the other CDC syndrome groups. The advice calls falling into the fever, upper GI, and rash syndromes were too infrequent to be useful as a predictor of the encounter data (Table 2). Potentially useful advice guideline groupings were found for the lower GI, total GI, and respiratory syndrome groups (Tables 3–5). R-squared values calculated for these groupings and for the various subpopulations described previously fell in the range 63%–95% (Table 6). These values were determined by correlating the counts of a given syndrome group constructed on the nurse advice guidelines with the (same day) counts of the same syndrome obtained from physician encounters, as explained in detail in the methods section.

Discussion

Certain syndrome group definitions contain entries that seem to be spurious coincidences (e.g., "ADMINISTRATIVE NOTE" in the infant GI and Respiratory syndromes or "FEEDING, NEONATAL" in Adult GI). These advice guidelines were rare and have negligible effect. They were left in the list to keep the methodology clear; no *ad hoc* editing of the syndrome groups was performed.

Certain entries seem inconsistent with respect to age groups (e.g., some guidelines labeled "adult" are included in the infant syndrome groups, and some labeled "peds" are included in the adult groups). This might indicate some inconsistent use of the age designations in the guideline names, but the actual age of the patient is recorded in the nurse advice call records; therefore, these ambiguities do not interfere with the ability in practice to stratify guideline syndrome groups by age. These age-inconsistent guideline names should not pose a problem.

TABLE 2. Average daily counts observed in Kaiser Permanente of the Mid-Atlantic States data, January 2003–November 2004*

Encounter	Fever	GI†-Lower	GI-Total	GI-Upper	Rash	Respiratory
Infant advice	0.06	8.84	17.63	0.00	0.02	5.77
Infant encounters	15.79	8.21	10.15	1.93	0.85	43.54
Pediatric advice	0.20	8.47	29.64	0.04	0.02	162.73
Pediatric encounters	38.76	12.75	18.62	5.87	2.54	132.60
Adult advice	0.10	36.23	36.31	0.07		163.12
Adult encounters	27.47	29.23	45.76	16.53	8.19	235.66

*No advice syndrome groups were formed for adult/rash.
†Gastrointestinal.

TABLE 3. Selected advice guideline syndrome group definitions for infants, determined by a linked analysis of Kaiser Permanente of the Mid-Atlantic States advice and encounter data, 2002

Syndrome groups/ Advice protocol name	Total	Positive predictive value (%)
Lower gastrointestinal (GI)		
Administrative note, advice supervisor	1	50
Diarrhea, 0–24 months, peds	1,103	54
Diarrhea, >2 years, peds	7	50
Diarrhea, peds	154	52
Pinworms, contact exposure	2	100
Reye's syndrome, peds	1	100
Total GI		
Acute GI, gastroenteritis, adult	1	33
Administrative note, advice supervisor	1	50
Diarrhea, 0–24 months, peds	1,141	56
Diarrhea, >2 years, peds	8	57
Diarrhea, peds	162	55
Pinworms, contact exposure	2	100
Reye's syndrome, peds	1	100
Vomiting, peds	940	41
Respiratory		
Administrative note, advice supervisor	1	50
Bronchiolitis, peds	362	52
Bronchitis, acute, adult	3	75
Bronchitis, chronic, adult	1	100
Croup, peds	238	60
Dehydration, adult	1	50
Epiglottitis, peds	2	50
HIV pneumonia, adult	1	100
Hypothermia, adult	1	100
Influenza, peds	15	37
Kaiser role, NOR/SOR operations	1	50
Neonatal sleep position	2	50
Phenylpropanolamine	1	100
Respiratory distress, adult	1	50
Sore throat, adult	1	100
Upper respiratory infection, long term	2	50

One type of finding in these results consists of advice guidelines that are unexpectedly missing from these groupings. For example, adult nausea and adult vomiting are not included because it was discovered that these complaints most often result in diagnoses falling outside of the CDC syndrome groups. However, if a particular public health threat of concern could present itself as adult nausea and vomiting only, then it would be important to monitor this channel also, but perhaps not aggregated with the other complaints, which are more specific to the GI syndrome as it usually occurs. Other similar examples are adult upper respiratory infection and respiratory distress in adults. The absence of acute laryngitis in the adult list has a different type of explanation; 4-digit ICD-9 codes were sometimes recorded in the database, while the CDC syndrome definitions only include the 5-digit codes for acute laryngitis. For methodological clarity, no attempt was made to "fix" this type of problem.

TABLE 4. Selected advice guideline syndrome group definitions for pediatrics, determined by a linked analysis of Kaiser Permanente Mid-Atlantic States advice and encounter data, 2002

Syndrome groups/ Advice protocol name	Total	Positive predictive value (%)
Lower gastrointestinal (GI)		
Diarrhea, 0–24 months, peds	26	48
Diarrhea, >2 years, peds	772	54
Diarrhea, adult	7	64
Diarrhea, long-term care	1	100
Diarrhea, peds	104	48
Diarrhea-prenatal, OB/GYN	2	67
ECM script	2	67
Pinworms, adult	1	100
Stool occult blood test, adult	1	50
Urgent/bun, creatinine, intact pth l	1	50
Urgent/Co_2 urgent values	1	33
Urgent/stat lab values	2	50
Total GI		
Acute GI, gastroenteritis, adult	5	50
Anxiety attack, adult	1	100
Dehydration, adult	2	50
Diarrhea, 0–24 months, peds	28	52
Diarrhea, >2 years, peds	807	56
Diarrhea, adult	7	64
Diarrhea, long-term care	1	100
Diarrhea, peds	110	51
Diarrhea-prenatal, obgyn	2	67
ECM script	2	67
Nausea-prenatal, OB/GYN	1	50
Pinworms, adult	1	100
Stool occult blood test, adult	2	100
Urgent/bun, creatinine, intact pth l	1	50
Urgent/Co_2 urgent values	2	67
Urgent/stat lab values	3	75
Vomiting, peds	2,229	44
Respiratory		
Bronchiolitis, peds	265	50
Bronchitis, acute, adult	17	49
Chicken pox, adult	1	50
Croup, peds	421	61
Epiglottitis, peds	8	50
Fever, adult	3	38
Fever, neonatal	2	67
Fever, peds	2,249	34
Influenza, peds	406	36
Lab results	348	41
Laryngitis, adult	2	67
Learning disabilities, peds	1	100
Medications-prenatal, OB/GYN	6	55
Meningitis, peds, 3 months–2 years	3	43
Mononucleosis, adult	2	67
Overdose, adult	1	100
Pediatric OTC chart (36–59 lbs.)	82	48
Pharmacy questions	3	43
Reye's syndrome, peds	3	50
Sore throat, adult	54	53
Sore throat, peds	6,561	57
Throat culture, positive	47	58
Throat culture, positive results	84	58
Upper respiratory infection, long term	1	100
Upper respiratory infection, peds	8,594	46

TABLE 5. Selected advice guideline syndrome group definitions for adults, determined by a linked analysis of Kaiser Permanente Mid-Atlantic States advice and encounter data, 2002

Syndrome groups/ Advice protocol name	Total	Positive predictive value (%)
Lower gastrointestinal (GI)		
Acute GI, gastroenteritis, adult	1,076	49
Diarrhea, >2 years, peds	1	50
Diarrhea, adult	2,534	59
Diarrhea-prenatal, OB/GYN	50	55
Diet change, long-term care	1	50
HIV diarrhea, adult	15	47
Pinworms, contact exposure	1	50
Total GI		
Acute GI, gastroenteritis, adult	1,224	56
Diarrhea, >2 years, peds	1	50
Diarrhea, adult	2,614	61
Diarrhea-prenatal, OB/GYN	54	59
Diet change, long-term care	1	50
Drug ingestion, peds	1	100
Feeding, neonatal	1	50
HIV diarrhea, adult	16	50
Pinworms, contact exposure	1	50
Respiratory		
Bronchiolitis, peds	1	50
Bronchitis, acute, adult	10,786	59
Bronchitis, chronic, adult	306	55
Influenza, adult	1,855	40
HIV dyspnea, adult	11	61
HIV pneumonia, adult	14	47
Medications-prenatal, OB/GYN	224	49
Meningitis, peds, >2 years	1	50
Sore throat, adult	7,456	60
Sore throat, peds	26	54
Throat culture, positive	15	79
Throat culture, positive results	40	63

The most marginal case is in respiratory syndrome for infants; this case has a relatively low number of identifiable nurse advice calls (Table 2).

Conclusion

A patient-linked mapping of nurse advice guidelines to physician diagnoses has yielded advice syndrome groups that accurately track the temporal behavior of the CDC diagnosis-based groups for respiratory, lower GI, and total GI syndromes separately for infants, pediatrics, and adults in approximately 2 years of data collected in the National Capital Region. This same technique indicates that advice syndrome groups that might be created for other single CDC-defined syndromes will be substantially less specific than the diagnosis-based syndrome groups; more cases would probably be determined to fall outside of the syndrome groups than in them on examination by a physician.

References

1. Henry JV, Magruder S, Snyder M. Comparison of office visit and nurse advice hotline data for syndromic surveillance—Baltimore–Washington, DC, metropolitan area, 2002. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:112–6.
2. Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the framework for evaluating syndromic surveillance systems. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:159–65.
3. Baker M, Smith G, Cooper D, et al. Early warning and NHS Direct: a role in community surveillance? J Public Health Med 2003;25:362–8.
4. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Available at http://www.bt.cdc.gov/surveillance/syndromedef.

TABLE 6. Percentage of physician encounter variance modeled by same-day nurse advice call fluctuations in Kaiser Permanente Mid-Atlantic States data, January 2003–November 2004

Days of week	Age	Respiratory	Gastrointestinal-lower	Gastrointestinal-total
Weekends and holidays	All ages	80	73	81
Mondays and holidays plus 1 day	All ages	88	78	82
Other weekdays	All ages	87	76	82
All days, smoothed	Infants	63	92	94
	Children	92	87	94
	Adults	95	86	82
	All ages	94	94	94

Simulation and Other Evaluation Approaches

Simulation for Assessing Statistical Methods of Biologic Terrorism Surveillance

Ken P. Kleinman,[1] A. Abrams,[1] K. Mandl,[2] R. Platt[1,3]

[1]*Harvard Medical School, Harvard Pilgrim Health Care, and CDC Eastern Massachusetts Prevention Epicenter and HMO Research Network Center for Education and Research in Therapeutics, Boston, Massachusetts;* [2]*Children's Hospital Boston, Boston, Massachusetts, and Harvard Medical School, Boston, Massachusetts;* [3]*Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts*

Corresponding author: Ken P. Kleinman, Harvard Medical School, Department of Ambulatory Care and Prevention, 133 Brookline Ave., 6th Floor, Boston, MA 02215. Telephone: 617-509-9935; Fax: 617-859-8112; E-mail: ken_kleinman@harvardpilgrim.org.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Multiple systems have been developed that use surveillance of health-care encounters to provide early warning of a terrorist attack. Limited practical experience and the absence of adequate theoretical assessments have precluded determining which alarm-generating algorithms should be preferred. In the absence of practical and theoretical results, choosing a particular statistical algorithm can be difficult. One way to evaluate algorithms is through simulation.*

Objectives: *This report describes conceptual features of an example simulation based on the dispersal of anthrax spores and presents results based on the example simulation.*

Methods: *A simulation was implemented based on the dispersal of anthrax spores from a crop-dusting plane. Simulated cases were then included into an observed data stream. Detection approaches included SaTScan™ and small area regression and testing (SMART) scores. An evaluation metric was developed for comparison of results. In addition, a simulation of a separate data stream was added; and then separate and combined surveillance data were compared.*

Results: *In the simulation in which a single data stream was used, the two statistical approaches were substantially similar in performance. The combined surveillance based on two data streams is superior to surveillance based on either stream separately.*

Conclusion: *The other potential uses of such a system are considered. These uses include the comparison of different data sources (e.g., outpatient versus emergency department and evaluating the impact of potential changes to the surveillance system, and increasing the population under surveillance). Simulation is a valuable technique for evaluating and planning for syndromic surveillance.*

Introduction

Since the attacks on the World Trade Centers on September 11, 2001, increasing attention has been paid to early detection of biologic terrorism attacks (*1,2*). Assuming biologic agents are not detected directly in the environment, the first indicator of a terrorist attack might be persons becoming ill. Surveillance systems designed to detect increased numbers of ill persons, consistent with exposure to these agents, have been developed (*3–7*). Key to these systems is the technique used to decide whether "too many" ill persons have been observed. However, despite an increasing number of statistical (*8,9*) and data mining (*5*) techniques, limited comparative work has been conducted to assess their relative strengths.

Theoretical comparisons are probably not possible in realistic settings, and practical experience is limited by small numbers of attack-like events. Thus, simulation is an important route to evaluation. Unfortunately, most simulations published to date have been too simple for meaningful analysis, except for a previously published analysis of a system generated by an epidemic simulation model (*10*).

The objective of this report is to present a simulation structure developed in the setting of an outpatient surveillance system (*11,12*). Simulation is only of cases generated by attacks, as opposed to cases arising both from attacks and natural disease. In contrast to the system generated by the epidemic simulation model (*10*), the outpatient system is simple in that it can be replicated using data likely to be available in many contexts. Thus, the simulation described in this report can be transferred to another data setting with ease.

Simulations can be used to determine the cost-benefit ratio of the surveillance system, to compare the relative value of different data sources and other features, and to evaluate statistical methods. Thus, it can be used to help determine the value of syndromic surveillance.

Methods

Outpatient Surveillance System

Simulation was performed by using a surveillance system that relies on data collected as part of outpatient visits near Boston, Massachusetts (*11,12*). When insured patients visited a clinic, their health-care provider created an electronic medical record. At the close of the visit, the provider entered *International Classification of Diseases, Ninth Revision* (ICD-9) (*13*) codes that described the patient's condition.

Each outpatient visit was classified into broad groups or syndromes of ICD-9 codes (e.g., lower gastrointestinal or neurologic symptoms). The intended effect of these groupings is that a patient's condition will likely be classified as a given syndrome, despite variability in coding practices among providers. Next, a census location and ZIP code, based on the patient's health maintenance organization record, were attached to the encounter. Finally, for privacy reasons, patient identities were removed and only the number of patients in each geographic region (e.g., census tract or ZIP code) was recorded. The count of each syndrome in each region was used for analysis.

For example, the respiratory syndrome includes 171 ICD-9 codes, including those for bronchitis, cough, and pneumonia. This syndrome is of particular interest because a case of inhalational anthrax identified in the prodromal phase would probably receive an ICD-9 code in this group (*11*).

Conceptual Simulation

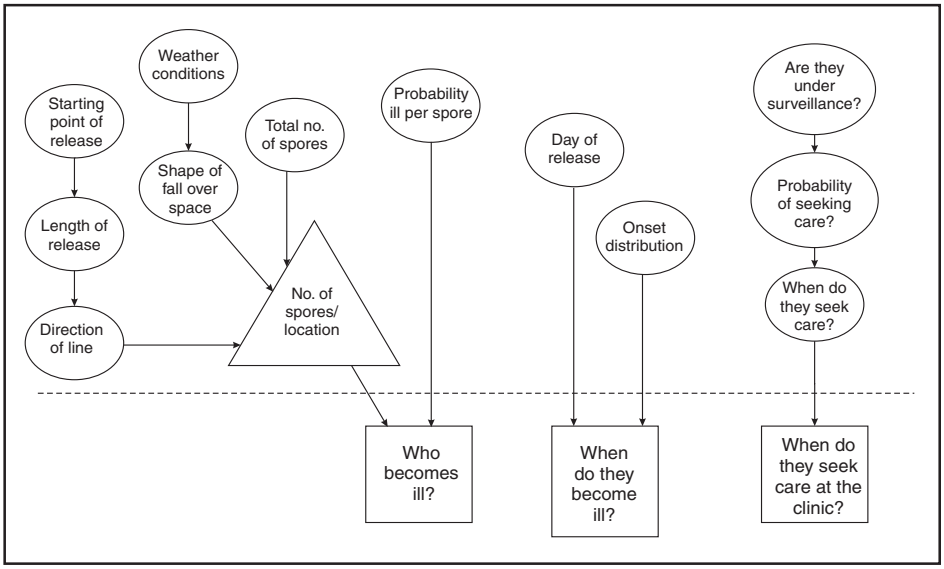
In concept (Figure 1), the first step of the simulation is to determine who becomes ill. For simulation of an anthrax attack, each person is exposed to a number of spores; this number is a function of the total number of spores released by the terrorist and what proportion of spores fall where the person is located. For each spore to which a person is exposed, a certain probability exists that they will become ill. The second step is to determine when they become ill. The time from infection to initial symptoms is variable. For each simulated infected person, the day on which their symptoms appear must be simulated. This day of initial symptoms depends on the day of the release of spores and on the distribution of symptom onset times. The third step is to determine, among persons eligible for surveillance, the probability of seeking care at a clinic and the time that care was sought. For infected persons who are eligible for surveillance, some probability exists that they will go to the clinic and therefore become part of the observed count for respiratory illness. They also visit the clinic at a simulated time after symptom onset.

Implementation of the Conceptual Simulation

In our implementation of the conceptual simulation, anthrax spores were assumed to have been dropped from the height of a crop-dusting plane (*14*). Exposures occurred only in the ZIP code of residence and to all residents of each ZIP code. Time was treated as discrete in days. These choices make the simulation simpler and mirror the nature of the underlying surveillance system.

To approximate the total number of persons living in each area, U.S. Census data were used. Persons in each area were randomly chosen to be included in the surveillance system. Because the exact location of the real persons under surveillance or of the other persons within each area was unknown,

FIGURE 1. Conceptual description of the simulation*



* Ovals represent modifiable assumptions, which can be fixed or drawn from random distributions. Triangles are deterministic functions of choices. Rectangles represent functions of the inputs that are applied to generate random features of each simulated person.

locations were chosen for them. Approximately 100 locations were chosen within each ZIP code. The population under surveillance and the remaining persons were distributed among these points equally for the purposes of calculating the number of spores to which persons were exposed.

Simulating Illness

Whether a simulated person becomes ill depends on how many spores they are exposed to and on the probability that each spore makes them ill. This step is the most complicated part of the simulation.

The number of spores that fall over any particular area on the map is a function of multiple parameters. The first parameter is the point from which the dispersal starts. The simulated release could begin at any point on the map. Reasonable distributions across space for this point include a uniform distribution, distributions based on population patterns and wind patterns, and distributions based on perceived levels of surveillance sensitivity. The map was stratified into an urban region around Boston and a suburban region; within each region, release points were generated from a spatially uniform distribution.

The spores could all be released at a single point on the map or along a line or curve mimicking a flight path. These cases are described as point-source versus line-source exposure. However, defining a point-source release as a release with length 0 would be more flexible. The length of release could be chosen from any random function with positive value, if one had some reason to expect various lengths. A length of 0 is used in this implementation. Finally, the number of spores falling at a given spot depends on how they fall. In this implementation, the anthrax spores are assumed to fall according to a Gaussian plume (*15*) function. This process is a flexible function, which defines for a given point downwind the proportion of released spores that will fall at any given spot. This function has been used to describe the concentration of spores in the context of an accidental release of anthrax (*14,16*).

The Guassian plume in its most general form (*15*) is simplified in this implementation by making constant the breathing rate (an estimated value of 0.03), the windspeed (5 m/sec = 18 kph), and the height above the ground (0), and inserting these values into the general equations. Inserting these constants into the general form generates the simplified function

$$s(x,y)=\frac{0.03Q}{5\pi\sigma_y\sigma_z}e^{-\frac{y^2}{2\sigma_y^2}-\frac{h^2}{2\sigma_z^2}}$$

where *s(x,y)* is the number of spores inhaled by a person x meters downwind and y meters in the crosswind direction from the release point. Q is the number of spores released, and h is the height of the release.

Parameters also are available to represent the height of release and the weather conditions. These parameters can be chosen randomly. Instead, for the weather conditions, two different sets of parameters are used: those derived from the accidental release (*14,16*) and the parameters described as Class A (*17*). For the Sverdlovsk parameters (*14*), these are

$$\sigma_y=\frac{0.08x}{\sqrt{1+0.0001x}} \text{ and } \sigma_z=\frac{0.06x}{\sqrt{1+0.0015x}}.$$

For the Class A weather patterns (*17*), they are

$$\sigma_y=(0.029\times10)^{0.53x^{-0.22}}\,10^{-0.27819}\,x^{0.865} \text{ and}$$

$$\sigma_z=(0.029\times10)^{0.53x^{-0.22}}\,0.28x^{0.9}.$$

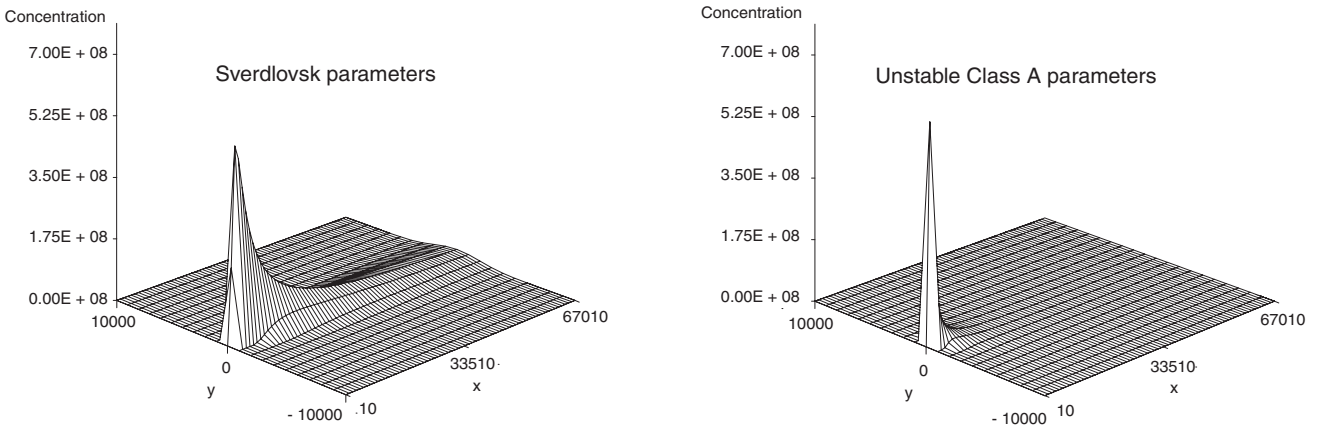
The height is set at the height of a typical crop-dusting plane, 100 meters (*14*). The resulting spore distributions are illustrated (Figure 2).

The number of spores falling at a given spot depends on how many spores were released. The approximate number of spores of anthrax per kilogram is 10^{15} (*14*). Without any specific information regarding the number of spores a terrorist can obtain, almost any positive-value distribution could be used to simulate the number of spores. The implementation described in this report includes 10^{15} spores.

The starting point, length and direction of dispersal, shape, and the number of spores together identify the number of spores that fall at any spot. This calculation was made for each point at which persons are assigned to live. The number of spores a person is exposed to is equal to the number of spores falling at their assigned location.

The other component describing the probability that each simulated person becomes ill is the probability of illness per spore. Estimates that result in probabilities of illness per spore ranging from 6.9×10^{-7} to 1.2×10^{-4} have been previously described (*18*). A distribution of probabilities per spore could also be chosen; a rescaled beta distribution might be an appropriate choice in this instance. In addition, the probability of illness might depend on individual characteristics (e.g., age) that are not available in the current application. The anthrax available to terrorists might be less functional than that indicated by the accidental release. Simulations were stratified to include five different values of probability of illness per spore: 10^{-10}, 5×10^{-10}, 10^{-9}, 5×10^{-9}, and 10^{-8}. This range was chosen to incorporate both cases in which few and most simulated attacks could be detected. An alternative approach to achieve this goal would be to use the derived probabilities (*18*) but to reduce the assumed quantity of anthrax spores released.

FIGURE 2. Number of spores above points on the map for two different weather conditions, Sverdlovsk parameters and Unstable Class A parameters*



* Corresponds to the Sverdlovsk parameters (Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proceedings of the National Academies of Science 2003;100:4346–51), and the Unstable Class A parameters (Spijkerboer HP, Beniers JE, Jaspers D, et al. Ability of the Gaussian plume model to predict and describe spore dispersal over a potato crop. Ecological Modeling 2002;155:1–18). The X and Y axes are measured in kilometers from the release; Z (height), in number of spores.

The probability of a person being affected was calculated by using the number of spores and the probability of illness per spore. The probability of illness per spore was treated as independent of the number of spores. The binomial distribution was used to calculate the probability of each person becoming ill as 1- the probability no spore makes them ill = 1 - (1 - probability of illness per spore)$^{\text{number of spores}}$. If data suggested that the probability of illness depended on the number of spores present, this individual-dependent probability could be incorporated.

Timing of Illness

Patterns exist across the year and weekdays for respiratory syndrome counts (*12*); a distribution of simulated releases across the calendar needs to be chosen. Some reasonable choices would be uniform across the calendar, uniform within month but different probabilities by month, uniform within weekday but different by day, or some continuous function of the day of the year. The choice could be made to mimic the expected behavior of terrorists. Simulations were stratified by day across the entire calendar, noting that other distributions can be constructed by bootstrap sampling of simulated attacks.

The timing of symptom onset has been described previously (*19*). The log time to onset has an approximately normal distribution, with the mean corresponding to 14.2 days; median, 11.0 days; and standard deviation, 0.713 (log scale). A time of onset was simulated according to this distribution, by using the largest integer included.

Timing of Entry into Surveillance System

In the evaluation in this report, the number of persons in each ZIP code included in the surveillance system was known. This number could be varied as well (e.g., to allow comparison of different potential systems or of improvements to an extant system).

Only those simulated persons eligible for surveillance who become ill and go to the clinic are added to the evaluation data set. Each person can be seen as having some probability for seeking care; this probability might vary according to individual characteristics (e.g., age, sex, and ethnic/cultural background). Obtaining individual characteristics of the patients under surveillance was not possible, so a fixed value of 0.2 was used for the probability of an eligible ill person entering the evaluation data set. This value is derived from the 2001 anthrax release in the United States, when two of the first 10 confirmed case-patients went to their physician before being admitted to the ED (*20*).

Persons might visit the clinic any time after the onset of symptoms. Increasing the probability of a clinic visit as the severity of symptoms increased would be feasible, possibly including probabilities that increase at different rates, depending on individual characteristics. The assumption was made that all patients visited the clinic on the day of symptom onset.

Extension to an Additional Data Stream

Historical data are available from a pediatric ED in Boston that also measures the count by ZIP code of respiratory syn-

drome visits (*21,22*). Extending the simulation to include simulated visits to the pediatric ED is straightforward. All persons aged <18 years are included in the surveillance; pediatric status was randomly assigned to simulated persons according to the proportion of children reported by the state. Going to the pediatric ED is simulated as a decreasing probability with distance from the facility. One less the inverse normal probability was used with mean 0 and variance 49 (km), implying that a child located 14 km from the pediatric ED had approximately a 2.5% chance of being examined. The day of the pediatric ED visit is simulated as 1 (p = 0.15), 2 (p = 0.55), or 3 (p = 0.30) days after symptom onset. Compared with the first 10 terrorism-related inhalational anthrax cases in 2001, this probability timing structure is generous to the ED; only four of 10 patients contacted the ED within 3 days of symptom onset (*20*). However, parents might bring children to the ED sooner than an adult would go.

Example Evaluation and Metrics

For the outpatient surveillance, two statistical methods were compared: the small area regression and testing (SMART) score and a SaTScan approach (*9,23,24*). In the SMART score, generalized linear models are used to establish the expected count per ZIP code per day, adjusting for seasonal, weekly, and temporal trends, and holiday status (*23*). Then, based on the theoretical distribution of case counts and after correcting for multiple testing, the SMART score generates a recurrence interval for each ZIP code each day; the recurrence interval is the length of follow-up required to expect one count as unusual as the observed count.

SaTScan identifies unusual clusters of ZIP codes each day (*9*). Every possible combination of ZIP codes within a circular area around each ZIP code is considered. Each possibility is ranked by likelihood to find the most unusual cluster. Then, a Monte Carlo step determines whether that cluster is unusual in the absolute sense among clusters expected by chance allocation of the cases. This step results in a recurrence interval previously described. The input to the SaTScan is adjusted by using the SMART scores to account for the trends described previously (*24*).

To illustrate the dual-stream outpatient and pediatric ED simulation, only SMART scores were used. Results for the outpatient and pediatric ED data streams are illustrated separately, and the results of a combined surveillance that incorporates signals from each stream are demonstrated. A combined p value is calculated by summing the p value from each stream, squaring the sum, and dividing by two, assuming that the sum is ≤1 (*25*). This combined p value was calculated for each ZIP code and converted to a recurrence interval as previously described.

The conditional receiver operating characteristic (ROC) curve was developed for use in this evaluation and is introduced in this report. The conditional ROC is defined as the ROC curve with respect to attacks generated by the simulation. The sensitivity, which is the probability of detecting an attack, is calculated as the proportion of simulated attacks detected at a given detection threshold. The probability of a false anthrax alarm is calculated as the proportion of days with alarms at that threshold when a simulated attack was not added. One less this quantity is used as the specificity in the ROC curve. Note that the different sources of information regarding the sensitivity and the specificity make the conditional ROC curve somewhat different from the usual (unconditional) ROC curve. In addition, the proportion of detected events, by day of detection, is illustrated. Detection requires that the signal, which is made up of some set of ZIP codes, includes at least one ZIP code with at least one simulated case.

Results

An example of the effects of a single simulated release in the outpatient data is illustrated (Figure 3). In this example, the release was in the urban area, the spores had a 10^{-8} probability of causing illness, and the shape was from the accidental release. Affected ZIP codes indicate no or little response on the day of or day after release, but by day 3 or 4, a noticeable difference was observed between the underlying data and the data with the added cases.

The two methods are similar in their diagnostic value with respect to the simulated attacks. The area for the SMART score is 0.93 and for SaTScan, 0.94, which suggests that the SaTScan is slightly superior in this context (Figure 4).

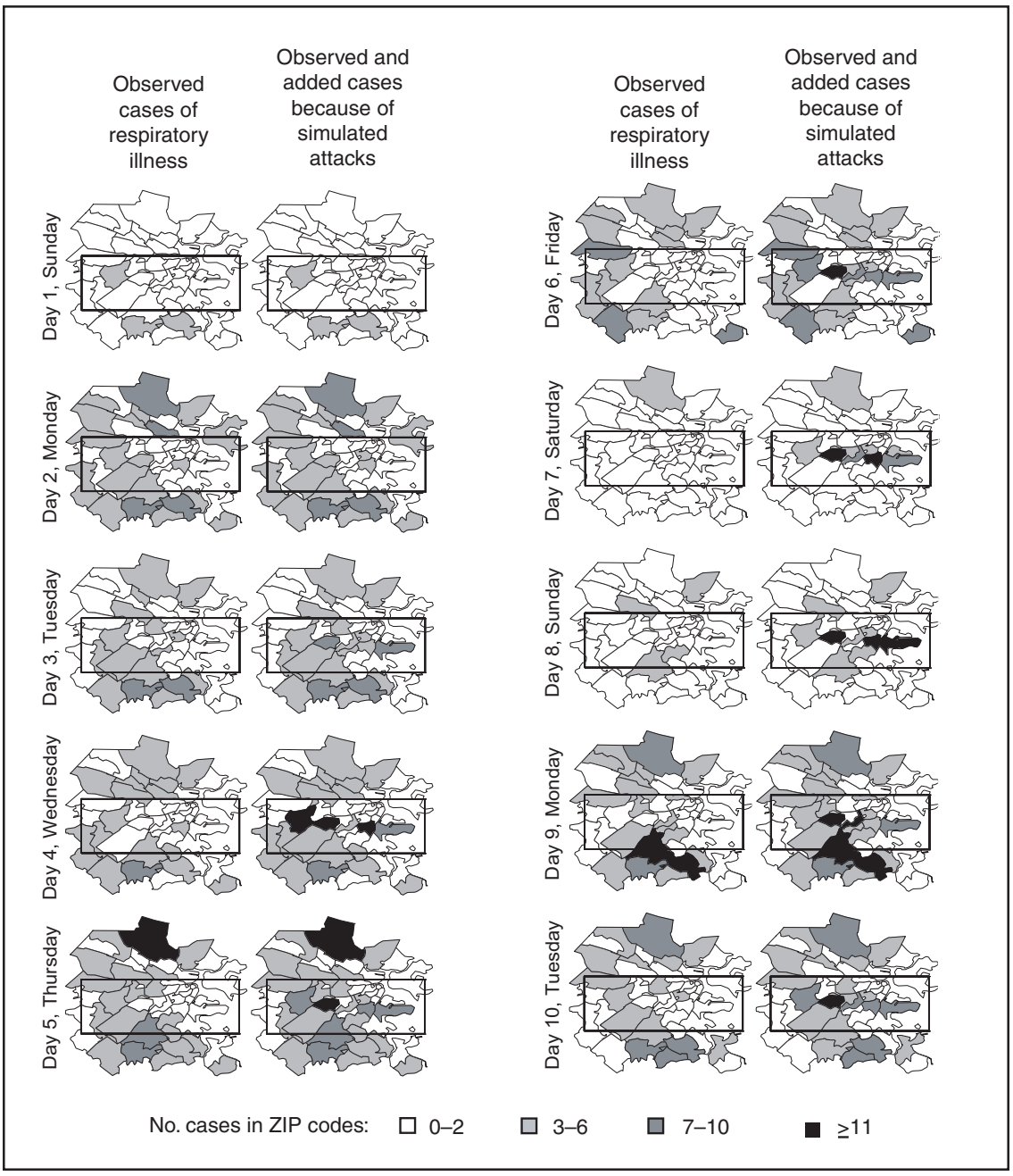
With 1,095 total simulations, SaTScan identified 85% of attacks within 9 days, and the SMART score approach identified 83% (Figure 5). SaTScan identified 57% within 4 days, and the SMART score approach identified only 50% in that time frame.

The pediatric ED surveillance detects more attacks than the outpatient surveillance for any proportion of false positives (Figure 6). The combined surveillance is slightly better than the pediatric ED. The areas under curves are 0.98, 0.93, and 0.99 for the pediatric ED, outpatient, and combined surveillance, respectively.

Discussion

The results of this report demonstrate how the simulation can be used to compare statistical methods and different data sources. With a focused spore dispersal and a high probability of infection, SMART scores are slightly inferior to SaTScan

FIGURE 3. Maps illustrating the outpatient data set before and after simulated cases are added*

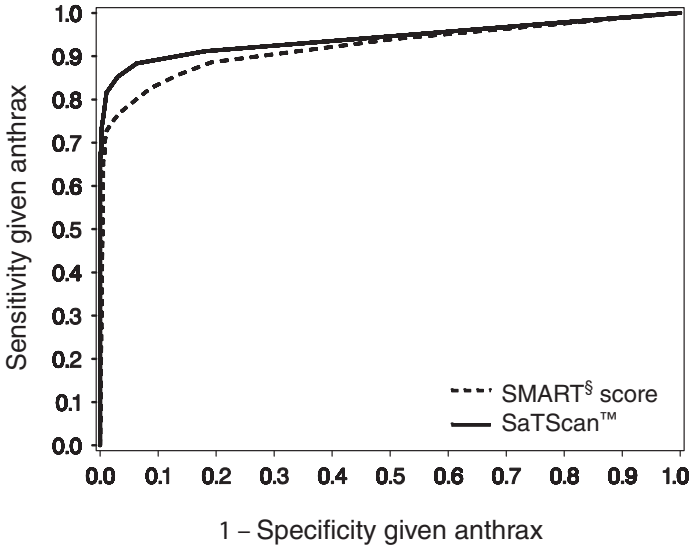


* Map area represents ZIP codes in and around Boston, Massachusetts. Black box superimposed over each map indicates a region in which the simulated release occurred.

with respect to detecting anthrax attacks, although both methods perform effectively. Similarly, outpatient surveillance performs slightly less effectively than pediatric ED or combined surveillance (Figure 6).

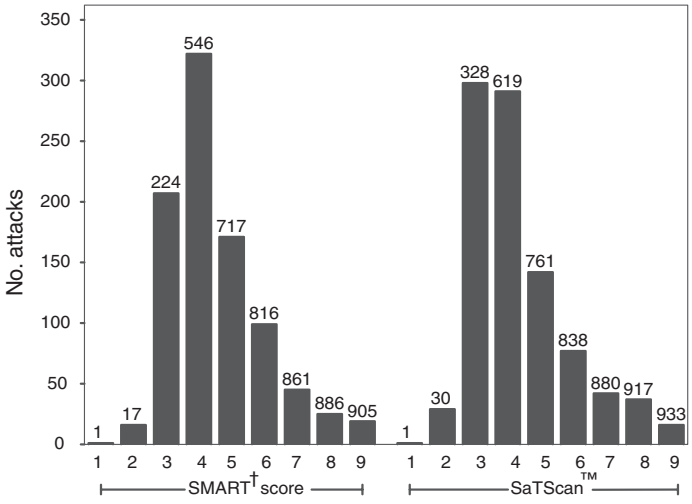
A potential use for the simulation is to assess the value of different levels of coverage in the surveillance system. For example, approximately 10% of residents are included in the outpatient-based surveillance described in this report. Increasing that proportion to determine whether the return would be worth the investment might be possible. Running the simulation with different levels of coverage would allow a comparison to be made. And coverage could be a parameter in the model.

FIGURE 4. Conditional ROC* curves for two statistical methods as detection algorithms employed to detect biologic terrorism†



* Receiver operator characteristic.
† Based on simulated anthrax attack generating outpatient visits.
§ Small area regression and testing.

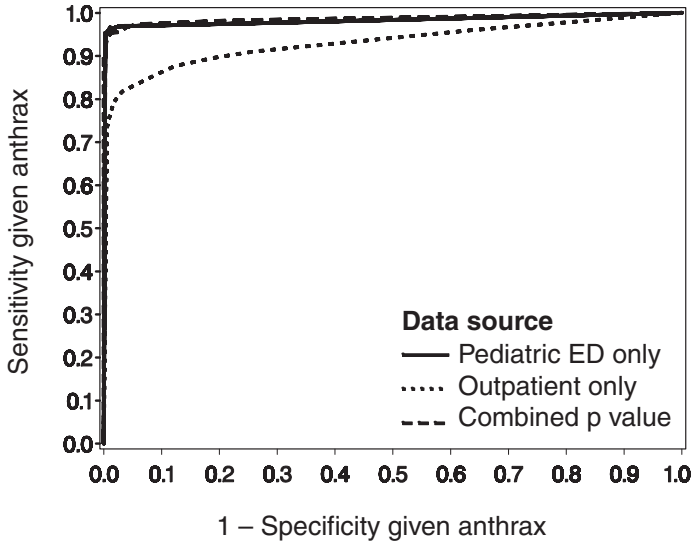
FIGURE 5. Cumulative number of anthrax attacks detected under two different algorithms, by day*



* Small number above each bar indicates the cumulative number detected through each day. Total number of simulations = 1,095.
† Small area regression and testing.

As with the ED example, hospital admissions, prescriptions, or over-the-counter pharmaceutical sales could be added. Including them in the simulation would allow comparison of the statistical methods for the multistream surveillance as well as an assessment of their relative utility.

FIGURE 6. Conditional ROC* curves for SMART† scores employed to detect biologic terrorism for outpatient, pediatric emergency department (ED), and a combination of p values from the two data streams§



* Receiver operator characteristic.
† Small area regression and testing.
§ Based on simulated anthrax attack generating outpatient visits.

Although the variable latency of the disease onset includes times as late as 60 days after exposure, differences in statistical algorithms a substantial period after exposure are not useful; a clinician will generate a laboratory-based diagnosis of anthrax substantially sooner than 60 days after release. The date of clinical detection might be modeled as a probability per patient contact or as an increasing probability with the number of affected persons. The number of ill persons each day was recorded by area through the 10th day; the assumption was that a clinical diagnosis would be initiated by that time. However, the number of ill persons overall were retained from ≥11 days.

Conclusion

This report presented a conceptual discussion and an implementation of simulating an anthrax attack. The simulation will allow various algorithms used to detect such attacks to be compared, which is otherwise difficult. Although the conceptual framework is general, implementation choices can simplify the simulation to suit different purposes (e.g., statistical comparisons or emergency planning).

Simulations can be a valuable part of evaluating surveillance systems. They can be used to compute comparisons of statistical and data mining algorithms and can help guide

public health officials in evaluating and planning surveillance systems.

Acknowledgments

The authors thank M. Kulldorff and K. Yih, both of Harvard Medical School and Harvard Pilgrim Health Care, for their comments and insights in the work reported.

References

1. Bravata DM, McDonald KM, Smith WM, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Annals Intern Med 2004;140:910–22.
2. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecurity and Bioterrorism 2003;1:77–81.
3. 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:i43–9.
4. 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:i32–42.
5. Wong WK, Moore A, Cooper G, Wagner M. WSARE: What's strange about recent events? J Urban Health 2003;80:i66–75.
6. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80:i89–96.
7. Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. J Urban Health 2003;80:i57–65.
8. 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.
9. Kulldorff M. Prospective time periodic geographic disease surveillance using a scan statistic. J Royal Stat Society, Series A 2001;164:61–72.
10. Buckeridge D, Burkom H, Moore A, Pavlin J, Cutchis P, Hogan W. Evaluation of syndromic surveillance systems: design of an epidemic simulation model. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:137–43.
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. 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.
13. International Classification of Diseases, 9th Revision, Clinical Modification. Ann Arbor, MI: US Department of Health and Human Services, CDC, National Center for Health Statistics; 1980.
14. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proc Nat Acad Sci USA 2003;100:4346–51.
15. Pasquill. Atmospheric diffusion. 2nd ed. New York, NY: Wiley; 1974.
16. Meselson M, Guillemin J, Hugh-Jones M, et al. The Sverdlovsk anthrax outbreak of 1979. Science 1994;266:1202–8.
17. Spijkerboer HP, Beniers JE, Jaspers D, et al. Ability of the Gaussian plume model to predict and describe spore dispersal over a potato crop. Ecological Modeling 2002;155:1–18.
18. Brookmeyer R, Johnson E, Bollinger R. Modeling the optimum duration of antibiotic prophylaxis in an anthrax outbreak. Proceedings of the National Academies of Science U S A 2003;100:10129–32.
19. Brookmeyer R, Blades N. Prevention of inhalational anthrax in the U.S. outbreak. Science 2002;295:1861.
20. 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.
21. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci U S A 2003;100:1961–5.
22. 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.
23. 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.
24. Kleinman K, Abrams A, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. Epidemiol Infect 2005;119:409–19
25. Edgington E. An additive method for combining probability values from independent experiments. J Psychol 1972;80:351–63.

An Evaluation Model for Syndromic Surveillance: Assessing the Performance of a Temporal Algorithm

David L. Buckeridge,[1,2] P. Switzer,[3] D. Owens,[1,2,4] D. Siegrist,[5] J. Pavlin,[6] M. Musen[2]
[1]*Palo Alto Veterans Health Care, Palo Alto, California;* [2]*Stanford Medical Informatics, Stanford University, Stanford, California;* [3]*Department of Statistics, Stanford University, Stanford, California;* [4]*Departments of Medicine and Health Research and Policy, Stanford University, Stanford, California;* [5]*Potomac Institute, Arlington, Virginia;* [6]*Walter Reed Army Institute of Research, Silver Spring, Maryland*

Corresponding author: David L. Buckeridge, Stanford University, Medical School Office Building, Rm. X-215, 251 Campus Drive, Stanford, CA 94305-5479. Telephone: 650-723-6979; Fax: 650-725-7944; E-mail: david.buckeridge@stanford.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Syndromic surveillance offers the potential to rapidly detect outbreaks resulting from terrorism. Despite considerable experience with implementing syndromic surveillance, limited evidence exists to describe the performance of syndromic surveillance systems in detecting outbreaks.*

Objectives: *To describe a model for simulating cases that might result from exposure to inhalational anthrax and then use the model to evaluate the ability of syndromic surveillance to detect an outbreak of inhalational anthrax after an aerosol release.*

Methods: *Disease progression and health-care use were simulated for persons infected with anthrax. Simulated cases were then superimposed on authentic surveillance data to create test data sets. A temporal outbreak detection algorithm was applied to each test data set, and sensitivity and timeliness of outbreak detection were calculated by using syndromic surveillance.*

Results: *The earliest detection using a temporal algorithm was 2 days after a release. Earlier detection tended to occur when more persons were infected, and performance worsened as the proportion of persons seeking care in the prodromal disease state declined. A shorter median incubation state led to earlier detection, as soon as 1 day after release when the incubation state was ≤5 days.*

Conclusion: *Syndromic surveillance of a respiratory syndrome using a temporal detection algorithm tended to detect an anthrax attack within 3–4 days after exposure if >10,000 persons were infected. The performance of surveillance (i.e., timeliness and sensitivity) worsened as the number of persons infected decreased.*

Introduction

Syndromic surveillance offers the potential to rapidly detect a change in the health status of a population. The main motivation for conducting syndromic surveillance has been to detect disease outbreaks resulting from an act of terrorism (*1*). Despite implementation of syndromic surveillance (*2*) and promulgation of evaluation guidelines (*3*), limited evidence exists that describes the performance of syndromic surveillance systems in detecting outbreaks from terrorism (*4*).

Published evaluations of syndromic surveillance have focused on the ability of a system to detect influenza outbreaks compared with traditional means of influenza surveillance (*5,6*). These evaluations provide useful information, but this proxy disease approach to evaluation has limitations. Most notably, the performance of syndromic surveillance in detecting influenza outbreaks might not be generalizable to detecting other types of outbreaks. Outbreaks resulting from intentional releases will likely have different characteristics than influenza outbreaks, including a different disease agent and means of introducing the agent into the population. These characteristics can change the presentation of an outbreak and the ability of a surveillance system to detect an outbreak. For example, an outbreak attributed to inhalational anthrax might have a faster increase in cases than an influenza outbreak, and the anthrax cases can also occur in tighter spatial clusters.

Another option for evaluating a syndromic surveillance system is to superimpose outbreak cases (i.e., an epidemic curve) directly onto an existing data source and then use the combination of the superimposed outbreak data and real baseline data to evaluate a surveillance system. One method for producing the outbreak cases is to specify an epidemic curve, which is the distribution of time until patients seek health care after exposure. Specifying the epidemic curve is appealing because of its simplicity, but the epidemic curve is essentially a black box. Predicting how changes in the many factors that might influence an outbreak would affect the epidemic

curve is not possible. Such factors include the progression of disease within infected persons and the propensity for symptomatic persons to seek medical care. The ability to examine the influence of these factors on the evolution and detection of an outbreak is important because uncertainty exists about these factors in medical literature (*7*), and debate continues about the influence of these factors on surveillance (*8*).

This report describes the implementation and application of a model for simulating cases that might result from an exposure to inhalational anthrax. Design of the model has been described previously (*9*); in this report, detailed components of the model for simulating disease progression and health-care–service use by persons with symptoms are described. The model is then used to evaluate the ability of syndromic surveillance to detect an outbreak of inhalational anthrax after an aerosol release.

Methods

Simulation Model

A model of the processes underlying the biological and medical sequelae of an anthrax attack was developed (*9*). This model included components for the dispersion of spores after a release, the infection of persons, disease progression within infected persons, and health-care use by symptomatic persons.

Evaluation of syndromic surveillance was conducted by using a temporal algorithm; thus, the dispersion and infection components of the model, which are necessary only for evaluating spatial algorithms, were not used. Simulation, therefore, began with disease progression and the number of infected persons. The components of the simulation model used were the disease and the health-care–seeking components. The disease component simulates a path through three disease states (i.e., incubation, prodromal, and fulminant) for each infected person. The health-care–seeking component then simulates the occurrence and timing of health-care visits in each disease state and the syndrome assigned to a person who seeks care.

Disease Component

Using previous studies modeling anthrax (*10*), disease progression was modeled through the three disease states: incubation, prodromal, and fulminant. Disease progression was modeled as a semi-Markov process (*11*), with lognormal holding time functions. The incubation times for inhalational anthrax fit a lognormal distribution (*12–14*), and the parameter's values for the lognormal holding time functions used in the base case are indicated (Table 1) (*10*). The log time in a state is normally distributed with mean μ and variance s^2, $\log(t) \sim N(\mu, s^2)$. Similar to previous studies (*12*), the parameter $d = e^s$ is referred to as the dispersion factor, and the median and dispersion factor are used to describe lognormal distributions. In simulating disease, an infected person begins in the incubation state, progresses to the prodromal state, and then to the fulminant state. A distinct path was simulated through the disease model for each infected person.

Health-Care–Seeking Component

Because detailed data on the probability of seeking care and the delay to seeking care given an illness are not readily available in medical literature, a single state semi-Markov process was used to model the probability of and time to seeking care. Consumer panel research was used for the probability of seeking care (A. Kress, Surveillance Data Inc., personal communication, 2004), and a right triangular distribution (*15*) fit to the time spent in the disease state was used to model the time to seeking care. For persons that sought care, the instantaneous probability of seeking care increased linearly over the time in a state. Persons were limited to a single episode of care in each state, and care-seeking was modeled independently for the prodromal and fulminant states. When a person sought health care, the syndrome assigned was simulated using the probabilities (Table 1) that reflect the distribution of clinical presentations for inhalational anthrax (*16,17*). The time to seeking care was modeled as a continuous variable, and surveillance analysis was performed daily.

Simulation Study

The simulation study examined the ability of syndromic surveillance conducted with a temporal algorithm to detect a

TABLE 1. Parameters used in simulation model

Parameter name	Base value	Source
Disease		
Incubation duration, median	10.95 days	14
Incubation duration, dispersion	2.04 days	14
Prodromal duration, median	2.50 days	10
Prodromal duration, dispersion	1.44 days	10
Fulminant duration, median	1.50 days	10
Fulminant duration, dispersion	1.44 days	10
Health-care seeking		
Probability of seeking care, prodromal state	0.4	Estimate
Probability of seeking care, fulminant state	0.8	Estimate
Probability of respiratory syndrome, prodromal state	0.7	16
Probability of gastrointestinal syndrome, prodromal state	0.2	16
Probability of fever syndrome, prodromal state	0.1	16
Probability of shock syndrome, fulminant state	0.7	16
Probability of neurological syndrome, fulminant state	0.3	16

simulated anthrax outbreak. A temporal algorithm follows the aggregate incidence of events throughout the surveillance region and does not examine the spatial distribution of events. The impact of the size of the outbreak on detection performance was examined and the sensitivity of findings to assumptions about disease progression and health-care–seeking behavior was also determined.

The study design simulated outbreaks and then combined data from each outbreak with real health-care use data to form multiple test data sets. Each test data set was the union of authentic data with a set of records from a simulated outbreak. A detection algorithm was then applied to each test data set, and outbreak-detection performance was measured. The authentic surveillance data and simulation region, the approach to generating simulated outbreaks, the outbreak detection algorithm, and the technical implementation are described below.

Surveillance Data and Region

Records of ambulatory-care visits were acquired in Norfolk, Virginia, from the TRICARE health maintenance organization (HMO). This HMO finances health care for active duty military personnel and their dependants. Syndromic surveillance systems, including the ESSENCE (*18*) and BioSense systems, routinely use these data. Data were available for 2001–2003. The simulation region includes 17 clinical facilities that services approximately 158 ZIP codes from two states. During 2001–2003, a total of 427,634 persons made approximately 7 million visits for syndromes routinely followed in syndromic surveillance. Records were classified into syndromes by using the *International Classification of Diseases, Ninth Revision, Clinical Modification* codes to syndrome mapping defined by the ESSENCE system.

Generation of Simulated Outbreaks

In generating simulated outbreaks, four parameters in the simulation model were varied: number of infected persons, median time in the incubation state, median time in the prodromal state, and the proportion seeking care in the prodromal state. For each parameter, eight settings were used and ten runs were performed at each setting, resulting in 320 simulated outbreaks. Eight dates were then randomly selected for the beginning of an outbreak, and each simulated outbreak was superimposed onto the authentic data beginning on each of the outbreak dates, resulting in 2,560 test sets (Table 2).

TABLE 2. Parameters varied in the sensitivity analysis*

Parameter	Settings evaluated
No. of infected persons (in thousands)	5, 10, 20, 30, 40, **50**, 60, 70
Incubation duration, median (in days)	3, 5, 7, 9, 10, **11**, 12, 13
Prodromal duration, median (in days)	1.5, 2.0, **2.5**, 3.0, 3.5, 4.0, 4.5, 5.0
Probability of seeking care, prodromal state	0.1, 0.2, 0.3, **0.4**, 0.5, 0.6, 0.7, 0.8
Date of anthrax release	March 9 (Sun), April 19 (Sat), June 14 (Sat), July 8 (Tue), July 25 (Fri), August 3 (Sun), September 7 (Sun), November 7 (Fri)

* Settings in bold are for the base case.

Outbreak Detection Method

An autoregressive seasonal integrated moving average (SARIMA) model (*19*) was used to calculate one-step-ahead daily forecasts of respiratory syndrome counts, and a cumulative sum (*20*) was applied to detect positive deviations in the forecast residuals. Researchers have employed this approach to outbreak detection in a surveillance setting (*21*). The respiratory syndrome was used because this is the syndrome an anthrax attack will likely affect and the temporal surveillance algorithm only considers a single syndromic category).

To fit the SARIMA model, the first 2 years of data for respiratory syndromes were used, and a procedure published previously was followed (*22*). This entailed subtracting the overall mean, day-of-week means, month means, and holiday means from the original count data to give a series centered on zero. Trimmed means (alpha = 0.1) were used for both day-of-week and month to minimize the influence of outliers. The temporal autocorrelation in this series was assessed, and a SARIMA model was fit to the series by using a standard approach to model specification (*19*). The fit of the SARIMA model was evaluated to the training and test data by using the mean absolute percentage error (MAPE), which is the average of the absolute difference between the daily forecast and the visit count, divided by the visit count. The standardized residual was calculated for each day as the observed count minus the one-step-ahead forecast from the SARIMA model, divided by the standard error of the forecast. To fit the cumulative sum, the parameters of the test were adjusted to achieve an alarm rate on the training data that was approximately one per month, which is intended to reflect a reasonable workload for a public health agency.

To evaluate outbreak detection, sensitivity and timeliness were calculated at a set alarm rate. Sensitivity was defined as the proportion of simulated outbreaks detected before or at the peak of the epidemic curve; timeliness was defined as the number of days until an alarm, given that an outbreak was detected.

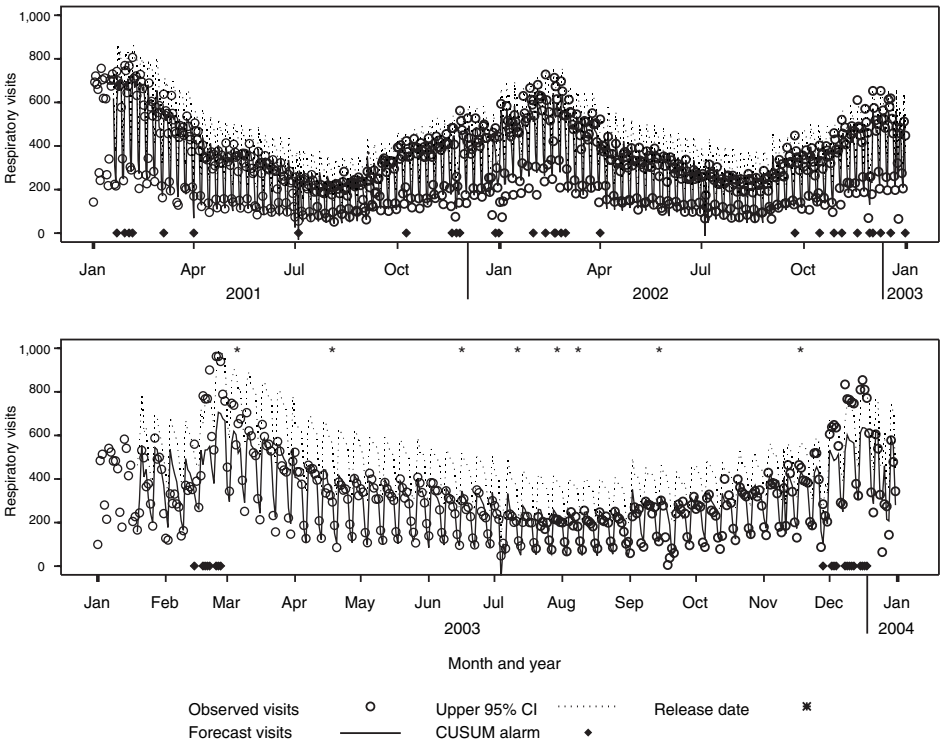
Results

During 2001–2003, the syndrome mapping classified 351,749 (6.6%) of the 5,319,347 visits as respiratory syndromes. The average was 321 respiratory visits per day (range: six to 963 visits), with lower counts on weekends (mean: 168 visits) than on weekdays (mean: 382 visits). Examination of the time series of respiratory counts revealed systematic variation by day-of-week, month, and holidays (Figure 1).

After subtracting the overall mean and means for day-of-week, month and holiday, the zero-centered series exhibited temporal autocorrelation at short lags on the order of days and cyclical lags of order seven. A SARIMA model (2,0,1) x $(2,0,1)_7$ had the best fit to the zero-centered series. One-step-ahead forecasts from this model resulted in a MAPE of 15.8% on the training data (2001–2002), indicating that the forecast values were, on average, within 15.8% of the true value. This fit is similar to or better than the fit reported in previous research using the same algorithm and similar data (*23*).

The cumulative sum was calibrated to a specificity of 97%, by setting the shift parameter to 1 and the threshold to 1.5. This is an alarm rate of approximately 1 in 4 weeks (3.5%), which other researchers have used to evaluate outbreak detection through syndromic surveillance (*24*).

FIGURE 1. Time-series of training (2001–2002) and test (2003) data demonstrating observed counts, forecast counts, and forecast confidence intervals, the alarms from the cumulative sum (CUSUM) applied to the forecast residuals in the absence of injected outbreaks, and the randomly selected release dates in the test data



Detection by Number Infected

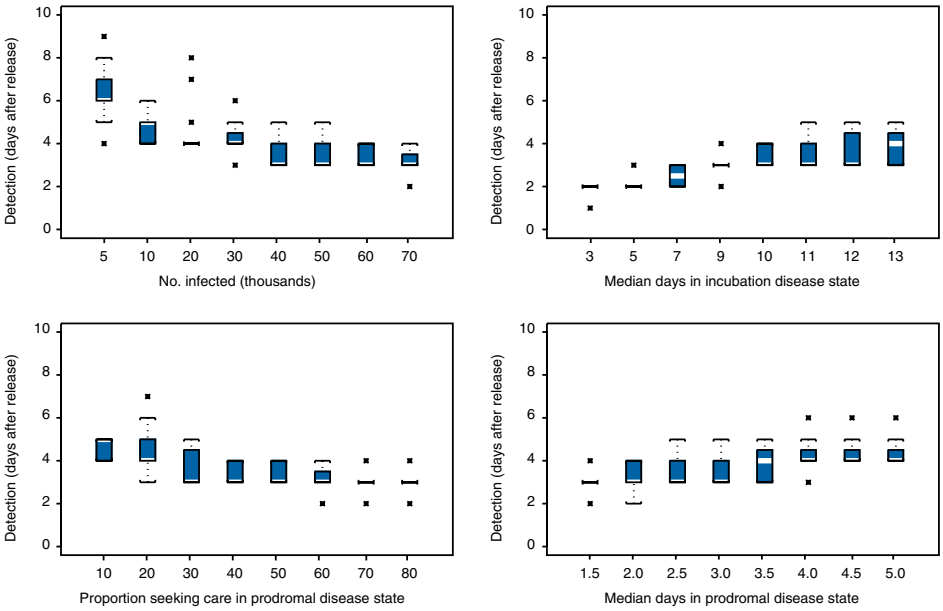
Surveillance with a temporal algorithm detected an outbreak 2 days after release at the earliest, with a trend towards earlier detection as the number of persons infected increased (Figure 2). When ≥40,000 persons were infected, the median time to detection was 3 days, and the maximum time to detection was 5 days. At lower release amounts, the median time to detection increased, reaching a maximum of 6 days when 5,000 were infected. Sensitivity was 100% when ≥30,000 persons were infected, decreased to 90% (95% confidence interval [CI] = 82%–95%) when 20,000 were infected, to 75% (CI = 65%–84%) when 10,000 were infected, and to 56% (CI = 45%–67%) when 5,000 were infected.

Sensitivity to Disease and Behavior Assumptions

In the base case, 40% of persons sought care in the prodromal disease state. Increasing the proportion of persons who sought care in the prodromal state resulted in temporal surveillance detecting the outbreak faster, but the change was minor. In contrast, the performance of temporal surveillance declined as the proportion seeking care decreased. With 10% of persons seeking care, syndromic surveillance detected the outbreak on median in 5 days. Varying the proportion seeking care had a similar effect on detection to varying the number infected (Figure 2).

In the base case, the median duration of the incubation state was 11 days. Some researchers have suggested that the performance of temporal surveillance is likely to decline as the incubation period becomes shorter (*8*); however, the findings in this report suggest the opposite. As the median duration of the incubation state decreased, temporal surveillance tended to detect the outbreak faster, with detection after 2 days for a median incubation of ≤5 days. Increasing the median duration of the incubation period to >11 days tended to increase the time to detection, but not markedly. In the base case, the median duration of the prodromal disease state was 2.5 days. A shorter median duration did not change considerably the time to detection for

FIGURE 2. Time to detecting an outbreak through syndromic surveillance using a temporal algorithm as a function of four parameters



syndromic surveillance. As the median duration of the prodromal state increased above 2.5 days, temporal surveillance tended to take longer to detect the outbreak, but this trend was not pronounced.

Implications for Performance of Syndromic Surveillance

To determine the implications of findings for syndromic surveillance, the timing of detection was considered with a release of anthrax that infected 20,000 persons. Within the simulation area, the population that was covered by the TRICARE HMO was approximately 400,000, thus 20,000 infected persons represented 5% of the covered population. If 40% of those infected sought care in the prodromal state, that would amount to 8,000 visits, representing 2% of the covered population. These visits would be spread over the duration of the outbreak, and of greater interest than the total number is the number seeking care each day over the first few days of the outbreak. For an outbreak that infects 20,000 persons, the distribution of the additional visits is illustrated (Figure 3).

The median time to detection with 20,000 infected persons was 4 days after release (Figure 2), which corresponds to an additional 56 visits, on average, on the day of detection. The daily average number of visits in this test data was 321, thus the additional 56 visits represented an increase of 17% over baseline for an average day. If the fourth day of an outbreak occurred on a weekend day, when the average visit count was 168, then the increase in visits would be 33% over baseline. Variation in the background occurs by season, and the maximum number of daily visits in the 1 year of test data (2003) was 963. An additional 56 visits is an increase in visits of 6% over baseline on a day with a baseline count of 963.

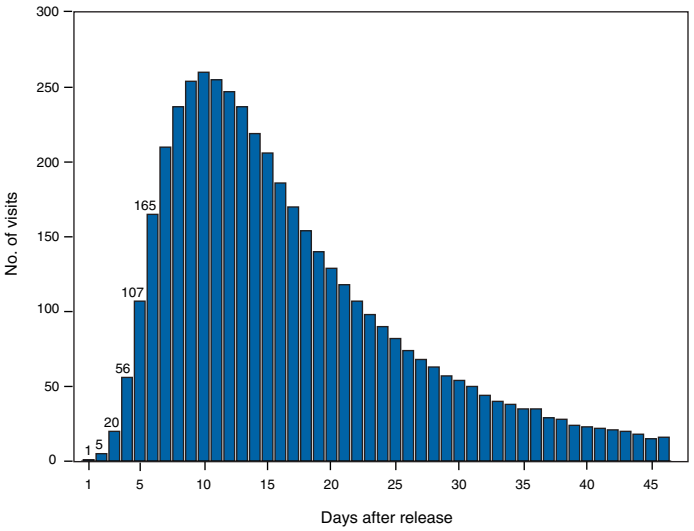
Discussion

This report describes a simulation model of the processes underlying the biological and medical sequelae of an anthrax attack; the model was used to evaluate the ability of syndromic surveillance to detect an outbreak of inhalational anthrax. In the base case (i.e., 50,000 infected, 11 days incubation period, 2.5 days prodrome, and 40% seeking care in prodromal stage), the earliest detection using a temporal algorithm was 3 days after a release. Earlier detection tended to occur when more persons were infected, and surveillance tended to detect an aerosol anthrax attack in a median of 3 to 4 days when >10,000 persons were infected. Median time to detection increased to 6 days when 5,000 persons were infected.

The sensitivity analysis demonstrated that the proportion of persons seeking care in the prodromal disease state affected the performance of syndromic surveillance. As the proportion seeking care declined, temporal surveillance took longer to detect an outbreak. This was similar to the effect observed when the number of persons infected in an attack was decreased. The median duration of the prodromal disease state had little influence over the performance of surveillance, but the median duration of the incubation state affected the performance of surveillance. Syndromic surveillance detected an outbreak sooner as the median duration of the incubation state decreased, with detection occurring 2 days after release when the median duration of the incubation state was ≤5 days.

In examining the implications of these findings for syndromic surveillance, for a release that infected 20,000 persons, detection through surveillance occurred 4 days after release when 56 persons in the outbreak used health-care services. Health-care providers could have possibly identified these 56 additional cases with nonspecific symptoms as an extraordinary increase in the absence of syndromic surveillance. Sev-

FIGURE 3. Average number of daily outpatient physician visits after an aerosol anthrax release that infects 20,000 persons



enteen clinical sites were in the simulation area, with a number of physicians at each site. If the cases distributed themselves across multiple sites and physicians, then an additional 56 cases in 1 day might not appear unusual. If persons reported to a small number of locations and physicians, then the same number of additional cases might raise concern. The location of seeking care was not modeled for this study, but will be considered for future research. Another factor to consider is whether some infected persons are likely to progress rapidly to a fulminant state and be diagnosed with inhalational anthrax before the fourth day after the release. This is another factor that was not considered in this study, but will be in future research.

The evaluation method described in this report has advantages over other approaches. This method models disease and behavior processes at a level of detail sufficient for examining the influence of assumptions about these processes on outbreak detection performance, thus allowing thorough examination of how characteristics of disease and health-care–seeking behavior influence the performance of syndromic surveillance. Even with the few assumptions examined, the findings offer insight into the role of syndromic surveillance in a disease-control strategy. Another advantage of this approach is that authentic surveillance data were used as the basis for the evaluation, and simulated outbreaks were superimposed onto these authentic data. Hence, detection algorithms must be robust to vagaries in real data, and these results are to some extent influenced by the characteristics of the underlying population and the authentic data.

In this report, syndromic surveillance conducted with a univariate temporal algorithm and operating on a single syndrome was examined. Following multiple syndromes with spatial algorithms might improve performance. The sensitivity of these results in relation to four parameters varied one at a time. A multi-way sensitivity analysis of a greater proportion of the parameters might prove more informative; this study is now under way. The evaluation described here looked at outbreak detection assuming that data are available in real time. In practice, this is rarely the case and the methods used to correct for reporting delay might influence detection. If the distribution of reporting was known, the simulation model could be modified to evaluate methods that account for reporting delay. To assess the true impact of syndromic surveillance on time until intervention, it will be necessary to extend the model to encompass outbreak detection through other routes such as clinical case-finding and to assess the timing of intervention decisions with information from syndromic surveillance.

Conclusion

Evaluation of outbreak detection through syndromic surveillance is difficult for many reasons, including the limited amount of data for outbreaks of interest. Multiple evaluation approaches exist, and the simulation method described in this report provides useful insight. Syndromic surveillance of a respiratory syndrome using a temporal detection algorithm tends to detect an anthrax attack within 3 to 4 days after exposure if >10,000 persons are infected. The performance of surveillance (i.e., timeliness and sensitivity) worsens as the number of persons infected decreases, and as the proportion seeking care in the prodromal stage declined.

References

1. Wagner MM, Tsui FC, Espino JU, et al. The emerging science of very early detection of disease outbreaks. J Public Health Manag Pract 2001;7:51–9.
2. 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.
3. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR 2004;53(No. RR-5).
4. 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.
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. 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.
7. Inglesby TV, O'Toole T, Henderson DA, et al. Anthrax as a biological weapon, 2002: updated recommendations for management. JAMA 2002;287:2236–52.
8. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis 2003;9:1197–204.
9. Buckeridge DL, Burkom H, Moore AW, Pavlin JA, Cutchis PN, Hogan WR. Evaluation of syndromic surveillance systems: Development of an epidemic simulation model. In: Sydromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):137–43.
10. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proceedings of the National Academy of Sciences of the United States of America 2003;100:4346–51.
11. Shedler G. Regenerative Stochastic Simulation. New York, NY: Academic Press; 1992.
12. Sartwell PE. The distribution of incubation periods of infectious diseases. American Journal of Hygiene 1950;51:310–8.
13. Meselson M, Guillemin J, Hugh-Jones M, et al. The Sverdlovsk anthrax outbreak of 1979. Science 1994;266:1202–8.
14. 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.
15. Law AM, Kelton WD. Simulation Modeling and Analysis. 3rd ed. Boston, MA: McGraw-Hill; 2000.
16. 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.
17. Rickmeier GL, McClellan GE, Anno GA. Biological warfare human response modeling. Military Operations Research 2001;6:35–47.
18. 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–i42.
19. Box GEP, Jenkins GM. Time Series Analysis, Forecasting and Control. Revised ed. Englewood Cliffs, CA: Prentice Hall; 1976.
20. Page E. Continuous inspection schemes. Biometrika 1954;41:100–15.
21. Williamson GD, Weatherby HG. A monitoring system for detecting aberrations in public health surveillance reports. Stat Med 1999;18:3283–98.
22. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. PNAS 2003;100:1961–5.
23. Noy NF, Sintek M, Decker S, Crubezy M, Ferguson RW, Musen MA. Creating semantic web contents with Protege-2000. IEEE Intelligent Systems 2001;16:60–71.
24. CDC. BioALIRT biosurveillance testbed evaluation. In: Sydromic surveillance: reports from a national conference, 2003. MMWR 2004; 53(Suppl):152–8.

Evaluation of Syndromic Surveillance Based on National Health Service Direct Derived Data — England and Wales

Alexander Doroshenko,[1] D. Cooper,[1] G. Smith,[1] E. Gerard,[2] F. Chinemana,[3] N. Verlander,[4] A. Nicoll[4]
[1]*Health Protection Agency West Midlands, Birmingham, England;* [2]*NHS Direct National Team, England;* [3]*NHS Direct Hampshire and Isle of Wight, Southampton, England;* [4]*Health Protection Agency, Center for Infections, London, England*

Corresponding author: Duncan 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.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Syndromic surveillance systems might serve as an early warning to detect outbreaks of infectious diseases and chemical poisoning, including those caused by deliberate release. In England and Wales, data from National Health Service (NHS) Direct, a national telephone health advice service, were used for surveillance of 10 syndromes commonly occurring in the community.*

Objectives: *The objective of this study was to evaluate NHS Direct syndromic surveillance using the "Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks", published by CDC.*

Methods: *Quantitative and qualitative assessments were performed. Examination of daily data flow was used to determine the timeliness and data quality. Validity was determined by comparing NHS Direct surveillance with a well-established clinical-based surveillance system using a time series analysis. Semistructured interviews of main stakeholders were conducted to determine usefulness, flexibility, acceptability, portability, stability, and system costs.*

Results: *NHS Direct syndromic surveillance has representative national coverage, provides near real-time recording and data analysis, and can potentially detect high-risk, large-scale events. Direct costs are low and variable costs are unpredictable. Flexibility depends on urgency of the need for change, and portability relies on the existence of infrastructure similar to NHS Direct. Statistically significant correlation exists between NHS Direct surveillance and a surveillance system based on the Royal College of General Practitioners data for influenza-like illness.*

Conclusion: *The CDC framework is a useful tool to standardize the evaluation of syndromic surveillance. NHS Direct syndromic surveillance is timely, representative, useful, and acceptable with low marginal costs and borderline flexibility and portability. Cross-correlation time series modeling might represent an appropriate method in the evaluation of syndromic surveillance validity.*

Introduction

Emphasis has been placed on the improvement of existing surveillance systems and developing innovative new surveillance systems around the world. Commitments to improve surveillance for health protection have been made in the United Kingdom (UK) (*1*). Because certain emerging infections and chemical poisonings, including those caused by deliberate release, might first appear as ill-defined syndromes, rapid outbreak detection is a challenge. Suspicious patterns of patient presentations might be apparent at the community level well before laboratory data raise an alarm. Syndromic surveillance might serve as an early warning to detect such occurrences (*2,3*).

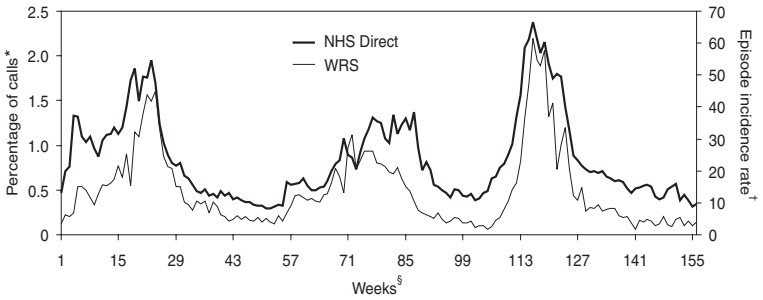
In 2004, an evaluation of the usefulness of 35 detection and diagnostic decision support systems for biologic terrorism response was performed. Most evaluations were critically deficient (*4,5*). The need for more detailed evaluation of syndromic surveillance projects culminated in the publication of the "Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks" by CDC in May 2004 (*6*). This guidance aims to standardize frequently fragmented evaluation efforts. The CDC framework is designed for the evaluation of relatively mature, fully operational syndromic surveillance systems (*7*). This report expands on the existing work on the NHS Direct syndromic surveillance system in England and Wales on the basis of call data from the national telephone health advice helpline operated by NHS (*8,9*). This report presents a preliminary evaluation of NHS Direct syndromic surveillance according to the CDC framework.

Methods

Both quantitative and qualitative assessments using CDC guidance were performed. Information was gathered with respect to the construct and utility of the NHS Direct syndromic surveillance. Comprehensive semistructured qualitative interviews with eight main stakeholders were conducted to determine usefulness, flexibility, acceptability, portability, stability, and costs of the system. Respondents were selected on the basis of their knowledge and experience of NHS Direct syndromic surveillance and included consultants in communicable disease control (CCDC), regional epidemiologists (RE), NHS Direct managerial and scientific staff, and national experts from the Health Protection Agency (HPA). Interviews were conducted by the same investigator using a devised standard questionnaire, and all answers were recorded and transcribed in a standard way. Examination of daily electronic NHS Direct surveillance data and weekly NHS Direct syndromic surveillance bulletins were used to determine timeliness and data quality. Qualitative estimates of numbers of outbreaks detected by NHS Direct syndromic surveillance were also obtained through interviews. This estimate was based on professional judgement and supplemented by the quantitative analysis.

Quantitative analysis included an evaluation of the system's validity by comparing NHS Direct syndromic surveillance for influenza-like illnesses (ILIs) with a well-established national clinical surveillance system (the Royal College of General Practitioners Weekly Returns Service [WRS]). WRS is a broadly representative network of 78 general practices that voluntarily participate in a scheme to collect information on consultations and episodes of illness diagnosed in general practice. Weekly incidence rates per 100,000 population for common illnesses are calculated. On the basis of historical trends, robust thresholds for ILI activity have been developed by WRS. These thresholds determine four levels of ILI activity in England and Wales: baseline activity, normal seasonal activity, higher than average seasonal activity, and epidemic activity. Weekly surveillance data on ILI syndromes were compared between NHS Direct and WRS systems during August 2001–August 2004. NHS Direct surveillance began collecting data on ILIs in August 2001, so all NHS Direct data available at the time of study were analyzed. NHS Direct call data were aggregated from daily to weekly to conform to WRS data format and two time series were constructed (Figure 1). Data from both sources were compared by calculating Spearman rank correlation coefficient and fitting time-series models and estimating a cross-correlogram between two time series at different lags (weeks of observations). For the time series models, both data sets were transformed and detrended by differencing to ensure that transformed series were stationary. Then appropriate autoregressive moving average models were fitted to the differenced, transformed time series so that each set of residuals were white noise. The models were determined by examining the autocorrelation and partial autocorrelation functions to determine autoregressive and moving average parts of the models. Residuals were determined from models and checked for normality and against the fitted values. Residuals were also checked for white noise by the Portmanteau test. Cross-correlation was estimated for residuals at different lags with the limit for statistically significant correlation being $2/\sqrt{(N-1)}$ in either direction, where N represented the number of data points.

FIGURE 1. National Health Service (NHS) Direct and Royal College of General Practitioners Weekly Return Service (WRS) time series for influenza-like illnesses, by week — England and Wales, August 2001–August 2004



* Percentage of calls relates to NHS Direct time series.
† Episode incidence rate per 100,000 population relates to WRS time series.
§ Week 1 corresponds to week 35 in 2001; week 156 corresponds to week 34 in 2004.

Results

System Description

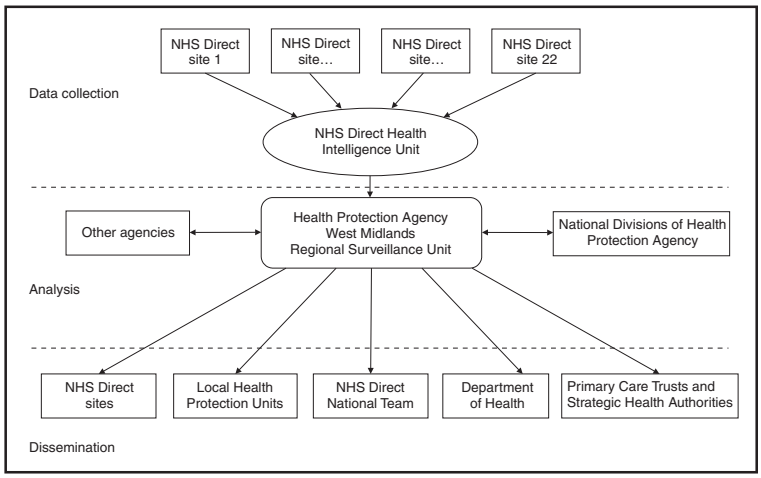
The initial purpose of NHS Direct syndromic surveillance was to augment other surveillance systems in detecting outbreaks of influenza. The aim was to facilitate the early implementation of preventative measures. In December 2001, the surveillance of 10 syndromes began, and the purpose of the system was expanded to provide an early warning for potential deliberate release of harmful biologic and chemical agents. The system is more likely to detect large-scale events or outbreaks and a rise in symptoms with no clear cause evident. NHS Direct syndromic surveillance is an example of how a system, initially designed for the clinical assessment (by telephone) of common conditions presenting in communities, has been used for surveillance purposes. Because the surveil-

lance process is fully operational, the validation of the system is considered a priority.

At the time of the inception of the NHS Direct syndromic surveillance system, the list of stakeholders was limited to NHS Direct central management, NHS Direct sites, the NHS Direct Health Intelligence Unit (HIU), and the Regional Surveillance Unit (RSU) of the Health Protection Agency West Midlands. The greater public health community has taken a greater interest in the activities of NHS Direct syndromic surveillance, and regional and national networks have been established. These networks include other divisions of the Health Protection Agency, the Faculty of Public Health, acute hospital NHS trusts and primary-care organizations. The need for additional expertise to interpret trends detected by NHS Direct surveillance resulted in collaboration with the UK Meteorological Office. Distribution of data within and across regional boundaries improved knowledge sharing and networking among epidemiologists and physicians working in public health.

Operations of the NHS Direct syndromic surveillance system have been previously described (*10*). NHS Direct is a nurse-led telephone helpline that provides health information and health advice to callers with symptoms, including directing them to the appropriate NHS service. NHS Direct handled 6 million calls per year (*11*). Nurses at 22 NHS Direct sites use a computerized clinical decision support system (CAS) containing approximately 200 clinical algorithms, each with series of questions relating to symptoms. The NHS Direct syndromic surveillance system provides surveillance of 10 syndromes (i.e., cold/influenza, cough, diarrhea, difficulty breathing, double vision, eye problems, lumps, fever, rash, and vomiting) commonly occurring in the community and requiring telephone health advice. An increase in the number of callers with these syndromes might be caused by a naturally occurring outbreak (e.g., influenza) or early stages of illnesses caused by biologic or chemical weapons. At RSU, information derived from the call data is initially analyzed using confidence interval and control chart methodology (stage 1 investigation). Any statistical aberrations from historical trends (i.e., exceedances) are further investigated by the team of scientific and medical staff (stage 2 investigation). A public health alert (stage 3 investigation) is issued if no plausible explanation can be found for the exceedance (*10*). An alert is usually triggered by close geographic clustering of calls and/or sustained high level of calls for the same syndrome (Figure 2).

FIGURE 2. National Health Service (NHS) Direct syndromic surveillance operational flowchart



Outbreak Detection

The NHS Direct syndromic surveillance system captures an event instantly when a caller contacts the NHS Direct helpline. Every weekday, approximately 5 minutes are required to process the previous day's data at each NHS Direct site and transmit them to the Health Intelligence Unit (HIU). HIU collates these 22 files from the 22 NHS Direct sites and transmits them to the RSU. Application of the pattern recognition tools including confidence intervals and control charts methodologies is normally complete by midday. Further (stage 2) investigation is completed within 2 hours of the detection of an exceedance and, if necessary, a stage 3 investigation is initiated on the same day. Public health interventions usually include communications and alerts to local public health professionals. These are normally implemented by the end of the working day and, depending on the severity and urgency of the situation, very prompt public health responses can be initiated. Other public health interventions include enhanced analysis of call data until the exceedance abates. During the weekends, data are collected but not analyzed until the following Monday. A similar lag exists during public holidays in England and Wales, although emergency surveillance and epidemiologists can be provided if necessary. NHS Direct syndromic surveillance is the only system producing daily surveillance data for England and Wales and has the ability to record an increase in syndromes 12–36 hours after the calls have been made.

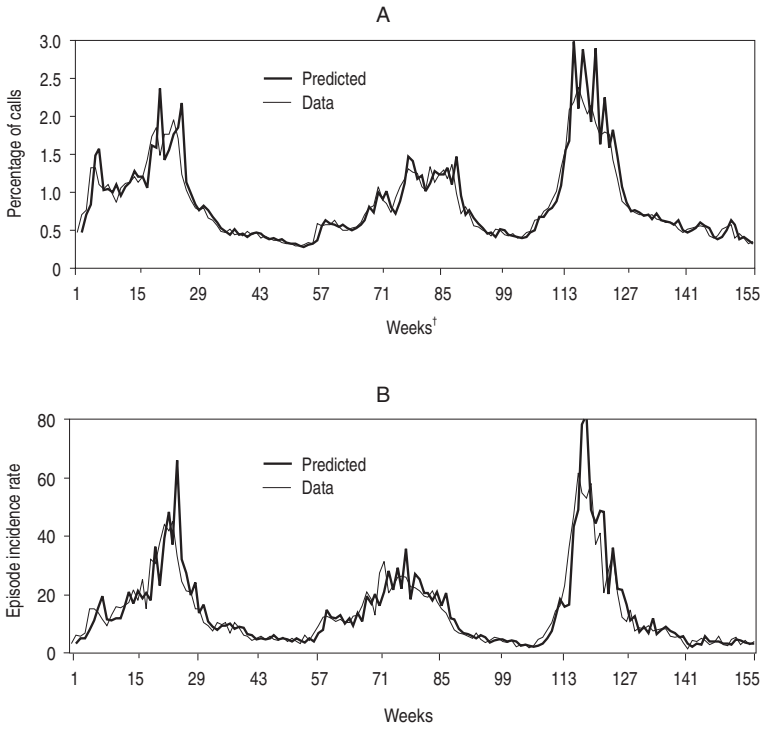
Data quality is determined by its completeness and representativeness of the coverage. The system is designed to capture all events from the population of England and Wales. The volume of calls is disproportionately low for the elderly

(aged ≥65 years) and high for young children (aged <5 years), suggesting that the surveillance system might have potential for the surveillance of common viruses predominantly affecting children (e.g., rotavirus). Regional variations in call rates are not substantial. Calls represent the ethnic mix of the UK population; however, 65% of callers are female (*12*). During the preceding 3 years, the volume of calls to NHS Direct has increased, indirectly improving representativeness of the surveillance system. Completeness of data transmitted from 22 NHS Direct sites to the Regional Surveillance Unit consistently approaches 100%; completeness of data collection at NHS Direct sites is more difficult to evaluate. Intuitively, because of the simplicity of use of NHS Direct software systems, data collection should be complete; however, separate audit is necessary to support this assumption.

Validation of the surveillance systems performance was conducted using both qualitative and quantitative approaches. The majority of interviewed stakeholders indicated that they perceived that the NHS Direct syndromic surveillance system registered an increase in calls about diarrhea and vomiting at the times when traditional public health surveillance systems indicated a national increase in Norovirus. Similarly, an increase in calls about colds and fever coincided with the increase of influenza incidence nationally. Although this is a subjective view, it was recognized that NHS Direct surveillance augmented data from other surveillance systems.

Traditionally, a quantitative approach to determine the validity of a surveillance system involves the calculation of sensitivity, specificity, and positive predictive value (*6,13*). In the context of syndromic surveillance, this is difficult to achieve. The unit of analysis is the detection of an outbreak or trend, but not an individual illness. Such a detection is frequently based on drawing information from various sources and ultimately on professional judgement. The standard needed for calculations is rarely available and frequently represents a variable itself. Another approach is to determine the correlation between data derived from different surveillance systems. The calculation of the Spearman rank correlation coefficients was used (*2*). However, this approach represents a historic evaluation over a prolonged period of time and does not take into consideration natural trends and seasonality. Therefore, while using the same principle of comparing NHS Direct syndromic surveillance to other robust surveillance systems, time series analyses were used to determine the cross-correlation between NHS Direct and WRS time series for ILIs. A comparison with the laboratory-based surveillance was considered less appropriate because only a small proportion of influenza samples are collected and tested in the laboratory in the UK. Data from real time series and those predicted by the models demonstrated a satisfactory fit (Figure 3). The Portmanteau test results indicated that two sets of residuals were white noise. Statistically significant but weak correlations were detected at lag (week) 0, 1, 2, and 3 between NHS Direct and WRS time series (Table 1). This indicates that an increase in consultations for ILIs recorded by WRS is preceded by the increase in calls to NHS Direct for ILI by 1–3 weeks and that increases recorded by both systems can occur simultaneously. The Spearman rank correlation coefficient was calculated to be 0.85, but the time series modelling approach takes into account the timing of observations and indicates how NHS Direct data are correlated to WRS data by giving the correlations at different lags. The conclusions of this report are dependent on the fit of the time-series models, normality of the residuals, and the number of observations (156 in the model). Time-series models with altered parameters were fitted but results remained similar. NHS

FIGURE 3. National Health Service (NHS) Direct and Royal College of General Practitioners Weekly Return Service (WRS) influenza-like illness data and prediction from time series model — England and Wales, 2001–2004*



* Graph A presents NHS Direct influenza-like illness data and prediction from time series model for 2001–2004. Graph B presents WRS influenza-like illness data and prediction from time series model for 2001–2004.
† Week 1 corresponds to week 35 in 2001; week 156 corresponds to week 34 in 2004.

TABLE 1. Results of time series analysis — England and Wales

Variable	Transformation	Portmanteau test p-value	Statistically significant correlation*
NHS Direct surveillance percentage of calls for influenza-like illnesses	Reciprocal	0.99	0.196, 0.224, 0.214 and 0.168 at lag 0, 1, 2 and 3
WRS episode incidence rate per 100,000 for influenza-like illnesses	Reciprocal square root	0.99	

*Cut-off for statistical significance is 2/√N-1, where N is the number of observations (156). It equals 0.1606.

Direct syndromic surveillance offers an additional benefit of having data available daily in contrast to the WRS operated by RCGP.

Experience

Qualitative interviews indicated that analysis and interpretation of data from the NHS Direct syndromic surveillance system resulted in outbreak detection, public health actions in response to alerts generated by the system, and research and development work. The majority of stakeholders agreed that NHS Direct surveillance detected national (England and Wales) outbreaks of ILI and increases in diarrhea and vomiting. An increase in callers reporting difficulty breathing was documented at a regional (countywide) level. Mapping of calls to NHS Direct by residential postcode was also feasible (*10*). However, it is unclear whether NHS Direct syndromic surveillance can be used to detect small-scale local outbreaks (e.g., neighbourhood or small town). However, on the basis of modelling work, evidence exists that the potential of the NHS Direct surveillance system to detect local outbreaks will be improved by the predicted rise in NHS Direct call rates in England and Wales (*14*).

A pilot study to investigate the feasibility of influenza self testing by NHS Direct callers was conducted during the winter of 2003 and 2004. A total of 22% of the callers involved in this pilot study tested positive for the influenza virus strain known to be prevalent during that season's influenza epidemic (*15*). NHS Direct syndromic surveillance output has also been used to track epidemics, for example, by identifying the age-groups most affected during the influenza season (*9*), and to reassure the public during periods of increased perceived risk that illness in the community has not increased. The majority of stakeholders agreed that the system contributed to a better understanding of trends and baselines of the syndromes under surveillance. In addition, the system has also promoted collaborative work between public health professionals and led to the formation of professional networks.

NHS Direct syndromic surveillance examines call data for 10 syndromes commonly presented in the community. All interviewed stakeholders believed that the expansion of the system to capture syndromes other than the 10 under surveillance is feasible because the clinical assessment software used by NHS Direct staff includes approximately 200 algorithms. Additional input will be required in terms of professional time and funds. If a strong need exists, such changes can be implemented quickly. Additional algorithms to handle potential deliberate release events can also be added to the NHS Direct clinical assessment software through negotiation with NHS Direct. These algorithms can be switched on in an emergency situation. The limitation is that when new data are available, time is needed to form a meaningful baseline to interpret new trends. The system potentially can aid the management of an outbreak and detection by examining local or regional trends when an outbreak is declared.

The NHS Direct surveillance system is embedded into operations of the NHS Direct service; therefore, duplicating such a surveillance system in different settings or jurisdictions is dependent on an existing service similar to NHS Direct. Although disseminating NHS Direct surveillance experience within the United Kingdom (i.e., to Scotland and Northern Ireland) might be easier, national coverage and reliance on the NHS Direct infrastructure for operations might preclude the system's replication elsewhere.

Most data acquisition is conducted by staff of NHS Direct sites who contribute indirectly to the operation of the surveillance system. The majority of stakeholders consider NHS Direct syndromic surveillance important and acceptable. Regional epidemiologists who staff the on-call roster to help interpret surveillance output and initiate public health actions accept the additional workload. The system is easy to operate because it does not require extensive computer or programming training for most of the front-line staff. The clinical assessment software (CAS) used at NHS Direct sites is under constant review, and the proven service robustness of NHS Direct ensures data are not lost if there is a technical problem at an individual site. Data transmission modalities are also regularly serviced and upgraded. Although occasional personnel shortages are recorded, this has never resulted in the loss of data. The NHS Direct syndromic surveillance system is funded by continuous appropriations and by research grants ranging from short to long term.

System Costs

The direct annual cost of operating the NHS Direct surveillance system is an estimated $280,000. This includes salaries and benefits for one fulltime scientist, one fulltime information analyst, and four parttime professionals funded by the surveillance project. Medical epidemiologists are funded by the Health Protection Agency. Surveillance activity is embedded into wider NHS Direct operations. Therefore, no additional costs are accrued for the use of NHS Direct software, maintenance of facilities, and data transmission. Overall, the marginal cost of operating the system is low. Estimating variable cost is more difficult because it depends on the frequency of additional analyses and initiated public health actions. Given the current workload, no extra cost was incurred as a result of acting on genuine alerts and screening out false alarms. In addition to personnel considerations, variable costs might increase if further testing (i.e., laboratory testing) is initiated to validate trends detected by the syndromic surveillance system. No information is available on clinical outcomes to estimate benefits resulting from decreases in the morbidity caused by precise outbreak detection or costs resulting from missed outbreaks or excessive false alarms.

Conclusion

The NHS Direct syndromic surveillance system is the only national syndromic surveillance system in England and Wales. Since the start of its operations in 1999, the capabilities have been expanded from augmenting data from other surveillance systems to detecting a variety of syndromic trends, forming historical baselines, and using it for the potential detection of deliberate release of harmful agents. Dissemination of the NHS Direct syndromic surveillance output has prompted interagency collaborations between medical, scientific, and public health professionals. NHS Direct syndromic surveillance is regarded as timely, representative, useful, and acceptable with low marginal costs. More work is needed to improve its portability and flexibility. It has the potential to detect high-risk, large scale events, but in its current state is less likely to detect smaller, localized outbreaks.

The CDC framework is a benchmark tool to evaluate well-established syndromic surveillance systems. The greatest challenge is to develop consistent techniques to assess whether syndromic surveillance systems provide an early warning of outbreaks of disease in the community. In the future, this needs to be considered at the stage of planning and purpose formulation of new systems. The creation and maintenance of an international database of evaluation projects can be beneficial for further development of research on syndromic surveillance.

Acknowledgments

The authors thank all stakeholders who volunteered their time to participate in this research, staff of NHS Direct sites and Douglas Fleming (Director of the Birmingham Research Unit of the Royal College of General Practitioners for Weekly Return Service) for WRS data.

References

1. UK Department of Health. Getting ahead of the curve: a strategy for combating infectious diseases. Available at http://www.doh.gov.uk/cmo/idstrategy.
2. 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;23:229–30.
3. Fleming DM, Barley MA, Chapman RS. Surveillance of the bioterrorist threat: a primary care response. Commun Dis Public Health 2004;7:68–72
4. Bravata DM, McDonald KM, Smith WM, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Intern Med 2004;140:910–22.
5. Bravata DM, Sundaram V, McDonald KM, et al. Evaluating detection and diagnostic decision support systems for bioterrorism response. Emerg Infect Dis 2004;10:100–8.
6. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V; CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR 2004;53(No. RR-5).
7. CDC. ESSENCE II and the framework for evaluating syndromic surveillance systems. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:159–65.
8. Baker M, Smith GE, Cooper D, et al. Early warning and NHS Direct: a role in community surveillance? J Public Health Med 2003;25:362–8.
9. Cooper DL, Smith GE, Hollyoak VA, Joseph CA, Johnson L, Chaloner R. Use of NHS Direct calls for surveillance of influenza—a second year's experience. Commun Dis Public Health 2002;5:127–31.
10. CDC. National symptom surveillance using calls to a telephone health advice service—United Kingdom, December 2001–February 2003. MMWR 2004;53:179–83.
11. Directorate of Access and Choice. Developing NHS Direct: a strategy document for the next 3 years. London, England: Department of Health; 2003.
12. Cooper DL, Arnold E, Smith GE, et al. The effect of deprivation, age and gender on NHS Direct call rates. Br J Gen Pract 2005;5:287–2915.
13. Murakami Y, Hashimoto S, Taniguchi K, Osaka K, Fuchigami H, Nagai M. Evaluation of a method for issuing warnings pre-epidemics and epidemics in Japan by infectious diseases surveillance. J Epidemiol 2004;14:33–40.
14. Cooper DL, Verlander NQ, Smith GE, et al. Can syndromic surveillance data detect local outbreaks of communicable disease? A model using a historical cryptosporidiosis outbreak. Epidemiology and Infection; In press.
15. Cooper DL. Can we use self-testing to augment syndromic surveillance? A pilot study using influenza. Presented at 2004 Syndromic surveillance conference, Boston, Massachusetts. November 3–4, 2004.

Initial Evaluation of the Early Aberration Reporting System — Florida

Yiliang Zhu,[1] W. Wang,[1] D. Atrubin,[2] Y. Wu[1]
[1]*College of Public Health, University of South Florida, Tampa, Florida;*
[2]*Hillsborough County Health Department, Tampa, Florida*

Corresponding author: Yiliang Zhu, Department of Epidemiology and Biostatistics, University of South Florida, 13201 Bruce B. Downs Blvd., MDC 56, Tampa, FL 33612. Telephone: 813-974-6674; Fax: 813-974-4719; E-mail: yzhu@hsc.usf.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *In recent years, many syndromic surveillance systems have been deployed around the United States for the early detection of biologic terrorism–related and naturally occurring outbreaks. These systems and the associated aberration detection methods need to be evaluated.*

Objective: *This study evaluated several detection methods of the Early Aberration Reporting System (EARS) under serially correlated syndromic data and to demonstrated the need for calibrating these methods.*

Methods: *In an initial evaluation of the Syndromic Tracking and Reporting System in Hillsborough County, Florida, serially correlated syndromic data were simulated using statistical models in conjunction with real syndromic data. The detection methods were tested against two patterns of simulated outbreaks. They were compared using a conditional average run length and a receiver operating characteristic curve under defined patterns of detection.*

Results: *Increasing serial correlation inflates the false alarm rate and elevates sensitivity. Among the detection methods in EARS, C2 and P-chart have the best overall receiver operating characteristic curve within the context of the simulations. C2 is least affected by the serial correlation, the outbreak type, and the defined patterns of detection signal.*

Conclusion: *Evaluation of the detection methods needs to be adaptable to the constantly changing nature of syndromic surveillance. Deployment of EARS and other methods requires adjusting the false alarm rate and sensitivity in accordance with the syndromic data, the operating resources, and the objectives of the local system. For timely detection, C2 is superior to other methods, including C3, under the simulation conditions. P-chart is the most sensitive when the serial correlation is negligible.*

Introduction

The risk for biologic terrorism attacks has promoted the development and deployment of syndromic surveillance systems in the United States and around the world (*1*). The majority of these systems use patient or consumer encounter data from multiple sources (e.g., hospital emergency departments (ED), military facilities, or theme parks). Workplace absenteeism and over-the-counter drug sales are also being monitored for statistical aberrations. The data are converted to specified syndrome categories and analyzed to detect significant temporal or spatial aberrations that deviate from the expected baseline trends. Although some systems conduct spatial analysis (*2*), most use trend analysis to detect temporal aberrations.

Developed by CDC, the Early Aberration Reporting System (EARS) consists of a class of quality-control (QC) charts, including Shewhart chart (P-chart), moving average (MA), and variations of cumulative sum (CUSUM) (*3*). Many syndromic surveillance systems use EARS for temporal aberration detection (*4*); some also use other QC charts such as exponentially weighted moving average (EWMA) (*5*,*6*). A common characteristic in adopting these QC charts for syndromic data analysis is the use of a sample estimate for the baseline mean and standard deviation (SD). This approach circumvents the difficulties associated with the modeling of the baseline trend of the syndrome, a process complicated by the discreteness, serial correlation, seasonality, and daily fluctuation of the syndromic data. At present, understanding of these methods within the context of syndromic data is limited, and systematic evaluations of syndromic surveillance have not been conducted.

In QC settings, serial correlations can substantially affect the time length to the first aberration signal (Average Running Length [ARL]) when using such methods as CUSUM and P-chart (*7*,*8*). How serial correlation also could affect other performance measures such as the false alarm rate (one minus specificity) or the sensitivity (defined generally as the probability of successful detection [PSD] associated with a pattern

of signals) is unclear. Certain signal patterns or event detection (e.g., signaling three days in a row) might provide more information than a single day signal about the strength or duration of the outbreak, thus guiding public health agencies in designating the necessary follow-up investigation according to the strength of the signals (*9*). Simulation studies were conducted to evaluate the sensitivity and specificity of a single day signal for three EARS' variations of CUSUM (C1, C2, and C3) (*10*), a seasonally adjusted CUSUM, and an historic limits method (*11*). However, these simulations assumed serially independent data, with the magnitude of aberrations ranging between 1.65 to 51.0 times the baseline mean (*10*) or 2–3 times the baseline standard deviation. Seasonality was added to simulated baseline data in the second study (*11*), but was not modeled in the baseline mean. The simulation indicated that C3 is superior to C2, without controlling for specificity.

This study focuses on the impact of serial correlation on the PSD of selected patterns of signals and compares the performance of five common detection algorithms (P-chart, C2, C3, MA [employed in EARS], and EWMA [a general statistical QC chart not included in EARS]). Two different outbreak patterns are considered here. The comparisons are based on two criteria. The first criterion is a conditional average run length (CARL) given successful detection in a given time window. The second is the receiver operation characteristic (ROC) curves. The purpose is to provide guidance for selection of detection methods and to illustrate the need for calibrating the methods to attain a required level of specificity and sensitivity.

Methods

Detection Algorithms

The QC charts in EARS use daily syndromic counts or incidences (daily counts of a specific syndrome divided by total ED volume for the day) y_t between day t−K+1 and day t (current day) to derive a monitoring statistic m_t (4). The monitoring statistic is

$$m_t = y_t,\ m_t = \frac{1}{K}\sum_{k=1}^{K} y_{t-k+1},\ \text{and}\ \ m_t = (1-\lambda)\sum_{k=1}^{\infty}\lambda^{k-1} y_{t-k+1}\ ,$$

for P-chart, MA, and EWMA, respectively. P-chart uses the current day only; MA is the average of K days before and including the current day; EWMA is a weighted average of all previous days with an exponentially decreasing weight given to days further away from the present day. The system generates a signal if m_t exceeds a threshold $c\,\sigma_m$ above the expected level μ_m. The PSD of a single alert (sensitivity) is the probability $\Pr(m_t - \mu_m > c\sigma_m)$ given an outbreak. The constant c determines the threshold in multiples of the standard deviation σ_m. Because the distribution of m_t is complex, EARS uses a sample estimate of μ_m and σ_m on the basis of data in a baseline window of B days: $y_{t\text{-}B\text{-}g}, y_{t\text{-}B+1\text{-}g}, \cdots, y_{t\text{-}1\text{-}g}$ with a gap of g days before the present day t. Specifically,

$$\mu_m = \frac{1}{B}\sum_{b=1}^{B} y_{t-b-g}\ ,\ \text{and}\ \ \sigma_m^2 = \frac{1}{B-1}\sum_{b=1}^{B}(y_{t-b-g}-\mu_m)^2.$$

EARS also employs three variations of the CUSUM method (*4*) (C1, C2, and C3). C1 uses data from the current day only and a baseline window of the preceding 7 days: day t−7 to t−1 (B = 7 and g = 0). If C1 generates a signal on day t, day t will become a part of the baseline for day t+1, which might inflate the corresponding baseline mean μ_m, and reduce the PSD for that day. C2 differs from C1 by shifting the 7-day baseline window to left with a gap of g = 3 days. As a result, its PSD of signaling on day 2 and 3 is not affected by a signal on day 1. Analytically, C1 and C2 are nearly equivalent in the absence of outbreaks, but C2 is more sensitive than C1 in signaling a continued outbreak past its onset. For this reason, C1 was not evaluated.

C3 differs further from C2 by using a partial sum of positive daily deviations for the current and 2 previous days (t−1 and t−2):

$$m_t = \frac{(y_t - \mu_m(t) - \sigma_m(t))^+}{\sigma_m(t)} + \frac{(y_{t-1} - \mu_m(t-1) - \sigma_m(t-1))^+}{\sigma_m(t-1)} I_{t-1} + \frac{(y_{t-2} - \mu_m(t-2) - \sigma_m(t-2))^+}{\sigma_m(t-2)} I_{t-2}\ .$$

The superscript $^+$ truncates the quantity in the parentheses to zero if its value is negative, and I assumes the value 0 if $y > \mu_m + 3\sigma_m$ for (t−1 and t−2) or 1 otherwise. Thus, C3 includes only the deviations that are 1–3 standard deviations above its mean and will generate a signal if m_t is >2, the default threshold in EARS. C2 exceeding the threshold of 3 implies the first component of m_t (C3) exceeds the threshold of 2. In EARS, the sample estimate of mean μ_m does not adjust for seasonality (e.g. the day-of-the-week effects) of the baseline. Ideally, seasonality can be filtered out before applying these methods to the data.

Performance Measures

QC charts are traditionally evaluated with respect to the sensitivity and false alarm rate (one minus specificity) of single day detections (*10*,*11*). Because disease outbreaks probably are associated with temporal patterns, corresponding patterns of aberration signals should be considered. For example, with a

disease outbreak that persists at a high level of incidence or count for a number of days after onset, consecutive signals might alert not only the onset but also the duration of the outbreak. Such a pattern of signals is called a detection event. More importantly, public health workers can use a detection event to estimate the duration or strength of an outbreak and respond accordingly. Therefore, detection events of composite detection signals (e.g., the first or the first pair of consecutive signals) are used to define sensitivity and false alarm rates, which are in turn converted to ROC curves and used as an overall performance measure.

Average run length (ARL) also provides information on the distribution of time to first signal. Analytical results on run-length distribution are difficult to obtain for the monitoring statistic (*12*), so a conditional ARL is considered by assuming a detection event on the first day or the first two days (within the first 4–5 days of the outbreak).

Simulations

Daily ED visit data were collected from a local hospital during March 2002–December 2003. A common syndrome, respiratory infection (RI), was chosen for simulation. The data revealed an average daily visit of 155 patients (standard deviation = 32.6). Daily RI count was 5.96 on average, or 0.0319 in incidence. The patient volume and RI count showed a lag-one serial correlation between two consecutive observations of 0.768 and 0.438, respectively. The lag-one correlation for incidences was 0.323. The magnitude of these correlations is expected for common syndromes; rare syndromes are less correlated. Ignoring the serial correlation can yield a misleading level of sensitivity and false alarm rate. The data also indicated day-of-week-effects (lower values on Friday and peaks on Sunday and Monday). In evaluating the detection methods used in STARS, these simulations did not incorporate day-of-the-week effects in the baseline.

On the basis of these parameters derived from the RI data, multilevel, generalized linear mixed-effects models (*13*) were used to simulate the baseline of daily ED volume N_t and RI count X_t. Under the Poisson distribution for N_t, the daily mean λ_t fluctuates around a constant λ_0 with

$$\log(\lambda_t) = \log(\lambda_0) + \alpha_{-t}.$$

Random effects α_t characterize the variation, which are serially correlated through a classic first order autoregressive time series (AR1) model

$$\alpha_t = \varnothing_1 a_{t-1} + \varepsilon_{1t}.$$

The errors $\{\varepsilon_{1t}\}$ are independent, normal $N(0,\sigma_1^2)$. Presently, $\lambda_0 = 155$, $\varnothing_1 = 0.8$, and $\sigma_1 = 0.479$ are used to approximate the observed parameters.

The simulation model for count X_t uses a binomial distribution $B(N_t, p_t)$ given the volume N_t and incidence p_t. The incidence p_t fluctuates around the baseline average p_0 through the model

$$\text{logit}(p_t) = \text{logit}(p_0) + \beta_t$$

with the random effects β_t also following an AR1 time series model

$$\beta_t = \varnothing_2 \beta_{t-1} + \varepsilon_{2t}.$$

The errors $\{\varepsilon_{2t}\}$ were independent, normal $N(0,\sigma_2^2)$. The RI data suggested $p_0 = 0.0319$, $\varnothing_2 = 0.699$, and $\sigma_2 = 0.315$ (*9*).

Each simulation consists of 100 days of data $\{N_t, X_t\}$, and the study involved 5,000 replications. Simulated incidences yielded a sample serial correlation of 0.277 (SD = 0.175).

Outbreak data are simulated under two patterns: a slow-building and a sudden-surge trend. The slow-building trend is characterized by a constant-pace increase in incidence that peaks on day 4 as it rises to three times the incidence at baseline. It then decreases at a constant rate back to the baseline by day 8. A norovirus outbreak that is spread from person-to-person with a short incubation period of 12–50 hours could produce such a pattern. The sudden-surge trend describes an elevation of incidence to three times that of the baseline on day 1, remaining at the same level for the following 3 days, then dropping back to the baseline level on day 8 at a constant rate. A foodborne salmonellosis outbreak is a good example of an outbreak of such a pattern. If persons were exposed during a 1–2 day window, many cases would be expected in a 3–4 day period, with the cases declining over the next few days. Secondary cases could occur as a result of person-to-person transmission that could account for some of the cases occurring during the end of the outbreak.

The magnitude for the simulated outbreaks was chosen at three times that of the baseline; the actual size will depend on several factors, particularly the size of the exposed population. Syndromic cases were simulated using the same models for the baseline but under the incidence parameters of the designed outbreak pattern and size. Outbreak cases replaced baseline data between day 81 and day 87. Simulations were also conducted with zero serial correlation to further demonstrate the impact of correlation.

Detection methods were tested against 5,000 replications of the outbreak and nonoutbreak data during days 61–67. The time periods of the outbreaks and nonoutbreaks were chosen to ensure that the simulated data reached statistical stability. The false alarm rate was calculated by the number of detected events divided by the total possible number of detection events during the nonoutbreak period. PSD was estimated identically, but over the outbreak period.

Results

Conditional ARL

The probability of successful first detection on day one, two (no alert on day 1), three (no alert on day 1 and day 2), or four (no alert on day 1, day 2, and day 3) varies, along with the probability of first detection within the first four days of the slow-building outbreak (Table 1). The detection algorithms were calibrated by a varying threshold so that the probability of false alert was approximately 0.05. Given the PSD within 4 days, the conditional average run length (CARL) is the weighted average of the days to the first detection, with the weight being the conditional probabilities of detection on each day, given detection by day four. P-chart has the highest probability (0.97) of first signal within four days, and MA has the lowest (0.32). Given detection by day four, CARL is 2.60 for P-chart and 3.32 for EWMA (λ = 0.72). For a single day, C2 is more likely to signal (PSD = 0.18) than others on day 1, P-chart on day 2 and 3 (0.33 and 0.37), and EWMA on day 4 (0.45).

With the detection event of first pair of consecutive signals, CARL is the weighted average of PSD on days 1 and 2, 2 and 3 (but not day 1), up to days 4 and 5 (Table 2). The PSD decreases with time for C2, C3, and P-chart, and increases for EWMA.

With the sudden-surge outbreak, the PSD of first signal (Table 3) and of first pair of consecutive signals (Table 4) decreases with time, with the only exception being MA on days 1 and 2 (Table 3). For the event of day 1 signal, P-chart has the highest PSD followed by C2. For the event of first pair of consecutive signals, the five methods are comparable on days 1 and 2, but EWMA (λ = 0.464) retains a slightly higher PSD on days 2 and 3.

ROC Curves

For the serially correlated baseline (ρ = 0.277) with the slow-building outbreak, the ROC curves of first detection occurred on day 1 (Figure 1). Each marked point on an ROC curve

TABLE 1. Probability* of first alert within 4 days of a slow-building outbreak

Method	Threshold	Day 1	Day 2	Day 3	Day 4	Probability†	CARL§
C2	2.56	0.18	0.27	0.25	0.15	0.85	2.44
C3	3.59	0.14	0.25	0.27	0.18	0.84	2.58
P-chart	2.72	0.11	0.33	0.37	0.16	0.97	2.60
MA	1.87	0.05	0.02	0.05	0.20	0.32	3.25
EWMA¶	1.83	0.05	0.06	0.26	0.45	0.82	3.32

* False alarm rate is approximately 0.05.
† Probability of first alert within first 4 days.
§ Conditional average run length in day.
¶ Exponentially weighted moving average. Weight parameter λ = 0.72.

TABLE 2. Probability* of first pair of consecutive alerts within 5 days of a slow-building outbreak

Method	Threshold	Day 1 and 2	Day 2 and 3	Day 3 and 4	Day 4 and 5	Probability†	CARL§
C2	1.35	0.30	0.13	0.09	0.03	0.55	2.73
C3	1.44	0.30	0.12	0.10	0.06	0.58	2.86
P-chart	1.69	0.25	0.14	0.13	0.05	0.57	2.96
MA	1.70	0.06	0.01	0.02	0.11	0.20	3.90
EWMA¶	1.58	0.06	0.04	0.11	0.23	0.44	4.16

* False alarm rate is approximately 0.05.
† Probability of two consecutive alerts within first 5 days.
§ Conditional average run length in day.
¶ Exponentially weighted moving average. Weight parameter λ = 0.72.

TABLE 3. Probability* of first alert within 4 days of a sudden-surge outbreak

Method	Threshold	Day 1	Day 2	Day 3	Day 4	Probability†	CARL§
C2	2.68	0.68	0.13	0.06	0.03	0.90	1.38
C3	3.85	0.59	0.21	0.09	0.03	0.92	1.52
P-chart	2.90	0.71	0.15	0.06	0.04	0.96	1.40
MA	1.84	0.31	0.41	0.17	0.04	0.93	1.94
EWMA¶	1.89	0.57	0.23	0.09	0.04	0.93	1.57

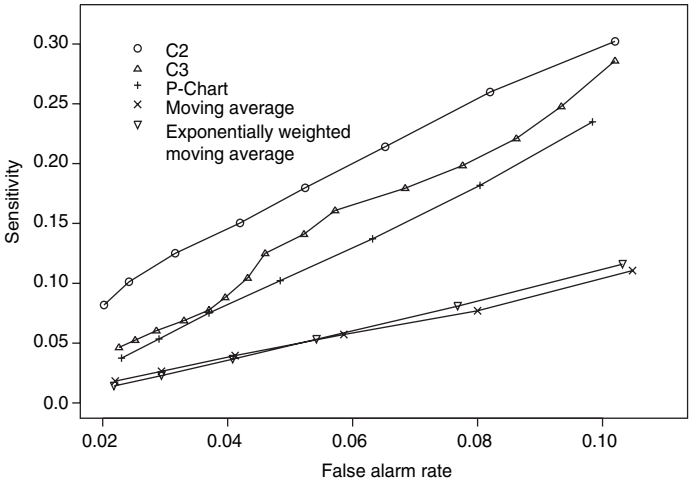
* False alarm rate is approximately 0.05.
† Probability of first alert within first 4 days
§ Conditional average run length in day.
¶ Exponentially weighted moving average .Weight parameter λ = 0.464.

TABLE 4. Probability* of first pair of consecutive alerts within 5 days of a sudden-surge outbreak

Method	Threshold	Days 1 and 2	Days 2 and 3	Days 3 and 4	Days 4 and 5	Probability†	CARL§
C2	1.41	0.72	0.05	0.01	0.00	0.78	2.09
C3	1.53	0.72	0.10	0.02	0.02	0.86	2.23
P-chart	1.74	0.72	0.06	0.02	0.01	0.81	2.16
MA¶	1.57	0.61	0.12	0.07	0.02	0.82	2.39
EWMA**	1.48	0.67	0.26	0.03	0.02	0.98	2.39

* False alarm rate is approximately 0.05.
† Probability of two consecutive alerts within first 5 days.
§ Conditional average run length in day.
¶ Moving average.
** Exponentially weighted moving average. Weight parameter $\lambda = 0.464$.

FIGURE 1. Receiver operation characteristic curves of detection on day 1 of a slow-building outbreak



indicates the PSD (sensitivity, vertical-axis) and false alarm rate (one minus specificity, horizontal-axis) associated with a given threshold. As the threshold increases, the point moves to the left along the ROC curve. The PSD can be read off of a ROC curve at its intersection with a vertical line given a fixed false alarm rate. The overall performance of a detection method is evaluated by the area under the curve. Thus, a method associated with a higher ROC curve is superior to one with a lower ROC curve. Ideally, ROC curves are plotted in an identical range of false alarm rates. However, because the false alarm rate is unknown for a given threshold, a sequence of threshold levels with 0.25 increments were used in determining the plotting range of false alarm rate for each curve. As a result of this approximation, ROC curves do not fill the maximum intended range in each plot. C2 has the best ROC for day 1 detection within the common range, followed by C3 and P-chart (Figure 1). This clarifies the misperception that C3 is more sensitive than C2 when one holds the threshold constant. C3 is only more sensitive if one ignores the higher false alarm rate associated with it. A threshold of 3 for C2 (third point from left) is similar to that of 4.75 for C3 (fourth point from left) in terms of false alarm rate (Figure 1). For consecutive signals on days 1 and 2 (Figure 2), C2 retains the highest PSD; MA and EWMA remain low. As the false alarm rate reaches a level above 5%, C3 performs nearly as well as C2 (Figures 1 and 2).

Under the sudden-surge outbreak, the ROC curves of day 1 detection (Figure 3) demonstrate a superior performance of P-chart compared to the other algorithms, with C2 performing nearly as well. At an approximately 2% false alarm rate, the PSD is approximately 60% for P-Chart (Figure 3) for the sudden-surge outbreak. In contrast, the PSD is less than 5% for the slow-building outbreak (Figure1). The PSD is generally higher with outbreaks of large magnitude. ROC curves for signals on both days 1 and 2 (Figure 4) indicate that C2 and P-chart perform the best. C3 is comparable only when the false alarm rate exceeds 5%. MA and EWMA gain in the PSD as the outbreak sustains itself (Figures 3 and 4).

ROC Associated with a Zero Correlation

To further demonstrate the impact of serial correlation on aberration detection, 5,000 replications of baseline and outbreak data with serial correlation ρ = -0.0176 (theoreti-

FIGURE 2. Receiver operation characteristic curves of detection on the first 2 days of a slow-building outbreak

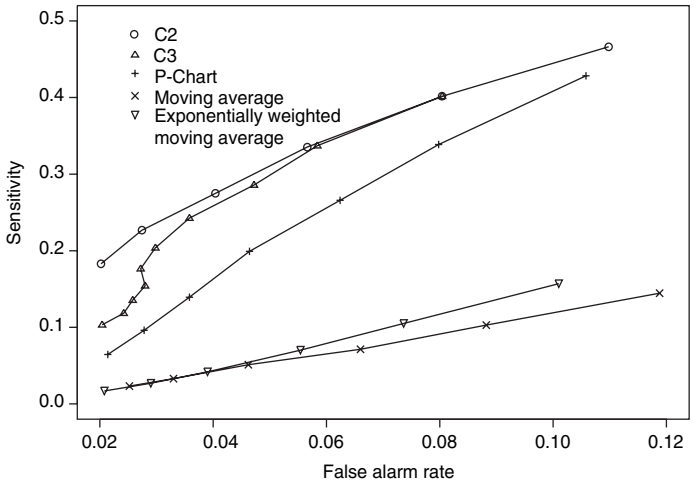


FIGURE 3. Receiver operation characteristic curves of detection on day 1 of a sudden-surge outbreak

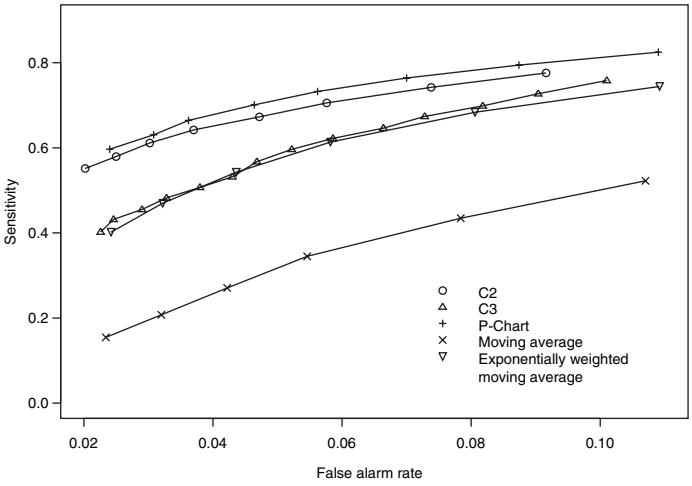


FIGURE 4. Receiver operation characteristic curves of detection on the first 2 days of a sudden-surge outbreak

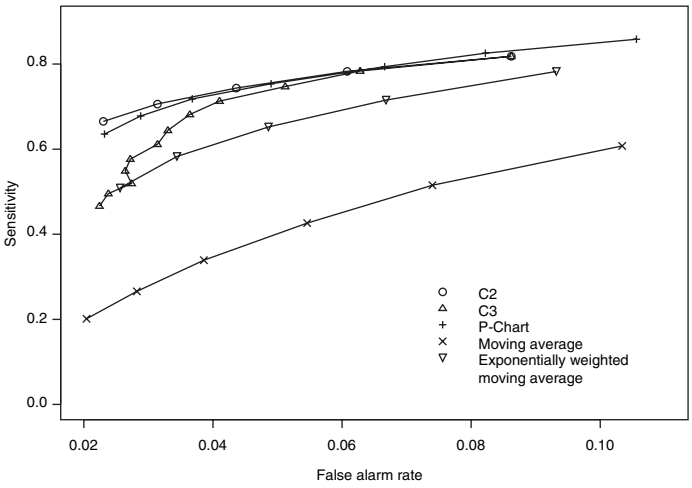
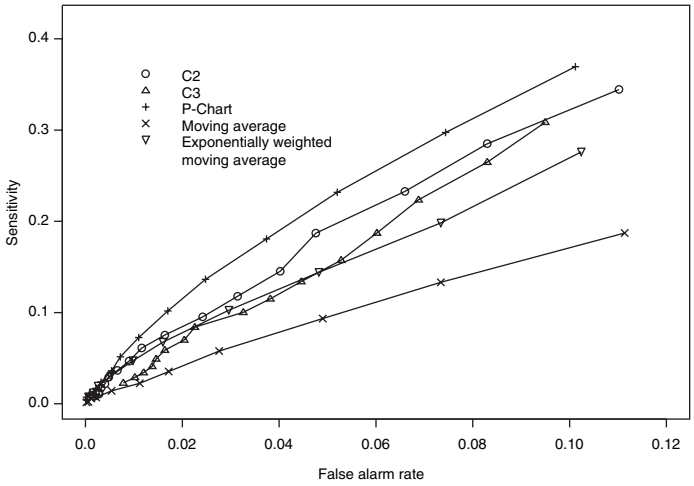
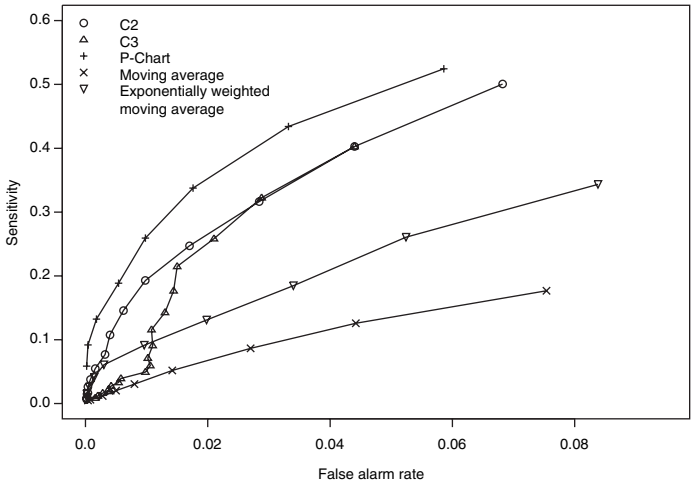


FIGURE 5. Receiver operation characteristic curves of detection on day 1 of a slow-building outbreak*



* Serial correlation = 0.

FIGURE 6. Receiver operation characteristic curves of detection on the first 2 days of a slow-building outbreak*

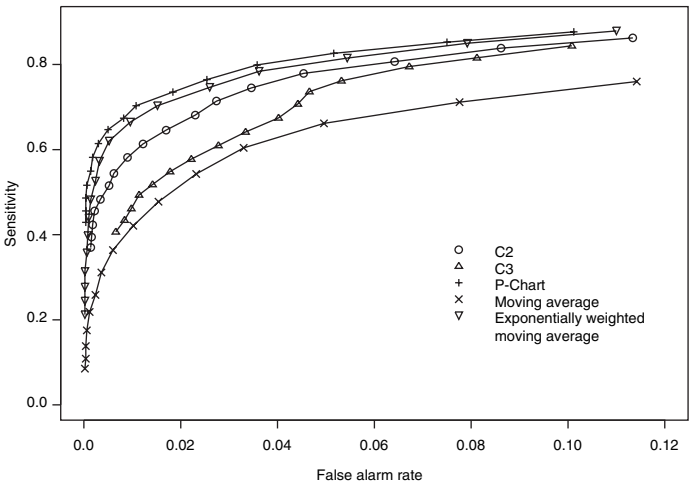


* Serial correlation = 0.

cally $\rho = 0$) were generated. The ROC curves of detection on day 1 of the slow-building outbreak (Figure 5) indicate that P-chart is most capable, followed by C2, C3, and EWMA; MA was the least capable. PSD increases for all methods, particularly P-chart, compared with the case of moderate correlation (Figure 1). Similar results are observed for the ROC curves of detection on both of the first two days (Figure 6) compared with correlated case (Figure 2). P-chart is superior to the other algorithms.

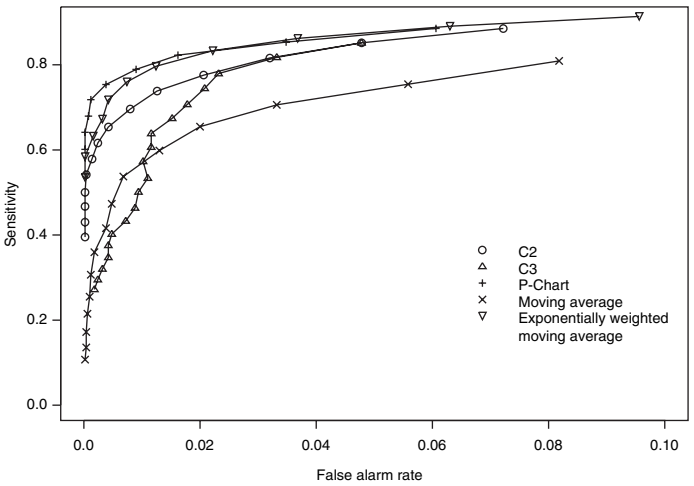
Evaluation under the sudden-surge outbreak with zero correlation yielded a similar conclusion. The PSD is generally higher in the absence of serial correlation. However, EWMA now places second in PSD, behind P-chart. PSD increases rapidly as the false alarm rate increases at the lower end (Figure 7). Compared with the nearly linear ROC curves under the slow-building outbreak (Figure 3), the PSD is larger under outbreaks of larger magnitude. The ROC curves of detection on both of the first 2 days (Figure 8) again depict a superior performance of P-chart, closely followed by EWMA. Unexpectedly, C3 has the lowest PSD when the false alarm rate falls below 0.01.

FIGURE 7. Receiver operation characteristic curves of detection on day 1 of a sudden-surge outbreak*



* Serial correlation = 0.

FIGURE 8. Receiver operation characteristic curves of detection on the first 2 days of a sudden-surge outbreak*



* Serial correlation = 0.

Conclusion

On the basis of the CARL and ROC, these simulations suggest that P-chart has the best overall performance when the data are serially independent under both the slow-building and sudden-surge outbreak (Table 5). This conclusion holds true for both detection on day 1 and detection on both day 1 and 2. Under the slow-building outbreak, C2 delivers the second best performance; under the sudden-surge outbreak EWMA performs well, closely following P-chart and clearly outperforming C3, especially at the lower false alarm rates.

With the moderate serial correlation (0.277), C2 has the best ROC under the slow-building outbreak, followed by C3. However, C3 is outperformed by C2 with respect to both day 1 detection and consecutive signals on day 1 and 2 at the lower false alarm rates. Under the sudden-surge outbreak P-chart and C2 outperforms the other methods. The difference between P-chart and C2 is generally small. Contrary to the common perception among the users of EARS, C3 is not more sensitive than C2 once the false alarm rate is held constant.

With an emphasis on timely detection of outbreaks within the first few days of onset, the findings of this report suggests the use of C2 and P-chart for surveillance purposes when the syndromic data are moderately correlated. Surveillance of rare syndromes also might benefit from EWMA because rare syndromes tend to be less correlated. EWMA also is expected to be more sensitive as an outbreak persists.

Serial correlation can considerably increase the false alarm rate and reduce the PSD. For example, the PSD of P-chart is approximately 0.20 with the false alarm rate of 0.04 for independent data; as the serial correlation increases to 0.277, the PSD decreases to approximately 0.075. In general, similar trends hold, although the magnitude of such impacts varies between detection methods and detection events. These trends were confirmed by an additional simulation with a higher serial correlation. The impact on MA and EWMA is particularly pronounced because these methods depend on an average of several days' data.

TABLE 5. Detection methods with the best ROC* under different outbreak, serial correlation, and detection objective

Serial correlation	0		0.277†	
Outbreak detection objective	Slow-building	Sudden-surge	Slow-building	Sudden-surge
Timely (day 1)	P-chart	P-chart; EWMA§	C2	P-chart; C2
Continuous (days 1 and 2)	P-chart	P-chart; EWMA	C2	C2; P-chart

* Receiver operation characteristic.
† Serial correlation among daily incidences.
§ Exponentially weighted moving average.

C2 appears to be least affected by serial correlation and is most robust. For example, the PSD of C2 declined from approximately 0.22 to 0.20 while holding the false alarm rate at 0.06; at the false alarm rate of 0.04, the PSD stayed almost constant. This offers another justification for favoring C2.

Another simulation was conducted to evaluate the use of count data (*9*). In comparison with daily incidence, using counts yielded slightly higher false alarm rates and PSDs. It seems practical to use counts for monitoring rare syndromes and proportions for common syndromes. The two approaches would be similar when the ED volume is stable and only specific syndrome cases fluctuate.

Syndromic data might demonstrate day-of-the week effects, which can be accounted for through modeling of the mean μ_m and the standard deviation σ_m. This is not done in EARS. Assessing the impact of ignoring such trend on the methods of EARS would be useful (*11*). Modeling syndromic data is complex, requiring sufficient amounts of historical data. Because syndromic data might vary by location, season, syndrome category, and type of outbreak, performance of aberration detection methods must be evaluated specifically. Appropriate use of any detection method requires evaluation and calibration of its operating characteristics. Although a higher level of PSD is generally desirable, false alarm rate needs to be controlled at a level according to the objectives of the local system and resources required to maintain the system.

References

1. Agency for Healthcare Research and Quality. Bioterrorism preparedness and response: use of information technologies and decision support systems. Available at http://www.ahrq.gov/clinic/epcsums/bioitsum.htm.
2. Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic. J Royal Stat Soc A 2001;164:61–72.
3. Bissell AF. CUSUM Techniques for Quality Control. Applied Statistics 1968;18:1–30.
4. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The Bioterrorism preparedness and response early aberration reporting system (EARS). J Urban Health 2003;80:89–96.
5. Hunter JS. Exponentially weighted moving average. J Qual Tech 1986;18:203–10.
6. Pavlin JA, Kelley PW. ESSENCE: Electronic Surveillance System for the Early Notification of Community-based Epidemics. Silver Spring, MD: U.S. Department of Defense, Global Emerging Infections Surveillance and Response System, 2001.
7. Johnson RA, Bagshaw M. The effect of serial correlation on the performance of CUSUM tests. Technometrics 1974;16:103–12.
8. Maragah HD, Woodall WH. The effect of autocorrelation on the retrospective X-chart. J Statistical Computation and Simulation 1992;40:29–42.
9. Wang W, Zhu Y. Evaluation report of Hillsborough County Health Department Syndrome Surveillance System. Tampa, Florida: University of South Florida; 2004.
10. Hutwagner LC, Thompson WW, Seeman GM, Treadwell T. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. Stat Med 2005; 24:543–50.
11. Hutwagner L, Browne T, Seeman GM, Fleischauer AT. Comparing aberration detection methods with simulated data. Emerg Infect Dis 2005;11:314–6.
12. Gan FF. Computing the percentage points of the run length distribution of an EWMA control chart. Journal of Quality Technology 1991;23:359–65.
13. Zeger LS. A regression model for time series of counts. Biometrika 1988;75:621–9.

Practice and Experience

Deciphering Data Anomalies in BioSense

Leslie Z. Sokolow,[1,2] N. Grady,[3] H. Rolka,[2] D. Walker,[2] P. McMurray,[3] R. English-Bullard,[2] J. Loonsk[2]
[1]*Innovative Emergency Management, Inc., Atlanta, Georgia;* [2]*National Center for Public Health Informatics, CDC;*
[3]*Science Applications International Corporation, Oak Ridge, Tennessee*

Corresponding author: Leslie Z. Sokolow, CDC, 1600 Clifton Rd., NE, MS E-06, Atlanta, GA 30333. Telephone: 404-498-6331; Fax: 404-498-6145; Email: lsokolow@cdc.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Since June 2004, CDC's BioIntelligence Center has monitored daily nationwide syndromic data by using the BioSense surveillance application.*

Objectives: *The BioSense application has been monitored by a team of full-time CDC analysts. This report examines their role in identifying and deciphering data anomalies. It also discusses the limitations of the current surveillance application, lessons learned, and potential next steps to improve national syndromic surveillance methodology.*

Methods: *Data on clinical diagnoses (*International Classification of Diseases, Ninth Revision, Clinical Modifications *[ICD-9-CM]) and medical procedures (CPT codes) are provided by Department of Veterans Affairs and Department of Defense ambulatory-care clinics; data on select sales of over-the-counter health-care products are provided by participating retail pharmacies; and data on laboratory tests ordered are provided by Laboratory Corporation of America, Inc. All data are filtered to exclude information irrelevant to syndromic surveillance.*

Results: *During June–November 2004, of the approximately 160 data anomalies examined, no events involving disease outbreaks or deliberate exposure to a pathogen were detected. Data anomalies were detected by using a combination of statistical algorithms and analytical visualization features. The anomalies primarily reflected unusual changes in either daily data volume or in types of clinical diagnoses and procedures. This report describes steps taken in routine monitoring, including 1) detecting data anomalies, 2) estimating geographic and temporal scope of the anomalies, 3) gathering supplemental facts, 4) comparing data from multiple data sources, 5) developing hypotheses, and 6) ruling out or validating the existence of an actual event. To be useful for early detection, these steps must be completed quickly (i.e., in hours or days). Anomalies described are attributable to multiple causes, including miscoded data, effects of retail sales promotions, and smaller but explainable signals.*

Conclusion: *BioSense requires an empirical learning curve to make the best use of the public health data it contains. This process can be made more effective by continued improvements to the user interface and collective input from local public health partners.*

Introduction

CDC's BioSense application, which has been in use since November 2003, permits the early detection of intentional and natural infectious-disease outbreaks. The application has an Internet-based interface that enables public health officials in 86 geographic regions (50 states, two territories, and 34 major metropolitan areas) to access prediagnostic health data on a near real-time basis (*1*). Statistical algorithms and analytical visualizations are used to present multiple streams of nationwide public health data. The CDC BioSense Initiative provides a new surveillance tool for state and local health use and gives CDC the responsibility for informing its use. Since June 2004, an average of six full-time CDC BioSense monitors have examined and analyzed data daily. This report describes the role of the BioSense data monitors and discusses how they monitor and decipher data anomalies.

BioSense receives daily data that are delivered electronically from four sources. Ambulatory-care clinics within the Department of Veterans Affairs (VA) and Department of Defense (DoD) systems provide *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) coded diagnoses and CPT-coded medical procedures for 580 (98%) of 592 facilities nationwide. On an average day, BioSense received more than 1,000,000 records from VA and 500,000 from DoD. During June–November 2004, data on select over-the-counter (OTC) pharmacy sales were received

from >10,000 pharmacies, representing 23% of the national market share (*2*) and 20% of U.S. ZIP codes (CDC, unpublished data, 2005). Data on laboratory tests ordered were received from Laboratory Corporation of America, Inc. (LabCorp), representing 33,674 (77%) of 44,008 U.S. ZIP codes.

BioSense is a multijurisdictional data-sharing surveillance application that is nationwide in coverage and is available to state and local public health departments. Other applications implemented at local or regional levels use other data sources and user interfaces and cannot easily be used to compare different localities across the United States (*3–17*). The BioSense Initiative includes partnerships with other application groups. BioSense's national scope maximizes its ability to detect local events and those that cross jurisdictional boundaries (e.g., states or counties). Because the sources are consistent across the country, data monitors can compare multiple locations by using the same types of data. Data-sharing efforts could lessen the surveillance burden for state and local health departments and clinical personnel (*18*).

Methods

BioSense organizes incoming data into 11 syndromes that are indicative of the clinical presentations of critical biologic terrorism–associated conditions (Table 1). These syndromic categories and their associated codes are classified on the basis of definitions identified by multiagency working groups (*19,20*).

BioSense represents individual syndrome information in maps, graphs, and tables. Multiple options are available for customizing visual displays. Data can be displayed by using raw counts or transformed in multiple ways before graphing and mapping (e.g., normalized standard deviation, log, ratio, and proportional). Geographic extent can be defined as an entire state or metropolitan area, an individual or group of contiguous ZIP codes, or a cluster of ZIP codes based on the first three digits (ZIP3). Useful data transformations include standard deviation and day-of-week adjustments. A specific age group (e.g., children aged ≤3 years or persons aged ≥60 years) or sex also can be selected for all data sources. Historical data ranging from 1 to 365 days can be included in the visual displays. Within the tabular view, filtering is available to include or exclude key variables. All these customizations increase users' ability to detect data anomalies.

TABLE 1. Biosense syndromic categories

Syndrome	Biologic terrorism–associated condition
Botulism-like	Botulism
Fever	NA*
Gastrointestinal	Gastrointestinal anthrax
Hemorrhagic illness	Various viral hemorrhagic fevers
Localized subcutaneous lesions	Cutaneous anthrax, tularemia
Lymphadenitis	Bubonic plague
Neurological	NA
Rash	Smallpox
Respiratory	Inhalational anthrax, tularemia, pneumonic plague
Severe illness/ Death	NA
Specific infection	Represents specific illnesses

* Not applicable.

A data anomaly is a change in distribution or frequency in data compared with geographic or historic context. Change is quantified by using the elevated scores of two different algorithms. An adaptation of the CUSUM (*9*) is useful for identifying anomalies on a regional level, whereas SMART Scores (*21*) detect changes at the ZIP-code level. Changes are also identified by abrupt departures in the visual presentation of the data, including changes in the daily data volume or in the types or combinations of clinical diagnoses and procedures.

BioSense displays data in multiple presentations (e.g., algorithm scores, line graphs, maps, and tabular summaries, and detail) that can reveal data aberrations clearly. However the application does not draw conclusions for the user; human interpretation is required in using the technology to enhance the ability of analysts to detect and understand nuances in the data.

Function of Data Monitors

The primary function of CDC data monitors is to support broad use of the system among state and local public health partners. To this end, CDC monitors gather and provide feedback to improve the BioSense interface by troubleshooting problems, increasing user friendliness, and generating ideas for application enhancement. Monitors also conduct inquiries of data anomalies to better understand the causes of data anomalies and develop procedures to follow in accomplishing inquiries quickly. With experience, monitors have recognized repeat patterns and recommended changes in program logic to eliminate certain kinds of routine records demonstrated not to be related to a health event (e.g., vaccination-associated diagnoses and obvious miscoding). Finally, data monitors help refine and encourage the use of BioSense at the state and local level by providing training support.

Steps in Data Anomaly Inquiries

Because CDC BioSense monitors examine the interface presentation daily, they frequently detect anomalies that trigger an in-depth data inquiry (Figure 1). Each data anomaly is examined and analyzed to determine if it is of public health interest. For BioSense to be useful as an early-detection system, prompt inquires are necessary. Initial inquiries are usually completed in hours or days, depending on the circumstance and estimated degree of urgency. Urgency is estimated qualitatively by the number of records, the specific diagnostic or procedural codes, and the strength of the spatial pattern.

Inquiries of BioSense data anomalies take a predetermined pathway (Figure 2). Before a data anomaly is detected, the first step is to check the quality of the data. Data usually are loaded into BioSense within 2–4 days of the event, but the data occasionally take longer to arrive (e.g., on rare occasions, up to several weeks late). The initial page of the BioSense interface displays the estimated completeness of the data for the 5 previous days (Table 2). These estimations are available for each state or metropolitan area. If data receipts are considered complete for a given range of dates, the number of records must also constitute a sufficiently large numerator for analysis. If a data anomaly is detected at the ZIP-code level, sufficient records might not be available from which to draw conclusions.

FIGURE 1. Example of a data anomaly in BioSense

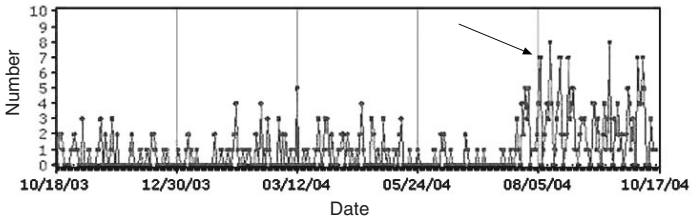


FIGURE 2. BioSense inquiry road map

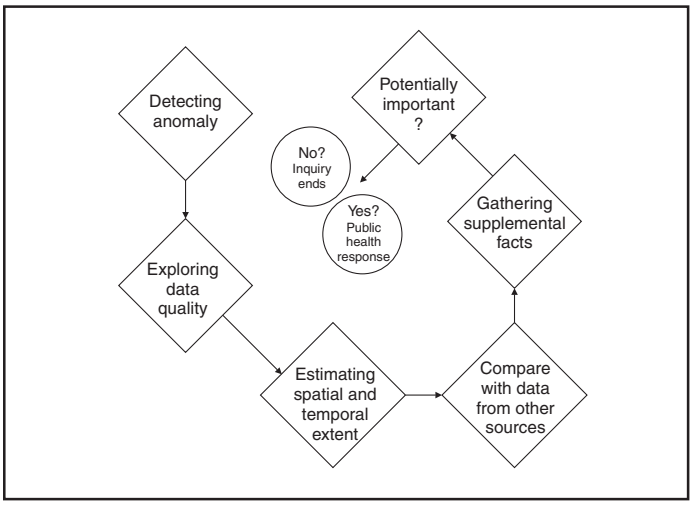


TABLE 2. Percentage of historic records received, by day of week, source, and type of record — BioSense application, October 27–31, 2004

	Wed 10/27	Thu 10/28	Fri 10/29	Sat 10/30	Yesterday Sun 10/31	
OTC*	96%	97%	94%	98%	3%	Store reports
DoD†	77%	69%	46%	3%		Encounters
VA§	101%	92%	32%	21%		Encounters

Sufficient for monitoring

Not sufficient for monitoring

* Over the counter.
† Department of Defense.
§ Department of Veterans Affairs.

The second step is to determine the geographic, temporal, and demographic extent of the data anomaly by using the different visualization features. Key questions to answer include the following:

- How widespread is the anomalous pattern?
- Are similar patterns found in adjacent regions?
- For how many days has the anomaly lasted?
- Has the geographic spread changed with time?
- Does the pattern have a day-of-week or cyclical nature?
- Did a similar pattern exist during the same period last year?
- Does the anomaly affect primarily one sex or age group?

The third step is to look for similar data anomalies in the other data sources. For example, if an anomaly is detected in the VA clinical data, monitors will assess whether a similar pattern exists in the DoD clinical data or in the LabCorp laboratory tests ordered. Although these data sources might represent different segments of the population, a correlation can exist between one data set and another. Within the BioSense data, additional information (e.g., other codes associated with visit) is available that is not viewable through the interface. BioSense monitors at CDC use SAS® EG (SAS Institute, Cary, North Carolina) to extract this information from the master files. Important information can be gathered by extracting all the available data associated with each individual visit. Monitors assess whether all visits associated with a given anomaly have other diagnostic codes in common. Although the majority of these codes are not associated with a given syndrome and are excluded from the interface, these codes are still useful when trying to decipher a detected anomaly. Displaying records by weekly or monthly totals, rather than daily, can make patterns more sharply discernable.

Once the BioSense data have been analyzed, the next step is to gather data from outside sources. State and local public health partners are a source of primary importance because they are familiar with the region in question, have numerous personal contacts, and often are the only persons who can rule out a data anomaly. Relevant information can be gleaned from online newspapers, public health bulletin boards (e.g., *Epi-X*[*] or ProMed[†]), and state/local public health department web sites. Weather information (e.g., temperature changes, pollen counts, ozone reports, cold fronts, and low-pressure zones) can be relevant to analyzing certain syndromes (e.g., respiratory).

The final step is to decide if a public health explanation exists for the data anomaly. If no public health explanation exists, the inquiry ends. If a public health explanation appears to exist, then state and local public health BioSense administrators are notified. Any further investigation, contact with DoD or VA clinics, or other follow-up response is determined by the state or local public health officials.

Results

During June–November 2004, of the approximately 160 data anomalies examined, no events involving disease outbreaks or deliberate exposure to a pathogen have been detected. Although monitors have contacted local and state BioSense administrators for their input on 12 occasions, no inquires have required a public health response. All inquiry results have been classified into one of eight categories (Table 3).

Sample Data Anomalies

The following section describes four in-depth inquiries in response to syndromic data anomalies in BioSense.

Example 1

Data Anomaly

In early September 2004, increased record counts associated with lymphadenitis syndrome were noted in DoD clinical data from one major metropolitan area. No similar increase was noted in local VA clinical data. On examination, 67 (72%) of 93 records during August 1–September 30, 2004, were for infectious mononucleosis (ICD-9-CM 075.0), all from the same DoD clinic. The visit/discharge disposition code information indicated that 37 (59%) of 63 patients with mononucleosis were sick enough to have work restrictions or be sent to bed.

Inquiry

At the end of the month, monitors compared monthly totals. Totals for August and September were 59 and 41, respectively, approximately 2.5 times higher than any of the previous 6 months (Figure 3). A clinician on location was contacted to obtain further information.

Findings

The clinician reported that the 59 records noted in BioSense for August actually represented 19 patients; 24 of the records were premature diagnoses whose diagnostic tests were later demonstrated to be negative. The actual number of laboratory-confirmed cases was 15 in August and four in September (B. Feighner, MD, Johns Hopkins Applied Physics

[*] *Epidemic Information Exchange* is a web-based communications network (available at http://www.cdc.gov/epix) that enables the secure exchange of information among authorized epidemiologists, laboratorians, and other public health professionals at CDC and state and local agencies.

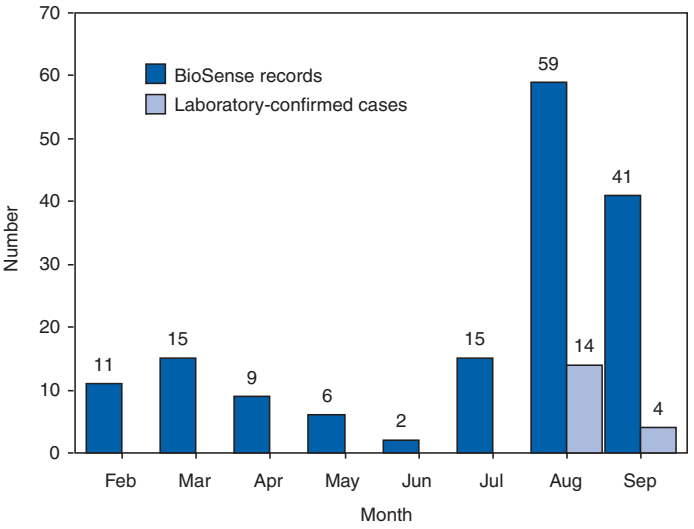
[†] Program for Monitoring Emerging Diseases, a global electronic reporting system for outbreaks of emerging infectious diseases and toxins (available at http://www.promedmail.org), is a program of the International Society for Infectious Diseases.

TABLE 3. Interpretation categories

Interpretation category	Examples
1 Predictable periodic trends	Seasonal Monthly Day-of-week Daily Weekly Clinic hours
2 Anomalies of limited duration	Associations with cold weather, high ozone, or pollen counts
3 Effects of data influxes	New clinics New retail store chains Temporary increase in military deployment
4 OTC* surges because of sales promotions	Merchandise sold at a loss to draw customers 2-for-1 sales
5 Duplicate or missing data	Data sent multiple times but not recognized by system as duplicates Days with missing data that make surrounding days appear as spikes
6 Vaccination-associated indications	Vaccination codes often paired with disease codes (diphtheria, anthrax, influenza, etc.)
7 Miscoded diagnoses	Congestive heart failure (CHF) coded as crimean congo hemorrhagic fever (CCHF) Prediagnostic determinations that are laboratory-confirmed negative later
8 Multiple follow-up visits by a few patients	Illnesses often require frequent follow-up visits, each appearing as a different record

* Over the counter.

FIGURE 3. Monthly count of mononucleosis cases at a single Department of Defense clinic, 2004



Laboratory, personal communication, 2004). The decreased count was attributable, in part, to multiple visits by patients for referral and follow-up treatment. Because DoD records do not include individual identifiers, the ability to distinguish between multiple patients and multiple visits by a limited number of patients is limited.

In addition, four patients coded as having mononucleosis were laboratory confirmed to be negative. Furthermore, August and September experienced an increase in the newly recruited population in this area; thus, the incidence for mononucleosis in this clinic was actually less than the same-type population incidence (0.15% versus 1.1%–4.8% for active-duty military personnel and college students annually) (B. Feighner, MD, Johns Hopkins Applied Physics Laboratory, personal communication, 2004). No escalation was required. Interaction with local public health providers was vital to gathering supplemental information because the circumstances would not have been understood without the information proved by the resident clinician.

Example 2

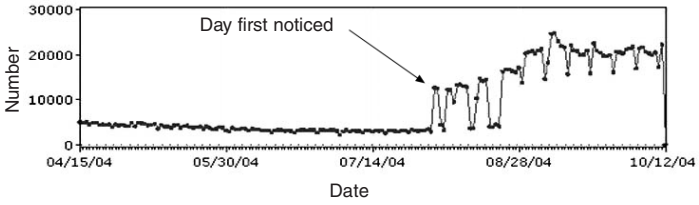
Data Anomaly

In early August, a sharp overnight increase was observed in OTC pharmacy sales volume for a single large state (Figure 4). Nearly all metropolitan areas within this state were affected similarly, and all OTC product classes had the same magnitude of change. No correlation was noted with DoD or VA data.

Inquiry

The OTC sales volume remained at elevated levels thereafter, and the data anomaly was not of limited duration. Within

FIGURE 4. Initial data anomaly for example 2: respiratory over-the-counter pharmacy sales for a single state



the state, the average number of units sold daily increased from approximately 3,100 in July 2004 to 11,500 in August and to 20,500 in September 2004. An examination of neighboring regions revealed that three adjacent states experienced similar surges in OTC sales volume.

A review of *Epi-X* and ProMed did not indicate any public health event that matched the scale of the anomaly. Monitors suspected that the anomaly reflected an influx of additional retail stores or pharmacy chains to the data-sharing program. An examination of the number of ZIP codes reporting data indicated that nationwide, the geographic range of localities reporting OTC sales increased by 2,156 new ZIP codes on a single day in early August 2004. A second increase of approximately 1,000 additional new ZIP codes nationwide occurred in early September 2004.

Findings

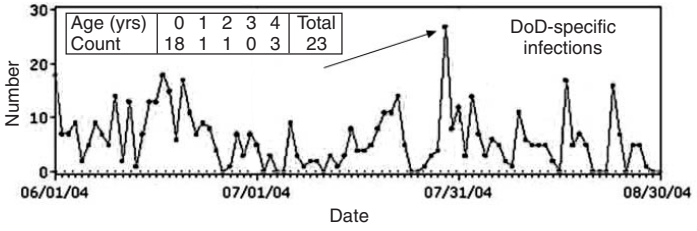
Although monitors could not confirm their suspicions, they hypothesized that the anomaly reflected the addition of new localities reporting data. One metropolitan area in that state previously had no pharmacies reporting. No response was required in this case because no public health threat was involved. This inquiry was completed in 5 business days.

Example 3

Data Anomaly

In late July 2004, monitors noticed an elevated number of records associated with the specific infections syndromic category (Figure 5). When the data on the tabular detail were examined, monitors identified 23 records with a diphtheria

FIGURE 5. Diphtheria-related records for a single clinic, July 29, 2004



diagnosis (ICD-9-CM code 032.9) on a single day at a single clinic. All the patients were children aged ≤4 years. By using visit/discharge disposition codes, monitors learned that all patients were released without any restrictions. If these records had really represented patients with a serious disease, these children would not have been released without restrictions.

Inquiry

No mention had been made of recent diphtheria cases in the United States in either *Epi-X* or ProMed. By examining the historic BioSense data, monitors identified 138 other visits associated with diphtheria in 2004, which was improbable when compared with nationwide data indicating that no laboratory-confirmed cases of diphtheria were reported in 2004 (*22*). SAS EG was used to extract all ICD-9-CM codes associated with these visits. In every instance, an associated ICD-9-CM code was identified that indicated that a DTaP vaccine had been received during the visit; certain visits also included CPT codes, indicating the visit had been a routine well-baby examination.

Findings

All of the diphtheria records represented children receiving routine childhood vaccinations. In response to this finding, BioSense now excludes all records associated with vaccinations.

Example 4

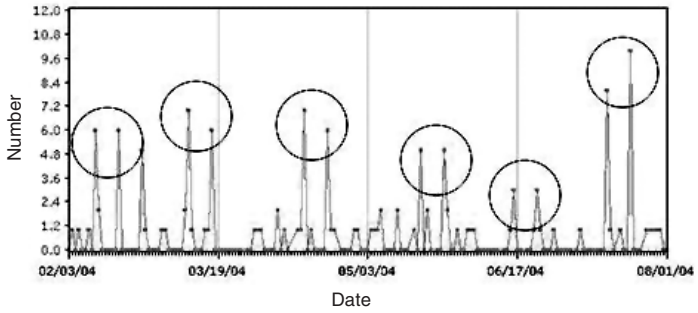
Data Anomaly

In mid-July 2004, monitors noticed five records associated with the severe illness/death syndrome. All patients were children aged 5–13 years who have been examined at the same DoD clinic on the same day. All had records with an ICD-9-CM code of 799.9, labeled in BioSense as "Mortality, cause unknown."

Inquiry

Monitors initially suspected that a violent event such as a house fire or motor-vehicle crash had occurred, but no mention was found of five children dying on the same day in the archived local or military newspapers. Monitors then noticed elevated levels of "child mortality" occurring on a regular basis at this particular clinic, all on Wednesdays during February–July 2004 (Figure 6). On multiple Wednesdays, the children matched in age and sex, and the visit/discharge disposition code of 1 indicated that these children had been released without any school or day limitations. Monitors also determined that these were not duplicate records; each record represented a unique visit.

FIGURE 6. Number of reported "child mortality" for a single Department of Defense clinic, February–July 2004



Findings

At this point, monitors contacted the local BioSense administrator, who contacted the DoD clinic directly. The administrator was informed that the repeating records represented the same cohort of children who had psychotherapy each Wednesday. The local administrator also reported that the correct description for ICD-9-CM 799.9 was "Mortality or Morbidity, cause unknown." This error was corrected immediately. No further escalation was necessary. Further logic strategies will be implemented to eliminate those visits that are psychotherapeutic in nature from the severe illness/death syndrome category.

This example underscores the importance of using information gleaned from other ICD-9-CM and CPT codes associated with the visit to clarify the nature of an individual event, especially when the key diagnostic code is vague, as is the case with 799.9. This case also demonstrates that the effectiveness of BioSense depends on the partnership of local, state, and federal public health agencies.

Conclusion

The BioSense Initiative represents a new paradigm for public health surveillance. BioSense makes secondary use of nationwide data sources, using data collected for other purposes aggregated without individual identifiers. An empirical learning process is required for BioSense users to understand how to make effective use of these data for public health purposes. Although the BioSense application was developed for use by local public health departments nationwide, CDC is responsible for developing an understanding of how to use it most effectively. To this end, CDC BioSense monitors conduct inquiries of data anomalies to rule out their potential threat to public health, provide feedback on their experiences for development of continued improvement to the user interface, perform a system troubleshooting function, generate and

collect input from local public health partners, develop new ideas for system enhancements, and most important, offer support and guidance to local users on the basis of their experience. Each CDC monitor has surveillance responsibility for a public health region but also works side by side with other monitors. Access to national scope information gives monitors the capability to coordinate information about adjacent regions. For example, in holding a telephone conversation with local public health officials, monitors have quick access to data in the surrounding regions to consider and characterize a potential event. As increasing use is made of additional secondary data sources, the learning requirements for their complementary use will continue to evolve. For new surveillance capabilities to be improved, procedures for investigating observed anomalies should be shared, information and knowledge compared, and automated means to account for data artifacts developed.

Acknowledgments

The findings in this report are based, in part, on contributions by Matt Miller, Chaital Patel, Duane Zomer, IEM Inc., Atlanta, Georgia; Shirley Willson, Northrop Grumman, Atlanta, Georgia; Yukun Jin, Business Computer Applications, Inc., Atlanta, Georgia; Barbara Wasely, Steve Bloom, Nick August, and Colleen Bradley, Science Applications International Corporation, Atlanta, Georgia; and Howard Burkom, Brian Feigner, and Martin Sikes, Johns Hopkins Applied Physics Laboratory, Laurel, Maryland.

References

1. Loonsk J. BioSense—a national initiative for early detection and quantification of public health emergencies. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):53–5.
2. Wagner M, Robinson J, Tsui F-C, Espino J, Hogan W. Design of a national retail data monitor for public health surveillance. J Amer Med Inform Assoc 2003;10:409–18.
3. Bravata DM, Rahman MM, Luong N, Divan HA, Cody HA. A comparison of syndromic incidence data collected by triage nurses in Santa Clara County with regional infectious disease data [Abstract]. J Urban Health 2003:80:i122.
4. Cho S, Hackmyer S, Li J, et al. Establishing an automated surveillance system [Abstract]. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):50–2.
5. Corwin A, McCarthy M, Larasati W, et al. Developing regional outbreak response capabilities: Early Warning Outbreak Recognition System (EWORS). Navy Med 2000;(September–October):1–5.
6. Dembek ZF, Carley K, Siniscalchi A, Hadler J. Hospital admissions syndromic surveillance—Connecticut, September 2001–November 2003. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):50–2.
7. Doyle TJ, Bryan RT. Infectious disease morbidity in the US region bordering Mexico, 1990–1998. J Infect Dis 2000;182:1503–10.
8. Heffernan R, Mostashari F, Das D, et al. New York City syndromic surveillance system. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):25–7.
9. Hutwager L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80:i89–i96.
10. Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity of bioterrorism and influenza surveillance. Ann Emerg Med 2003; 41:447–52.
11. Lombardo J, Burkom H, Elber 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:i32–i42.
12. Paladini M. Daily emergency department surveillance system—Bergen County, New Jersey. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl): 47–9.
13. Tsui FC, Espino JU, Wagner MM, et al. Data, network, and application: technical description of the Utah RODS Winter Olympics Biosurveillance System. Proc AMIA Symp 2002:815–9.
14. Wagner M, Tsui F-C, Espino J, et al. National retail data monitor for public health surveillance. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):40–2.
15. Yih WK. Caldwell B, Harmon R, et al. National Bioterrorism Syndromic Surveillance Demonstration Program. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004; 53(Suppl):43–6.
16. Yuan CM, Love S, Wilson M. Syndromic surveillance at hospital emergency departments—Southeastern Virginia. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):56–8.
17. Zelicoff A, Brillman J, Forslund DW, et al. The rapid syndrome validation project. Albuquerque, New Mexico: Sandia National Laboratories; 2001.
18. Schable CA. CDC bioterrorism preparedness efforts: protecting the public health. Testimony before the Committee on Science, United States House of Representatives; May 3, 2004.
19. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Available at http://www.cdc.gov/epo/dphsi/syndromic/framework.htm.
20. Ma H. Lab order data implementation in BioSense early outbreak detection system. In: Syndromic surveillance: reports from a national conference, 2004. MMWR 2005;27–30.
21. 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.
22. CDC. Table I. Summary of provisional cases of selected notifiable diseases, United States, cumulative, week ending December 24, 2004 (50th Week). MMWR 2004;53:1185.

Syndromic Surveillance on the Epidemiologist's Desktop: Making Sense of Much Data

Kathy J. Hurt-Mullen,[1] J. Coberly[2]
[1]*Montgomery County Department of Health and Human Services, Silver Spring, Maryland;*
[2]*Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland*

Corresponding author: Kathy J. Hurt-Mullen, Montgomery County Department of Health and Human Services, 2000 Dennis Avenue, Silver Spring, MD 20902. Telephone: 240-777-1643; Fax: 240-777-4750; E-mail: Kathy.Hurt-Mullen@montgomerycountymd.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Syndromic surveillance systems are becoming increasingly common in health departments. These systems represent a substantial improvement in the timeliness of ascertainment of community health status. For the value of such systems to be realized, protocols are needed for review and analysis of the findings that these systems produce.*

Methods: *A workgroup of experienced syndromic surveillance users and developers was convened to discuss approaches to data review and analyses. The discussion was structured to include general principles of the use of syndromic surveillance; how and why specific data are reviewed; integration of multiple data sources; daily versus research uses of systems; how data anomalies are identified by users and surveillance systems; the relative merits of anomalies; how a data anomaly is investigated to determine if it warrants a public health response; and how such a public health response should be framed.*

Results: *From this discussion, a generalized and more detailed process was documented that describes the common elements of analysis used by the workgroup participants.*

Conclusion: *Establishment of a framework for evaluation and response to syndromic surveillance data will facilitate the implementation of these systems and standardization of procedures for validation of system findings. Careful development of an evaluation and response framework should be undertaken to assess whether use of syndromic surveillance systems requires excess work to distinguish between statistical anomalies and important public health events.*

Introduction

Although initially conceived to assist public health officials in detecting occurrence of intentional disease outbreaks (i.e., those caused by a biologic terrorist attack), syndromic surveillance systems are becoming a basic tool for public health epidemiologists. The majority of these systems employ multiple data streams (including data from hospital emergency departments [EDs] or other emergency encounters, physician office visits, over-the-counter [OTC] pharmaceutical sales, and school absenteeism records) to detect potential disease clusters in the community. The increased sensitivity provided by multiple data sources requires users to review and summarize an unprecedented amount of data daily. Clear guidelines for using these systems are needed to help epidemiologists 1) quickly identify and disregard statistically significant but epidemiologically unimportant events, 2) distinguish true disease clusters from groups of unrelated cases, 3) determine which true disease clusters warrant further evaluation or public health response, and 4) perform these tasks quickly and cost effectively. The Montgomery County (Maryland) Department of Health and Human Services (MCDHHS) has been using the Electronic Surveillance System for Early Notification of Community-Based Epidemics (ESSENCE) syndromic surveillance system continuously since spring 2001. This report describes a framework for daily evaluation of ESSENCE data that was developed jointly by staff from MCDHHS and the Johns Hopkins University Applied Physics Laboratory (JHU/APL). This framework can be generalized for use with other electronic syndromic surveillance systems.

Methods

The framework for using the ESSENCE system presented in this report is based on the experience of the authors gained through daily use of the MCDHHS ESSENCE system and on a structured discussion with 10 syndromic surveillance systems users and persons with expertise that was designed to collect qualitative information on how public health profes-

sionals use ESSENCE. Participants included representatives of state health agencies in Maryland, Virginia, and the District of Columbia; civilian and military users of ESSENCE; and members of the JHU/APL ESSENCE development team. Topics discussed included general principles of the use of syndromic surveillance; how and why specific data are reviewed; integration of multiple data sources; daily versus research uses of the systems; how data anomalies are identified by users and by the surveillance systems; the relative merits of anomalies; how a data anomaly is investigated to determine if it warrants a public health response; and how such a public health response should be framed. Comments on how and why syndromic surveillance is used were also provided by representatives of the New York City Department of Health and Mental Hygiene after a review of the meeting notes. Syndromic surveillance systems such as ESSENCE are intended to identify higher-than-expected counts of visits to EDs or physicians' offices, retail sales of pharmaceutical products, or other similar events grouped into broad syndromic categories. These increases are assumed to represent increases in disease incidence.

Syndromic surveillance systems can identify certain health events that are of sufficient concern that a single occurrence warrants a public health response (e.g., the collection of additional data when certain rashes appear among persons in particular age groups or with certain neurologic complaints). Those procedures and public health responses are not considered in this report.

Results

Use of Syndromic Surveillance Data

Epidemiologists use syndromic surveillance systems for multiple purposes. Because these systems collect and store longitudinal disease incidence data, epidemiologists can use them to trace disease patterns over time, describe patterns of disease in the community geographically and demographically at any given time, and determine the impact of specific targeted health interventions. These systems most commonly are used for early detection of changes in a community's health status that might represent a public health emergency. A protocol is followed to detect and analyze the importance of anomalies in the data (Figure 1).

Anomaly Detection

All syndromic surveillance systems use a statistical algorithm to determine whether the number of reports for a specific syndrome exceeds the norm for the community and then to alert the user that a statistically significant increase has occurred. However, important health events can be detected in other ways. Syndromic surveillance systems enable epidemiologists to systematically monitor disease trends and identify suspicious clusters of disease. Epidemiologists also receive information about unusual disease clusters or incidence from local health-care providers, which they then can evaluate in the surveillance system. Both system- and operator-generated alerts are considered to be of equal value.

Anomaly detection presents jurisdictional challenges that should be accommodated in developing response guidelines. Although public health officials have obligations associated with defined geographic boundaries, neither the movement of persons nor the distribution of illness is so limited. Anomaly detection should be conducted at multiple levels (local, regional, state, and possibly federal) to capture anomalies that might remain undetected within any single jurisdiction.

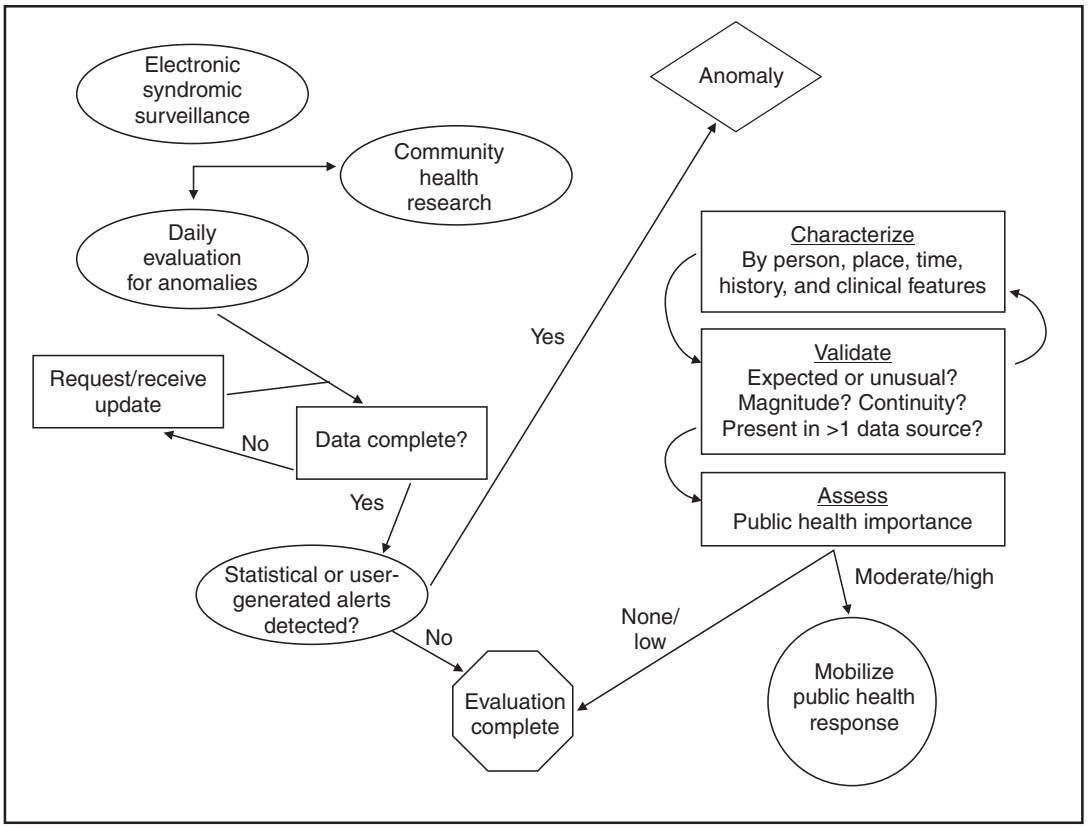
Characterization of Anomalies

Anomalies should be described fully to determine whether they are likely to represent an important public health event that requires a public health response. The anomaly should be described with respect to person, place, and time by using whatever demographic and geographic information is available. In addition, available clinical data (e.g., chief complaint and discharge diagnosis and disposition) should be evaluated and summarized. In certain cases, additional information (e.g., laboratory test requests or results or detailed ED admission data) might be available to further characterize the anomaly.

Validation of Anomaly

Once the epidemiologist is convinced that an anomaly represents a true cluster of similar health events, whether the anomaly is expected or unusual should be determined. Normal seasonal and temporal syndrome and disease trends should be reviewed. Influenza-like illness is easily spotted in the majority of syndromic surveillance systems. An increase in disease causes regular statistical alerts in established syndrome groups as illness spreads through the community. Because this increase is expected each winter, the anomaly does not require further evaluation efforts. Similarly, environmental factors should be considered. For example, seasonal increases in pollen generate increases in respiratory illness, which might cause statistical alerts in surveillance systems even though these events are expected. If more than one source of data is available within a system, then part of the validation effort should be to ascertain whether corroboration is expected in those sources and whether it exists.

FIGURE 1. Theoretical framework for response protocols in use of syndromic surveillance systems



Assessing Public Health Importance of Findings

Once an anomaly is fully characterized, its public health importance should be considered. First, the magnitude and continuity of the increase generating the anomaly should be evaluated in the context of the particular syndrome group in question. Regardless of statistical significance, a substantial 1-day increase warrants more scrutiny than a limited one; similarly, a relatively modest increase during multiple days that deviates from known seasonal and historic patterns also should be evaluated closely. In each of these instances, the size of the actual increase is characterized by the nature of known patterns of the data source and syndrome being evaluated; these considerations require an understanding of the usual frequency distribution for the particular event of concern.

Certain signals can be expected and, when detected, are of less concern, especially when the public health response is well established (e.g., the beginning of the influenza season and winter increases in cases of viral gastroenteritis). However, observations of such anomalies at other times of the year, or when frequency is much different than expected or presentations more severe, are more likely to represent important public health events.

Other Factors of Importance

The majority of syndromic surveillance systems collect multiple streams of data to be evaluated as indicators of changes in the health status of the populations they monitor. Each data stream has unique attributes that must be understood to estimate the value of the results they produce. Chief among these attributes is the lag time between the occurrence of an event and the time it is available to the system for anomaly detection. Another critical attribute is the clinical value of the data. ED data, for example, convey important information about clinical encounters when chief complaint data are coded into syndromes. However, ED data are more clinically valid when the syndromes are coded on the basis of discharge diagnoses, as physician office visit data are coded. In determining

the attributes of a system with these data sources, the user should be aware of the timing of the coding efforts and the change in lag time introduced by the coding effort. Data on OTC pharmaceutical sales can be available quickly but have less clinical value than encounter data sources. Whether the data source provides individual level data to allow scrutiny at that level to characterize and validate any changes detected also should be considered.

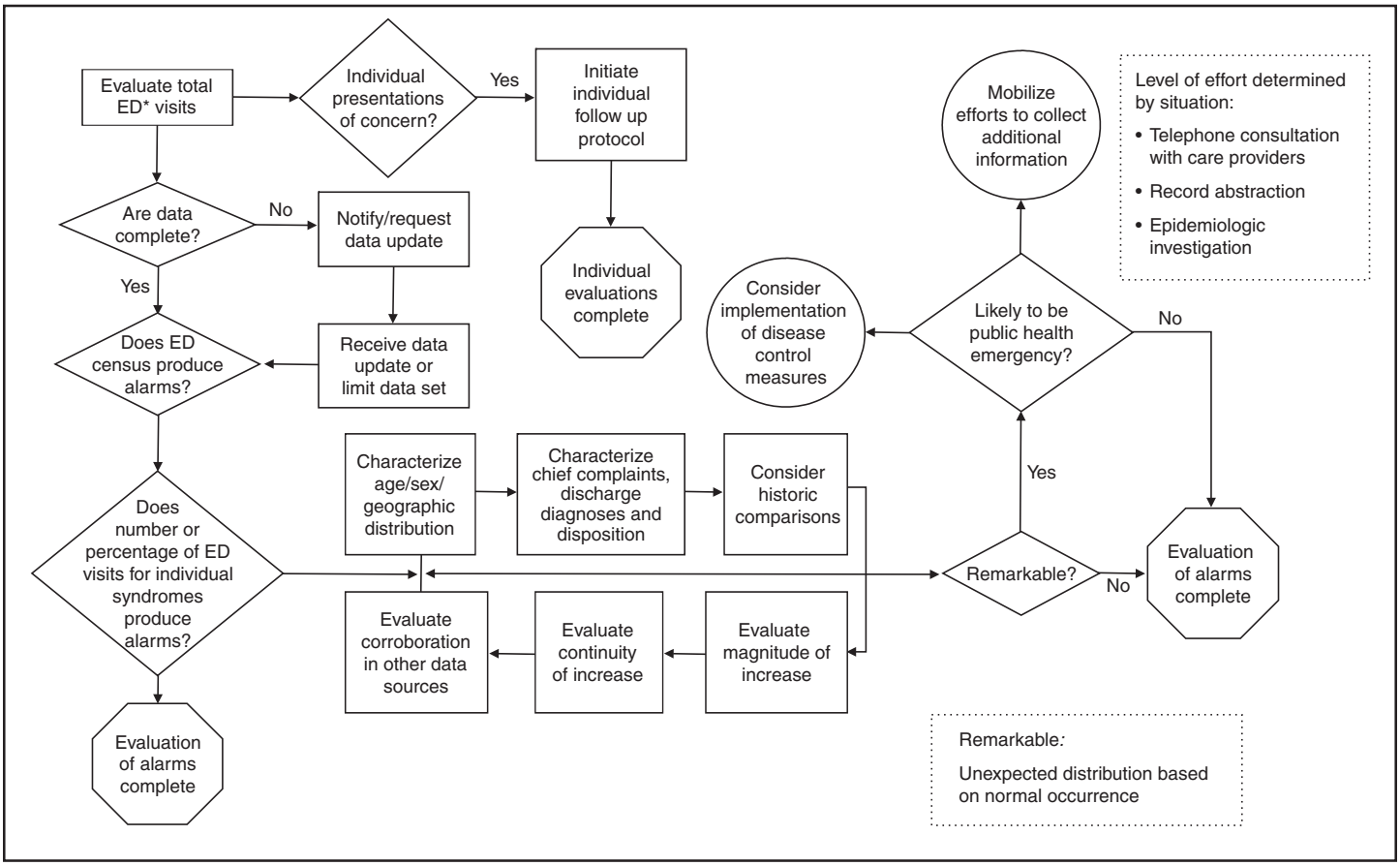
On the basis of these attributes, identifying a single data stream within the system to serve as the principal indicator of changes in community health status might be desirable; remaining data sets can be relied on as secondary sets that are used to corroborate findings in the primary stream or to assist in refining hypotheses when indications of important changes are observed in the primary data set. A process flow chart (Figure 2) can be used to illustrate how evaluation and response might proceed with a system that has designated a primary, individual encounter–based data source.

Discussion

Syndromic surveillance is emerging as a practical tool for public health epidemiologists. Procedures for evaluation of data provided to public health practitioners through syndromic surveillance systems will necessarily vary based on the system in place and the jurisdiction in which it operates. However, to facilitate identification and response to important public health events, users believe that standardized evaluation and response strategies should be developed and adopted by all public health practitioners.

A step-by-step evaluation strategy (Figure 2) can be used for evaluation of any individual-encounter based surveillance data source (i.e., ED chief complaints, hospital discharge codes, and call-center encounters). The process should begin with an evaluation of data completeness and with efforts to either complete the data set or limit the analytical data set to one which is as complete as possible. Proceeding then from general evaluations of incidence and aberration patterns to more

FIGURE 2. Evaluation of data from syndromic surveillance systems



* Emergency department.

specific ones (e.g., from the all-encounter level to age-group syndrome evaluations), all anomalies should be analyzed to identify usual and expected patterns for consistency of presenting complaints; age, sex, and geographic-distribution; and severity of illness as indicated by discharge dispositions and diagnoses (when data are available). When possible, historic comparisons (e.g., with the same period in previous years or seasons) should be evaluated. Finally, consideration should be made of whether an unusual but modest change in patterns has persisted for longer than can be explained or whether an increase is simply too large to dismiss as a random occurrence.

These information-gathering tasks are the same as those conducted in the early stages of conventional outbreak investigations. Although this framework is intended to assist public health practitioners in distinguishing between statistical anomalies and anomalies of public health concern, the work required parallels that of a typical outbreak investigation. Both processes have the same objective: to determine whether the appearance of an outbreak is, in fact, an outbreak (Table).

The majority of the tasks in the proposed evaluation framework overlap, which is often the case in outbreak investigation as well. Ideally, the surveillance system will allow for the routine and automated collection of the data required for completion of these tasks (e.g., clarification of clinical complaints, determination of existence of the anomaly, creation of a case definition, scrutiny for other similar cases, and descriptive epidemiology work). This will serve to ease both the burden and costs of the efforts. The majority of anomalies will require no more consideration than can be made systematically by using information routinely available to users through the system. As such, the distinction between evaluation and response becomes less clear, and response can be thought of as something that, in most cases, can be determined easily from the epidemiologist's work station. The evaluation of the data is, in effect, a limited investigative response that either can support the need to proceed to field work or indicate that the anomaly is not important in a public health sense. If the results of this descriptive work fail to rule out the existence of an event of public health concern, syndromic surveillance systems can provide data to support epidemiologic studies (e.g., case-control studies) to evaluate findings further. Depending on the implementation of particular systems, cases and controls can be distinguished from one another with respect to time of the event, distribution of particular chief complaint keywords, or discharge diagnoses within interesting clusters.

The move to field work also requires a considered, step-by-step approach. The level of effort required depends on the particular features of the anomaly. The response could be as simple as a telephone consultation with a health-care provider or a review of ED record face sheets. Rarely, a response might require the full effort of an outbreak investigation.

Evaluation work should be conducted by staff members who have sufficient experience with the data to be familiar with fluctuations in incidence attributable to common variation and who have training to support their interpretation of the statistics employed for anomaly detection. Furthermore, evaluation work should be assigned to staff members who understand the demographic features of the community, its habits, and current activities or events because these factors can influence the appearance of anomalies in the absence of a real shift in the health status of the community. Often, responsibility for community-level evaluations should be assigned to health officials in local health departments, and responsibility for region- or statewide evaluations should be assigned to officials in those jurisdictions. Evaluation and response strategies need not be materially different at these levels, but communication of findings and required response should be communicated and monitored across all levels.

Conclusion

Establishment of a framework for evaluation and response to syndromic surveillance data will facilitate the implementation of these systems and standardization of procedures for validation of system findings. The framework presented in this report was developed on the basis of the experience of a substantial number of users of multiple systems and may be generalized for use in other systems and jurisdictions. Evaluating data at the local, regional, and state levels might ensure

TABLE. Comparison of outbreak investigation and syndromic surveillance evaluation tasks

Outbreak investigation task	Syndromic surveillance response framework task
Establish existence of anomaly	Evaluate data quality and completeness; evaluate corroboration of other data sources; evaluate magnitude and continuity of increase; make historical comparisons
Verify the diagnosis	Clarify the chief complaints and diagnoses causing the anomaly
Construct a working case definition	Clarify the chief complaints and diagnoses causing the anomaly; evaluate corroboration of other data sources
Find all related cases; develop line list	Evaluate continuity and magnitude of increase
Complete descriptive epidemiology	Characterize age, sex, and geographic distribution

that officials with local awareness will evaluate changes detected in the context of current events and demographics and that increases that occur across jurisdictional boundaries are detected and coordinated by appropriate officials. Finally, careful development of an evaluation and response framework should be undertaken to assess whether use of syndromic surveillance systems requires excess work to distinguish between statistical anomalies and important public health events.

Acknowledgments

The findings in this report are based in part on contributions by the following ESSENCE users in the National Capital Region: Sheryl Happel Lewis, Rekha Holtry, Steve Babin, and other JHU/APL staff. Financial support for portions of this work was provided by the National Association of City and County Health Officials Advanced Practice Center program.

Connecting Health Departments and Providers: Syndromic Surveillance's Last Mile

James B. Daniel,[1] D. Heisey-Grove,[1] P. Gadam,[1] W. Yih,[2] K. Mandl,[3] A. DeMaria, Jr.,[1] R. Platt[2,4]
[1]*Bureau of Communicable Disease Control, Massachusetts Department of Public Health, Boston, Massachusetts;*
[2]*Harvard Pilgrim Health Care and Harvard Medical School, Boston, Massachusetts;* [3]*Children's Hospital, Boston, Massachusetts;*
[4]*Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts*

Corresponding author: James B. Daniel, Massachusetts Department of Public Health, State Laboratory Institute, 305 South Street, Jamaica Plain, MA 02130. Telephone: 617-983-6808; Fax: 617-983-6840; E-mail: James.Daniel@state.ma.us.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *A critical need exists for mechanisms to identify and report acute illness clusters to health departments. The Massachusetts Department of Public Health (MDPH) works with partner organizations to conduct syndromic surveillance. This effort is based on CDC's Health Alert Network program and includes automated generation and notification of signals and a mechanism to obtain detailed clinical information when needed.*

Methods: *Syndromic surveillance partners collect emergency department and ambulatory care data. The principal communications platform between syndromic surveillance partners and MDPH is the Massachusetts Homeland and Health Alert Network (HHAN). This Internet-based application serves as a portal for communication and collaboration and alerts predefined groups of users involved in emergency response. Syndromic surveillance partners' systems report to HHAN by using Public Health Information Network Messaging System events that meet thresholds selected by MDPH. Cluster summaries are automatically posted into a document library. HHAN notifies users by electronic mail, alphanumeric pager, facsimile, or voice communications; users decide how they want to be notified for each level of alert. Discussion threads permit real-time communication among all parties.*

Results: *This automated alert system became operational in July 2004. During July–December 2004, HHAN facilitated communication and streamlined investigation of 15 alerts.*

Conclusion: *The system allows rapid, efficient alerting and bidirectional communication among public health and private-sector partners and might be applicable to other public health agencies.*

Introduction

A critical need exists for mechanisms to report acute illness clusters and for public health personnel to obtain timely clinical information about persons who are part of these clusters. Timely identification of newly emerging pathogens and syndromes, as well as unusual clusters of illness, is difficult when health departments rely solely on traditional methods of surveillance (e.g., laboratory reporting). Identifying illnesses early facilitates treatment, prevention, and control of disease. Syndromic surveillance systems access and analyze data sources that are not normally accessible to departments of public health (e.g., symptoms and signs of illness captured by *International Classification of Diseases, Ninth Revision* [ICD-9] codes and chief-complaint data). As a result, syndromic surveillance might detect unusual clusters of illness before definitive diagnoses are made and thus potentially earlier than traditional disease reporting allows.

In Massachusetts, two syndromic surveillance systems are in place, one capturing visit data from ambulatory-care settings and one using chief-complaint data from hospital emergency departments (EDs). Both systems were established in partnership with the Massachusetts Department of Public Health (MDPH). The Harvard Pilgrim Health Care/Harvard Vanguard Medical Associates (HPHC/HVMA) system collects ambulatory care data from an electronic medical record system at 14 clinic sites in eastern Massachusetts (*1–4*). The Children's Hospital Boston system (AEGIS) utilizes chief-complaint data from eight Massachusetts hospital EDs (*5,6*). In both systems, new visits are grouped into syndromes defined previously by a CDC-led working group (*7*) and aggregated by ZIP code. Statistical models are used to assess whether each day's syndrome counts are unusual.

The first challenge to response to syndromic surveillance signals is to establish bidirectional communication among

partners. MDPH uses the Massachusetts Health and Homeland Alert Network (HHAN) to automate communication of alerts. HHAN is an Internet-based application that serves as the principal platform permitting communication among Massachusetts' syndromic surveillance partners. This report describes the experience gained and outlines remaining challenges.

Methods

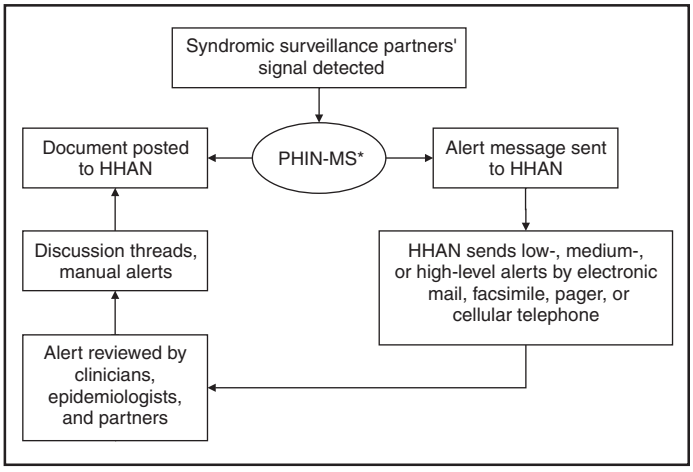
Data on HPHC/HVMA patient encounters (i.e., visits or calls), including demographic information and diagnostic codes, are recorded electronically as part of routine patient care, usually on the same day. Every 24 hours, encounters with codes of interest are extracted automatically from the clinical data system. The data are deidentified and aggregated by syndrome and ZIP code of residence; the resulting aggregate counts of illness are automatically uploaded to a data-coordinating center. During processing of the daily data file, a line list of the day's encounters is generated and kept at HPHC/HVMA. This list contains demographic information and the text of the diagnostic codes assigned during each encounter and allows a first-level epidemiologic assessment (short of consulting the full medical record). These methods have been described previously (*1*–*4*).

The Small Area Regression and Testing (SMART) scores method (*8*) is used. This method uses generalized linear mixed models that adjust for day of the week, holidays, seasonal patterns, and any secular trends on the basis of historic data to determine the degree of statistical aberration associated with each date–syndrome–ZIP code count. The signal detection is automated and occurs at the data coordinating center. If the number of cases of a syndrome detected in a particular ZIP code on a particular date is higher than expected, an automatic alert is generated and sent to designated recipients at MDPH through HHAN by means of an interface between the data-coordinating center and HHAN. MDPH created three alert levels (low, medium, and high), corresponding to recurrence interval (i.e., the number of days of surveillance that would normally elapse between the chance occurrence of counts as unusual as the one observed) thresholds. MDPH uses recurrence-interval thresholds of 2 months (low), 6 months (medium), and 2 years (high), except for respiratory syndrome, for which the chosen thresholds are 6 months, 1 year, and 2 years, respectively. Syndrome-specific alert levels are defined by MDPH staff and can be changed readily; for example, the alert threshold can be lowered during periods of heightened concern (e.g., during a political convention).

HHAN functions as a secure collaboration portal that allows role-based alerting and access to a document library (Figure). Once a syndromic surveillance partner's system detects a signal, a document is posted in the library on HHAN through the Public Health Information Network Messaging System (PHIN-MS). Simultaneously, an automatic alert message is sent from PHIN-MS to HHAN, which then uses built-in functionality to distribute the alert further by electronic mail, facsimile, alphanumeric pager, or voice (e.g., cellular telephone). Each user decides how to be notified for each alert level (low, medium, or high). Simultaneously, the system automatically posts a document providing details about the cluster that generated the alert. Once alerts and documents have been sent, bidirectional communication between MDPH and clinical responders is facilitated by using discussion threads associated with the alert document in question.

After the HPHC/HVMA system generates an alert, MDPH staff contact a designated clinical responder on call. Responders have been trained for this purpose and are available at all times according to an established schedule. Within the line list for the day in question, the clinician reviews the cases responsible for the alert, and, if a cluster of illness of public health importance is suspected (e.g., a substantial cluster within a single ZIP code of lower gastrointestinal illness that includes family members), MDPH staff proceed with further investigation. A response protocol guides MDPH epidemiologists in contacting on-call responders to obtain further details about cases contributing to suspect clusters of illness. Before full implementation of the HHAN alert system, this response protocol was pilot tested with clinical responders from HPHC/HVMA.

FIGURE. Protocol followed by Massachusetts Homeland and Health Alert Network (HHAN) to generate and respond to a syndromic surveillance alert



* Public Health Information Network Messaging System.

Results

Pilot testing of HPHC/HVMA clinical response began in June 2004, and the automated alert system became operational in July 2004, in time for use during the Democratic National Convention held in Boston July 26–29. During July–December 2004, HHAN received 15 alerts from the HVMA/HPHC system. Two alerts required investigation beyond consultation of line lists (e.g., chart review), but no public health intervention was necessary.

Case Scenario #1: Medium-Level Neurologic Alert

In September 2004, MDPH received an alert through HHAN about 23 persons who reported symptoms consistent with a neurologic syndrome. These 23 cases occurred in multiple areas with ZIP codes beginning with 021 (HVMA/HPHC population: 55,866 persons; U.S. Census population: 1,183,247 persons). The estimated recurrence interval was 405 days. Review of the line list determined that 20 (86.9%) of 23 patients had headaches. The clinical responder reviewed the patient medical records and determined that the patients' clinical presentations did not suggest a genuine cluster. The clinician posted the results of the chart review in the discussion thread, allowing MDPH to halt the investigation.

Case Scenario #2: Medium-Level Lower Gastrointestinal Illness Alert

In October 2004, MDPH received an alert through HHAN about 37 persons (age range: <1–>80 years) who reported lower gastrointestinal illness. The cases occurred in multiple areas with ZIP codes beginning with 021 (HVMA/HPHC population: 55,866 persons; U.S. Census population: 1,183,247 persons). The estimated recurrence interval was 296 days. MDPH staff contacted the clinical responder, who posted a deidentified line list of cases in the discussion thread of the alert. Review of the ICD-9 codes associated with the visit indicated multiple symptoms and diagnoses, including abdominal pain, diarrhea, gastroenteritis, and *Clostridium difficile*. In the discussion thread, the clinical responder noted that among the 18 towns included in the alert, symptoms generally did not appear to be similar within each town. In one town, four patients had abdominal pain, but the characteristics of the pain varied. Epidemiologic review of the line-list information ruled out the need for further clinical response, and MDPH closed the investigation.

Case Scenario #3: High-Level Respiratory Alert

In September 2004, MDPH received a high-level respiratory alert through HHAN involving five cases in a single ZIP code (HVMA/HPHC population: 56 persons; U.S. Census population: 7,480 persons). The estimated recurrence interval was >200 years. The clinical responder posted the deidentified line list in the discussion thread and noted that two of the five patients were members of the same family and that three patients had asthma. A second clinical responder viewed the discussion thread and concurred with this assessment. Epidemiologic review of the line-list information ruled out the need for further clinical response, and MDPH closed the investigation.

Discussion

The HHAN alert system meets multiple needs. It allows routine, timely, and automated aggregation of clinical information from a large, defined population; identification of unusual clusters of illness; and communication about these events to designated health department epidemiologists. It also establishes a repository of limited clinical information (line lists) about each case that contributes to a cluster and creates a formal protocol for obtaining additional information about cases at any time from a clinician in the delivery system who has access to full-text medical records.

In implementing this system, MDPH recognized a need for a more timely response to HHAN alerts. To address this critical need, MDPH epidemiologists need access to the limited line-list information about cases that contribute to clusters. In each of the scenarios described in this report, initial clinical response time for review of line lists was approximately 2–4 hours, whereas epidemiologists were available to review HHAN alert information within 30 minutes. To address this delay in response time, the system is being revised so that deidentified line lists of cases will be provided automatically to HHAN along with the alert. This will allow MDPH staff to perform an initial epidemiologic evaluation of the cluster without contacting the clinical responder. The line list will include a unique identifier, age range, sex, ZIP code, or town of residence, ICD-9 codes, and a family identifier (to identify multiple cases occurring in the same household). Receiving the information in this way will reduce the need to involve clinical responders so responders can focus on performing chart reviews and conferring with health department personnel about additional follow-up (e.g., contacting patients or advising clinicians to be alert for additional cases). Only the first scenario described

in this report required medical record review; epidemiologic review of the line-list information in the other two scenarios described ruled out the need for further clinical response.

Although the HVMA/HPHC catchments encompass densely populated urban areas, implementing this system in rural areas might present a problem with regard to line lists being fully deidentified. Submitting patient-identifiable information only when the health department and clinical responders agree that a cluster is potentially of public health concern minimizes the total amount of personal health-care information provided routinely. Avoiding routine transfer of information for health-care encounters that are not part of clusters helps the delivery system assure patients about the confidentiality of their health-care data. In addition, data are protected by a secure web portal and by MDPH privacy and confidentiality standards.

Once a potential cluster of public health concern is identified, obtaining a comprehensive line list that includes identifiable patient information remains problematic. Lists cannot be posted on HHAN because HHAN does not have a second-tier identification mechanism; second-tier authentication (e.g., a secure token or digital certificate) would give HHAN an additional layer of security and permit posting of identifiable information. MDPH is also exploring using PHIN-MS to securely transfer these data to a proposed Internet-based disease reporting system.

Given the investments that have been made and the effort involved in responding to alerts, the still-unanswered question of the usefulness of syndromic surveillance to public health can and should be addressed even before the issues discussed in this report are fully resolved. As part of a multistate effort, MDPH and its partners are undertaking an evaluation of the sensitivity, predictive value-positive, timeliness, and cost-benefit of these alerts and establishing databases of alerts and outbreaks that include nontraditional data elements (e.g., person-time spent on investigation and interventions and assessment of costs and benefits of receiving and responding to each alert).

Conclusion

The HHAN alert system allows rapid, efficient alerting and bidirectional communication among public health and private-sector partners. Automatic generation of alerts saves time because epidemiologists do not have to manually review data each day to define clusters. Issues identified in the implementation of the system include the need to generate and make accessible to public health signal-specific line lists, a problem that is being addressed. Future MDPH plans include building the ability to send alerts to local boards of health to gain local public health participation earlier in the investigation process.

This experience with HHAN might be applicable to other public health agencies, including those with access to syndromic surveillance data. Evaluation of the effectiveness and utility of this surveillance and reporting system in improving public health is needed and is currently under way.

Acknowledgment

This project was funded in part by two grants from CDC cooperative agreements (UR8/CCU115079 and U90/CCU116997).

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. BioMed Central 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. 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(Suppl 1):i25–31.
4. Yih WK, Caldwell B, Harmon R, et al. National Bioterrorism Syndromic Surveillance Demonstration Program. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):43–9.
5. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak 2003;3:2.
6. 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.
7. 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.
8. 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 Epid 2004;159:217–24.

Comparison of Syndromic Surveillance and a Sentinel Provider System in Detecting an Influenza Outbreak — Denver, Colorado, 2003

Debra P. Ritzwoller,[1] K. Kleinman,[2] T. Palen,[1] A. Abrams,[2] J. Kaferly,[1] W. Yih,[2] R. Platt[2,3]
[1]*Kaiser Permanente Colorado, Denver, Colorado;* [2]*Harvard Pilgrim Health Care and Harvard Medical School, Boston, Massachusetts;*
[3]*Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts*

Corresponding author: Debra P. Ritzwoller, Clinical Research Unit, Kaiser Permanente Colorado, 580 Mohawk Dr., Boulder, CO 80301. Telephone: 303-554-5045; Fax: 303-554-5043; E-mail: debra.ritzwoller@kp.org.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Syndromic surveillance systems can be useful in detecting naturally occurring illness.*

Objectives: *Syndromic surveillance performance was assessed to identify an early and severe influenza A outbreak in Denver in 2003.*

Methods: *During October 1, 2003–January 31, 2004, syndromic surveillance signals generated for detecting clusters of influenza-like illness (ILI) were compared with ILI activity identified through a sentinel provider system and with reports of laboratory-confirmed influenza. The syndromic surveillance and sentinel provider systems identified ILI activity based on ambulatory-care visits to Kaiser Permanente Colorado. The syndromic surveillance system counted a visit as ILI if the provider recorded any in a list of 30 respiratory diagnoses plus fever. The sentinel provider system required the provider to select "influenza" or "ILI."*

Results: *Laboratory-confirmed influenza cases, syndromic surveillance ILI episodes, and sentinel provider reports of patient visits for ILI all increased substantially during the week ending November 8, 2003. A greater absolute increase in syndromic surveillance episodes was observed than in sentinel provider reports, suggesting that sentinel clinicians failed to code certain cases of influenza. During the week ending December 6, when reports of laboratory-confirmed cases peaked, the number of sentinel provider reports exceeded the number of syndromic surveillance episodes, possibly because clinicians diagnosed influenza without documenting fever.*

Conclusion: *Syndromic surveillance performed as well as the sentinel provider system, particularly when clinicians were advised to be alert to influenza, suggesting that syndromic surveillance can be useful for detecting clusters of respiratory illness in various settings.*

Introduction

The 2003–04 influenza season in the Denver metropolitan area began earlier, was more severe than in recent years, and included reports of pediatric mortality (*1*). Influenza outbreaks occur each year, but uncertainty exists with respect to the timing and severity of these outbreaks. Effective surveillance is critical for tracking the spread and severity of disease and for determining the types and subtypes of viruses that circulate during the influenza season.

Most public health organizations that monitor influenza use the U.S. Influenza Sentinel Provider Surveillance Network, a collaborative effort between CDC, state and local health departments, and health-care providers (*2*). This system monitors influenza activity in the general population. Traditionally, the sentinel provider surveillance system operates from October to mid-May each year. Each week, sentinel providers report the total number of patient visits during the preceding week and the total number of patient visits for influenza-like illness (ILI) (*2*), stratified by age categories.

In recent years, substantial investments have been made in syndromic surveillance systems. These systems allow rapid detection of natural infectious disease clusters and of intentional acts of terrorism (*3–5*). Previous studies have demonstrated that syndromic surveillance can be useful in detecting ILI (*3,6,7*). Investment in these systems might enhance public health organizations' ability to identify and react to infectious disease outbreaks. This report compares the dates on which a syndromic surveillance system and a Sentinel Provider Network, both in a single health-care delivery system, identified unusual ILI activity associated with the onset of the

fall 2003 influenza outbreak in Denver, as determined by reported laboratory-confirmed cases of influenza.

Methods

A retrospective comparison of data collected by three ILI detection systems (i.e., two ambulatory care–based and one laboratory-based) was conducted during fall 2003. The two ambulatory-care surveillance systems were situated in Kaiser Permanente Colorado (KPCO), a closed panel, group-model health maintenance organization serving approximately 380,000 members in the Denver-metropolitan area.

Laboratory-Based Surveillance

During the 2003–04 influenza season, laboratories in Colorado reported positive influenza tests to the Colorado Department of Public Health and Environment (CDPHE), either via the Colorado Electronic Disease Reporting System (CEDRS) or by fax or telephone; test results were displayed graphically by week of report on CDPHE's website (*8*). Weekly electronic newsletters that included the county specific counts of laboratory-confirmed influenza cases were generated and distributed to providers and public health officials. Laboratory-confirmed cases were from the seven-county Denver-metropolitan area, consistent with KPCO's service area; however, test results were obtained by clinicians outside and within KPCO*. Laboratory confirmation of influenza cases was based on direct fluorescent antibody and viral culture results. Information about other viruses in the community was obtained from The Children's Hospital (TCH) in Denver, Colorado, which published counts of confirmed cases of respiratory syncytial virus (RSV), adenovirus, parainfluenza, rhinovirus, and pertussis (*9*). The source of these laboratory specimens was pediatric patients who sought medical care at the TCH emergency department with respiratory illness during October–May.

Syndromic Surveillance System

The CDC-sponsored National Bioterrorism Syndromic Surveillance Demonstration Program, in which KPCO participates, has been described previously (*4,7,9–11*). This syndromic surveillance system is based on diagnostic codes entered in patients' electronic medical records (EMR) by providers during the routine delivery of ambulatory care. Diagnostic codes are mapped to 13 syndromes. To be counted as a case of ILI, the encounter must have at least one of a set of respiratory illness codes and have measured fever of at least 100°F (37.8°C) in the temperature field. If no value is present in that field, an *International Classification of Diseases, Ninth Revision* (ICD-9) primary, secondary, or tertiary code of fever (code 780.6) must be provided. These data are extracted daily, and counts by ZIP code are reported on a secure website in both a graphical and a map format and on a daily basis. Signals of unusual clusters of ILI are identified by three statistical models: small area method, spatio-temporal method, and purely temporal method. These models are estimated daily, and signals are reported based on pre-determined thresholds.

The small area method has been described previously (*12*). The historical series of counts in each small area are used to create a regression estimate of the count to be expected in each area on each day, adjusting for seasonal, weekly, and secular trends, as well as holiday effects. The results are used to create p-values for statistical significance. These estimates are then corrected to account for multiple ZIP codes and used to create recurrence intervals (RIs). RIs are defined as the number of surveillance days required to expect a count as unusual as the one observed to occur exactly once by chance. This method, also called the SMART (Small Area Regression and Testing) scores method, is advantageous in that large values imply more unusual results, and the multiple tests are adjusted for in the same step. For this analysis, ZIP codes were used as the small areas.

The spatio-temporal method is a space-time scan statistic, implemented by using the public domain SaTScan software (*13*). Day of the week, holidays, season, secular trends, and the unique characteristics of each ZIP code area (e.g., the health-seeking behavior of the population) were adjusted for by using the results of the regression needed in the SMART scores method (*14,15*). The space-time scan statistic searches all circular areas incorporating one or more ZIP codes for the most unusual grouping, as measured by a likelihood ratio statistic. A p-value is calculated by using Monte Carlo methods (*16*). The maximum geographic size was set at 50% of the adjusted population at risk (*14*) and the temporal length at 3 days.

The temporal method implemented the space-time scan statistic by also using SaTScan, but required the area to include 100% of the surveillance area, effectively removing the spatial aspect of the test. In all other respects, it was identical to the space-time scan statistic used. For the space-time and temporal scan statistics, the RI was calculated, although no correction for multiple comparisons was required, and the RI was the inverse of the p-value.

The full-text medical records of patients with ILI who were counted as part of signals in September and October with RI >30 days were reviewed by a clinician to assess mention (present, absent, not mentioned) of ILI clinical characteris-

* Counties include Adams, Arapahoe, Boulder, Broomfield, Denver, Douglas, and Jefferson.

tics. These characteristics included headache, myalgia, or muscle aches; malaise; cough; sore throat; ocular pain; photophobia; dyspnea; and/or fever.

Sentinel Provider System

The sentinel provider system is overseen by CDPHE and is part of the CDC funded Colorado Influenza Surveillance Project (*2*). This ongoing project recruits providers each season to report weekly ILI activity from early-October through mid-May. In 2003, KPCO submitted ILI data from its EMR for approximately 250 primary care providers. The surveillance system relies on a report of the total number of patients evaluated and the number of those patients with ILI stratified by age group and reported weekly. From these data, the percentage of patient visits for ILI is calculated. During the 2003–04 influenza season, KPCO used an EMR system that employed a controlled-medical-terminology vocabulary from SnoMed (*17*) for the documentation of diagnoses. A patient visit was reported to the sentinel provider system as ILI if the physician actively selected either of the SnoMed terms "influenza or influenza-like illness" within the diagnosis section of the patient's EMR. Data were extracted weekly, based on the specific SnoMed terms selected rather than on the specific ICD-9 codes. SnoMed terms were extracted from the patient's electronic chart for analysis because in KPCO's data warehouse both "influenza-like illness" and "influenza" were mapped only to ICD-9 code 487.1, which is "influenza, not otherwise specified." In addition to reporting the percentage of all visits for ILI, this analysis stratified data by the specialty of the provider (i.e., pediatrics, family practice, internal medicine, and urgent care).

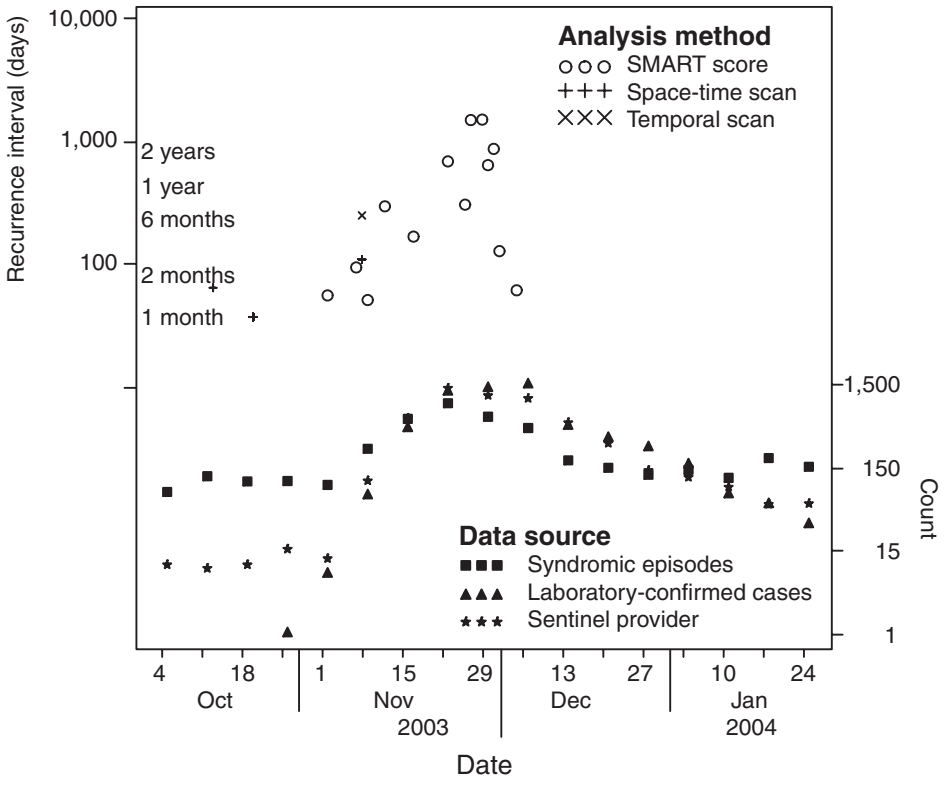
Although many visits captured by one KPCO provider-based system were also captured by the other, the two sets were not completely overlapping. The sentinel provider system did not explicitly require the patient to meet the fever criterion of the syndromic surveillance system. In addition, the syndromic surveillance ILI system could capture visits for which the provider did not assign the influenza or ILI diagnosis, either because the provider believed the cause was not influenza or because the provider simply chose a different diagnosis, such as cough, upper respiratory infection, or pneumonia.

Results

A single, positive laboratory-confirmed influenza case was reported before October 25, 2003, in Denver. On November 10, CDPHE reported a substantial increase in reported laboratory-confirmed cases, from seven during the week ending November 1 to 69 during the week ending November 8. A total of 447 laboratory-confirmed cases were reported during the week ending November 15; cases peaked at 1,504 during the week ending December 6 (Figure). During that week, TCH reported <15 positive tests each for RSV, parainfluenza, rhinoviruses, or pertussis.

When the daily syndromic surveillance data were aggregated into comparable weekly units, the number of episodes that met the syndromic surveillance definition for ILI exceeded the number of sentinel provider reports until the week ending

FIGURE. Recurrence intervals from daily syndromic surveillance conducted at Kaiser Permanente Colorado using three statistical alogirthms, weekly counts of syndromic surveillance and a sentinel provider system both conducted at Kaiser Permanente Colorado, and laboratory-confirmed influenza cases — Denver, Colorado, October 2003–January 2004



November 15. ILI episodes identified by the syndromic system increased from 89 (week ending November 1) to 242 (week ending November 8) and to 556 (week ending November 15). For the sentinel system, identified cases increased from 11 (week ending November 1) to 100 (week ending November 8) and to 567 (week ending November 15). For both the syndromic and sentinel systems, the number of episodes and cases peaked the week ending November 22, with 859 episodes and 1,304 cases, respectively (Figure).

The number of new clinical episodes identified in the ambulatory-care setting that met the syndromic surveillance system definition of ILI are illustrated, as well as signals with RIs >30 days for the three signal detection algorithms (Figure). Signals of at least this magnitude occurred on 1 day in September and 2 days in October; during this 2-month period, six events would have been expected by chance (1 per method per 30 days). Signals of at least this magnitude were then observed on November 1 and every day during November 6–December 1. RIs exceeded 10,000 (expected to occur by chance no more often than once in 27 years) on every day during November 8–30.

Among the three signal detection algorithms, the SMART score generated RIs of >30 days on 3 days during November 1–8. The SMART scores also generated RIs of 200–9,000 on 8 days and RIs of ≥10,000 on 9 days during November 8–December 6. Both the spatio-temporal and purely temporal SaTScan generated RIs of >30 days beginning November 7, with RIs of >10,000 days for 23 days during November 8–December 6.

Manual review of the medical records for the 20 patients who were part of the September and October ILI signals generated by either SMART scores or SaTScan indicated that all 20 patients had fever (100°F [37.8°C]), 15 had cough or nasal congestion, and 17 had cough, nasal congestion, or sore throat. Chart abstraction also indicated that more than half of the patients associated with the early signals (i.e., 12 of 20 patients in signals before November 1, 2003, had cough and fever, the combination most predictive of influenza (*18*). Two or more of the four signs that are most commonly reported among patients with confirmed influenza A diagnoses and consistent with CDC's case definition (i.e., fever, cough, nasal congestion, and sore throat), were reported in 15 of 20 patients (*19–21*).

The average weekly sentinel surveillance data from visits from all KPCO primary care providers first indicated an increase in ILI above 1% of visits during the week ending November 15. During the dominant weeks of the outbreak, weeks ending November 8–December 6, substantial variation was observed in the sentinel provider group by primary care specialty. During the week of November 22, the Pediatrics department providers assigned an ILI diagnosis for >8% of visits, compared with <2% of visits in the Internal Medicine department.

Discussion

Although the ambulatory-care–based syndromic surveillance system described in this report was designed principally to detect terrorism events, it would have automatically generated ILI alerts at the same time a meaningful number of laboratory-confirmed cases were reported. The sentinel provider also demonstrated increased activity. The syndromic surveillance and sentinel provider systems shared important features. Both obviated the need for clinicians to participate directly in reporting because the information recorded as part of routine documentation of clinical encounters was extracted from an EMR. For this reason, this is a best-case implementation of sentinel provider surveillance. A difference between the syndromic and sentinel systems is that the syndromic surveillance system is more standardized it does not require clinicians to make an explicit diagnosis of influenza or ILI. Instead, influenza can be defined by signs and symptoms the provider might not recognize or code as influenza. Because of this, clinicians were not reminded to use the influenza codes; outreach to clinicians was a prominent feature of the sentinel provider program. This difference would be particularly important for surveillance of conditions that are not expected at a particular season or that are not readily recognized by clinicians; examples include the early phases of many terrorism-related illnesses or severe acute respiratory syndrome.

Another difference is that the sentinel provider system did not explicitly require the patient to meet the fever criterion of the syndromic surveillance system. In addition, the syndromic surveillance ILI system could capture visits for which the provider did not assign the influenza or ILI diagnosis, either because the provider believed the cause was not influenza or because the provider simply chose a different diagnosis, such as cough, upper respiratory infection, or pneumonia.

The findings of this report suggest that clinicians' likelihood of choosing an "influenza" diagnosis might have been subject to external information about the presence of influenza in the community (i.e., media reports and public health alerts). During the week ending November 8, the number of ILI episodes identified by KPCO's syndromic surveillance system increased by 153, whereas the number of cases identified by the sentinel provider system increased by 89. Given the increase in reported laboratory-confirmed cases and the lack of evidence that other respiratory viruses were circulating in

the metropolitan area during that week, a substantial fraction of the difference represents cases that were missed by the sentinel provider system. In contrast, after an outbreak was recognized, the number of sentinel provider reported cases (1,304) substantially exceeded the number of new syndromic surveillance episodes (859). This is also the week that the Infectious Disease Department sent reminders to KPCO primary care providers asking them to use the KPCO specific ILI coding terms for suspected influenza cases. Most of this difference in the counts of new episodes versus cases (445) is likely to be a result of clinicians' assigning an influenza diagnosis without documenting fever.

The findings also demonstrate the utility of the signal detection algorithms that were used to analyze the syndromic surveillance data. They provided unequivocal signals, despite the syndrome definition being nonspecific, as evidenced by the baseline rate of nearly 100 new episodes per week before influenza became widespread in the community. These methods might also be useful for detecting unusual clusters of other endemic infectious diseases, despite being designed to ignore typical seasonal increases in ILI episodes.

Theoretical considerations suggest that the spatio-temporal approach has the best combination of sensitivity and specificity for detecting events that occur in more than one adjoining small area (e.g., more than one ZIP code that is under surveillance), whereas a purely temporal approach is best when the events are scattered throughout all regions under surveillance. The size of the 2003 influenza outbreak overwhelmed these theoretical differences among the algorithms, and all three—SMART score, spatio-temporal, and purely temporal SaTScan—provided strong signals that coincided with the increase in laboratory-confirmed cases of influenza.

All of the signal detection algorithms used in the syndromic surveillance adjusted for typical seasonal fluctuations in illness, including ILI, to be able to detect a terrorism event against a background of normal patterns of morbidity. Because the influenza season arrived earlier than usual facilitated its identification in November. If identification of seasonal respiratory illness is a goal of such a syndromic surveillance system, it will be necessary to develop signal detection algorithms that are optimized for this purpose.

Conclusion

Automated syndromic surveillance identified unusual ILI activity early, as did the traditional sentinel provider surveillance and reported laboratory-confirmed influenza cases, despite being designed to detect terrorism rather than natural outbreaks of diseases such as influenza. The syndromic surveillance system's ability to use uniform criteria for case identification might be an advantage in situations in which clinicians are not alerted to the potential presence of a problem. Because the syndromic surveillance system is a passive system, this might limit bias in the data collection that might be associated with external factors such as media reports and public health alerts in contrast to sentinel provider recognition. In addition, the three different signal detection algorithms used by the syndromic surveillance system proved useful and might have broader applicability for surveillance of other infectious diseases.

Acknowledgments

This study was funded by CDC cooperative agreement UR8/CCU115079. The findings in this report are based, in part, on contributions by Ken Gershman and Barbara Stone from the Colorado Department of Public Health and Environment. Additional assistance was also provided by Ray Bem, Mike Bodily, Capp Luckett, and Elizabeth Newsom from the Kaiser Permanente Colorado Clinical Research Unit, and Adam Jackson from Kaiser Permanente Department of Infectious Disease, Denver, Colorado.

References

1. CDC. Update: influenza activity—United States and worldwide, 2003–04 season, and composition of the 2004–05 influenza vaccine. MMWR 2004;53:547–52.
2. CDC. Preliminary assessment of the effectiveness of the 2003–04 inactivated influenza vaccine—Colorado, December 2003. MMWR 2004;53:8–11.
3. 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.
4. 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.
5. Bravata DM, McDonald KM, Smith WM, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Intern Med 2004;140:910–22.
6. Miller B, Kassenborg H, Dunsmuir W, et al. Syndromic surveillance for influenza like illness in ambulatory care network. Emerg Infect Dis 2004;10:1806–11.
7. 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. BioMed Central Public Health 2001;1:1–9.
8. Colorado Department of Public Health and Environment. Summary of 2003–04 influenza season, reported cases of laboratory confirmed influenza. Available at http://www.cdphe.state.co.us/dc/Influenza/lab_chart_03_04.pdf.
9. The Children's Hospital Departments of Epidemiology and Pathology. Contagious comments 'bug watch'. Available at http://www.thechildrenshospital.org/pro/publications/bug.pdf.
10. 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(Suppl 1):i25–i31.

11. Yih WK, Caldwell B, Harmon R, et al. The National Bioterrorism Syndromic Surveillance Demonstration Program. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004; 53(Suppl):43–6.
12. 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 Epid 2004;159:217–24.
14. Kulldorff M; Information Management Services Inc. SaTScan v4.0: software for the spatial and space-time scan statistics. Available at http://www.satscan.org.
15. Kleinman K, Abrams A, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. Epid Infect 2005; In press.
16. Kulldorff M. Prospective time periodic geographic disease surveillance using a scan statistic. J Royal Stat Soc [Ser A] 2001;164:61–72.
17. SNOWMED International. SNOWMED clinical terms. Available at http://www.snomed.org.
18. Dwass M. Modified randomization tests for nonparametric hypotheses. Ann Math Stat 1957;28:181–7.
19. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. Arch Intern Med 2000;160:3243–7.
20. Besag PJ, Newell J. The detection of cluster/signals in rare diseases. J R Stat Soc [Ser A] 1991;154:143–55.
21. Boivin G, Hardy I, Tellier G, Maziade J. Predicting influenza infections during epidemics with use of a clinical case definition. Clin Infect Dis 2000;31:1166–9.

Ambulatory-Care Diagnoses as Potential Indicators of Outbreaks of Gastrointestinal Illness — Minnesota

W. Katherine Yih,[1] A. Abrams,[1] R. Danila,[2] K. Green,[2] K. Kleinman,[1] M. Kulldorff,[1] B. Miller,[2] J. Nordin,[3] R. Platt[1]
[1]*Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, Massachusetts;*
[2]*Minnesota Department of Health, Minneapolis, Minnesota;* [3]*HealthPartners, Minneapolis, Minnesota*

Corresponding author: W. Katherine Yih, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Ave., 6th floor, Boston, MA 02215. Telephone: 617-509-9822; Fax: 617-859-8112; E-mail: katherine_yih@hphc.org.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *Syndromic surveillance's capability to augment existing surveillance for community-acquired gastrointestinal disease is unknown.*

Objective: *The objective of this study was to evaluate the capability of a syndromic surveillance system to detect outbreaks of gastrointestinal disease.*

Methods: *A retrospective analysis was conducted comparing ambulatory care data from a health plan with a set of 110 gastrointestinal-disease outbreaks identified by the Minnesota Department of Health during 2001–2002. Unusual clusters of illness (i.e., signals) in the health-plan data were identified by analyzing daily counts of gastrointestinal illness using an adjusted space-time scan statistic. Concordance was defined as ≤5 km between outbreak and signal and the signal occurring within 1 week of the outbreak.*

Results: *During 104 weeks, the number of signals was roughly what would have been expected by chance, suggesting that the modeling did a good job of estimating the expected counts of illness and that false alarms would not have occurred much more often than the number predicted at the various thresholds. During the same period, the health department identified 110 eligible gastrointestinal outbreaks. Apparent associations of the three statistically most unusual concordant signals with outbreaks of viral or bacterial gastrointestinal illness were ruled out by the health department on the basis of detailed knowledge of the circumstances and low numbers of affected persons seeking medical care.*

Conclusion: *No previously known gastrointestinal outbreaks were identified by this surveillance system. However, relatively few recognized outbreaks resulted in patients seeking medical care, and the sensitivity of this system to detect outbreaks of real significance to public health remains to be determined. Prospective evaluation probably will be required to understand the usefulness of syndromic surveillance systems to enhance existing disease surveillance.*

Introduction

The increasing availability of electronic health data and interest in biologic terrorism preparedness have accelerated the development of new public health-related disease surveillance systems, including systems intended to provide early detection of unusual clusters of illness before the etiology of the cases is known (*1–4*). This approach is sometimes referred to as syndromic surveillance. These systems could possibly augment public health departments' traditional systems of surveillance for naturally occurring disease, both through early detection and by providing better overall understanding of illness patterns. However, their usefulness for this purpose has not been convincingly demonstrated (*5*), and their performance needs to be evaluated (*6*), particularly in light of the resources required to establish and maintain these new surveillance systems and to respond to false or uninteresting alarms they might generate.

Ultimately, these systems will be evaluated on the basis of their ability to provide useful information to guide public health practice. One way to assess this ability is to compare historic information from novel surveillance systems to actual morbidity reported to public health departments. In this report, information about unusual clusters of gastrointestinal illness obtained in the ambulatory-care setting was compared with information about outbreaks that a state health department obtained through routine disease surveillance. Gastrointestinal illnesses were selected because these are the most common discrete outbreaks that health departments investigate.

Methods

Surveillance System Overview

The surveillance system tested here monitors and analyzes daily counts of new cases of illness assigned by primary-care providers during the routine delivery of care. All participating health-care organizations use electronic medical records, allowing near real-time extraction of diagnosis information in a manner that is transparent to clinicians. This system has been described previously (*7–10*). The system includes data from health-care organizations in Colorado (Denver metropolitan area), eastern Massachusetts (Boston metropolitan area), Minnesota (Minneapolis-St. Paul metropolitan area), and central Texas (Austin area), and information from a nurse telephone triage company that operates in all 50 states. Together, these systems include approximately 20 million persons. Typically, health-care organizations identify encounters with *International Classification of Diseases, Ninth Revision*-coded diagnoses of interest that have occurred during the preceding day. They use open-source software to assign these encounters to syndromes defined by the CDC–Department of Defense working group (*11*) to identify and ignore repeat visits and to assign new episodes of illness to the ZIP code where the affected persons live.

Health Plan Data Source and Processing

This report uses information about gastrointestinal illnesses in Minnesota, where the state health department provided data on a substantial number of recognized outbreaks for comparison. Ambulatory-care information during February 2001–January 2003 originated from HealthPartners, a 240,000-member health plan that services approximately 8% of the population in the Minneapolis-St. Paul metropolitan area in ZIP codes beginning with 550, 551, 553, 554, or 563.

Signals of unusual clusters of gastrointestinal illness were identified by analyzing the daily counts of gastrointestinal illness with a space-time scan statistic (*12*), adjusting for day of the week, holidays, season, secular trends, and the unique characteristics of each ZIP code area (e.g., its population's health-seeking behavior) by means of a generalized linear mixed model (GLMM) (*13,14*) on the basis of past data. The space-time scan statistic is a likelihood-ratio test statistic that can detect clustering of cases in space and time. The "window" through which it "looks" can be visualized as a cylinder of variable size that moves across both space (with its circular base) and time (with its height). The maximum geographic size of the window was set at the size covering 25% of the adjusted population at risk (*14*) and the temporal length at 1 day.

To reflect the degree to which characteristics of clusters (i.e., size for syndrome, location, and date) deviated from the expected, an index of statistical aberration called the "recurrence interval" was used (*13–15*). This recurrence interval is the expected number of days of surveillance needed for one such cluster of at least the observed magnitude to occur in the absence of any actual outbreaks and is the inverse of the nominal p value from the space-time scan statistic. Therefore, the larger the recurrence interval, the more unusual the cluster of illness. Because the maximum possible p value is 1, the minimum possible recurrence interval is 1 day, meaning one would expect clusters of that magnitude every day.

The period analyzed was February 1, 2001–January 31, 2003. The statistical evaluation emulated a prospective surveillance system with data collection beginning on January 1, 2001. Statistical evaluation started on February 1, 2001, using data from January 2001 to calculate the expected counts. For every succeeding month, the expected counts were recalculated for each ZIP code area with at least one case as of then, incorporating the additional month's worth of data. The analyses in March 2001 used data from January–February 2001, the analyses in April 2001 used data from January–March 2001, and so on until the final month in the observation period (January 2003), whose analyses used data from January 2001–December 2002. The statistical evaluation probably improved over time.

Information on Outbreaks From Health Departments, Exclusion Criteria

Information about gastrointestinal illness outbreaks from 2001 through early 2003 was provided by the Minnesota Department of Health. These outbreaks had been detected by the department's foodborne and other surveillance systems and have been described separately (*16*). Most of these outbreaks involved foodborne transmission. Information included the ZIP code of the outbreak, the number of recognized cases, the date of presumed first exposure to the pathogen, the date of first onset of illness, the date the illness or outbreak was reported, and the date on which the health department initiated its investigation.

From a list of 206, excluded were 1) outbreaks occurring outside of the catchment area of HealthPartners (i.e., not occurring in ZIP code areas beginning with 550, 551, 553, 554, or 563); 2) all institutional outbreaks (e.g., those in universities, prisons, and long-term care facilities) because affected persons would not have obtained care from HealthPartners;

and 3) those events accompanied in the database by a comment of "no investigation" or "not an outbreak." This remaining list included 110 eligible outbreaks.

When specific dates were unavailable, other dates were substituted for purposes of the comparison with the HealthPartners data: three outbreaks lacked first exposure date, so date of first onset was substituted; seven outbreaks lacked the date the investigation was initiated, so the report date was used for three and the date of first onset of symptoms was used for the remaining four. All of these substitutions would have reduced the chances of a finding a match with syndromic surveillance signals because they all shortened the time-window for comparison.

Comparison of Syndromic Surveillance Signals and Known Outbreaks

The correspondence between outbreaks of gastrointestinal illness reported by the health department and signals detected in the ambulatory-care data were examined. Signals were defined as clusters with recurrence intervals greater than or equal to each of six threshold values ranging from 2 weeks through 2 years. The date and ZIP code areas of outbreaks identified by public health authorities were compared with those signals detected in the ambulatory-care data. Signals and known outbreaks were considered concordant if they satisfied both geographic and temporal proximity requirements. The geographic requirement stipulated that the closest point in the significant ambulatory-care signal area and the centroid of the outbreak ZIP code area be within 5 km of each other. This distance was selected to allow some chance that affected persons, effectively located at the centroid of their ZIP code of residence, would be linked to an outbreak placed at the centroid of the ZIP code where exposure was thought to occur (e.g., a restaurant). The timing requirement, set *a priori*, was that the signal in the ambulatory-care data occur in the period from 1 week before the first known exposure of a case in the outbreak determined by the health department to 1 week after the investigation was initiated. The rationale for this interval was that some outbreaks might have started before the first exposure known to the health department and might have continued for at least several days beyond the start of the health department's investigation. This temporal criterion might have been overly generous. For example, in point-source outbreaks where first-exposure dates were known with certainty, a more appropriate approach might have been to eliminate from consideration any putative matching signals occurring during the week before the first exposure.

If a signal was concordant with more than one outbreak, only its association with the geographically closest outbreak was kept. All the substitutions for missing exposure or investigation dates would have reduced the potential for concordance by decreasing the eligible period scanned and would have reduced the timeliness of any concordant signal by using an earlier comparison date in lieu of the investigation start-date.

Statistical Analysis of "Hits"

To determine whether more hits (signals concordant with outbreaks) occurred than would be expected by chance in the analysis of the Minnesota outbreaks, a permutation-based test (*17*) was used to ascertain the distribution of the number of hits to be expected by chance alone. The null hypothesis in such an approach is that no relation exists between the outbreaks found by the health department and the signals found by the syndromic surveillance system, except by chance co-occurrence. Under the null hypothesis, it was not assumed that the signals would be either evenly distributed across the map (because of variations in density of the population at risk) or evenly distributed over time (because of known seasonal patterns in illness). Therefore, the null hypothesis is that the space-time scan statistic signals occurred randomly (not evenly) in space-time, conditioned on the purely spatial and purely temporal empirical distributions. The alternative hypothesis is that signals occur in close proximity to the known outbreaks in both space and time.

The temporal and spatial components of the health-plan–based syndromic surveillance signals were randomly permuted (i.e., each random data set included the same collection of signal days and the same collection of signal locations and sizes, but the pairing of the temporal and spatial attributes was randomized or "permuted"). Therefore, the temporal and spatial elements were no longer associated; however, the two spatial attributes, location and size, always stayed together. For each of 999 randomly permuted data sets, the number of hits with the health department outbreaks was calculated in exactly the same way as for the real signals, and then all 1,000 numbers were ordered from both the real and random data sets. If hits were random, then the rank of the number of hits from the real data would be equally likely to be any number from one to 1,000, and the rank of the number of hits from the real data would be in the top 5% (or *x*%) 5% (or *x*%) of the time. This provides a p value for the null hypothesis that the observed number of concordant signals arose solely by chance, defined as p = rank / (number of random datasets + 1) (*18*). The test is not meaningful where the number of signals is <5.

Results

Syndromic Surveillance Signals

During two calendar years covering approximately 1.3 million person-years, one gastrointestinal syndrome signal was identified with a recurrence interval of at least 2 years and 58 signals with a recurrence interval of at least 2 weeks (Table 1). The number of signals was only marginally more than what would have been expected by chance and not statistically significantly more than expected for any of the thresholds (one-sided test, p = 0.22 for RI = 2 weeks, p>0.05 for all), raising the question of whether any of these signals reflected a true outbreak. The median number of health-plan cases in these signals was four to nine; the median radius ranged from 0 km (with one ZIP code area) to 10 km (with a median of 15 ZIP code areas).

Known Outbreaks

During the same 2-year period, the health department identified 110 eligible gastrointestinal outbreaks, with a median number of seven recognized cases (Table 2). Foodborne outbreaks were most common, followed by person-to-person and environmental sources. Approximately half of the outbreaks were caused by viral pathogens such as caliciviruses (or noroviruses, a genus within the calicivirus family), which generally cause self-limiting illness. The patients in most of the outbreaks of gastrointestinal illness investigated by the Minnesota Department of Health rarely seek professional health care.

TABLE 1. Number and size of signals of gastrointestinal illness detected in HealthPartners patients, by recurrence interval, based on a 1-day adjusted space-time scan statistical technique — Minnesota, February 2001–January 2003

Recurrence interval	No. signals	No. signals expected by chance	Median no. cases in health plan	Median radius (km)	Median no. ZIP codes
≥2 years	1	1	4	0.0	1
≥1 year	3	2	5	0.0	1
≥6 months	6	4	9	10.0	15
≥2 months	18	12	9	7.3	14
≥1 month	30	24	9	8.9	14
≥2 weeks	58	52	9	8.2	9

TABLE 2. Selected characteristics of the eligible outbreaks of gastrointestinal illness identified by the Minnesota Department of Health, February 2001–January 2003

Mode of transmission	Etiology	No. outbreaks	Range and median of recognized cases	Range and median number of days between symptomatic case and initiation of evaluation
Food	Calicivirus, norovirus, astrovirus, other viral gastroenteritis	55	Range: 2–76 Median: 8	Range: 0–33 Median: 4
	Clostridium perfringens	6	Range: 3–6 Median: 4	Range: 0–5 Median: 1.5
	Bacillus cereus or *Staphylococcus aureus*	6	Range: 4–17 Median: 10.5	Range: 1–18 Median: 2
	Salmonella	3	Range: 2–46 Median: 29	4, 23 (one missing)
	Escherichia coli O157:H7	2	3, 5	13 (one missing)
	Campylobacter	1	4	7
	Scrombrotoxin	1	2	8
	Unknown	18	Range: 2–10 Median: 3	Range: 0–16 Median: 2.5
Person-to-person	Calicivirus or norovirus	6	Range: 9–720 Median: 23	Range: 0–30 Median: 2
	E. coli O157:H7	3	Range: 3–27 Median: 25	3, 5 (one missing)
	Salmonella	1	59	5
	Unknown	5	Range: 3–21 Median: 11	Range: 0–10 Median: 1
Recreational water	*E. coli* O157:H7	1	20	16
	Unknown	2	4, 11	1, 78

Comparison of Syndromic Surveillance Signals and Known Outbreaks

The number of hits increased with more inclusive thresholds. One (of one) signal was concordant with an outbreak at the 2-year threshold; 17 (29%) were concordant with outbreaks at the 2-week threshold (Table 3). Of the 110 outbreaks, one had a concordant signal at the 2-year threshold; this number increased to 14 (13%) at the 2-week threshold. The probability that the observed number of concordant signals occurred by chance alone was in no instance significant.

The three unique instances of concordance between signals and outbreaks at the more restrictive thresholds of 6 months involved foodborne outbreaks of *Bacillus cereus* and probable calicivirus (with 17 and 13 ill, respectively) and a person-to-person outbreak of suspected viral gastroenteritis (17 ill). However, all three were ruled out as true associations by the health department on the basis of the circumstances and the low numbers of affected persons seeking medical care.

Discussion

When evaluating a syndromic surveillance system, the two most important features are the number of false signals and the number of true outbreaks detected. With respect to the former, the number of signals was roughly what would have been expected by chance alone, suggesting that the modeling did a good job of estimating the expected counts of illness and that false alarms would not have occurred much more often than the number predicted at the various thresholds. This implies that health departments need not be concerned about unexpectedly high numbers of false alarms from this surveillance system and the consequent waste of resources to investigate them.

Of the previously known outbreaks of gastrointestinal illness, none were found with certainty in this retrospective study. The three most unusual signals (with recurrence intervals of at least 6 months) that appeared linked to outbreaks were ultimately deemed unrelated by the health department. Instances of concordance between signals and outbreaks where the signals where less strong (those with recurrence intervals of <6 months) were not investigated in depth, so the plausibility of a true connection in those cases is unknown, although it could be clarified through additional investigation of outbreak data and patient-level data held at the health plan.

Sensitivity of the system was low at all thresholds. Among the possible reasons for this finding is that approximately two thirds of foodborne outbreaks in Minnesota are caused by caliciviruses, which lead to a self-limiting illness for which the affected typically do not seek medical care. Other possible reasons for low sensitivity include the somewhat low proportion (about 8%) of the Twin Cities population in the surveillance system and the fact that points of exposure (e.g., a lunch-time restaurant) might be far from the ZIP codes of residence on which the signal detection method is trained.

Of the unlinked signals, although many might have been false alarms, others might have represented undetected outbreaks. For example, outbreaks where no laboratory specimens are submitted can elude Minnesota's enhanced surveillance systems. Distinguishing between these two was not possible in this study because such a determination requires case-by-case investigation, possibly including specimen collection and interviews, which would not likely be informative so long after the events.

Conclusion

Prospective evaluation is the best way to understand the usefulness of syndromic surveillance systems to enhance existing public health surveillance because it allows immediate case-by-case investigation of exposure histories and course of illness and the collection of clinical specimens. Performance probably will vary depending on the type of acute illness in question because of differences in such characteristics as predominant mode of transmission, speed of transmission, severity of illness, and whether the disease is reportable to state or local health departments. In any prospective evaluation, it will be important to determine not only how often signals

TABLE 3. Relation between signals and 110 known outbreaks of gastrointestinal illness, by recurrence interval, based on 1-day adjusted space-time scan statistical technique — Minnesota, February 2001–January 2003

Recurrence interval	No. signals concordant with an outbreak (apparent predictive value positive)	Probability that observed no. occurred by chance	No. of the 110 outbreaks with a concordant signal (apparent sensitivity)
≥2 years	1/1 (100%)	NA*	1 (1%)
>1 year	2/3 (67%)	NA*	2 (2%)
≥6 months	3/6 (50%)	0.63	3 (3%)
≥2 months	5/18 (28%)	0.87	5 (5%)
≥1 month	10/30 (33%)	0.96	9 (8%)
≥2 weeks	17/58 (29%)	0.98	14 (13%)

* The permutation test is not meaningful where the number of signals is <5.

and true outbreaks are associated with each other but how many detected outbreaks are of public health significance, how many of those are also picked up by existing public health surveillance systems, and how many of those doubly detected outbreaks are identified *earlier* by the syndromic surveillance system.

Acknowledgments

This work has been supported by CDC cooperative agreement UR8/CCU115079 and Minnesota Department of Health contract A57182. Daniel Neises tabulated the outbreaks in the realization of his Master of Public Health thesis, and Arthur Davidson and Ross Lazarus critiqued early drafts.

References

1. Gesteland PH, Gardner RM, Tsui FC, et al. Automated syndromic surveillance for the 2002 Winter Olympics. J Am Med Inform Assoc 2003;10:547–54.
2. Hogan WR, Tsui FC, 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.
3. 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.
4. 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.
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. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecur Bioterror 2003;1:1–5.
7. Lazarus R, Kleinman K, 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;1:1–9.
8. 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.
9. 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;80:25–31.
10. Yih WK, Caldwell B, Harmon R, et al. The National Bioterrorism Syndromic Surveillance Demonstration Program. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004; 53:43–6.
11. CDC. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Available at http://www.bt.cdc.gov/surveillance/syndromedef/index.asp.
12. Kulldorff M. Prospective time periodic geographic disease surveillance using a scan statistic. J R Stat Soc 2001;164:61–72.
13. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident cluster/signals of disease in small areas, with an application to biological terrorism (with invited commentary). Am J Epidemiol 2004;159:217–24.
14. Kleinman K, Abrams A, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. Epidemiol Infect 2005;119:409–19.
15. Besag PJ, Newell J. The detection of cluster/signals in rare diseases. J R Stat Soc 1991;154:143–55.
16. Neises D. Rate-limiting factors in foodborne outbreak detection and non-traditional foodborne illness surveillance in Minnesota. Minneapolis, MN: University of Minnesota; 2003.
17. Kleinman K, Abrams A, Yih WK, Kulldorff M. Assessing the utility of spatial surveillance: detection of known outbreaks in real data. Statistics in Medicine 2005; In press.
18. Dwass M. Modified randomization tests for nonparametric hypotheses. Ann Math Stat 1957;28:181–7.

Emergency Department Visits for Concern Regarding Anthrax — New Jersey, 2001

Paul C. Allegra,[1] D. Cochrane,[1,2] E. Dunn,[1] P. Milano,[1] J. Rothman,[2] J. Allegra[1,2]
[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 George Cochrane, 241 Brook Valley Rd., Kinnelon, NJ 07405. Telephone: 973-331-0069; Fax: 973-331-0228; E-mail: cochraned@verizon.net.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *In October of 2001, after letters processed in Trenton, New Jersey, resulted in multiple cases of anthrax, emergency departments (EDs) in New Jersey experienced an increase in visits from patients concerned about possible exposure to agents of biologic terrorism. Information about the effect of an actual biologic terrorism attack on the emergency department population might be useful in the design of biosurveillance systems, particularly with regard to their performance during the mitigation phase that occurs after an attack. In addition, such information might help identify issues that arise regarding the public health response in the ED setting.*

Objectives: *The objectives of this report were to identify and characterize ED visits, by patients concerned with exposure to biologic terrorism agents, in selected New Jersey hospitals after the anthrax attack in fall 2001.*

Methods: *A retrospective cohort design was used in this study. The setting was 15 New Jersey EDs within a 55-mile radius of Trenton. Participants were consecutive patients evaluated by ED physicians for the following four periods in 2001: 1 month before September 11; 1 month after September 11; 1 month after October 11; and for the second month after October 11. Percentages of visits were calculated with a concern for exposure (CE) visits by using* International Classification of Diseases, Ninth Revision *(ICD-9) descriptors: Feared Complaint–No Diagnosis (ICD-9 code v65.6) and Screening for Infectious Disease (ICD-9 code v75.9) for all hospitals and for Trenton versus non-Trenton hospitals as a percentage of ED visits. Charts were reviewed by using a structured data form.*

Results: *A total of 225,403 ED visits occurred during the 4 months, of which 698 were CE visits. The percentages of CE visits for the four periods were 0.06%, 0.06%, 0.92%, and 0.10%, respectively. For the peak third period, the percentage was increased for the two Trenton hospitals, 1.81%, versus 0.82% for the 13 non-Trenton hospitals. This report is a summary of the 508 visits associated with concern for anthrax exposure during the peak third period: 47% reported exposure to powder, 13% were postal workers, 4% received chest radiographs, 65% had a nasal swab for anthrax, 13% had ED decontamination, and 32% received antibiotics.*

Conclusion: *An increase in CE visits occurred during the 1-month period after October 11, 2001. During the peak month, a higher increase occurred in Trenton EDs. Considering the substantial variation in diagnostic evaluation and treatment, readily available guidelines are needed.*

Introduction

Media coverage after the September 11, 2001, attacks included warnings regarding possible biologic terrorism. On October 4, 2001, less than 1 month after the terrorist attacks on the World Trade Center and Pentagon, the condition of a man in Florida was diagnosed as respiratory anthrax; he had no known exposure risk factors (*1*). On October 12, a case of cutaneous anthrax was reported in New York City. At NBC News, a person was exposed to a letter containing a suspicious powder. The Federal Bureau of Investigation (FBI) reported that four recovered envelopes containing *Bacillus anthracis* spores were postmarked at the U.S. Postal Service Trenton Processing and Distribution Center in Hamilton Township, New Jersey. The investigation revealed that the Hamilton Township Postal Facility handled two envelopes containing *B. anthracis* that were mailed to news organizations in NYC on September 18, 2001, and two envelopes mailed to U.S. Senate offices in Washington, DC, on October 9, 2001.

Concerns increased in New Jersey on October 18, when two postal workers from the Trenton facility were confirmed to have cutaneous anthrax. These exposures were presumed to represent contact with letters or cross-contamination from letters containing *B. anthracis* mailed from this location. A total of 22 confirmed or suspect cases of anthrax infection occurred; 11 were inhalation cases, and 11 were cutaneous cases. Five persons died (*2*).

Multiple patients sought medical treatment at New Jersey emergency departments (EDs) with concerns of possible exposure (CE visits) after the media reports of October 12, 2001. Records of these visits, including physician notes, were documented in an electronic format.

Syndromic surveillance techniques have been previously reported by using a large existing computerized database of ED visits (*3*). Techniques included using *International Classification of Diseases, Ninth Revision* (ICD-9) groups to identify a subgroup of patient visits, examining time-series volume data for these groups, and using electronic chart reviews, to examine characteristics of the subpopulation identified. Limited information has been written regarding the use of such techniques in the aftermath (mitigation phase) of an actual biologic terrorist attack.

The objectives of this report were 1) to use these techniques to determine the temporal and geographic aspects of CE visits at 15 New Jersey EDs during the period surrounding the anthrax attacks of 2001 and 2) to characterize the clinical presentations, diagnostic evaluations, and treatments provided, based on a chart review of the electronic medical record.

Methods

A retrospective analysis of a computerized database was performed of ED visits from one emergency physician group staffing 15 New Jersey EDs within a 55-mile radius of Trenton. The annual ED volume was 20,000–65,000 patients. These hospitals included urban and suburban teaching and nonteaching hospitals and two hospitals in Trenton. These hospitals comprise approximately 30% of all northern New Jersey hospitals and receive an estimated 35%-40% of all ED visits in northern New Jersey. Consecutive patients evaluated by ED physicians August 11–December 11, 2001, were included. The ED physicians evaluate 85%–95% of the patients at these EDs. Private physicians evaluate the remainder of the patients. The physicians' billing department coded the physicians' charts according to the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) codes. All coders were located at one central facility and trained and supervised similarly. Patients were included as concern for exposure (CE) visits if the primary ICD diagnosis included the following ICD-9 descriptors: Feared Complaint–No Diagnosis (ICD code v65.6) and Screening for Infectious Disease (ICD-9 code v75.9).

Percentages of CE visits were calculated for all hospitals and for Trenton versus non-Trenton hospitals as a percentage of ED visits. The following four periods in 2001 were used in the calculation: 1 month before September 11; 1 month after September 1, 1 month after October 12; and the second month after October 12. These periods were chosen to offer sufficient time to establish a baseline percentage before September 11 and to include a similar period after the anthrax cases decreased to the baseline percentage. The student t test and the Chi-square were used to test for statistical significance with alpha set at 0.05. The Bonferroni correction was used when multiple comparisons were made.

A chart review was conducted by using a structured data form. The details were provided of the chart review for the peak third period for those visitors who had a concern for anthrax. The details characterized the mechanisms of exposure, symptoms, diagnostics, and treatments. The Internal Review Board at Morristown Memorial Hospital Residency in Emergency Medicine, Morristown, New Jersey, approved the study.

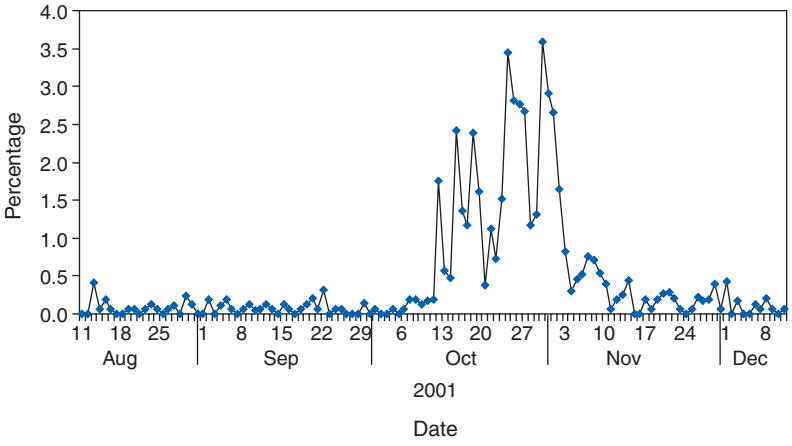
Results

For the four periods, 225,403 ED visits occurred, of which 698 (0.3%) were CE visits. The percentage of CE visits versus time has been illustrated (Figure 1). A sharp increase occurred on October 12, and a decrease occurred near the baseline 1 month later. The increase on October 12, coincides with the first media news reports of an anthrax case in NYC. Of the 698 CE visits for the four monthly periods, 561 (80%) occurred during the 1-month period after October 12.

The percentages of CE visits for the four periods, compared with total visits were 0.06%, 0.06%, 0.92%, and 0.10%, respectively. When the first period is used as the baseline, only the third period indicated a statistically significant increased percentage of CE visits (0.92%; $p<0.001$). For the peak third period, the percentage was significantly increased for the two Trenton hospitals, compared with the 13 non-Trenton hospitals (1.81% versus 0.82%; $p<0.001$).

Periods 1 and 2 represent periods before the anthrax attacks occurred and therefore indicate baseline characteristics for the patients seeking medical treatment for CE. The chart review in Periods 1 and 2 indicated that the majority of the visits (89%) involved concern regarding one of the following five categories: 1) fear of foreign bodies without findings; 2) motor-

FIGURE 1. Percentage of visits for ICD-9* descriptors "Feared Complaint–No Diagnosis" and "Screening for Infectious Disease," as a percentage of total daily visits† at 15 emergency departments — New Jersey, August 11–December 11, 2001



* *International Classification of Diseases, Ninth Revision.* The ICD-9 code for Feared Complaint–No Diagnosis is v65.6. The ICD-9 code for Screening for Infectious Disease is v75.9.
† N = 225,403 emergency department visits.

vehicle accidents; 3) falls without injury; 4) concern for sexually transmitted disease; and 5) possible toxic ingestion/exposure with none identified. The first case-patient who sought medical treatment for CE to anthrax occurred on October 9.

Of the 57,981 ED visits during the third peak period, 561 visits were identified for CE by the ICD-9 codes (CE visits), of which 508 were related to a concern for anthrax (CA) visit. The details of the chart review for the peak third period are summarized (Tables 1–4). These results summarize the 508 patient visits that were related to a concern for anthrax.

Demographic information is reported for the CA visits during the peak period, October 12–November 11, 2004 (Table 1). Females comprised 277 (55%) of the 508 CA visits during this period. The mean age in years was 39 ± 16. Employees of the post office accounted for 64 patients, and 32 were hospital employees.

TABLE 1. Age and sex of persons who sought medical treatment for concern for anthrax exposure at 15 emergency departments — New Jersey, October 12–November 11, 2004*

	Male		Female		Total	
Age group (yrs)	No.	(%)	No.	(%)	No.	(%)
0–10	11	(2)	11	(2)	**22**	**(4)**
11–20	15	(3)	18	(4)	**33**	**(7)**
21–40	105	(20)	126	(25)	**231**	**(45)**
41–65	85	(17)	109	(21)	**194**	**(38)**
≥65	15	(3)	13	(3)	**28**	**(6)**
Total	**231**	**(45)**	**277**	**(55)**	**508**	**(100)**

* N = 508.

Based on information charted by the ED physician, 19% of all patients sought help before going to the ED. Seeking help included contacting their primary physician, police department, FBI, postmaster, or job supervisor. A total of 6% reported their concern to the department of public health. In addition, 17% received assistance before going to ED, which included receiving an antibiotic prescription from their primary physician, having their suspected material sent for analysis, and being decontaminated.

The mechanisms of exposure for the CA patients during the peak period are presented (Table 2). Of the 508 cases, 120 had documentation of working in or visiting a post office. A total of 128 (25%) patients did not have a history suspicious for exposure to anthrax or the exposure status was not reported. These visits primarily consisted of complaints of fever, cough, chest pain, rash, insect bites, and myalgias. An additional primary complaint was from persons

TABLE 2. Mechanisms of concern for exposure, by anthrax visits to 15 emergency departments — New Jersey, October 12–November 11, 2004*

Concern for exposure	No.	(%)
Worked in post office with known anthrax	20	(4)
Visitor to post office with known anthrax	41	(8)
Worked in post office with no known anthrax	44	(9)
Visitor to post office with no known anthrax	15	(3)
Exposure to powder associated with a letter	136	(27)
Exposure to powder not associated with a letter	103	(20)
No exposure or status unknown	128	(25)

* N = 508.

TABLE 3. Symptoms and diagnostics associated with concern for anthrax visits to 15 emergency departments — New Jersey, October 12–November 11, 2004*

Symptoms/Diagnostics	No.	(%)
Respiratory symptoms	70	(14)
Skin symptoms	52	(10)
Gastrointestinal symptoms	8	(2)
Chest radiograph	20	(4)
Nasal swab for anthrax	330	(65)

* N = 508.

TABLE 4. Treatments of concern for anthrax visits to 15 emergency departments — New Jersey, October 12–November 11, 2004*

Treatment	No.	(%)
Patient decontaminated	56	(11)
Treatment with antibiotic started	161	(32)
Ciprofloxacin	114	(22)
Doxycycline	35	(7)
Treatment >14 days	30	(6)

* N = 508.

who worked near or traveled by a possible anthrax site and wanted to be checked for anthrax. Examples of other documentation include 1) a hospital employee who was sent by corporate health but did not think she was exposed, and 2) a person who opened several letters from Iraq over the previous year.

Respiratory symptoms were present in 70 (14%) patients, and skin symptoms were present in 52 (10%) patients. GI symptoms were present in eight (2%) case-patients. Chest radiographs were done in 20 (4%) of the CA visits, and nasal swab for anthrax was performed during 330 (65%) CA visits. The percentage of visits where nasal swabs were performed was similar until November 3 when a noticeable decline to 22% occurred for the rest of Period 3 (Figure 2).

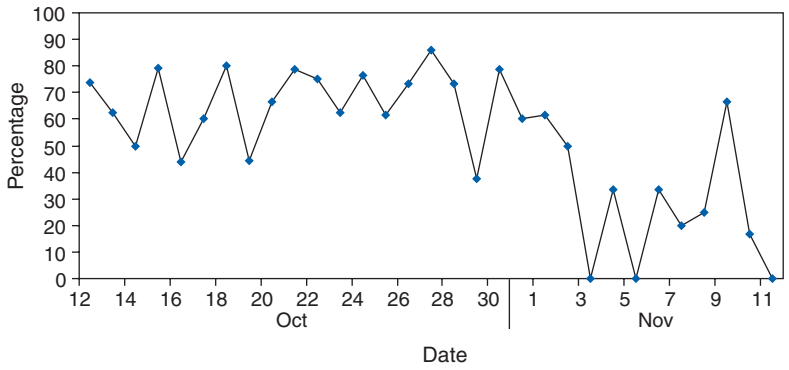
ED decontamination of the patient occurred during 56 CA visits (13%). Antibiotic treatment was initiated for 161 (32%) CA visits. Of these 161 visits, 114 (71%) received ciprofloxacin, and 35 (22%) received doxycycline. Of the 161 patients who received an antibiotic, only 30 (19%) received a prescription for an antibiotic course >14 days.

Although determining whether all calls were documented is difficult, calls made to health or police departments were made by the ED and documented for 90 (18%) of the visits.

Discussion

A marked increase in CE visits occurred after anthrax was identified in New York City and New Jersey. The CE visits returned to the baseline after 1 month. Despite initial warnings from the media immediately after September 11, 2001, a substantial increase did not occur in CE visits until the initial anthrax case in New York City and the attendant media coverage starting October 12. The number of visits increased substantially in the immediate vicinity of the contaminated postal facility in Trenton.

FIGURE 2. Percentage of visits for concern for anthrax exposure in which nasal swabs were performed at 15 emergency departments, by date — New Jersey, October 12–Novembver 11, 2001



These findings have implications for the design of biosurveillance systems. Although there were only 22 cases of anthrax, suspected or confirmed, 508 visits to the selected EDs for CE occurred during the peak period. This finding suggests that for detecting and trending of cases after a biologic terrorist attack, using syndromic surveillance might be obscured by patients with a concern rather than patients who are actually victims of the attack.

In the chart review of CA patients for the peak third period, a substantial variation occurred in the reasons patients sought treatment with a CE to anthrax. Postal workers in a facility that had known cases of anthrax were at maximum risk. Postal workers in other facilities were at lower risk. Members of the general public who had no history of exposure but still wanted to be evaluated were at the lowest risk. Only 4% of the patients received radiographs. This low rate indicates that the majority of the patients were considered to be at little or no risk for having active pulmonary anthrax.

A total of 65% of patients received a nasal swab for anthrax, a fairly routine procedure for these patients. Routine nasal swabs for anthrax detection were not recommended by CDC. Instead, CDC recommendations during that period indicated that culture of nasal swabs might be appropriate in the epidemiologic investigation of a known outbreak but would not rule out exposure to or infection with *B. anthracis* (*4*). Given the substantial volume of swabs, the minimal evidence of communication with public health departments, and the lower use of prophylactic antibiotics, these swabs probably were performed primarily for patient reassurance or in the mistaken belief that a culture of a nasal swab would rule out anthrax. There was a drop-off in the percentage of patients receiving nasal swabs for anthrax at the end of the one-month peak period which might represent the time at which physicians began following CDC recommendations. It might also represent a decrease in general concern about anthrax.

Because of the substantial percentage of patients who received a nasal swab for anthrax and because the nasal swabs were not recommended, these findings suggest that guidelines (*5*) were not effectively reaching emergency physicians. These findings also suggest that current systems need to be reinforced and further developed to communicate recommendations immediately to ED physicians to manage the influx of victims of possible biologic terrorism. Additional studies have examined the need for better communication systems (*6,7*), including the use of the Internet (*8,9*).

The percentage of patients who received antibiotics was 32%. The CDC recommendation at the time was that the basis for initiating antibiotic treatment

should be exposure or contact, not laboratory test results (*5*). Whether the initiation of antibiotics was consistent with the CDC recommendation cannot be determined from the data.

The same recommendations also called for 60 days of antibiotic prophylaxis. This report determined that only 19% of patients in the series who received an antibiotic prescription received one for >14 days. Sufficient data was not available to determine the extent that these prescriptions were written as "starter" prescriptions intended to be continued if deemed appropriate by the follow-up physician.

No other reports specifically examine the volume or character of ED visits during this period. An epidemiologic study by CDC in October 2002 sought to determine the extent of the anthrax outbreak in New Jersey, assess potential sources of *B. anthracis* exposure, and prevent additional cases by developing and implementing control measures (*10*). However, this study did not assess the effects of the outbreak and media exposure on New Jersey EDs.

In this study, methods were used that had been previously developed for syndromic surveillance to identify and characterize the effect of the anthrax attacks on EDs in the mitigation phase. The investigation was conducted on a large electronic database in which ICD-9 groups were used to identify patient visits of interest, time-series data to detect volume peaks, and a chart review of the affected patients. The level of detail available yielded information that might be valuable to public health preparations for the mitigation phase of a biologic terrorism attack. The data indicated that syndromic surveillance during the mitigation phase might have been affected by a substantial volume of patients concerned about exposure, although exposure was unlikely. The data also indicates that communication between emergency medicine practitioners and public health officials are vital during the mitigation phase and that communication needs improvement.

Limitations

All patients examined in the EDs were not included in this report, only those examined by ED physicians. However, this report encompassed the majority of these patients. Some variation in coding styles might have led to a variation in how patients were classified. However, this potential source of error was mitigated, because ICD-9 coding is performed centrally by a single group of coders who were trained in a similar manner. This report did not include all ED visits in northern New Jersey. However, because the volume represents approximately 30% of hospitals and 35%–40% of ED visits in northern New Jersey, the sample was representative of the total population of ED visits in northern New Jersey during this period.

The chart review is limited by the retrospective nature of the study. Certain characteristics that were present were possibly underestimated, because they were not documented in the chart.

Conclusion

Techniques originally developed for syndromic surveillance identified an increase in CE visits that occurred during the 1-month period after the identification and media coverage of anthrax cases in New York City and at a Trenton postal facility. During the peak month, a higher increase of CE visits occurred in the Trenton EDs than in other EDs in the state. A substantial variation occurred in the reasons that patients sought medical treatment. Although the suspicion for anthrax in these patients appeared to be low, the majority of patients received a nasal swab for anthrax in excess of the guidelines from CDC. This finding suggests the need for improved communications during a biologic terrorist attack and readily available diagnostic and treatment guidelines.

References

1. CDC. Update: investigation of anthrax associated with intentional exposure and interim public health guidelines, October 2001. MMWR 2001;50:889–93.
2. CDC. Investigation of bioterrorism-related inhalational anthrax—Connecticut, 2001. MMWR 2001;50:1049–51.
3. Cochrane D. Perspective of an emergency physician group as a data provider for syndromic surveillance. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):209–14.
4. CDC. Update: investigation of bioterrorism-related anthrax and interim guidelines for exposure management and antimicrobial therapy, October 2001. MMWR 2001;50:909–19.
5. CDC. Update: investigation of anthrax associated with intentional exposure and interim public health guidelines, October 2001. MMWR 2001;50:889–97.
6. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proceedings Natl Acad Sci U S A 2003;100:4346–51.
7. Williams AA, Parashar UD, Stoica A, et al. Bioterrorism-related anthrax surveillance, Connecticut, September–December, 2001. Emerg Infect Dis 1002;8:1078–82.
8. Kittler AF, Hobbs J, Volk LA, Kreps GL, Bates DW. The Internet as a vehicle to communicate health information during a public health emergency: a survey analysis involving the anthrax scare of 2001. J Med Internet Res 2004;6:e8.
9. Hobbs J, Kittler A, Fox S, Middleton B, Bates DW. Communicating health information to an alarmed public facing a threat such as a bioterrorist attack. Journal of Health Communication 2004;9:67–75.
10. Greene CM, Reehuis J, Tan C, et al. Epidemiologic investigations of bioterrorism-related anthrax—New Jersey, 2001. Emerg Infect Dis 2002;8:1048–55. Available at http://www.cdc.gov/ncidod/EID/vol8no10/02-0329.htm.

Hospital Admissions Syndromic Surveillance — Connecticut, October 2001–June 2004

James L. Hadler, A. Siniscalchi, Z. Dembek
Infectious Diseases Division, Connecticut Department of Public Health, Hartford, Connecticut

Corresponding author: James L. Hadler, Infectious Diseases Division, Connecticut Department of Public Health, MS#11FDS, Hartford, CT 06134-0308. Telephone: 860-509-7995; Fax: 860-509-7910; E-mail: james.hadler@po.state.ct.us.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: *The Connecticut Department of Public Health (CDPH) has continuously monitored daily nonelective hospital admissions through a syndromic surveillance reporting system (HASS) since September 2001. Admission diagnoses are categorized into 11 syndromes including one possible indicator of smallpox, fever with rash, and one possible indicator of influenza and pneumonia.*

Objectives: *The objectives are to describe findings from systematic investigation of individual admissions attributed to fever and rash and to determine the utility of monitoring pneumonia admissions as an indictor of severe influenza activity during the 2003–04 influenza season.*

Methods: *The incidence of admissions for fever and rash illness was determined for a 12-month period, and results of clinical discharge diagnoses were tabulated. Excess admissions for pneumonia by week during the influenza seasons beginning 2001–03 were determined and compared. Trends in admissions from the 2003–04 season were compared with trends from laboratory and sentinel physician surveillance.*

Results: *A total of 57 admissions for fever and rash illness were reported from 32 acute-care hospitals and verified for an incidence of 1.7 per 100,000 population. Specific clinical diagnoses were made for 29. Many were compatible with the initial clinical presentation of smallpox. Excess admissions for pneumonia during the 2003–04 season occurred concurrently with sharp increases in positive laboratory reports and percentages of visits to physician's offices attributed to influenza-like illness. The 2003–04 influenza season had many more excess admissions than the 2001–02 and 2002–03 seasons.*

Conclusion: *HASS is a useful surveillance tool for rapid detection of sentinel cases of smallpox. Monitoring excess pneumonia admissions during the influenza season appears to be an effective and specific method for determining levels of influenza activity and for quantification of influenza-related morbidity and impact on the hospital system.*

Introduction

Syndromic surveillance systems are being implemented and evaluated by federal, state, and local public health jurisdictions and by academic institutions to determine their ability to detect outbreaks of illness earlier than clinician and laboratory disease-specific reporting systems and their sensitivity to detect outbreaks that might otherwise be missed (*1–6*). In addition, they are being explored for their possible utility to help monitor and respond to potentially large-scale events with substantial morbidity (*1–3,6*).

In response to the September 11, 2001, World Trade Center attacks, the Connecticut Department of Public Health (DPH) initiated a hospital admissions syndromic surveillance system (HASS) to monitor for a possible concurrent biologic attack (*7*). The system has been continued since then with two main objectives. The first is to increase the sensitivity and timing of detection of initial cases of smallpox and severe acute respiratory disease syndrome (SARS). This objective is carried out by monitoring individual admissions with selected unusual syndromes of concern (i.e., rash illness and fever [possible smallpox] and pneumonia in a health-care worker [possible SARS]). The second objective is to have a system with readily available data to assess the magnitude and geographic distribution of severe illness requiring hospital admission brought to attention by other systems (e.g., to monitor the impact and geographical distribution of influenza; to assess whether evidence exists of wider activity if a case of anthrax were diagnosed).

Among current syndromic surveillance systems, the HASS is unique in two respects. First, it is intended to detect an

initial, sentinel case of smallpox before a larger outbreak occurs. Second, it monitors hospital admissions rather than outpatient visits.

The objectives of this study were twofold. The first is to describe the results of investigation of persons admitted with fever and rash illness to assess the utility and workload of HASS as a tool to monitor for possible smallpox. The second is to examine trends in hospital admissions for pneumonia during the 2003–04 influenza season to assess the utility of the HASS as an influenza surveillance tool. The 2003–04 influenza season was unusual in that influenza activity nationally and in Connecticut appeared early, became most intense during November and December, and was largely over by mid-January (*8,9*). The season also was associated with excess mortality compared with the preceding 3 seasons (*9*).

Methods

Statewide, each of 32 acute-care hospitals reviews unscheduled admissions for the previous 24 hours, manually categorizes them on the basis of admission diagnosis into 11 syndromes, and reports this information daily through an Internet-based reporting system to DPH. The 11 syndromes include fever with rash and pneumonia.

The public health response to the data received has several steps. First, active follow-up begins the day of report of admissions for rash illness and fever to determine the course of illness, evolving differential diagnosis, and whether testing is needed to rule out smallpox. Second, data is reviewed daily to weekly by syndrome to determine unusual levels of activity. Finally, a comprehensive review is conducted if issues are unresolved (e.g., examination of trends in pneumonia admissions during the influenza season to determine how long they remain elevated.

To describe the experience with fever and rash illness, admission and follow-up data were reviewed for the 12-month period, July 2003–June 2004. The number of cases and incidence overall and by county were determined. The results of verification of cases as to whether they presented with both fever and rash and the final clinical diagnoses and their potential to mimic the typical clinical presentation of smallpox are described.

To examine the utility of the HASS for monitoring influenza, weekly totals of admissions for pneumonia were examined for the state and three largest counties (each with 800,000–900,000 persons representing 77% of the state's population) for temporal trends during November 2003–March 2004. In addition, findings were compared with the weekly results from two other influenza monitoring systems operating at the same time and to weekly results of pneumonia admissions for the preceding 2 years for HASS. The two additional influenza surveillance systems in use included required reporting to DPH of the results of all positive laboratory tests for influenza and a sentinel physician surveillance system in which volunteer physicians reported daily the total number of visits and the percentage attributed to influenza-like illness (ILI, defined as temperature of >100°F [>38°C]) and either a cough or a sore throat. Finally, the burden of excess admissions for pneumonia each week and overall were tallied and compared over the 3 years. Excess admissions were defined as the number of admissions for pneumonia above the annual weekly average.

Results

Fever and Rash Illness

Overall, 78 cases of fever and rash illness were reported from 32 hospitals in HASS. Of these, 57 were verified as fever with rash. Cases excluded usually had either fever or rash, but not both.

The 12-month incidence was 1.7 per 100,000 population. The county-specific incidence for the three largest counties was consistent, ranging from 1.5–1.7. The incidence in the five smaller counties was wider, ranging from 0.8–3.5, with each having no more than 2–4 cases.

By the time patients were discharged, 29 (51%) had a specific clinical diagnosis (Table 1). The most common diagnoses were drug hypersensitivity (11%), varicella (7%), and urticaria (4%). Cases were diagnosed in which the initial presentation was similar to that of smallpox (e.g., fever with diffuse undifferentiated rash).

TABLE 1. Clinical diagnoses in persons hospitalized with fever and rash — Connecticut, July 2003–June 2004

Diagnoses	No.	%
Drug hypersensitivity	6	11
Varicella	4	7
Urticaria	2	4
Other infectious* (Ehrlichiosis, Kawasaki disease, parvovirus, meningococcemia, Rocky Mountain spotted fever, roseola, "tickborne disease," toxic shock syndrome, viral exanthem, cellulitis, sinusitis with rash, and staphylococcal urosepsis)	12	21
Other noninfectious* (Contact dermatitis, erythema nodosum, lymphoma with rash, psoriasis, Stevens-Johnson syndrome)	5	9
No certain diagnosis	28	49

* One case for each listed diagnosis.

Pneumonia

Throughout the 3 years, the average number of weekly admissions for pneumonia reported through the HASS was 305. For both the 2001–02 and 2002–03 influenza seasons, pneumonia admissions peaked in early January and then continued to be much higher than average either continuously (2001–02) or intermittently (2002–03) through most of February, with a decline to near baseline by early March (Figure 1). This pattern paralleled known influenza activity for these years (*10*), although it is unclear how many of the admissions were influenza-related.

For the 2003–04 influenza season, the pattern was different and was characterized by a sharp increase in mid-December to a peak in early January that was much higher than peaks in either of the two preceding years, then a rapid decline to near baseline. The pattern for each of the three largest counties was similar to the overall pattern for all 3 seasons.

On the basis of findings from two influenza surveillance systems and excess hospital admissions for the 2003–04 influenza season, both the number of laboratory reports positive for influenza and the percentage of acute office visits due to ILI show increasing activity beginning in mid-November, followed by sharp increases in mid-December and peaks in late December (Figure 2). Activity then fell nearly as sharply as it rose in both systems. The excess number of pneumonia admissions paralleled the activity of these influenza-specific systems with a slight lag in reaching its peak and a more rapid return to baseline.

For the 2003–04 season during the weeks of peak activity, approximately 350 extra pneumonia admissions were reported per week, an index of surge capacity needed by hospitals (Figure 3). In addition, the 2003–04 influenza season put an unusual seasonal stress on hospitals compared with the preceding two seasons.

FIGURE 1. Number of hospital admissions for pneumonia, by month and week — Connecticut, November–March, 2001–02, 2002–03, and 2003–04 influenza seasons

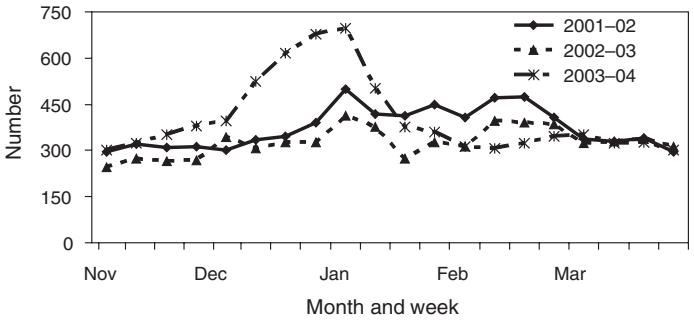


FIGURE 2. Percentage of physician visits for influenza-like illness (ILI), number of laboratory-positive reports, and excess hospital admissions for pneumonia, by week — Connecticut, October 2003–February 2004

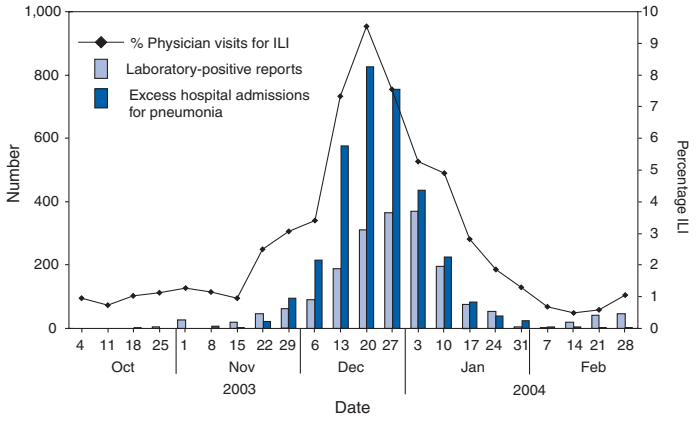
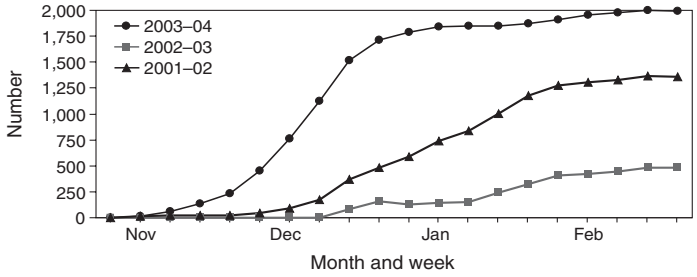


FIGURE 3. Cumulative number of excess hospital admissions for pneumonia, by month and week — Connecticut, November–March, 2001–2004 influenza seasons



Discussion

Syndromic surveillance systems have attracted the interest of public health officials and academia, particularly in the context of monitoring for biologic terrorism (*1,11,12*). In the absence of specific biologic terrorism events since the anthrax attacks of 2001, the sensitivity and timeliness of syndromic surveillance to detect biologic terrorism events compared with clinician and laboratory reporting systems have not been determined. The sustainability of syndromic surveillance systems will depend on whether their cost is worth their usefulness. If there are no biologic terrorism attacks, sustainable syndromic surveillance systems for biologic terrorism will have to have added value to the clinician and laboratory reporting systems or have demonstrated alternative utility.

On the basis of this analyses, HASS has demonstrated utility in monitoring for smallpox and has distinct added benefits to other forms of influenza surveillance. The initial manifestations of smallpox in a nonimmune person will probably be severe (i.e., prostrating febrile illness with generalized nondescript rash). The rash will not evolve to a more charac-

teristic form for up to a week, but the severity of the initial febrile illness probably will result in a medical visit and hospital admission for observation and supportive treatment. On the basis of the diagnoses of the 57 cases described in this report, HASS clearly detects such serious febrile rash illness. Reports to HASS are timely because they occur the day after admission and enable immediate follow-up by public health epidemiologists even before a clinician is likely to suspect smallpox. In addition, HASS detects single cases and can assure that each is evaluated for possible smallpox. Although completeness of reporting was not formally measured, the even geographic distribution of fever and rash admissions suggests reporting might be relatively complete. In addition, HASS is an enhancement to clinician reporting of suspected smallpox. Although it identified 57 suspect cases, only two cases of illness were reported to DPH as possible smallpox during this period. One was an inpatient also detected by this system and found to have viral exanthem, and the other was an outpatient with chickenpox. Finally, the amount of epidemiology staff work for follow-up of each reported case is readily manageable, averaging about 4 hours for initial hospital contact and follow-up calls. Most of the time is spent making initial contact with the hospital and the managing physician.

The 2003–04 influenza season provided an opportunity to examine the utility of HASS as an influenza monitoring tool. The findings of this report indicate that nearly all excess hospital admissions for pneumonia during the fall and winter of 2003–04 were related to influenza. Nearly all excess activity occurred during November and December. No sustained excess pneumonia activity was reported after early January when winter respiratory disease, including influenza, is usually expected. On the basis of this experience, it is reasonable to assume that most excess severe pneumonia activity during any fall-winter season is most likely related to influenza. Other notable characteristics of HASS as an influenza surveillance tool are that it appears to be sensitive to high levels of influenza activity, and it provides a timely measure of the changing and overall burden influenza puts on hospitals. Finally, HASS provides data for assessing the overall severity of an influenza season. Connecticut intends to use HASS data prospectively throughout future influenza seasons to determine when the influenza season becomes severe (continuing and sustained increase in excess pneumonia admissions), when it peaks (continuing decrease in excess admissions) and its overall impact.

HASS is meeting the objective to provide a context to evaluate levels of severe disease activity by syndrome, at least for pneumonia and in the context of influenza. Because of concerns about pandemic influenza and about hospital surge capacity, this system should continue to be worth the cost of operation. Although data collection is largely manual, it is simple, requiring ≤15 minutes per hospital per day.

The Connecticut HASS is a unique surveillance system in the United States. CDC's Biosense system and most state and local health departments that have such systems are conducting syndromic surveillance on the basis of outpatient visits with a primary objective to identify outbreaks earlier. Limitations of HASS include its inability to detect predominantly outpatient outbreaks of illness and a lag time of one or more days between initial patient presentation to the health care system and admission to the hospital. However, the latter limitation might not be of as much practical impact as it is in theory. Users of outpatient syndromic surveillance systems are finding that it is very difficult to effectively set thresholds to identify levels of disease activity that merit full investigation, and that investigation of all but the largest sustained signals is very labor intensive and has little added value in the absence of a biologic terrorism attack (*1,2*). As a result, increases in syndrome-specific visits are monitored for one or more days to determine if they are sustained before initiating an investigation. In addition, these outpatient systems are limited in their ability to facilitate follow-up of individual cases. Furthermore, although outpatient systems have the potential to enable monitoring for the outpatient burden of influenza, numbers of outpatient visits do not directly measure the burden of severe illness. A hospital admission-based system such as that in Connecticut is of sufficient value relative to its cost that it should be tried and evaluated elsewhere. HASS might be of more practical long-term utility in meeting some biologic terrorism and public health preparedness surveillance objectives than syndromic surveillance systems that focus on outpatient visits.

Acknowledgments

The authors acknowledge Stephanie Petrahai and Albert DeLoreto for maintaining the influenza laboratory reporting database and sentinel physician influenza surveillance system respectively during 2003–04 influenza season, and Kenneth Carley for his assistance with analysis of HASS data. In addition, the authors are indebted to the designated reporters to HASS at each of the 32 acute-care hospitals in Connecticut for their daily efforts to report and assist with follow-up of persons admitted with febrile rash illness.

References

1. CDC. Review of the 2003 National Syndromic Surveillance Conference—lessons learned and questions to be answered. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:18–22.
2. CDC. New York City syndromic surveillance systems. MMWR 2004;53:25–7.

3. Loonsk JW. Biosense: a national initiative for early detection and quantification of public health emergencies. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:53–5.
4. 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. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:28–31.
5. Yih WK, Caldwell B, Harmon R, et al. National bioterrorism syndromic surveillance demonstration program. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:43–6.
6. Paladini M. Daily emergency department surveillance system—Bergen County, New Jersey. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:47–9.
7. Dembek ZF, Carley K, Siniscalchi A, Hadler J. Hospital admissions syndromic surveillance—Connecticut, September 2001–November 2003. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53:50–2.
8. CDC. Update: influenza activity—United States and worldwide, 2003–04 season, and composition of the 2004–05 influenza vaccine. MMWR 2004;53:547-52.
9. CDC. 2003-04 U.S. influenza season summary. Available at http://www.cdc.gov/flu/weekly/weeklyarchives2003-2004/03-04summary.htm
10. CDC. Influenza activity: weekly surveillance reports, 2001–2002, 2002–2003. Available at http://www.cdc.gov/flu/weekly/fluactivity.htm.
11. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and biologic terrorism-related epidemics. Emerg Infect Dis 2003;9:1197–204.
12. CDC. Focus area B: surveillance and epidemiology capacity. Available at http://www.bt.cdc.gov/planning/CoopAgreementAward/CDC2BTEmergencySupplemental_2_19_2—2_MASTER-4-00.asp.

Three Years of Emergency Department Gastrointestinal Syndromic Surveillance in New York City: What Have we Found?

Sharon Balter, D. Weiss, H. Hanson, V. Reddy, D. Das, R. Heffernan
New York City Department of Health and Mental Hygiene, New York, New York

Corresponding author: Sharon Balter, New York City Department of Health and Mental Hygiene, 125 Worth Street, Box 22A, New York, NY 10013. Telephone: 212-788-9662; Fax: 212-676-6091; E-mail: sbalter@health.nyc.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Background: *Use of syndromic surveillance as a tool to detect outbreaks and potential biologic or chemical terrorist attacks is increasing. Evaluating health departments' use of syndromic surveillance is necessary to determine the value of this methodology.*

Methods: *Syndromic surveillance signals detected by the New York City Department of Health and Mental Hygiene (DOHMH) during November 2001–August 2004 were reviewed for diarrhea and vomiting syndromes, the methods used to investigate such signals, and results of these investigations to determine if any unreported outbreaks were detected. Gastrointestinal (GI) outbreaks reported to DOHMH also were reviewed to understand why they were not detected by DOHMH's Emergency Department (ED) syndromic surveillance system.*

Results: *During the study period, ED surveillance generated 98 citywide and 138 spatial GI signals. Multiple outbreaks suspected to be caused by norovirus and rotavirus were identified, as well as a citywide increase in diarrheal illness. Of 98 citywide signals detected, 73 (75%) occurred during seasonal outbreaks. During the same period, 49 GI outbreaks were reported to DOHMH; none was detected simultaneously by ED surveillance.*

Conclusion: *Only substantial, citywide syndromic signals were identified as outbreaks and routinely reported. GI outbreaks did not generate syndromic signals. Syndromic surveillance signals occur frequently, are difficult to investigate satisfactorily, and should be viewed as a supplement to, rather than a replacement for, well-maintained traditional surveillance systems that rely on strong ties between clinicians and public health authorities.*

Background

Syndromic surveillance is increasingly used as a tool to detect both naturally occurring outbreaks and potential biologic or chemical terrorist attacks (*1*). In the absence of etiologic information, these systems use constellations of symptoms, complaints, or diagnostic codes to group patients into syndrome categories. Data can be gathered from emergency department (ED) logs (*2–5*), hospital admissions records (*6*), ambulatory care center records (*7,8*), ambulance dispatch records (*9*), or clinical laboratory submissions (*10*). Other data sources can include over-the-counter (OTC) medication sales (*10–13*), nurse hotline calls (*14,15*), and work and school absenteeism records (*16*) to identify patients who have not sought medical care. Although syndrome-based surveillance has long been used to detect and track diseases for which etiologic diagnoses are made infrequently (e.g., influenza and poliomyelitis) (*17–20*), since the 2001 anthrax attacks, such systems have been used as early warning systems to detect biologic or chemical terrorism (*21–25*). Syndromic surveillance is based on the concept that illnesses caused by agents likely to be used in a biologic or chemical terrorist attack (e.g., plague or anthrax) will first manifest with nonspecific (prodromal) symptoms (*25*). In principle, syndromic surveillance systems should detect outbreaks of naturally occurring illness and those caused by intentional attacks. The theoretic ability of these systems to detect such attacks has been described (*1,2,25*); however, to date, these systems have been useful primarily to detect and monitor substantial seasonal outbreaks of influenza, rotavirus, and norovirus (*7,8*). Whether syndromic surveillance systems also can detect smaller, more localized outbreaks or identify outbreaks that are not reported through traditional surveillance is not known, and, despite their increasing use, few systems have been evaluated (*24*). Having a better understanding of the experiences of health departments that use syndromic surveillance systems might help to improve the usefulness of this methodology.

Methods

Since November 2001, the New York City Department of Health and Mental Hygiene (DOHMH) has operated an ED syndromic surveillance system (*2*). Every day, EDs transfer electronic data to DOHMH regarding the age, sex, home ZIP code, date and time of visit, and chief complaint of patients examined the previous day. A computer algorithm codes chief complaints into four syndromes: vomiting, diarrhea, fever, and respiratory; complaints that do not fit these categories are coded as "other." Data are analyzed daily for aberrations in time and space, which are reported as either citywide or spatial signals. Spatial signals indicate clustering in syndrome visits by either hospital or patient home ZIP code.

For this report, DOHMH reviewed ED gastrointestinal (GI) syndromic surveillance data collected during November 15, 2001–August 15, 2004, to determine whether GI syndromic signals represented real disease clusters and whether syndromic surveillance detected known GI outbreaks. During the study period, the data collection system increased from 28 (42%) of 67 EDs, representing approximately 57% of ED visits in NYC, to 48 (73%) of 66 EDs, representing approximately 90% of ED visits (one ED had closed during that period).

To determine whether GI syndromic surveillance signals represented real outbreaks, DOHMH reviewed 236 GI signals (e.g., vomiting or diarrhea) detected during November 15, 2001–August 15, 2004, together with any documented signal investigations conducted during this period. An analyst and physician jointly decided whether to begin, and how far to pursue, an investigation.

Multiple possible steps are involved in an investigation of citywide or spatial signals. For a citywide signal, hospital-level data are evaluated to determine whether one or multiple hospitals account for the majority of cases to focus the investigation, and, if so, ED clinical staff at these hospitals are asked whether they have noticed anything unusual and whether the trend is continuing. They are also asked to be aware of new patients reporting with the syndrome of concern and to notify DOHMH if they notice clusters of patients with similar symptoms or young and otherwise healthy patients with severe symptoms. For a spatial signal, the ED patient line list is reviewed to determine the age and chief complaints of patients in the cluster before the ED is called. For both citywide and spatial signals, other syndromic surveillance data sources (e.g., records of sales of OTC medications) are reviewed for corroboration. If concern persists after the line list has been reviewed and clinicians have been contacted, the midday 12-hour chief complaint log is requested from hospitals in the signal to determine if the trend is continuing. Eight hospitals can send midday logs electronically; other hospitals photocopy their paper logbooks for the period from midnight to midday, black out identifying information, and send them to DOHMH by facsimile. Faxed logs are then hand coded and the proportion of syndrome visits compared with the signal and 7-day baseline. Depending on the size and timing of the signal, whether or not it is sustained, or other information suggests that the signal indicates a true increase in illness, the signals might raise greater concern. For such signals, medical charts are abstracted by either hospital or DOHMH staff. Occasionally, patients have been called and asked whether they have improved; on one occasion, after a blackout in August 2003, a case-control study was conducted (*26*). Because a common microbial pathogen suggests a link among patients, an attempt was made to identify an etiologic diagnosis for signals of greater concern. However, obtaining specimens was challenging because the patient usually had been discharged from the ED by the time a signal was detected. Efforts to persuade EDs to augment their specimen collections have not succeeded because these laboratory studies typically do not affect clinical care and incur added effort, cost, and burden of tracking results. DOHMH staff members have occasionally gone onsite to collect specimens, but this activity is resource intensive, and despite arranging specimen transport, compliance has been low.

A total of 49 GI outbreaks investigated by DOHMH during November 15, 2001–August 15, 2004, were reviewed to determine whether syndromic surveillance can detect known GI outbreaks. All outbreaks involved ≥10 persons with vomiting or diarrhea symptoms. Outbreaks were reported to DOHMH from multiple sources (Table).

Results

Review of Syndromic Surveillance Signal Investigations

During the study period, investigations detected 236 GI signals, including 98 citywide and 138 spatial signals. Of these, 20 (8.4%) were documented in writing, including all investigations that determined microbial etiology. Signal investigations determined that annual citywide outbreaks of diarrheal illness were likely attributable to norovirus (typically during fall and winter) and rotavirus (typically during spring). Although no etiology was determined, one citywide increase in diarrhea after the August 2003 blackout was believed to have represented a true increase in diarrheal illness (*26*). No other citywide GI signals were linked to disease outbreaks. A total of 73 (75%) signals occurred during annual seasonal outbreaks of norovirus and rotavirus (Figure 1). No spatial signal was linked to an outbreak. However, chart reviews are often

TABLE. Sources of reported gastrointestinal illness outbreaks* — New York City, November 15, 2001–August 15, 2004

Source	No.
Patient self-report	24
Health-care provider	7
Department of Education	6
Other health department	2
Restaurant	2
Analysis of reportable diseases	4
Combined reports from patient and health-care provider	1
Combined analysis of reportable diseases and report by health-care provider	1
Other city agency	1
Unknown caller to a city telephone hotline	1
Total	**49**

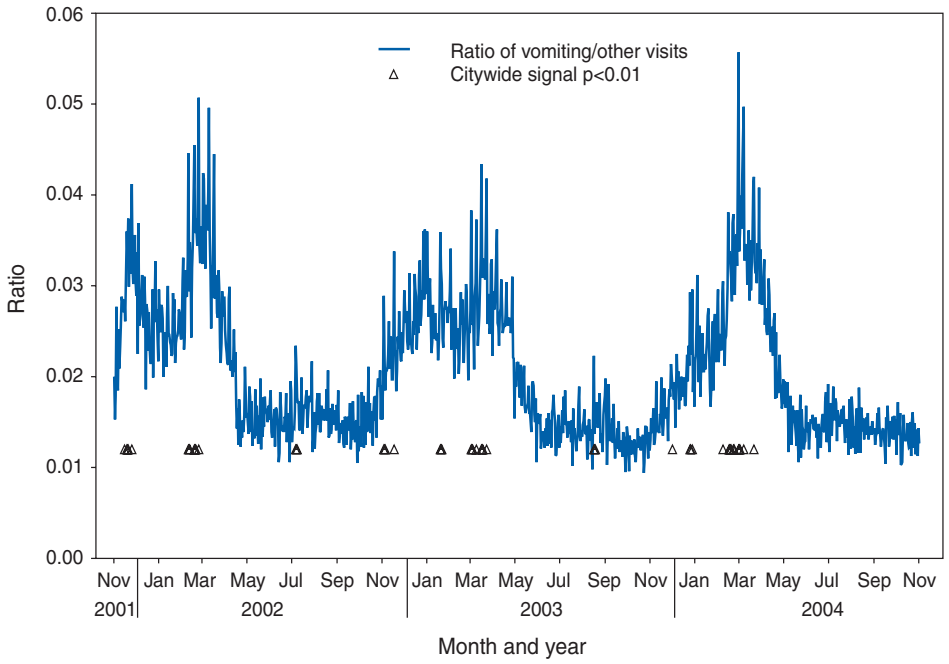
* Involving ≥10 persons with vomiting or diarrhea symptoms.

unrevealing because medical histories are only briefly documented, especially with regard to risk exposures, and laboratory work-ups performed during a typical ED visit are minimal, especially for GI illness. No onsite signal investigations have demonstrated a connection among cases.

On November 13, 2002, as a result of citywide diarrheal signals during October 27–28 and November 3–5 and citywide vomiting signals during November 7–12, a health alert was issued to physicians in the community. This alert noted that in addition to the citywide signals, stool specimens collected during October 2002, before the first citywide signal, were positive for calicivirus. The alert asked physicians to increase diagnostic testing to help DOHMH better understand these trends and prevent illness. Physicians also were asked to emphasize hand hygiene, proper cleaning of vomitus, and the need for persons to stay home when ill. A similar alert was issued on August 17, 2003, after the postblackout signal. This alert also reminded providers to advise patients to discard perishable food purchased before the blackout. In addition, a press release with this message was issued.

During February 16–17, 2004, a citywide signal occurred involving 1,803 observed cases of vomiting and 942 observed cases of diarrhea, compared with an expected 1,487 and 729 cases, respectively. DOHMH piloted having a chain of primary care clinics assist in the specimen collection. Because any citywide outbreak of diarrheal illness would likely have been detected in outpatient clinics, DOHMH sent specimen collection kits to five outpatient clinics with supplies for four children with a chief complaint of vomiting or diarrhea. Three specimens were requested per child: two rectal swabs* obtained in the clinic for ova and parasite and for culture and sensitivity testing, and a stool collection cup for viral pathogens, which was sent home with the family. Specimen transport from the patient home was arranged for viral specimens. Of the 20 distributed kits, specimens on 10 children were returned to DOHMH; however, a substantial proportion (percentages varied by test) were inadequate for testing. Three were tested for ova and parasites and were negative, nine were tested for culture and sensitivity, and five were forwarded to the New York State Department of Health's Wadsworth Laboratory for viral testing. Four tested positive for calicivirus, a norovirus. On March 24, 2004, results were received by DOHMH, 6 weeks after the citywide signal investigation began. Evidence from other outbreaks investigated at the time in schools, restaurants, and institutional settings suggested that norovirus was circulating in the community.

FIGURE 1. Emergency department visits for vomiting syndrome, all ages, by month and year — New York City, November 2001–October 2004



Review of GI Outbreak Investigations

Of 49 GI outbreaks investigated during November 15, 2001–August 15, 2004, none was detected by the ED syndromic surveillance system. In 36

* Although rectal swabs are not the optimal means for collecting a specimen for ova and parasite testing, this method was used because of the difficulty in obtaining more than one stool specimen from patients.

outbreaks, few or no patients went to the ED, and in two outbreaks, a substantial proportion of the patients were visitors to NYC who returned to other jurisdictions before the onset of symptoms. In 11 outbreaks, patients reported to NYC EDs, but, for multiple reasons, no signal occurred. In three outbreaks, patients reported to NYC EDs not in DOHMH's system; in three outbreaks, patients reported over several days or weeks; in two outbreaks, patients reported as a group and were coded in the triage log by a group code (e.g., "school incident") that did not translate to a syndrome (*27*); and in two outbreaks, a combination of these problems occurred. To further understand why outbreaks might not have been detected by the syndromic surveillance system, DOHMH conducted a detailed retrospective examination of an outbreak at a grade school that was reported by traditional means by the ED physician the day it occurred. A traditional outbreak investigation indicated that the outbreak involved 150 children, 65 of whom reported to the same hospital ED in which the reporting physician worked. However, this ED was not included in the DOHMH system. Of 79 case-patients who were interviewed, all reported vomiting, and 75% reported diarrhea. One stool culture grew a norovirus.

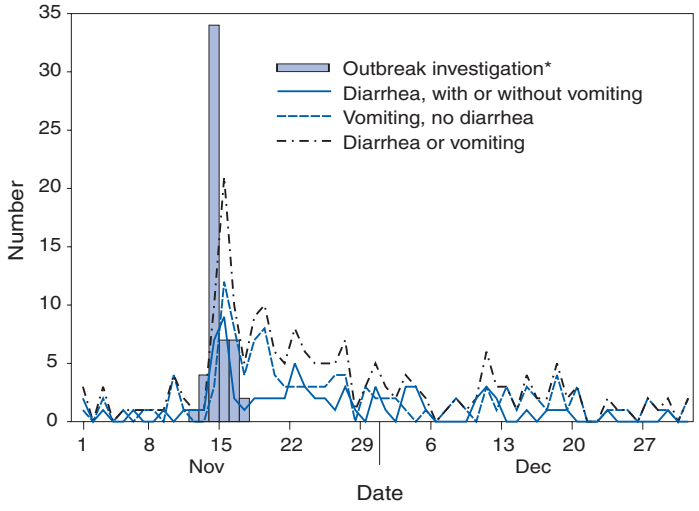
Data were obtained from this ED for a 2-week period before and during the outbreak, and routine daily analyses were run again to explore whether the outbreak could have been detected by the system. On the first day of the outbreak, a significant ZIP code signal (six observed compared with 0.2 expected; $p<0.001$) was detected for diarrhea syndrome in all age groups in the ZIP code in which the school is located. Borderline clustering also was observed at two hospitals in the hospital-based analysis (29 observed compared with 16 expected; $p = 0.08$). As this hospital cluster was of borderline significance, it likely would not have been investigated. The significant ZIP code cluster is also unlikely to have triggered an investigation because the number (six) of excess cases was not unusual (ranked 25 of 138 spatial clusters). No signals occurred the following day. More robust signals would have been generated if a different set of analyses had been employed. For routine analyses, ED visit records are examined separately for vomiting and diarrhea; for each syndrome, all ages are analyzed together. An analysis that combined diarrhea and vomiting into one syndrome would have been more appropriate for this outbreak because certain children were coded as having diarrhea, and others were coded as having vomiting. Examining diarrhea and vomiting separately diluted the signal into two smaller signals. In addition, analyses focused on children aged 5–17 years generated a stronger series of signals on the day of this grade school outbreak (hospital diarrhea signal: seven observed, compared with one expected [$p = 0.002$]; ZIP code diarrhea signal: eight observed, compared with one expected [$p = 0.009$]) and on the following day (hospital diarrhea signal: nine observed, compared with two expected [$p = 0.01$]; ZIP code diarrhea signal: nine observed, compared with two expected [$p = 0.007$]; and ZIP code vomiting signal: 13 observed, compared with four expected [$p = 0.01$]) (Figure 2). However, each analysis added to the daily routine would increase the total numbers of signals observed. For example, when daily analyses were simulated for four age categories (0–4, 5–17, 18–59, and ≥60 years) and three GI syndrome categories (diarrhea and vomiting alone or in combination), an additional 296 GI signals were generated annually.

Discussion

In NYC, syndromic surveillance has proven useful primarily for detecting and monitoring annual citywide outbreaks of norovirus, rotavirus, and influenza. However the utility of this information for preventing illness is uncertain (*28*). For norovirus and rotavirus, the only public health intervention that can be offered is outreach to child-care and school settings regarding the importance of hand washing and of ensuring that children stay home when ill. For norovirus, the etiologic diagnosis can be difficult to make because commercial testing is not readily available, so syndromic surveillance might permit a better description of the disease epidemiology. For influenza, syndromic surveillance might allow public health officials to recognize the start of the season in a more timely manner so vaccine prevention messages can be emphasized. However, surveillance systems that also include a laboratory component are paramount to confirm that influenza has arrived and to identify circulating strains.

Syndromic surveillance systems have also provided reassurance during times of concern (e.g., the 2001 anthrax attacks) and states of

FIGURE 2. Number of emergency department visits by persons aged 5–17 years, by syndrome code and date of visit — New York City, November 1–December 31, 2002



* N = 55 persons for whom date of onset was known.

elevated security alerts (e.g., during the 2004 Republican National Convention) that an excess number of patients citywide has not sought ED care for acute illnesses. Although envisioned as an early warning system, syndromic surveillance has thus far functioned more as a back-up system to traditional reporting. Constructing a syndromic surveillance system that detects statistical aberrations in the number of citywide ED visits has not been technically difficult. What has proven difficult is determining a rational, timely and resource-efficient response to signal investigation. By the time an increase in citywide ED visits is investigated by using existing methods and the etiology is determined to be either a natural or an intentional outbreak, the problem is likely to be widespread.

The NYC syndromic surveillance system originated in part from the need to perform enhanced surveillance for cryptosporidiosis because the NYC water supply is not filtered. The 1993 cryptosporidiosis outbreak in Milwaukee was detected by reports to the city health department of widespread absenteeism and substantial increases in sales of OTC antidiarrheal medications (*29,30*). Delay in detection of this outbreak has been attributed to multiple shortcomings of disease surveillance. Cryptosporidiosis was not a reportable disease at the time. Patients with mild symptoms, especially those who are immunocompetent, usually do not seek medical care for diarrhea, and the majority of persons affected recover without treatment. Diagnostic tests are seldom ordered for those who do seek medical care, or the diagnosis is delayed because *Cryptosporidium* is not considered in the differential diagnosis, and the specific test is not included in standard ova and parasite examinations (*30*). Considering these issues, retrospective analyses of syndromic surveillance systems, including surveillance of OTC medication sales, clinical lab submissions for any test on a stool specimens, nursing home surveillance, and ED surveillance have suggested that aberrations would have been noticeable weeks before detection of the waterborne outbreaks of cryptosporidiosis (*10,30,31*). Because no such waterborne outbreaks have occurred in NYC, whether such aberrations would have led to early detection and intervention cannot be determined.

Syndromic surveillance has not been useful in detecting acute localized GI outbreaks in NYC, in part because signal investigations to determine etiologic and epidemiologic links among patients are difficult and time consuming. The primary problem with using syndromic surveillance to prospectively detect outbreaks is that analyses that are sensitive enough to detect smaller outbreaks signal falsely so often that they generate too many signals from which to distinguish genuine outbreaks. Without diagnostic or epidemiologic data, whether the apparent cluster represents patients who are linked or even have the same etiologic cause for their symptoms cannot be determined easily. Multiple studies have analyzed whether ICD-9–coded discharge diagnoses yield better results for syndromic surveillance analyses (*32–34*), but little diagnostic work-up is performed on ED patients. Although less a problem during large-scale citywide outbreaks, misclassification obscures limited, localized signals caused by real outbreaks and can cause spurious signals composed of unrelated cases. Enhancing existing syndromic surveillance systems might improve their usefulness by increasing the specificity and positive predictive value of signal detection and investigation. Data streams containing laboratory and radiologic findings on patients could help rapidly determine if a cluster is more concerning to help prioritize signal investigations. The development of rapid, multiplex, point-of-care diagnostic assays that allow clinicians to rapidly include a natural cause such as influenza or exclude potential biologic terrorist agents would greatly improve the ability to determine whether an outbreak is occurring and its cause. Such improvements will only help if ED clinicians order the diagnostic tests; however, such tests might not be ordered for persons who have mild or moderate illness.

Health departments receive reports of disease clusters from multiple sources and set priorities regarding use of limited staff resources. If insufficient information exists to initiate an investigation, the decision is often made to observe whether the signal continues the next day, thereby losing syndromic surveillance's theoretical advantage of timeliness. Regardless of whether an investigation is begun immediately, the decision to launch a public health intervention (e.g., vaccination or antibiotic prophylaxis) requires that an etiologic diagnosis be determined. Syndromic surveillance might prove useful for detecting a problem and quantifying its magnitude but, by its very design, cannot determine the true etiology.

This evaluation is subject to at least two limitations. First, the analysis reviewed only the ED syndromic surveillance system and was focused on the system's ability to detect GI outbreaks. As localized respiratory outbreaks are reported less commonly by traditional surveillance systems, experience with fever and respiratory signals could not be evaluated. Second, other systems using different data might have a greater ability to detect outbreaks. A system using outpatient data is being piloted in NYC.

Conclusion

Concerns about biologic terrorism have generated substantial financial support for development of syndromic surveillance detection systems (*28,35*) at the local, state, and national levels. However, the utility of these systems has not been demonstrated. The rareness of biologic terrorism means that syndromic surveillance systems can be evaluated only by their ability to detect naturally occurring outbreaks in a timely manner. Given the increasing investments being made in

syndromic surveillance, communities should examine their systems and report their findings. Critical evaluations are needed to determine whether the resources spent by public health agencies conducting signal investigations, which cannot then be used elsewhere, are worth the theoretical benefits of detecting outbreaks more quickly.

References

1. Henning KJ. What is syndromic surveillance? In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53 (Suppl):5–11.
2. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorf M, Weiss D. Syndromic surveillance in public health practice, NYC. Emerg Infect Dis 2004;10:858–4.
3. Paladini M. Daily emergency department surveillance system—Bergen County, New Jersey. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):47–9.
4. Wagner M, 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. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):28–31.
5. Yuan CM, Love S, Wilson M. Syndromic surveillance at hospital emergency departments—Southeastern Virginia. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):56–8.
6. Dembek ZF, Carley K, Siniscalchi A, Hadler J. Hospital admissions syndromic surveillance—Connecticut, September 2001–November 2003. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):50–2.
7. 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.
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. 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):i143–9.
10. Miller JR, Mikol Y. Surveillance for diarrheal disease in New York City. Urban Health 1999;76:388–90.
11. Das D, Mostashari F, Weiss D, Balter S, Heffernan R. Monitoring over-the-counter pharmacy sales for early outbreak detection—New York City, August 2001–September 2003. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2003;53(Suppl):235.
12. Rodman JS, Frost F, Davis-Burchat L, Fraser D, Langer J, Jakubowski W. Pharmaceutical sales—a method of disease surveillance? Environmental Health 1997;60:8–14.
13. Wagner M, Tsui F-C, Hogmn W, et al. National retail data monitor for public health surveillance. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):40–2.
14. Rodman JS, Frost F, Jakubowski W. Using nurse hot line calls for disease surveillance. Emerg Infect Dis 1998;4:329–32.
15. Cooper DL, Smith G, Baker M, et al. National symptom surveillance using calls to a telephone health advice service. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):179–83.
16. Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection—New York City, 2001–2002 [Abstract]. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):230.
17. CDC. Update: influenza activity—United States, January18–24, 2004. MMWR 2004;53:63–5.
18. Robertson SE, Suleiman AJM, Mehta FR, Al-Dhahry SHS, El-Bualy MS. Poliomyelitis in Oman: acute flaccid paralysis surveillance leading to early detection and rapid response to a type 3 outbreak. Bull World Health Organ 1994;72:907–14.
19. Welliver RC, Cherry JD, Boyer KM, et al. Sales of nonprescription cold remedies: a unique method of influenza surveillance. Pediat Res 1979;13:1015–7.
20. Lenaway DD, Ambler A. Evaluation of a school-based influenza surveillance system. Public Health Rep 1995;110:333–7.
21. Loonsk JW. BioSense—a national initiative for early detection and quantification of public health emergencies. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):53–5.
22. 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.
23. Lober WB, Bryant TK, Wagner MW, et al. Roundtable on bioterrorism detection: information system–based surveillance. J Am Med Inform Assoc 2002;9:105–15.
24. Bravata DM, McDonald KM, Smith WM, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Int Med 2004;140:910–22.
25. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis 2003; 9:1197–204.
26. 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. Am J Public Health; In press.
27. Steiner-Sichel L, Greenko J, Heffernan R, Layton M, Weiss D. Field investigations of emergency department syndromic surveillance signals—New York City. In: Syndromic surveillance: reports from a national conference, 2003. MMWR 2004;53(Suppl):184–9.
28. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 2003;1:77–81.
29. Mac Kenzie W, 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.
30. 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.
31. Edge VL, Pollari F, Lim G, et al. Syndromic surveillance of gastrointestinal pharmacy over-the-counter sales. Can J Public Health 2004;95:446–50.
32. 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.
33. Fleischauer AT. The validity of chief complaint and discharge diagnosis in emergency department-based syndromic surveillance. Acad Emerg Med 2004;11:1262–7.
34. Ivanov O, Wagner MM, Chapmman WW, Olszewski RT. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. Proc AMIA Symp 2002:345–9.
35. Sosin DM. Syndromic surveillance: the case for skillful investment. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 2003;1:247–53.

Abstracts

Public Health Surveillance for World Youth Day — Toronto, Canada, 2002

Kate L. Bassil,[1] B. Henry,[2] E. Rea,[1] M. Varia,[3] D. Cole[1]
[1]*University of Toronto, Toronto, Canada;* [2]*Toronto Public Health, Toronto, Canada;* [3]*Health Canada, Ottawa, Ontario, Canada*

Corresponding author: Kate L. Bassil, New College, 45 Willcocks Street, Toronto, Canada, ON M5S 1C7. Telephone: 416-946-0353; Fax: 416-971-3072; E-mail: kate.bassil@utoronto.ca.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: World Youth Day (WYD) is a biannual international Catholic event for persons aged 17–35 years. During July 13–28, 2002, WYD activities took place in Toronto, Canada; 176,100 persons registered to participate, and 800,000 persons attended an overnight vigil and papal mass. Because of the potential for imported communicable disease and local disease outbreaks, Toronto Public Health, Health Canada, and WYD organizers developed a surveillance network to facilitate timely public health response.

Objectives: This study assessed the effectiveness of the syndromic surveillance network established for WYD 2002.

Methods: The surveillance network collected data from multiple sources, including 1) all WYD medical facilities, 2) emergency departments (EDs), 3) pharmacies, and 4) 9-1-1 emergency calls. Surveillance activities were coordinated from a downtown Toronto office staffed by physicians, epidemiologists, community medicine residents, and administrators. Findings were communicated daily to public health and WYD authorities. Ten syndrome definitions were developed on the basis of outbreak potential, public interest, and need for timely detection. Cumulative mean count, two standard deviations from the rolling 7-day mean, and CUMSUM methods were used for the analysis.

Results: Although no substantial outbreaks occurred, enough activity was noted to indicate that an event would have been detected if it had occurred. Activity included a case of malaria, a case of chickenpox, and an outbreak of foodborne illness involving 18 persons. In addition, 3,332 (21%) of 15,717 ED visits, 4,394 (39%) of 11,250 calls to 9-1-1, and approximately 35% of onsite clinic visits met syndrome definitions. Heat-related illness was the most prevalent event documented, with an increased proportion of 9-1-1 calls and the most common syndrome (n = 105) above two standard deviations from the rolling mean reported through EDs. Heat-related illness also was the most frequent onsite clinic diagnosis received by WYD participants among >5,000 visits during 6 days.

Conclusion: For an event-specific syndromic surveillance network to be effective, multiple data sources and redundancy are needed. A range of communication channels, back-up methods for data collection, and complementary surveillance components were employed for this event. Substantial time, resources, and planning were required for the implementation of this surveillance network. However, certain activities that are feasible for event-specific surveillance are difficult to sustain on an ongoing basis because of a lack of resources. For this reason, monitoring 9-1-1 calls appears to offer the greatest potential usefulness for ongoing public health surveillance.

Effect of Site of Care and Age on Timeliness and Accuracy of Syndromic Surveillance Data

John S. Brownstein,[1,2,3] K. Olson,[1,2,3] K. Kleinman,[4] K. Mandl[1,2,3]
[1]*Children's Hospital Informatics Program, Children's Hospital Boston, Boston, Massachusetts;*
[2]*Division of Emergency Medicine, Children's Hospital Boston, Boston, Massachusetts;*
[3]*Department of Pediatrics, Harvard Medical School, Boston, Massachusetts;*
[4]*Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, Massachusetts*

Corresponding author: John S. Brownstein, Division of Emergency Medicine, Children's Hospital Boston, 1 Autumn St., Rm. 542, Boston, MA 02115. Telephone: 617-355-6998; Fax: 617-730-0267; Email: john.brownstein@childrens.harvard.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Staff of emergency and ambulatory care departments encounter data that have unique advantages for timely outbreak detection. Patient age distribution might have an effect on timeliness and accuracy of prediction.

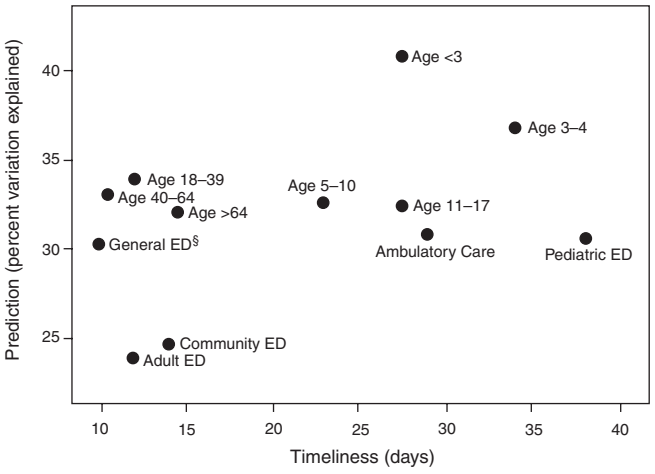
Objectives: This report identifies patient subpopulations on the basis of sites of care and patient age distribution, which signal influenza epidemics earliest and most accurately.

Methods: Analyses performed included cross-sectional, time series analyses of seven patient populations comprising a health maintenance organization providing ambulatory care, three emergency departments (EDs) at urban tertiary care, and three community-based hospitals in eastern Massachusetts. Except for two EDs (one adult and one pediatric), these hospitals serve all patient age groups. Patients having respiratory infection syndromes and who visited each health-care setting during January 1, 2000–September 30, 2004, were identified and categorized by age (Figure). Cross-spectral analyses and Poisson regression models were used to evaluate timelines and prediction for New England's influenza and pneumonia mortality, as reported to CDC.

Results: Patient age significantly influences timeliness of signal for influenza and pneumonia mortality ($p = 0.026$), with the pediatric ED patients presenting with influenza earliest in the season. In the cohorts, children aged 3–4 years consistently presented to sites of care first ($p<0.05$). By using regression models to predict mortality based on the time-shifted surveillance data, all cohorts were identified as significant predictors of influenza mortality. However, patient age also significantly influences level of prediction ($p = 0.036$). The age group of children aged <3 years can be used to predict significantly more of the variation than other age groups ($p<0.05$).

Conclusion: Patient age is a key determinant in the timing of visits for respiratory infections. Pediatric patients seek ambulatory and emergency care before adult patients. The earliest arriving group is preschool-aged children (aged 3–4 years), who are considered predominant vectors in household spread of influenza. The age group of children <3 years can be used to best predict influenza mortality, again highlighting the importance of treating pediatric patients as sentinels. Monitoring pediatric patient subpopulations separately from other subpopulations might enhance syndromic surveillance systems.

FIGURE. Prediction and timeliness* of patients with respiratory infection syndromes for signaling influenza and pneumonia mortality, by site of care and patient age† (years) — Eastern Massachusetts, 2000–2004



* Lead times to influenza mortality obtained by cross-spectral analysis are plotted against the proportion of variance explained by poisson regression.
† Data for both sites of care (blue circles) and age groups across sites (black circles) are displayed.
§ Emergency department.

Three Stages of Evaluation for Syndromic Surveillance from Chief-Complaint Classification — Pennsylvania and Utah

Wendy W. Chapman, J. Dowling, O. Ivanov, B. Olszewski, M. Wagner
Real-Time Outbreak Detection System (RODS) Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania

Corresponding author: Wendy W. Chapman, RODS, Forbes Tower, Suite 8084, Pittsburgh, PA 15213. Telephone: 412-647-7167; Fax: 412-647-7190; E-mail: chapman@cbmi.pitt.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Evaluations from multiple perspectives are needed to determine whether syndromic classification of chief complaints is useful for outbreak detection.

Objective: This study quantified the performance of a naïve Bayesian classifier, Complaint Classifer (CoCo), at syndromic classification from chief complaints by using a three-stage evaluation process.

Methods: First, CoCo was evaluated to determine its level of technical accuracy in answering the question, "Can we accurately classify a chief-complaint string into a syndromic category?" For example, the area under the ROC curve of CoCo classifications were calculated into eight syndromes for 28,990 chief complaints from 30 hospitals in Utah during a 1-month period (Olszewski RT. Bayesian classification of triage diagnoses for the early detection of epidemics. In: Proceeding of the Florida Artificial Intelligence Research Society Conference; May 12–14, 2003; St. Augustine, FL. Menlo Park, CA: AAAI Press; 2003:412–6). Standard classifications were made by a physician reading only the chief complaints. Second, CoCo was evaluated to determine its performance at case classification to answer the question, "Does the syndromic classification from the chief complaint accurately represent the patient's clinical state?" For example, the sensitivity and specificity of the CoCo classification of 527,228 patients over a 13-year period in a single hospital in Pittsburgh, Pennsylvania was measured (Chapman WW, Dowling JN, Wagner MM. Classification of emergency department chief complaints into seven syndromes: a retrospective analysis of 527,228 patients. Ann Emerg Med. In press 2005.). Reference standard classifications were assigned by syndromic groups of primary *International Classification of Diseases, Ninth Revision* (ICD-9) discharge diagnoses. Third, CoCo was evaluated to determine its performance at outbreak detection to answer the question, "How timely and accurately can we detect a public health outbreak by monitoring chief-complaint classifications?" For example, by using the Exponentially Weighted Moving Average (EWMA) detection algorithm, the factors measured were timeliness, sensitivity, and specificity of chief complaints classified by CoCo for predicting outbreaks of pediatric respiratory and gastrointestinal illness (Ivanov O, Gesteland P, Hogan W, Mundorff MB, Wagner MM. Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. In: Proceedings of the American Medical Informatics Association Annual Fall Symposium; November 8–12, Washington, DC. Bethesda, MD: American Medical Informatics Association; 2003: 318–22.). Reference standard classification comprised ICD-9 discharge diagnoses of pneumonia, influenza, and bronchiolitis for respiratory illness and rotavirus and pediatric gastroenteritis for gastrointestinal illness.

Results: For technical accuracy, areas under the ROC curve ranged from 78% for botulinic syndrome to 96% for respiratory syndrome. For case classification, sensitivity and specificity, respectively, were as follows: respiratory: 63%, 94%; botulinic: 30%, 99%; gastrointestinal: 69%, 95%; neurologic: 67%, 93%; rash: 47%, 99%; constitutional: 46%, 97%; and hemorrhagic: 75%, 99%. For outbreak detection, three respiratory and three gastrointestinal outbreaks were detected by CoCo with 100% sensitivity and specificity. Time series of chief complaints correlated with hospital admissions and preceded them by an average of 10.3 days for respiratory outbreaks and 29 days for gastrointestinal outbreaks.

Conclusion: Three stages of evaluation are useful in determining the performance of syndromic surveillance from chief complaints. CoCo was evaluated to determine its ability to classify patients into prevalent syndromes (e.g., respiratory and gastrointestinal) and into syndromes that are rare and difficult to characterize (e.g., hemorrhagic, botulinic, and constitutional). Chief-complaint classification might be useful for detecting moderate to widespread outbreaks; however, to increase sensitivity and specificity, the techniques in this report should be extended to other clinical information sources, including chest radiograph and emergency department reports.

Using Self-Testing to Augment Syndromic Surveillance — United Kingdom, December 2003–January 2004

Duncan L. Cooper,[1] G. Smith,[1] P. Loveridge,[1] F. Chinemana,[2] E. Gerard,[2] C. Joseph,[3] M. Baker,[4] D. Mant,[5] J. Watson,[3] R. Griffiths,[5] M. Zambon[6]
[1]Health Protection Agency, West Midlands, England; [2]NHS Direct, London, England; [3]Communicable Disease Surveillance Centre, Health Protection Agency, London, England; [4]Royal College of General Practitioners, London, England; [5]University of Oxford, England; [6]Health Protection Agency, London, England

Corresponding author: Duncan Cooper, Health Protection Agency, Regional Surveillance Unit, Floor 2, Lincoln House, Heartlands Hospital, Bordesley Green East, Birmingham, B9 5SS, United Kingdom. Telephone: 44-121-773-7077; Fax: 44-121-773-1407; E-mail: duncan.cooper@hpa.org.uk.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Effective community surveillance is needed for the rapid identification of outbreaks of serious illness. In the United Kingdom, surveillance of calls to NHS Direct, a national telephone health helpline, are used as a method of rapid outbreak detection. A limitation of this syndromic surveillance approach is the lack of laboratory confirmation of diagnosis after surveillance signals have been generated.

Objective: This report presents a pilot study to investigate the feasibility of virologic sampling conducted by NHS Direct callers.

Methods: During December 2003–January 2004, NHS Direct nurses in two regions of England were asked to recruit NHS Direct callers aged >12 years who had reported cold or influenza symptoms (i.e., "cold/flu" callers). Persons who agreed to participate in the study were mailed a specimen kit. Callers were asked to take a swab from each nostril and return the swabs by mail to the United Kingdom national influenza reference laboratory. Swabs were tested by multiplex polymerase chain reaction (PCR) for influenza viruses and, if identified as positive, were cultured for viable virus isolation.

Results: During the study period, 686 cold/flu callers were eligible for the study. Although 67 of these were recruited for the study, determining how many cold/flu callers were asked to participate but refused was impossible. Of the 67 specimen kits sent, 36 (54%) were returned to the laboratory. The mean time between the call to NHS Direct and laboratory analysis of the specimen was 7.5 days. This period was shorter for positive samples (mean time: 6.12 days) than negative samples (mean time: 7.9 days), although the difference was not significant ($p = 0.13$). Eight specimens (22%) were positive on PCR for influenza virus. Five were antigenically characterized as Fuijian/411/2002-like influenza A H3N2. Higher positivity rates might have been achieved if the sampling study had begun earlier in the year before the peak of the influenza season.

Conclusion: This study demonstrates the possibility of community-based clinical surveillance that does not require sampling by a health-care worker. This methodology will allow novel approaches to be developed to integrate syndromic surveillance with virologic sampling. Therefore, the rapid follow up of syndromic surveillance signals might provide confirmation of a specific common infection or provide evidence of a potentially more sinister cause.

Using Modified Spatial Scan Statistic to Improve Detection of Disease Outbreak When Exposure Occurs in Workplace — Virginia, 2004

Luiz Duczmal,[1] D. Buckeridge[2,3]
[1]*Universidade Federal de Minas Gerais, Brazil;* [2]*Veterans Affairs Palo Alto Healthcare System, Palo Alto, California;* [3]*Stanford University, Stanford, California*

Corresponding author: Luiz Duczmal, Statistics Department, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901. Telephone: 55-31-3499-5900; Fax: 55-31-3499-5924; E-mail: duczmal@est.ufmg.br.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Detecting a disease outbreak is more difficult when the exposure occurs in a workplace but only the patient's home address is available for analysis. In these situations, application of the customary spatial scan statistic designed by Martin Kulldorff does not account for possible differences between home and work addresses, thereby reducing the power of detection.

Objectives: This study examined whether modifying Kulldorff's spatial scan statistic to take into account the movement of persons between home and work can improve detection of disease outbreaks when exposure occurs in the workplace.

Methods: The study region was partitioned into m cells $Z(1),\ldots,Z(m)$. $L(k,i)$ is the proportion of the population living in cell $Z(k)$ that works at cell $Z(i)$. For each cell $Z(i)$, $i = 1,\ldots,m$, consider the r nearest cells from $Z(i)$, $r = 1,\ldots,R$ as the location of a possible outbreak that occurs during working hours. For each i and each r, build the m zones $Y(1),\ldots,Y(m)$, adjoining successively the residential cells indicating where the workers from the r nearest cells from $Z(i)$ live, in decreasing order of proportion of workers within these cells. The factors $L(k,i)$ are used to compute the observed cases in the residential zones attributable to the contamination from workers at the r nearest neighbors of cell $Z(i)$. This quantity, with the corresponding expected number of cases, is used to build the modified spatial scan statistic, similar to the usual spatial scan statistic. The modified scan statistic is computed m^2R times, and the maximum value obtained indicates the most likely pair of outbreak focus and associated residential area found. A Monte Carlo procedure is used to compute the p-value of the most likely pair. The study region consisted of 158 ZIP codes located near Norfolk, Virginia. The following three typical simulated clusters, with their corresponding ZIP codes, are representative of much more extensive simulations: 1) Cluster A: 23601, 23606, 23607, 23661, 23666, 23668, and 23669; 2) Cluster B: 23601, 23602, 23606, 23665, 23666, and 23693; and 3) Cluster C: 23666 and 23669.

Results: Power evaluations of 0.85 (A), 0.70 (B), and 0.53 (C) were obtained by using the modified scan statistic compared with 0.68 (A), 0.52 (B), and 0.42 (C) obtained by using Kulldorff's spatial scan statistic.

Conclusion: Using a modified scan statistic that takes into account the movement of persons between home and work might be a useful complementary tool for the early detection of outbreaks in the workplace. Through simulations, a statistically significant increase in power was observed compared with the usual spatial scan statistic.

Performance-Critical Anomaly Detection — United States, December 2002–March 2004

Colin R. Goodall,[1] A. Lent,[1] S. Halasz,[1] E. Koski,[2] D. Agarwal,[1] S. Tse,[1] G. Jacobson[1]
[1]*AT&T Labs (Research), Middletown, New Jersey;* [2]*Quest Diagnostics Incorporated, Teterboro, New Jersey*

Corresponding author: Colin R. Goodall, AT&T Labs, 200 S. Laurel Ave. D4 3D28, Middletown, NJ 07760. Telephone: 732-420-5816; Fax: 732-368-7201; E-mail: cgoodall@att.com.

Disclosure of relationship: The contributors of this report have disclosed that they are employees of AT&T Labs or Quest Diagnostics, Inc., and that their employment compensation may include ownership of company stock. This report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Performance-critical anomaly detection for biomedical surveillance requires 1) reliable data that are both geotemporally and demographically representative; 2) efficient, real-time, large-scale information-processing capabilities; 3) comprehensive, tunable anomaly-detection algorithms; 4) a flexible platform for investigation and management of anomalies; and 5) alert distribution and management.

Objectives: This study analyzed a reliable, high-performance, end-to-end, modular process for early event detection that included data loading and transformation, statistical anomaly detection, and tools for user interaction.

Methods: The process architecture and implementation included three components: 1) a data layer, including modules for data loading, cleaning, normalization, coding, and aggregation; 2) an anomaly-detection layer, including multiple methods for statistical anomaly detection and an anomaly case manager; and 3) a presentation layer, including dynamic visualization of data (geographically, temporally, and logically) used in case investigation, publication, and process monitoring. Specific statistical anomaly detection methods used included process-control techniques; SaTScan™ (a free software program used to calculate spatial, temporal, and space-time scan statistics); a square-root technique; and a new adaptation of Bayesian shrinkage estimation (Kalman Filter Gamma Poisson Shrinker [KF GPS]) used to monitor a stream of events organized into a periodic (daily) array of cross-classified counts with geographic and medical dimensions. Shrinkage estimates were obtained of ratios of observed counts to proportionally fit expected counts that update smoothly with time after allowing for changes in marginal totals. KF GPS was used to model spatial associations and dependencies among the medical measurements. The case manager was used to organize groups of related anomalies into cases and to support collaboration, by providing a set of functions and software linkages for persons with subject-matter, statistical, and analytic expertise to use to investigate and manage anomalies. Each case could be resolved as an alert, deferred, or dismissed. The case manager included a logic-rich engine and two feature-rich, configurable tools for case organization and dynamic data visualization. Similar technology used by AT&T for telecommunications monitoring and case management in an environment in which >300 million calls are received daily was adapted to health-care data, including laboratory test and emergency room data, with comparable performance.

Results: In collaboration with Quest Diagnostics, Inc. (QDI), AT&T used a subset of QDI's nationwide testing data for December 2002–March 2004 for three syndromic groupings (respiratory, gastrointestinal, and heavy metals [lead]) in the New York City (NYC) metropolitan area and nationwide (lead only). The system computed approximately 600,000 scores, resulting in approximately 400 anomalies and their cases. Certain anomalies included a spike in overall respiratory test requisitions in the area of Bensonhurst, Queens, NYC; a spike in mycobacteria requisitions in Orange County, New York; and a change in data coding affecting viral tests in Bergen County, New Jersey.

Conclusion: This analysis demonstrated 1) the importance of end-to-end process architecture; 2) the utility of multiple algorithms, especially KF GPS, for anomaly detection; and 3) the effectiveness of using a case manager to investigate anomalies and reduce the burden of false positives. The system can handle massive data streams and allows rapid anomaly detection through use of a suite of analytic, data management, and visualization tools.

Syndromic Surveillance for Early Location of Terrorist Incidents Outside of Residential Areas

Manfred S. Green,[1,2] Z. Kaufman[1]
[1]*Ministry of Health, Tel Aviv, Israel;* [2]*Tel-Aviv University, Tel Aviv, Israel*

Corresponding author: Manfred S. Green, Israel Center for Disease Control, Gertner Institute, Sheba Medical Center, Tel Hashomer, Israel 52621. Telephone: 972-3-737-1500; Fax: 972-3-737-1515; E-mail: m.green@icdc.health.gov.il.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: A terrorist attack might occur at places outside of residential areas (e.g., workplace, entertainment venues, or shopping centers). Cluster analysis using available data on residential addresses alone will not yield the probable place of exposure.

Objectives: The study aimed to facilitate early detection of terrorist incidents in areas outside of residential areas (e.g., workplaces and shopping centers).

Methods: An approach was suggested for estimating the probable location of exposure on the basis of the distribution of the residential addresses on the assumption that persons tend to live closer to their place of work, or visit entertainment venues or shopping centers closer to their place of residence than would be expected by chance. A two-stage process of implementing available spatial statistics programs was proposed. In the first stage, a cluster analysis program was employed by using residential addresses. The SaTScan™ software, developed at the National Cancer Institute, was used with its space-time permutation model specifically developed for conducting continuous surveillance. If more than one substantial cluster is identified, a possible incident outside of the residential area was considered. The mean center and standard deviation are computed for those census tracts included in the substantial clusters to identify the area where the exposure might have occurred. For this task, the CrimeStat II™ was used, a spatial statistic program originally developed for analysis of crime locations. To narrow potential places of exposure, the cluster analysis should be performed in the first stage by age groups such as infants, children, persons aged <19 years, the working age category (19–65 years), and older persons (>65 years). Substantial clusters in the working age category alone could focus the resulting investigation on workplaces. A geographical information system (GIS) program (ArcGIS™, ESRI™, Redlands, California) was used for geocoding addresses and for other procedures needed to prepare data for cluster analysis and for presenting the results. This approach was demonstrated through a series of simulations of deliberate dispersion of anthrax spores in a large hospital where residential addresses of all the staff members were available. Simulated cases among the staff were superimposed on background data of patient visits to community clinics for influenza-like illness.

Results: A possible place of exposure was identified at a distance of approximately 1 kilometer from the hospital

Conclusion: The combination of multiple spatial statistics tools demonstrates promising capabilities for identifying terrorist incidents outside of residential areas, even when only residential addresses are available.

Leveraging Syndromic Surveillance During the San Diego Wildfires, 2003

Jeffrey M. Johnson, L. Hicks, C. McClean, M. Ginsberg
San Diego County Health and Human Services Agency, San Diego, California

Corresponding author: Jeffrey M. Johnson, San Diego County Health and Human Services Agency, Community Epidemiology Division, 1700 Pacific Hwy., MS P511C-A, San Diego, CA 92186. Telephone: 619-531-4945; E-mail: Jeffrey.johnson@sdcounty.ca.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: On October 25, 2003, one of the largest fires in California history began in San Diego County. Over a period of three days, the air quality deteriorated to unhealthy and hazardous levels, prompting school cancellations and the general public to stay at home. In response to the fires, smoke, and circulating ash, San Diego County Public Health leveraged existing syndromic surveillance capabilities to assess impact on the county's emergency medical system.

Objectives: This surveillance capability was rapidly deployed to assess the impact of the fires on selected types of emergency department visits.

Methods: In response to the fire, two existing syndromic surveillance data sources were monitored: prehospital paramedic transport chief-complaint data and local over-the-counter (OTC) medication sales data acquired from the National Retail Data Monitor system. In addition, 15 emergency departments reported syndromic surveillance information including asthma, bronchitis, emphysema, or other respiratory symptoms with no fever, eye irritation, smoke inhalation, burns, chest pain, and diarrhea. Daily air-quality data was also acquired. The analytic methods included time-series and process-control charts (e.g., P-Chart, U-Chart, CUSUM, and EWMA).

Results: Information on 31,321 emergency department visits, 8,625 prehospital transports, and OTC data were analyzed. Respiratory indicators demonstrated substantial increases during the days of greatest fire burn and unhealthy air quality, with postfire levels approaching prefire levels when air quality improved. A marked increase in smoke inhalation and eye irritation visits was also observed. No noticeable increase was noted among visits for chest pain or diarrhea. The total number of emergency department visits initially declined during the fire period, which corresponded to the days that students and employees were asked to remain at home. Air quality in San Diego deteriorated substantially during the fires concurrent with substantial increases in asthma-related emergency department visits and increases in local OTC sales of bronchial remedies, cold/cough syrup, and nasal products.

Conclusion: Existing syndromic surveillance capabilities were used to monitor the immediate impact of the wildfires in San Diego County. These results demonstrated a real impact on selected medical services. Certain fire-related outcomes were expected, especially related to asthma and other respiratory health outcomes and increased sales of selected OTC products. In retrospect, this disaster served as an "outbreak," validating the importance of syndromic surveillance as a dual-use tool and highlighting the need for system flexibility. Syndromic surveillance is a useful tool during a natural disaster, assisting future disaster preparations and generating hypotheses for long-term follow-up studies.

Using Data on an Influenza B Outbreak To Evaluate a Syndromic Surveillance System — Israel, June 2004

Zalman Kaufman,[1] E. Cohen,[2] T. Peled-Leviatan,[1] C. Lavi,[2] G. Aharonowitz,[1] R. Dichtiar,[1] M. Bromberg,[1] O. Havkin,[2] Y. Shalev,[3] R. Marom,[4] V. Shalev,[4] J. Shemer,[4,5] M. Green[1,5]

[1]Israel Center for Disease Control, Ministry of Health, Tel Hashomer; [2]Central District Health Office, Ministry of Health, Rechovot; [3]Central Virology Laboratory, Ministry of Health, Tel Hashomer; [4]Maccabi Healthcare Services, Tel-Aviv; [5]Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

Corresponding author: Manfred S. Green, Israel Center for Disease Control, Gertner Institute, Sheba Medical Center, Tel Hashomer 52621, Israel. Telephone: 972-3-737-1500; Fax: 972-3-737-1515; E-mail: m.green@icdc.health.gov.il.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Since 2002, as part of a national biologic terrorism preparedness program, the Israel Center for Disease Control (ICDC) has been developing a syndromic surveillance system based on data from community clinics and hospital emergency departments. Selected analytic tools are being evaluated for possible integration within this system.

Objectives: This study evaluated the performance of the What's Strange About Recent Events (WSARE) algorithm (http://www.autonlab.org/autonweb/showSoftware/159) for anomaly pattern detection when applied to records of daily patient visits to clinics of a local health maintenance organization (HMO).

Methods: Data from an influenza B outbreak that occurred in June 2004 in an elementary school in a small (population: approximately 7,000 persons) Israeli town were used. WSARE searches for groups with specific characteristics (e.g., a recent pattern of place, age, and diagnosis associated with illness that is anomalous when compared with historic patterns). The data set used was limited to 1) patients living in the county where the outbreak occurred; 2) a 35-day period during May–June 2004; and 3) records containing *International Classification of Diseases, Ninth Revision* (ICD-9) codes for signs, symptoms, and syndromes associated with infectious morbidity. On average, the data set included 510 records/day. Besides ICD-9 codes, data included date of visit to clinic, day of week, city/town code, and patient's age.

Results: Two successive significant anomalies ($p<0.0001$) were detected in the HMO data set that could signal the influenza outbreak, both sharing three constituents: 1) the town code; 2) the age category of affected children; and 3) the ICD-9 code for viral infection, which was the most prevalent diagnosis assigned by HMO physicians identified in an investigation by the regional health department. Had the data been available for real-time analysis, the first anomaly could have been detected on day 2, when the outbreak was first reported to public health officials.

Conclusion: A centralized, comprehensive surveillance system can rapidly detect localized, fast-developing outbreaks. Although early detection is hard to achieve in this instance, timely and reliable information produced by syndromic surveillance is of great value in supporting outbreak management and placing it in the context of the background morbidity in the country. However, had the outbreak occurred in winter, detection would have been more complex. When the outbreak data were superimposed on winter background, only a single significant two-constituents anomaly was detected at day 2 of the simulated outbreak, lacking the information to target on the specific age group of the schoolchildren.

Daily Electronic Disease Surveillance System — Bergen County, Paramus, New Jersey

Brian La Forgia,[1] L. Fiorenza,[1] S. John,[1] M. Paladini[2]
[1]*Bergen County Department of Health Services, Paramus, New Jersey;* [2]*New York City Department of Health and Mental Hygiene, New York, New York*

Corresponding author: Brian La Forgia, Bergen County Department of Health Services, 327 E. Ridgewood Ave., Paramus, NJ 07652. Telephone: 201-634-2843; Fax: 201-986-1068; E-mail: blaforgia@co.bergen.nj.us.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Emergency Department Surveillance System: The Bergen County Daily Electronic Disease Surveillance System (DEDSS) is composed of three discrete components used to analyze public health data. The core of the system is the Emergency Department Surveillance System (EDSS), which monitors hospital emergency department (ED) data for syndromes based on chief complaint data. Data from six local hospitals are sent daily in batches by file transfer protocol (FTP) to the department's Internet webserver, loaded into a Microsoft™ structured query language (SQL) server, and analyzed with SAS® (SAS Institute, Cary, North Carolina) software by using algorithms developed by Bergen County and the New York City Department of Health. Substantial improvements have been made by using the SQL server rather than the previous SAS database storage format.

First Watch™ System: In 2004, First Watch, a commercially available product, was added to DEDSS. The system adds real-time monitoring of requests for advanced life support through the Mobile Intensive Care Communications (MICCOM) paramedic dispatch center. First Watch monitors MICCOM's computer-aided dispatch system for complaint types.

Alerts are generated based on complaint types of interest to public health. The system monitors specified complaint-type syndrome groupings (e.g., respiratory and gastrointestinal) for aberrations. Algorithms are based on a 1-year baseline and compare totals of 1) individual calls in a syndrome category, 2) the ratio of syndrome calls to all calls, and 3) the cumulative sum (CUSUM) calculation of syndrome triggers. The system also includes a geocluster signal that triggers when the number of calls for a syndrome exceeds eight calls in a 1-mile radius.

Provider Surveillance System (PROS): PROS monitors patient visit data from seven affiliated physician groups that are geographically dispersed throughout Bergen and Passaic counties. The medical groups use an integrated electronic clinical information system for all offices. The data collected include two levels of *International Classification of Diseases, Ninth Revision* (ICD-9), diagnosis coding. Data are sent 4 days after the date of the patient visit to allow for ICD-9 coding, which produces a more accurate syndrome grouping. Patients are grouped by syndromes based on the CDC-recommended syndromic surveillance ICD-9 groupings. The Department of Defense Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) ICD-9 groupings for influenza-like illness (ILI) is used for ILI surveillance. The combination of all three signaling systems will be evaluated and is expected to provide a more effective, real-time picture of disease activity in the county than the original stand-alone ED system.

Comparisons of Timeliness and Signal Strength for Multiple Syndromic Surveillance Data Types — San Diego County, July 2003–July 2004

Steven F. Magruder,[1] N. Marsden-Haug,[2] S. Hakre,[2] J. Coberly,[1] C. McClean,[3] J. Johnson,[3] A. Anderson,[2] J. Pavlin[2]
[1]*Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland;* [2]*Walter Reed Army Institute of Research, Silver Spring, Maryland;* [3]*San Diego County Health and Human Services Agency, San Diego, California*

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

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: San Diego County is the site of the BioNet Project, which seeks to improve the ability of the Navy Region Southwest and San Diego County to respond to a biologic attack on its population and its critical infrastructure by improving, integrating, and enhancing disparate military and civilian detection and characterization capabilities. BioNet is funded by the Department of Homeland Security. One component of this project is the comparison of data sources available in San Diego County to understand their relative strengths and weaknesses for syndromic surveillance purposes.

Objectives: This study quantitatively compared the different syndromic data sources (both military and civilian) available in San Diego County both in terms of signal strength and timeliness.

Methods: Multiple types of data were compared, including emergency medical services (EMS), school nurse, school absentee, physician outpatient encounters, over-the-counter (OTCs) pharmaceuticals, and prescription pharmaceuticals. Three major historical disease outbreaks are used as points of comparison. The specific outbreaks are respiratory disease caused by a major wildfire event in October 2003, influenza-like illness in December 2003, and a surge of gastrointestinal illness in February 2004. Each data source is separately filtered to bring out the types of symptoms associated with each of the outbreaks. The sources are compared both before and after smoothing with a moving 7-day average, designed to eliminate certain idiosyncratic effects and to reduce noise. Finally, the data sources were compared on the basis of timing and signal-to-noise ratio for their ability to capture these outbreaks. Additional time-series comparisons were also used to determine whether the data sources trend together during nonoutbreak periods.

Results: The disease outbreaks are each observable in multiple data sources, but the most useful data source varies with the event. EMS, military ambulatory encounters, OTCs, and school nurse reports were especially useful for different illness events. For example, EMS data indicate the strongest signal-to-noise ratio for disease caused by wildfires; the school nurse data give an early indication of influenza; and the military ambulatory encounter data provide the strongest indication of an outbreak of gastrointestinal illness.

Conclusion: These results indicate that a system that integrates multiple syndromic data streams into a single prospective surveillance tool might enhance the ability of military and civilian authorities in San Diego County to detect biologic terrorist or other disease outbreaks in a timely fashion.

Evaluation of Joint Services Installation Pilot Project and BioNet Syndromic Surveillance Systems — United States, 2004

Nicola Marsden-Haug, V. Foster, S. Hakre, A. Anderson, J. Pavlin
Walter Reed Army Institute of Research, Silver Spring, Maryland

Corresponding author: Nicola Marsden-Haug, Walter Reed Army Institute of Research, Division of Preventive Medicine, 503 Robert Grant Avenue, Silver Spring, MD 20910. Telephone: 301-319-7355; Fax: 301-319-9104; E-mail: nicola.marsden-haug@na.amedd.army.mil.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Key elements that can be measured objectively to evaluate the effectiveness of a surveillance system include sensitivity, specificity, and timeliness. Statistical algorithms and data sources have been evaluated frequently; however, factors that affect whether public health professionals will use a system (e.g., acceptability, generalizability, flexibility, representativeness, and reliability) are more difficult to assess and are reported less frequently. Through the Joint Services Installation Pilot Project (JSIPP), the Department of Defense provided nine military installations with enhanced capabilities to detect and respond to weapons of mass destruction. The Department of Homeland Security is funding BioNet to improve outbreak management in San Diego, California, by integrating military and civilian information. Both programs implemented versions of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) IV for medical surveillance and required evaluation within these environments.

Objectives: This study measured JSIPP and BioNet users' perspectives of syndromic surveillance and the use of ESSENCE IV.

Methods: For JSIPP, registered ESSENCE users were surveyed regarding system usage, utility of features, user friendliness, and suggested improvements. For BioNet, potential ESSENCE users were identified by program administrators and surveyed before implementation to identify qualities and features they viewed as important to syndromic surveillance.

Results: The JSIPP survey response rate was only 34% (17 of 50), probably because registered but infrequent users were included in the survey population. The majority of respondents found the system easy to use and valued having access graphics and summary statistics of disease trends. Seven (41%) of 17 reported concern about the inability to obtain patient identifiers in a timely manner, which diminished their ability to investigate suspicious alerts. Barriers to system use included dislike of the layout and difficulties in interpreting nonclinical data sources. The BioNet survey response rate was 59% (13 of 22). Respondents ranked the usefulness of four elements of syndromic surveillance: 1) usefulness during an outbreak investigation; 2) ability to detect outbreaks rapidly; 3) reassurance that no ongoing outbreaks are occurring; and 4) capability to generate summary reports. Major weaknesses included 1) difficulty in interpreting and responding to alerts; 2) uncertainty in outbreak detection capability; 3) use of abstract data sources; and 4) difficulty in sustaining a system for long-term use. Additional concerns included high false-positive rates, timeliness, reliability, and cost.

Conclusion: Users often continue to have reservations about the utility of these systems. Feedback on which data sources and system features users value can help system developers direct resources for development.

Comparison of Outpatient Visit and Emergency Department Data for Use in Syndromic Surveillance — New York City, 2001–2004

Kristina Metzger, F. Mostashari, M. Kendall
New York City Department of Health and Mental Hygiene, New York City, New York

Corresponding author: Farzad Mostashari, New York City Department of Health and Mental Hygiene, 125 Worth St., Box 6, New York, NY 10013. Telephone: 212-788-5384; Fax: 212-788-4473; Email: fmostash@health.nyc.gov.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Monitoring outpatient visits might enable more timely and sensitive syndromic surveillance than emergency department visits because of higher daily volumes and the potential for capturing illness at an earlier stage.

Objectives: Data from all 11 public hospitals in New York City with outpatient and emergency departments (EDs) were evaluated to compare the usefulness of these two data sources for monitoring communitywide respiratory and gastrointestinal illness.

Methods: Historic data were obtained on outpatient and ED visits during November 1, 2001–May 31, 2004. Demographic characteristics of patients were compared. The seasonal and temporal trends of comparable syndrome categories (respiratory, fever/viral, asthma, and gastrointestinal) were examined and coded according to the *International Classification of Diseases, Ninth Revision* (ICD-9), diagnosis code for outpatient visits and chief complaint for ED visits. For each syndrome, timing and frequency were assessed for 1-, 2-, and 3-day temporal clusters with a 14-day baseline period by using temporal scan statistics with SaTScan™ software.

Results: On weekdays, more patients visited outpatient clinics (mean: 3,727) than EDs (mean: 2,906). On weekends, limited outpatient visits occurred (mean: 95); EDs had a slightly lower volume than on weekdays (mean: 2,492). Compared with the ED population, the outpatient population included more patients aged <12 years and >65 years, more females, and more minority patients and those on Medicaid. The temporal trends of the respiratory syndrome from outpatient clinics and EDs were strongly correlated ($r = 0.67$), as were the fever/viral ($r = 0.57$) and asthma ($r = 0.60$) syndromes, but less correlated for the gastrointestinal syndrome ($r = 0.36$). Citywide temporal clusters were occasionally detected on the same day.

Conclusion: This evaluation of outpatient visits indicates that this data source might be potentially useful for syndromic surveillance, particularly in conjunction with ED data. The demographic characteristics differ, allowing for the examination of complementary populations, and might explain certain differences in observed temporal clusters. Whereas few outpatient visits occur on weekends, both data sources have comparable overall daily volumes. These data sources could be examined simultaneously by using multivariate methods, which might increase the power to detect outbreaks. Outpatient visit data include clinician and patient information as well as diagnoses, which substantially increases the feasibility of cluster investigations. Using both outpatient and ED visit data as part of syndromic surveillance might enhance the ability to detect and validate outbreaks.

Emergency Department Surveillance for the 2003 Rugby World Cup — New South Wales, Australia

David J. Muscatello, T. Churches, J. Kaldor, W. Zheng, C. Chiu, P. Correll, T. Mannes
New South Wales Department of Health, Sydney, Australia

Corresponding author: David J. Muscatello, NSW Department of Health, 73 Miller St., North Sydney, NSW 2060 Australia. Telephone: 61-2-9391-9408; Fax: 61-2-9391-9232; E-mail: dmusc@doh.health.nsw.gov.au.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: The Department of Health in New South Wales, Australia, expanded public health surveillance for the 2003 Rugby World Cup and for its ongoing counterterrorism response. Cup games were played in and around the Sydney region.

Objectives: This study explains the development of an automated, near real-time, syndromic surveillance system in which data are used that are already being collected in emergency department (ED) databases for routine patient administration.

Methods: Demographic, triage, physician-assigned provisional diagnosis, and disposition information from 12 of 49 public hospital EDs in the greater Sydney metropolitan area was frequently and automatically transmitted, analyzed, and reported in daily statistical summaries on Intranet websites beginning October 10, 2003. Diagnoses were categorized by syndrome, disease, and injury. Presenting problem and nursing-assessment free text routinely entered by nurses during patient triage were automatically classified into >30 syndrome categories by using automated preprocessing techniques and naïve Bayesian automatic text classification methods. The diagnosis-based categories were used to train the automatic classifier to associate words in the free text with syndrome categories. An adjusted cumulative sum (CUSUM) accumulating day-of-week differences in daily counts was used to assess the statistical significance of disease and injury trends. Public health personnel monitored the reports daily and notified the Rugby World Cup Public Health Committee of unusual trends.

Results: During the tournament, October 10–November 22, health trends identified by the system were not sufficient to cause concern among public health personnel but did provide reassurance that the health of the population was not adversely affected. Data collection did not add to the work load of clinical staff in EDs, and surveillance downtime was negligible. Since the games, this now ongoing surveillance system has rapidly identified a community-based epidemic of gastrointestinal illness, an increase in recreational drug misuse, the annual influenza epidemic, and an increase in episodes of acute asthma. Variation by syndrome occurred in the degree of correlation between daily visit counts from automatically classified nurse text and the equivalent diagnosis groups used to train the classifier.

Conclusion: During the Rugby World Cup, the surveillance system complemented traditional public health surveillance to provide a comprehensive assessment of health trends in the population. A substantial advantage of the system has been its ongoing sustainability. Physician-assigned diagnoses are more specific than free text for some syndrome and diagnostic categories. However, diagnoses are not available for analysis until at least the end of a patient's ED visit, and they are sometimes incomplete. Free text from patient triage is available for analysis early in an ED visit, and depending on the scope of the text description, multiple text-based syndromes can be automatically assigned to a visit.

Detecting Elongated Disease Clusters

Daniel B. Neill, A. Moore, M. Sabhnani
Carnegie Mellon University, Pittsburgh, Pennsylvania

Corresponding author: Daniel B. Neill, Carnegie Mellon University, Department of Computer Science, 5000 Forbes Avenue, Pittsburgh, PA 15213. Telephone: 412-621-2650; Fax: 412-268-5576; E-mail: neill@cs.cmu.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: When pathogens are dispersed by wind or water, the resulting disease clusters can be highly elongated in shape, and tests for circular or square regions will have lower power to detect these high-aspect ratio clusters. One possible solution is to search for rectangular clusters by using a variant of Kulldorff's (1997) spatial scan statistic to find the most significant rectangular region and by computing the region's statistical significance (p value) by randomization. However, when data are aggregated to an N x N grid, an exhaustive search would require searching over all $O(N^4)$ gridded rectangular regions (both for the original grid and for each Monte Carlo replication). Such a search is computationally infeasible for certain large, real-world data sets.

Objectives: This study attempted to accelerate the spatial scan statistic, enabling rapid detection of the most significant rectangular cluster (and its p value) without a loss of accuracy.

Methods: A fast spatial scan algorithm was presented that allowed computation of the same region and p value as the exhaustive search approach, but hundreds or thousands of times faster. The algorithm divides the grid into overlapping regions (using a novel overlap-kd tree data structure), bounds the maximum likelihood ratio of subregions contained in each region, and prunes regions that cannot contain the most significant region. The resulting effect was searching over all rectangular regions while only examining a fraction of these. The fast spatial scan was also extended to multidimensional data sets, enabling the application of spatial scan statistics to other domains with more than two spatial dimensions; in addition, these extra search dimensions allowed incorporation of temporal information (allowing fast spatio-temporal cluster detection) and demographic information (e.g., patients' age and sex).

Results: The fast spatial scan achieves speedups from 20–2,000 times compared with the exhaustive search approach on real and simulated data sets, including data from emergency department records and over-the-counter (OTC) drug sales. For example, elongated clusters were detected in national OTC data in 47 minutes, compared with 2 weeks for an exhaustive search. Theoretical and empirical results, including preliminary comparisons to Kulldorff's SaTScan software, indicate that the fast spatial scan makes the detection of elongated clusters computationally feasible.

Conclusion: In collaboration with the RODS Laboratory at the University of Pittsburgh, the fast spatial scan is being applied to prospective disease surveillance nationwide, using daily OTC drug sale data from the National Retail Data Monitor.

Use of Modeled Anthrax Attacks on the Mall of America To Assess Sensitivity of Syndromic Surveillance — Minnesota, 2003–2004

James D. Nordin,[1] M. Goodman,[1] M. Kulldorff,[2] D. Ritzwoller,[3] A. Abrams,[2] J. Donahue,[4] J. Vest[5]
[1]*HealthPartners Research Foundation, Minneapolis, Minnesota;* [2]*Harvard Medical School/Harvard Pilgrim Health Care, Boston, Massachusetts;*
[3]*Kaiser Permanente Colorado, Boulder, Colorado;* [4]*Marshfield Clinic Research Foundation, Marshfield, Wisconsin;*
[5]*Austin County Public Health Department, Austin, Texas*

Corresponding author: James D. Nordin, HealthPartners Research Foundation, P.O. Box 1524 MS 21111R, Minneapolis, MN 55440-1524. Telephone: 952-967-5087; Fax: 952-967-5022; E-mail: james.nordin@healthpartners.com.

Disclosure of relationship: One contributor to this report, D. Ritzwoller, has disclosed that she was a co-investigator on the grant noted in the report. The other contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: The goal of syndromic surveillance systems is to detect and define biologic terrorism releases earlier than previously possible. These systems have rarely been evaluated.

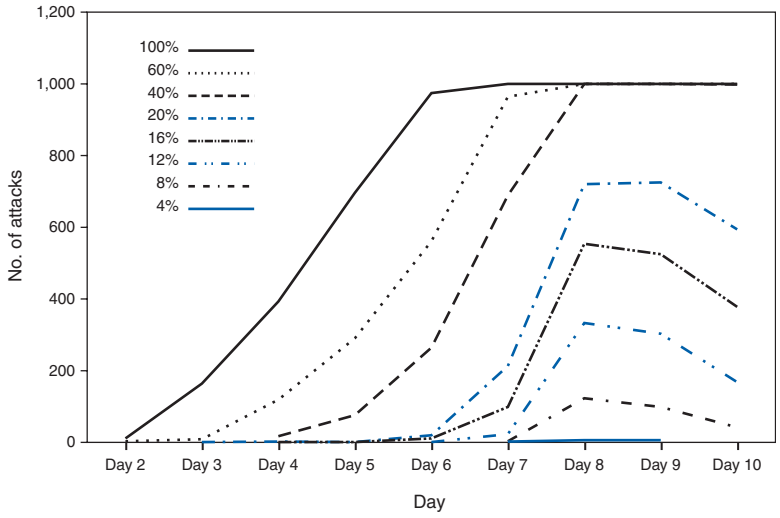
Objectives: This study measured a biologic terrorism surveillance system's sensitivity and timeliness.

Methods: Models of an anthrax release in the Mall of America (Bloomington, Minnesota) were developed by using ZIP code data from the U.S. Census Bureau; data on mall visitors; and data provided by HealthPartners Medical Group, which covers 9% of the Twin Cities metropolitan population. For each infection level from 4% to 100%, 1,000 random dates during July 1, 2003–June 30, 2004, were selected with replacement for simulated releases. Timing of symptoms after release was based on data from the 1979 Sverdlovsk anthrax release. Cases from the simulated outbreak were added to respiratory visits recorded for those dates in HealthPartners' data. Analysis was performed by using the SaTScan™ space-time scan statistic (available at http://www.satscan.org) and Kleinman's generalized mixed model.

Results: Timeliness and completeness of detection of events varied by infection rate (Figure). At a 40% infection rate, first events were detected by day 2; 25% by day 6; 75% by day 7; and 100% by day 8. Sensitivity decreased to ≤20% and timeliness increased to >40%. The system was most sensitive in summer, intermediate in fall and spring, and least in winter because of increased background rates of illness in winter. Sensitivity is better if a greater portion of the population is covered by the system.

Conclusion: This biologic terrorism surveillance system can detect a modeled anthrax release in the majority of instances at a 20% infection rate and in all instances at a 40% infection rate.

FIGURE. Number of attacks detected, by days after attack and percentage of mail customers infected, by number of days — Minnesota, 2003–2004



Clinician Syndromic Surveillance Consequent to Participation in the National Bioterrorism Syndromic Surveillance Demonstration Project (NBSSDP) — Central Texas, 2003

Janet L. Pichette,[1] E. Sherwood,[1,2] A. Valadez,[1] J. Jackson,[2] N. Neighbors,[3] J. Schultz[1]
[1]*Austin/Travis County Health and Human Services Department, Austin, Texas;* [2]*Williamson County and Cities Health District, Georgetown, Texas;* [3]*Hays County Health Department, San Marcos, Texas*

Corresponding author: Janet L. Pichette, Austin/Travis County Health and Human Services Department, 15 Waller, Austin, TX 78702. Telephone: 512-972-5486; Fax: 512-972-5772; E-mail: janet.pichette@ci.austin.tx.us.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: The implementation of automated syndromic surveillance of physician office visits in three counties in central Texas resulted in the addition of active clinician-based syndromic surveillance and a structured approach to reactive surveillance. Historically, mandated notifiable disease reporting has been passive in these counties. Compliance was inconsistent and triggered by final laboratory confirmation as opposed to first clinical suspicion.

During January–March 2003, three physician groups in these counties agreed to participate in the National Bioterrorism Syndromic Surveillance Demonstration Project (NBSSDP). During project implementation, three local health departments (LHDs) and the physician groups recognized limitations of mandated disease reporting and automated syndromic surveillance. The LDHs and infection-control practitioners (ICPs) applied the concept of syndromic surveillance to clinician-initiated reporting.

Objective: This abstract describes how active clinician-based reporting and reactive syndromic surveillance enhanced automated syndromic surveillance.

Methods: Active syndromic surveillance for specified concerns to public health (CPH) was initiated to include reporting of patients with signs or symptoms consistent with exposure to biologic or chemical agents, previously healthy persons with acute onset of a severe undiagnosed illness, or any condition warranted by ICP. All ICPs agreed to report immediately and to respond to a weekly e-mail verifying that they were "alive, well, and on-the-job."

LHD staff and ICPs also formed the Surveillance Emergency Response Group, providing a structured approach to reactive surveillance. Positive signals from surveillance triggers a request for ICPs to identify all inpatients with illnesses or symptoms compatible with an implicated agent. When a severe cryptic illness is reported, this procedure is also used to validate the absence of similar cases.

Results: During a 6-month period, three severe cryptic illnesses in previously healthy patients were reported. A reactive surveillance drill, which simulated a biologic release of *Yersinia pestis,* was conducted to test the ICP response in identifying and reporting patients in their facility with possible plague or pneumonia-like symptoms. For the 11 hospitals reporting, the average response time in identifying and reporting patients was 36.4 minutes (range: 14–88 minutes).

Conclusion: The implementation of NBSSDP facilitated improvements needed in clinician-based syndromic surveillance. By applying NBSSDP's approach, ICPs could quickly survey their hospitals and identify patients meeting a specified syndrome. Reporting CPH places minimal burden on ICPs and might contribute to the early detection of disease outbreaks. Reactive surveillance can provide timely information in response to positive surveillance signals.

Using Categorization of Reason-for-Visit Strings as the Basis for an Outbreak Detection System — Minnesota, 2002–2003

Sivakumaran Raman, J. Levin, D. Hall, K. Frey
Children's Hospitals and Clinics, Roseville, Minnesota

Corresponding author: Sivakumaran Raman, Children's Hospitals and Clinics, 2900 Centre Pointe Drive, Roseville, MN 55113. Telephone: 651-855-2055; Fax: 651-855-2075; Email: sivakumaran.raman@childrenshc.org.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: The syndromic surveillance system used by Children's Hospitals and Clinics (CHC) to detect disease outbreaks uses reason-for-visit (e.g., chief complaint) information recorded when patients register at CHC facilities.

Objectives: A machine-learning approach that employs text categorization of reason-for-visit fields from patient encounters was used to assess whether daily counts of disease syndromes based on *International Classification of Diseases, Ninth Revision* (ICD-9) diagnostic codes correlated with daily counts of text categorizer-assigned syndromes.

Methods: Reason-for-visit text strings collected from CHC emergency departments and general pediatrics clinics were used to define a set of pediatric-focused disease syndromes. Text-categorization programs that were based on common algorithms available as modules in the Perl programming language (open source software available at http://www.perl.org) were given previously categorized data from 2002. Data for 2003 were used to evaluate the agreement between CHC syndromes assigned by ICD-9 codes and syndromes assigned by text categorizers. Spearman's rank correlation coefficients were calculated to permit examination of the association between daily counts of ICD-9–assigned syndromes and categorizer-assigned syndromes. Receiver operating characteristic (ROC) curves were plotted for certain categorized data to examine the performance of the text-categorizer.

Results: From 2003 data, 102,435 reason-for-visit strings were classified into syndromes by using the associated principal ICD9 codes and running Perl programs for the Support Vector Machine (SVM) and naïve Bayesian categorizers. Spearman's rank correlation coefficient values for daily counts of categorizer-assigned syndromes and ICD-9–based syndromes demonstrated a correlation between the two. Spearman's coefficients for the counts for SVM versus ICD-9 syndromes were 0.754 for the EENT (eyes, ears, nose, and throat) syndrome, 0.722 for the FEVER syndrome, 0.843 for the GASTROINTESTINAL syndrome, 0.923 for the INJURY syndrome, and 0.913 for the RESPIRATORY syndrome. Correlation was also seen between EENT-ICD9 and RESPIRATORY-SVM syndromes and EENT-ICD-9 and FEVER-SVM syndromes. Similar correlation results were obtained for the naïve Bayesian categorizer. ROC curves drawn for the naïve Bayesian categorizer-assigned scores (used as the test) against the ICD-9–assigned scores (used as the standard) provide evidence of the categorizer's high performance. Areas under ROC curves for the 13 syndromes ranged from 0.966 (for the INJURY CHC syndrome) to 0.701 (for the EENT CHC syndrome).

Conclusion: Text categorization of reason-for-visit strings gives robust results and can be combined with a statistical or algorithmic method of detection of extraordinary events to create an outbreak detection system. The code used is available for public use at http://www.childrenshc.org/downloads/syndromesurveillance/syndromesurveillance.asp.

Evaluating a Syndromic Surveillance System for the Detection of Acute Infectious Gastroenteritis Outbreaks — North Carolina, 2004

Emily E. Sickbert-Bennett,[1] M. Scholer,[2] J. Butler,[3] D. Travers,[2] J. MacFarquhar,[4] A. Waller,[2] G. Ghneim[5]

[1]University of North Carolina Healthcare System, Chapel Hill, North Carolina; [2]University of North Carolina at Chapel Hill, Department of Emergency Medicine, Chapel Hill, North Carolina; [3]Orange County Health Department, Hillsborough, North Carolina; [4]North Carolina Statewide Program for Infection Control and Epidemiology, Chapel Hill, North Carolina; [5]North Carolina General Communicable Disease Control Branch, Division of Public Health, Raleigh, North Carolina

Corresponding author: Emily E. Sickbert-Bennett, UNC Health Care System, Hospital Epidemiology, 1001 West Wing CB #7600, 101 Manning Drive, Chapel Hill, NC 27514. Telephone: 919-843-4165; Fax: 919-966-1451; E-mail: esickber@unch.unc.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: During January 21–February 9, 2004, a norovirus outbreak occurred among University of North Carolina (UNC) students, some of whom sought care at the UNC hospital's emergency department (ED). Despite an established ED-based syndromic surveillance system using CDC's Early Aberration Reporting System (EARS), no increases in gastrointestinal illness (GI) were detected during the outbreak period.

Objectives: This study used outbreak data to evaluate North Carolina's syndromic surveillance system and GI case definition.

Methods: The hospital ED electronically sent data for all visits to a state data repository. Recorded chief complaints, vital signs, and triage nurse notes were searched electronically for key terms to assign ED visits to CDC syndrome classifications. The GI case definition criteria required at least one constitutional symptom and one GI symptom. The outbreak line-listing of 429 cases as determined by the local health department was reviewed, patient age distribution examined, and hospital records used to identify ED patients by name. Data on these case-patients were reviewed and each symptom recorded. On the basis of the symptom frequency, a modified GI case definition was drafted and tested on the ED data from the time of the outbreak. The data were stratified by the age distribution of the known outbreak cases. The number of GI cases that met the modified definition was examined for aberrations by using EARS.

Results: Of the 11 case-patients seen in the ED, one was identified with syndromic surveillance by using the original GI case definition. Of the 42 remaining ED visits during the outbreak period that were classified as GI syndrome, eight (19%) were misclassified as a result of lack of recognition of negation terms (e.g., no fever), and 34 (81%) were classified correctly. Frequency analysis of the 11 known ED case-patients' symptoms indicated nausea and vomiting, 11 (100%); diarrhea, nine (82%); abdominal pain, eight (73%); fever, two (18%); and body aches, one (9%). When a modified GI case definition that did not require a constitutional symptom (e.g., fever) was used and syndromic cases were stratified by age distribution for persons aged 17–22 years, all 11 cases were captured, and an aberration was detected on January 21, the first day of the outbreak.

Conclusion: Automated systems must be monitored to ensure proper syndrome classification. For syndromic surveillance to be used to detect both biologic terrorism–related and community outbreaks, case definitions must be constructed with careful consideration of different clinical presentations with different etiologies and illness severities.

Syndromic Surveillance System Evaluation — District of Columbia, 2001–2004

Michael A. Stoto,[1] A. Jain,[1] A. Diamond,[2] J. Davies-Cole,[3] A. Adade,[3] S. Washington,[3] G. Kidane,[3] C. Glymph[3]
[1]RAND, Arlington, Virginia; [2]Harvard University, Cambridge, Massachusetts; [3]District of Columbia Department of Health, Washington, DC

Corresponding author: Michael A. Stoto, RAND, 1200 South Hayes St., Arlington, VA 22202. Telephone: 703-413-1100, ext. 5472; Fax: 703-413-8111; E-mail: stoto@rand.org.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: In September 2001, the District of Columbia Department of Health began a syndromic surveillance program based on hospital emergency department (ED) visits. ED logs are faxed daily to the health department, where staff code them by chief complaint and record the number of patients, in each hospital who die or experience sepsis, rash, respiratory complaints, gastrointestinal complaints, unspecified infection, and neurologic or other complaints.

Objectives: This study evaluates the completeness, usefulness, and effectiveness of the syndromic surveillance system.

Methods: Data were received from nine hospitals in the first 32 months of the operation of the system (September 2001–May 2004). These data were used to describe the operation of the completeness of the system (whether reports were sent to health departments daily), by hospital, season and day of the week, and variability in patterns of symptom groups across hospital and season. Three statistical detection algorithms also were applied retrospectively to identify departures from normal patterns associated with the beginning of the winter influenza season and other disease outbreaks.

Results: Completeness varied by calendar quarter and hospital, ranging from no missing data for some hospitals and quarters to 100% missing data. Data were missing primarily in weekly patterns and stretches of time that varied across hospitals, which might reflect staff availability to fax data to the health department. In seven of nine hospitals from which the data were more than 75% complete, with limited exceptions, the number and proportion of cases in each symptom group were constant over time. The distribution of symptom groups were similar in all except one hospital, possibly reflecting a different patient population. Day-of-the-week effects were apparent in certain hospitals but varied substantially by symptom, group, and hospital. Application of various detection algorithms indicated that, particularly when pooling data across seven hospitals, the syndromic surveillance data can be used to identify the onset of the influenza season within 2–3 days. The data also can be used to determine indications of the "worried well" who sought care during the 2001 anthrax attacks and a previously undetected series of gastrointestinal illness outbreaks that occurred during a 4-month period in five different hospitals. No single symptom group or detection algorithm consistently signaled each of the gastrointestinal events.

Conclusion: If problems with completeness of the data can be improved through a planned automatic electronic reporting system, syndromic surveillance data might offer the potential for early detection of influenza and other disease outbreaks. Additional research is needed, however, to characterize normal patterns in the data, identify the most effective detection algorithms and symptom groups for various purposes, and characterize their sensitivity and specificity when used prospectively in real time.

Automated Anomaly Detection Processor for Biologic Terrorism Early Detection* — Hampton, Virginia

D. Michael Thomas,[1] S. Arouh,[1] K. Carley,[2] J. Kraiman,[1] J. Davis[1]
[1]*Dynamics Technology, Inc., Arlington, Virginia;* [2]*Carnegie Mellon University, Pittsburgh, Pennsylvania*

Corresponding author: D. Michael Thomas, Dynamics Technology, Inc., 1555 Wilson Blvd, Ste 703, Arlington, VA 22209. Telephone: 703-841-0990; Fax: 703-841-8385; E-mail: mthomas@dynatec.com.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Disease surveillance databases can range in size into the terabytes, making rapid, meaningful analysis and conclusions about the data impracticable and expensive. Robust, automated, nontemplate-based real-time processing techniques capable of monitoring large-scale disease, health-care, and environment tracking and surveillance data sets are needed to discriminate between naturally occurring events and emergent diseases or biologic terrorist attacks.

Objectives: This study evaluated the ability of an automated anomaly detection processor to detect a simulated anthrax attack during influenza season.

Methods: This report describes the application of data-mining techniques in developing an Automated Anomaly Detection Processor (AADP), which uses the Self Organizing Map clustering algorithm in conjunction with a Gaussian Mixture Model and a Bayesian Analyzer probabilistic model to detect anomalous occurrences in health data sets.

The test case for the model is a data set from the BioWar Model (Carnegie Mellon University) and is based on real-life census data, medical records, and social and behavioral patterns in Hampton, Virginia. The data files include sales from 16 pharmacies in 25 product categories, absenteeism records from 34 school and 982 work sites, and medical insurance records of the residents. The BioWar data are unique because they contain a simulated biologic weapons attack and human response against a background of naturally occurring illness. In addition, two data files contain time series of nonquantitative observables (e.g., *International Classification of Diseases, Ninth Revision* codes).

Results: AADP identified a simulated biologic terrorism attack occurring during the influenza season. First detection of the anthrax outbreak occurred approximately 4.7 days after the attack. AADPs for any pharmacies deemed to be anomalous yielded a drill-down table identifying the relative contributions of the variables causing the anomaly. The population of the pharmacy AADPs yielded an excessive number of anomalous pharmacies simultaneously after the simulated attack began. Pharmacies began to turn anomalous initially adjacent to the attack site. Drill-down of the anomalies indicated shared patterns of sales of several categories of pharmaceutical products. These results indicate a systematic cause rather than a random correlation of probability of anomaly. In addition, anomalous periods of extended absenteeism were detected soon after the attack.

Conclusion: Development of AADP for biosurveillance adds a complementary method to extant surveillance systems and can improve real-time alerting so assets can be vectored for further epidemiologic investigation and early intervention. These results are prompting additional query of the anomalous events and inclusions of additional data streams to improve early warning and response.

* This work is supported by the U.S. Army Medical Research and Materiel Command under Contract No. DAMD17-03-C-0061. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Department of Defense.

Early Detection of Outbreaks Using the BioDefend™ Syndromic Surveillance System — Florida, May 2002–July 2004

Kristin B. Uhde,[1] C. Farrell,[1] Y. Geddie,[2] M. Leon,[3] J. Cattani[1]
[1]*University of South Florida, Tampa, Florida;* [2]*MacDill Air Force Base, Tampa, Florida;*
[3]*Celebration Research Institute, Celebration, Florida*

Corresponding author: Corey L. Farrell, University of South Florida, 3602 Spectrum Blvd., Tampa, FL 33612. Telephone: 813-974-1473; Fax: 813-974-1479; E-mail: cfarrell@bt.usf.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: The threat of biologic terrorism (BT) requires the public health infrastructure to focus attention on challenges posed by emerging and re-emerging infectious diseases. Validating real-time approaches to surveillance that can provide timely alerts of epidemics is critical whether the epidemics occur naturally or through a BT attack.

Objectives: This research study implemented and evaluated the BioDefend™ syndromic surveillance system to determine if outbreaks and potential BT attacks could be detected 24–36 hours before routine surveillance and if syndromic surveillance is a feasible approach for BT preparedness and early detection of infectious disease.

Methods: The study was conducted in central Florida high-risk facilities (i.e., theme parks, hospitals, and a military base). A 6-month period of baseline data was collected to identify normal illness trends and seasonality patterns and to serve as the comparison for the test period. Internet-based data entry of provider-identified syndromes was linked to an automated analysis tool that provided alerts through an e-mail–enabled device when substantial increases of syndromes exceeded the pre-established thresholds. Thresholds are based on a 30-day rolling mean, and alerts were generated when any syndrome exceeded three standard deviations above the 30-day rolling mean.

Results: Outbreaks of public health importance were detected by comparing the BioDefend™ data with regional, state, and national surveillance data. Two epidemics, one of gastroenteritis and one of influenza-like illness, were detected by BioDefend™ >1 month before identification through routine surveillance. In addition, a small cluster of cases (five) of fifth disease was identified among South American children visiting central Florida theme parks.

Conclusion: This study indicated that health events can0 be recognized in near real-time through the use of automated analysis and notification. Early detection of events allowed for timely interventions, including vaccination campaigns at military and civilian hospitals, and it demonstrated one way that syndromic surveillance can be an effective tool for early detection of outbreaks in addition to its potential role in BT preparedness. The system continues to be used in central Florida and was implemented for the 2005 Super Bowl.

Syndromic Approach to West Nile Virus — The Netherlands, 2002–2003

Cees van den Wijngaard, L. van Asten, B. Rockx, W. van Pelt, G. Godeke, M. Koopmans
National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Corresponding author: Cees van den Wijngaard, RIVM, Pb 1, 3720 BA, Bilthoven, the Netherlands. Telephone: 31-0-30-274-29-10; Fax: 31-0-30-274-44-09; E-mail: kees.van.den.wijngaard@rivm.nl.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: In October 2002, after the West Nile virus (WNV) epidemic in the United States and reports on a changing epidemiology for flaviviruses in Europe, the National Institute for Public Health and the Environment (RIVM) launched a project to study trends in neurologic infectious disease in the Netherlands.

Objectives: This study used a prospective syndromic approach to initiate surveillance for WNV by monitoring patients with unexplained meningoencephalitis and conducting additional laboratory testing for WNV during June–November, when mosquitoes are active in the Netherlands.

Methods: Because neurologic illness of possible viral etiology other than acute flaccid paralysis is not notifiable in the Netherlands, RIVM 1) examined medical registration data from 103 (99%) of 104 Dutch hospitals covering 16 million persons to identify all discharge diagnoses for unexplained meningoencephalitis; 2) examined data from 11 laboratories covering 2.5 million persons to study trends in submissions of cerebrospinal fluid (CSF) for virologic testing for common neurotropic viruses (e.g., herpes and enteroviruses); and 3) actively collected CSF samples from six virologic laboratories for further exclusion of WNV infection.

Results: Hospital surveillance for 2002 and 2003 indicated that approximately 500 patients per year had meningitis or encephalitis (unspecified viral or unexplained) diagnosed during June–November. In 2002, of 158 CSF submissions, 137 (87%) tested negative for common viruses; none of these samples had been tested for WNV. Samples that were subsequently collected by RIVM for further WNV testing (150 in 2002, 294 in 2003, and 337 in 2004) tested negative for antibodies to WNV. Because WNV can cause meningoencephalitis, a patient with unexplained meningoencephalitis might be infected by WNV. In 2003, a total of 500 hospital patients received such a diagnosis, but only 294 CSF samples were further tested for WNV. At this level of testing, the probability of detecting WNV meningoencephalitis would have been 0.99 if five WNV-caused meningoencephalitis cases had occurred among the 500 hospital patients but only 0.59 if one WNV-caused case had occurred.

Conclusion: No endemic WNV transmission has been detected in the Netherlands since 2002. CSF submission data for 2003 and 2004 and hospital discharge data for 2004 are not yet available. On the basis of available data, no substantial endemic transmission of WNV occurred. However, a limited outbreak of WNV meningoencephalitis might not be detected. Ruling out WNV as an etiologic agent in all CSF samples when no common pathogen has been detected will improve surveillance.

Constant Specificity Surveillance for Real-Time Outbreak Detection

Shannon C. Wieland,[1,2] B. Berger,[1,3] K. Mandl[2,4]
[1]*Massachusetts Institute of Technology, Cambridge, Massachusetts;* [2]*Children's Hospital, Boston, Massachusetts;*
[3]*Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts;* [4]*Harvard Medical School, Boston, Massachusetts*

Corresponding author: Shannon C. Wieland, Massachusetts Institute of Technology, 290 Massachusetts Avenue, Cambridge, MA 02139. Telephone: 617-645-6875; E-mail: shann@mitt.edu.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: Certain modeling techniques have been implemented to detect unusually high emergency department (ED) visit rates on the basis of historical data, including parametric regression, autoregression, multiresolution wavelet analysis, and additive modeling. An outbreak-detection strategy is informative only to the extent that its specificity is known because the probability of a false alarm in the absence of an outbreak should be understood to allocate resources appropriately in the event of an alert that triggers a public health investigation.

Objectives: This study demonstrated that the specificity of outbreak detection using current methods varies substantially on multiple timescales. A modeling approach that provides constant specificity surveillance was developed.

Methods: Autoregressive, Serfling, trimmed seasonal, and wavelet-based outbreak detection models were evaluated for changes in specificity over time. All model simulations used 12 years of historical respiratory syndrome ED visits at a major pediatric hospital in an urban setting. Changes in specificity were detected by using error analyses modified for binomial data and chi-squared analysis. Sensitivity was evaluated by adding synthetic 1-day outbreaks among 10 patients to the historical data. A new outbreak-detection method was developed that used generalized additive models of both the ED visit mean and variance.

Results: The specificity of four previously published models (i.e., autoregressive, Serfling, trimmed seasonal, and wavelet-based) was a nonconstant function of the day of the week, the month of the year, or the year of the study ($p<0.05$). The seasonal changes in specificity led to a paradoxical increase in sensitivity to simulated outbreaks during winter months when compared with summer months. The new method had constant specificity over all three time scales ($p<0.05$) without a loss of sensitivity compared with previous models.

Conclusion: Interpretation of alarms using outbreak detection strategies is difficult because the specificity is extremely variable. The fluctuations in specificity are caused by changes on the same time scales in the variance of the ED use signal. For example, false alarms are more frequent during winter months when signal variance peaks. Previous models adjust for changes with time of the expected number of ED visits but fail to adjust for the changing variance of visits with the season, day-of-week, and long-term trend. The developed model accounts for changes with time of both the expected number of ED visits and the variance of the number of visits. This enables surveillance of known, constant specificity, enhancing the ability of public health practitioners to interpret the meaning of an alarm.

Visual Portrayal of Syndromic Surveillance Data by Using High-Low Charting — Maine, 2004

Gwen M. Rogers, A. Valenti
Department of Epidemiology and Infection Prevention, Maine Medical Center, Portland, Maine

Corresponding author: Gwen M. Rogers, Department of Epidemiology and Infection Prevention, Maine Medical Center, 22 Bramhall St., Portland, ME 04102. Telephone: 207-662-2550; Fax: 207-662-6785; E-mail: rogerg@mmc.org.

Disclosure of relationship: The contributors of this report have disclosed that they have no financial interest, relationship, affiliation, or other association with any organization that might represent a conflict of interest. In addition, this report does not contain any discussion of unlabeled use of commercial products or products for investigational use.

Abstract

Introduction: The Maine Medical Center (MMC) Department of Epidemiology began conducting syndromic surveillance on January 1, 2002, by using computerized chief complaints from the emergency department (ED).

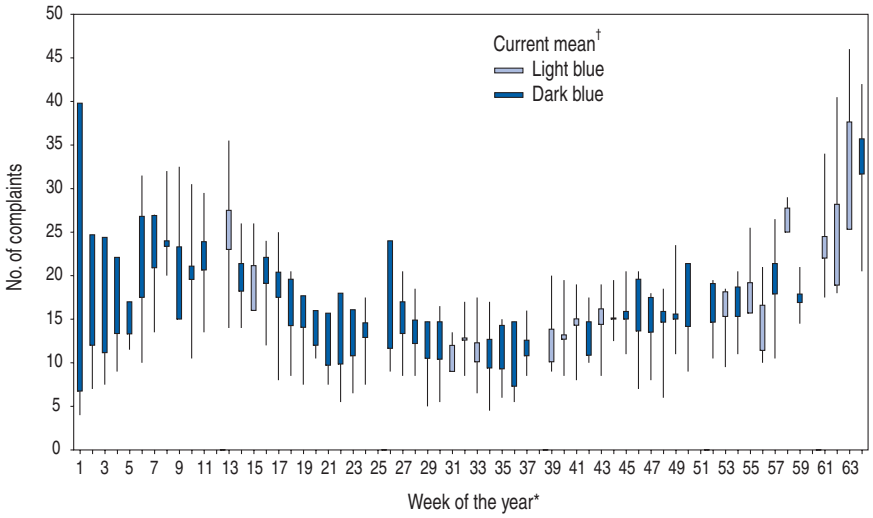
Objective: Using an electronic spreadsheet and corresponding high-low graph, current patterns of chief complaints are measured against historic rates in real time.

Methods: Chief complaints are classified into four syndromic surveillance categories: gastrointestinal, dermatologic, neurologic, and respiratory. They are manually totaled by category, day, and week. Historic high, low, mean, and the mean for the current week are entered into a computerized spreadsheet and graphed. The mean of the current week is visually portrayed within the historic high, low, and mean (Figure). Graphs are returned to EDs monthly, and reviews are conducted when historic rates are exceeded for 2 consecutive weeks.

Results: During fall 2002, the Maine Bureau of Health began receiving reports of gastrointestinal illness from sick callers. A foodborne outbreak was confirmed by syndromic surveillance. Syndromic surveillance demonstrated that the 2003–04 influenza season extended longer and peaked at a much higher rate than in previous years. These are just two examples of the usefulness of the system in detecting public health events in an ED with >80,000 visits each year.

Conclusion: The syndromic surveillance in this report has the potential to enhance public health surveillance by detecting abnormalities in the volume of chief complaints that are of public health concern. The flexibility and utility of the system is enhanced by its simplicity. Measurement of individual chief complaints allows for variation in case definitions and minimal data manipulation. Use of existing software limits operating costs to person time of approximately 40 minutes daily. Manual calculation is required because of a lack of computerized automation and Monday through Friday staffing; weekend data are not calculated until Monday.

FIGURE. Respiratory complaint visits to emergency departments in 2004 compared with the same weeks in 2002–2003*



* Because the last week of each month comprises <7 days, the number of weeks in the year is more than usual, because every week is not 7 days in this model. However, year-to-year, the number of days in each week is the same to allow for comparison.

† The vertical lines are historic highs and lows. The bars comprise the average for that particular week to the average of the same week in the previous years. If the current week is higher than the usual mean, the bar goes up and the color is dark blue. If the current mean is less, the bar goes down and the bar is colored light blue. The length of the bars indicate how far away the usual mean is from the current mean. If the bar extends beyond the line in either direction, it is beyond the historic high or low. The dark blue bars illustrate that the 2004 weekly average was higher than the 2002–2003 average. The light blue bars illustrate that the 2004 weekly average was lower than the 2002–2003 average.

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 K-95, CDC, 1600 Clifton Rd., N.E., Atlanta, GA 30333; telephone 888-232-3228.

All *MMWR* references are available on the Internet at http://www.cdc.gov/mmwr. Use the search function to find specific articles.

Use of trade names and commercial sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

References to non-CDC sites on the Internet are provided as a service to *MMWR* readers and do not constitute or imply endorsement of these organizations or their programs by CDC or the U.S. Department of Health and Human Services. CDC is not responsible for the content of these sites. URL addresses listed in *MMWR* were current as of the date of publication.

☆U.S. Government Printing Office: 2005-733-116/00102 Region IV ISSN: 1546-0738