

Review Paper ■

## Roundtable on Bioterrorism Detection:

### Information System–based Surveillance

---

WILLIAM B. LOBER, MD, MS, BRYANT THOMAS KARRAS, MD,  
MICHAEL M. WAGNER, MD, PhD, J. MARC OVERHAGE, MD, PhD,  
ARTHUR J. DAVIDSON, MD, MSPH, HAMISH FRASER, MBChB, MSc,  
LISA J. TRIGG, BSN, RN, KENNETH D. MANDL, MD, MPH,  
JEREMY U. ESPINO, MD, FU-CHIANG TSUI, PhD

**Abstract** During the 2001 AMIA Annual Symposium, the Anesthesia, Critical Care, and Emergency Medicine Working Group hosted the Roundtable on Bioterrorism Detection. Sixty-four people attended the roundtable discussion, during which several researchers discussed public health surveillance systems designed to enhance early detection of bioterrorism events. These systems make secondary use of existing clinical, laboratory, paramedical, and pharmacy data or facilitate electronic case reporting by clinicians. This paper combines case reports of six existing systems with discussion of some common techniques and approaches. The purpose of the roundtable discussion was to foster communication among researchers and promote progress by 1) sharing information about systems, including origins, current capabilities, stages of deployment, and architectures; 2) sharing lessons learned during the development and implementation of systems; and 3) exploring cooperation projects, including the sharing of software and data. A mailing list server for these ongoing efforts may be found at <http://bt.cirg.washington.edu>.

■ *J Am Med Inform Assoc.* 2002;9:105–115.

---

Affiliations of the authors: University of Washington, Seattle, Washington (WL, BK, LT); University of Pittsburgh, Pittsburgh, Pennsylvania (MW, JUE, FCT); Indiana University School of Medicine, Indianapolis, Indiana (JMO); Denver Public Health Department, Denver, Colorado (AD); Children's Hospital, Boston, Massachusetts (HF, KDM).

This work was supported in part by grants GO8 LM06625-01, T15 LM/DE07059, N01-LM-4-3510, and N01-LM-6-3546 from the National Library of Medicine; contracts 290-00-0009 and 290-00-0020 from the Agency for Healthcare Research and Quality; a contract from the Air Force under the DARPA Biosurveillance Program; and cooperative agreement U90/CCU318753-01, Cooperative Agreement U90/CCU817608-02, Health Alert Network/Training Exemplar Projects, and State Bioterrorism Preparedness grant (B2 section) U90/CCU017010-02 from the Centers for Disease Control and Prevention.

The contents of this paper are solely the responsibility of the authors and do not represent the official views of the agencies.

Correspondence and reprint requests: William B. Lober, MD, MS, Box 357420, 1959 NE Pacific St, Seattle, WA 98195; e-mail: <[lober@u.washington.edu](mailto:lober@u.washington.edu)>.

Received for publication: 11/12/01; accepted for publication: 11/21/01.

Bioterrorism has quickly become a new and frightening part of life in America. A host of potential agents, with varying degrees of virulence and a confusing array of nonspecific symptoms, are now household words. The field of medical and public health informatics has long concerned itself with developing methods to represent, store, and analyze data that describe the complexities of individual and population-based health.<sup>1</sup> Now, informatics tools such as knowledge representation, controlled vocabularies, heterogeneous databases, security and confidentiality, clinical decision support, data mining, and data visualization are being applied with a new urgency to the task of early detection of intentional outbreaks of disease.

In November 2001, as part of the activities of the Anesthesia, Critical Care, and Emergency Medicine Working Group, investigators from several research groups took part in the "Roundtable on Bioterrorism Detection" at the AMIA Annual Symposium. The session was subtitled "Information System–based Sentinel Surveillance." These researchers, and others,

are developing public health surveillance systems that make secondary use of data gathered during normal clinical workflow or that facilitate electronic case reporting by clinicians. These surveillance strategies are intended to enhance early detection of changes in the health of the community. This paper combines brief case reports of a number of existing systems with a discussion of some commonly employed techniques and approaches.

Several bioterrorism-related posters and papers were presented at the Symposium.<sup>2-7</sup> A handful of systems, all in active development, are currently deployed. The utility of these systems in detecting bioterrorism events is unproven, and it is hoped that their full capabilities will never need to be tested directly. However, the value of monitoring and aggregating disease indicators across a population is clear, if intuitive, and such surveillance has a strong precedent in public health practice.<sup>8-10</sup>

There are strategies for indirectly measuring the performance of these systems and for improving their diagnostic accuracy and timeliness, even in the absence of bioterrorism cases. These strategies include measuring the accuracy of detection of components of case definitions, as opposed to detection of outbreaks. Other strategies involve the detection of surrogate diseases, such as influenza, whose symptoms are similar to the initial symptoms of inhalational anthrax. Espino et al.<sup>4</sup> showed a 44 percent sensitivity and 97 percent specificity in detection of cases of acute respiratory illness, a common symptom prodrome of many illnesses spread by bio-aerosol agents. A companion study<sup>3</sup> showed that time-series analysis of such cases in a population could detect an outbreak of influenza. McClung et al.<sup>11</sup> found relatively similar sensitivity and specificity (37 and 97 percent, respectively) in a system detecting asthma visits, based on chief complaint on presentation to an emergency room.

A number of federal and other agencies have funded the work on these surveillance systems. These include the Centers for Disease Control and Prevention (CDC), through State Bioterrorism Preparedness grants, the Health Alert Network program, and cooperative agreements; the Agency for Healthcare Research and Quality (AHRQ); the Defense Advanced Research Projects Agency (DARPA); the National Library of Medicine (NLM), both directly through grant funding and indirectly through support of NLM Fellowships in Informatics and Integrated Advanced Information Management System sites; and by state and local public health agencies using CDC and other funds.

## AMIA Roundtable Discussion

The overall goal of the roundtable was to foster communication and cooperation among researchers in an effort to increase the pace of research and system deployment. The specific aims of the roundtable were to 1) share information about the systems, including their origins, goals, current capabilities, stages of deployment, and architectures; 2) share lessons learned during the development and implementation of these systems; and 3) explore cooperation between projects, which may include the sharing of designs, software, test data sets, operational data, and algorithms.

Representatives of systems being developed at Public Health–Seattle and King County/University of Washington, the University of Pittsburgh, Children’s Hospital Boston, Denver Public Health, and Stanford University all spoke. Investigators from University of North Carolina, Nebraska Public Health, George

Table 1 ■

### Centers for Disease Control and Prevention (CDC) Description of Surveillance Systems<sup>12</sup>

Public health importance of the health-related event under surveillance:

- Indexes of frequency, severity, disparities, associated costs, preventability potential clinical course and public interest

Purpose and operation of the system:

- Purpose and objectives of the system
- Planned uses of the data
- Case definition/event under surveillance
- Legal authority for data collection
- Organizational home of system
- Level of integration with other systems
- Flowchart
- Description
  - Population
  - Interval of data collection
  - Data collected
  - Reporting sources
  - Data management
  - Data analysis and dissemination
  - Patient privacy/data confidentiality/systems security
  - Records management program

Resources used to operate the surveillance system:

- Funding sources
- Personnel requirements
- Other resources

SOURCE: “Task B. Describe a Surveillance System to be Evaluated,” which is part of the CDC methodology for evaluating public health surveillance systems.<sup>12</sup>

Table 2 ■

## Comparison of Data Types of Several Surveillance Systems

Surveillance System	CC Free Text	Symp Survey	CC ICD-9	Dx ICD-9	Age	Gender	Date Time	Visit Site	Arrival Mode*	Disposition†	Geo Code or Address	Other‡
Bergen County, New Jersey	X		X	X	X		X		X		City	
Children's Hospital Boston	X	X			X		X				X	1
Denver Public Health	X			X	X	X	X	X	X		X	2
Dept. of Defense/GEIS				X	X	X	X	X		X	X	3
Los Alamos/U. New Mexico <sup>5</sup>		X			X	X	X	X			Zip	4
Regenstrief/Indiana U.	X		X	X	X	X	X	X				5
Seattle-King County/ U. Washington	X			X	X	X	X	X	X	X	Zip	6
U. Pittsburgh	X		X	X	X	X	X	X			Zip	7

NOTES: The surveillance systems are identified by the organization(s) responsible for their development—Bergen County Department of Health Services, New Jersey; Children's Hospital Boston, Massachusetts; Denver Public Health Department, Department of Defense, Global Emerging Infections System, Johns Hopkins Applied Physics Laboratory, George Washington University, and Carnegie Mellon University; Los Alamos National Laboratories, University of New Mexico, and New Mexico State Department of Health; Regenstrief Institute and Indiana University School of Medicine; Public Health—Seattle & King County and University of Washington; University of Pittsburgh.

ABBREVIATIONS: CC indicates chief complaint; Dx, diagnoses; H, home; W, work.

\* Paramedic or walk-in, for example.

† Admitted to hospital or returned home, for example.

‡ Numbered notes: 1, Presenting complaint coded internally, patients' symptoms in free text. 2, Structured symptoms, computer-assisted phone triage protocols. 3, Provider seen, home zip, site zip, date only (no time included), longitudinal follow-up capability (SOURCE: Kelley P, Walter Reed Army Institute of Research, personal communication; Feb 5, 2002.) 4, Home and work zip codes, provider seen, occupation. (ADDITIONAL SOURCE: Brillman J, University of New Mexico, personal communication; Feb 6, 2002.) 5, Variable by location—laboratory results, radiographic study reports, vital signs, encounter data, procedures, images, electrocardiograms, notes, adding inpatient medications, surgical notes, surgical pathology, tumor registry. 6, Emergency medical services dispatch, hazardous material calls. 7, Orders, cultures, x-rays, laboratory results, dictations, home and work zips.

Washington University, and CDC identified themselves. The authors have also spoken to investigators at Vanderbilt University, Regenstrief Institute, and Bergen County Department of Health Services (New Jersey). A total of approximately 64 people attended the roundtable, which lasted 90 minutes. We have certainly, if inadvertently, failed to mention others who contributed valuable information and insight, and we apologize for this.

## Methods

Investigators described a number of systems, representing several approaches and various stages of implementation. For purposes of this report, we have focused on systems that are currently operational.

The CDC has updated its guidelines for evaluating public health surveillance systems.<sup>12</sup> The description criteria for surveillance systems, taken from these guidelines, comprise 12 major and 8 minor categories and are listed in Table 1. This framework helped us organize this overview of surveillance systems developed by roundtable participants and others, as summarized in Table 2. In the interest of brevity, we combined certain categories, included others only when we judged the available information to be significant, and reported certain categories as common to all systems.

Several CDC criteria, although important, were difficult to apply to these systems, because they are in early stages of system development; others could be applied equally to all systems, given the common focus on bioterrorism. For instance, these systems share a common purpose and intended utility for early detection of intentional outbreaks and to varying extents address the need for tactical communication once an outbreak is identified. Issues of case definition and patient privacy/data confidentiality remain largely unresolved as our society at large grapples with the implications of this new threat.

## System Descriptions

These systems differ in many respects, including history, funding, implementation, and methodologies used to collect and analyze data. Although they have evolved independently, these systems display striking similarities in the types of data they collect and in their overall system architectures.

Each project description begins with a discussion of the organizational home of the project, funding, and legal issues or operational agreements. If applicable, "Scope" addresses the covered population and reporting sources. "Data Collection" describes data elements and timeliness of data. "Design" includes descriptions

of how data are collected, where they are stored, and how security and reliability issues are addressed. "Data Analysis" describes the user interface for accessing the data, the algorithms to interpret the data, and dissemination of the results of the algorithms. For each system, there is also a "Lessons Learned" section.

### **Bergen County, New Jersey**

The Bergen County (New Jersey) Department of Health Services has implemented a county-wide system. System development has taken place for about 1 year, with substantial acceleration since Sep 12, 2001.

All five acute care facilities in Bergen County report emergency department census data daily, and two facilities also transmit visit-level data electronically. Two of the three remaining facilities are scheduled to begin transmission shortly. The visit-level data include patient's age, date of visit, chief complaint/reason for visit, method of transport, and zip code. Historical data have been back-loaded for approximately a year. Emergency department census data are reported through a Web-based reporting system or by fax transmission. An epidemiologist monitors these data.

### **Children's Hospital Boston**

The Children's Hospital project, which is funded by the AHRQ, has concentrated on four areas of bioterrorism surveillance: collection of data through a Web-based reporting tool, real-time analysis of existing emergency department data, development of an online diagnosis and treatment manual, and development of decision support systems tailored to early detection.

#### Scope

The Children's Hospital surveillance system incorporates both a Web-based reporting system and a sentinel syndrome database-mining surveillance system. Similar data are available from the emergency department of Beth Israel Deaconess Medical Center. Additional collaborators include the Boston Medical Center and four community hospitals, and discussions with other hospitals in the area are taking place.

#### Data Collection

Routine data recorded from visits to the emergency department of Children's Hospital include date and time of assessment, age, address, presenting complaint (coded internally), and free text describing symptoms. Seven years of data are available, allowing annual trends to be ascertained. Data are available in real time and are augmented by survey data. The survey data

are collected via a Web-based form over an SSL connection and include a detailed list of symptoms and probable diagnoses.

#### Design

The simple Web-based reporting tool to collect data about suspected cases is intended for use similar to that of the "drop-in surveillance" performed by the CDC, but the Children's Hospital system is more automated and includes more detailed data. Details of patients' diagnoses and some types of symptoms can be recorded in a secure fashion on the server. Researchers at Children's Hospital have recently begun a trial of the system. This approach may make it possible to collect very specific data in a structured form. However, it requires extra time of physicians, which limits its use.

#### Data Analysis

Current analyses of visit-level data from the Children's Hospital emergency department include temporal patterns over days, weeks, and years and geographic information system models. Researchers from Harvard University are developing novel detection methods using spatial clustering.

Coworkers at the Massachusetts Institute of Technology have developed a Web-based decision support tool to assist in the identification of illness caused by bioterrorism agents. Input forms record symptoms, signs, and possible syndromes for patients admitted to the emergency department. Two inference mechanisms are used. In the first, the diagnoses from the manual are just linked to tables of findings (divided into early and late stages), allowing the user to link to relevant parts of the treatment manual. In the second system, a Bayesian belief network is used to combine prior probabilities of potential bio-agents with the odds ratios from patient findings. The output is a list of possible agents ranked by probability. Currently, anthrax, smallpox, and West Nile virus are included as nodes in the model, and other nodes are being added.

#### Lessons Learned

Although presenting complaints and diagnoses are useful, it will be important to examine the symptom level. This extension may require additional reporting to augment existing data sets. In addition to the work described above, researchers at Children's Hospital are developing a diagnosis and treatment manual. This manual will shortly be published on the Web and in print.

## Denver Public Health

The Denver Center for Public Health Preparedness, an exemplar site in the CDC-funded Health Alert Network,<sup>13</sup> is housed at the Denver Public Health (DPH) Department. This collaborative center includes participation by the emergency department of Denver Health Medical Center and the Rocky Mountain Poison and Drug Center.

### Scope

The Denver Center for Public Health Preparedness is currently developing a syndromic surveillance system to detect, in near real time, unusual symptom patterns or syndrome incidence in the City and County of Denver. Denver Health annually serves nearly 130,000 (25 percent) of Denver's population. It is a unique, vertically integrated public health care system that includes a public hospital, level-1 trauma emergency department, the county emergency medical system, a network of nearly two dozen community- and school-based clinics, the Rocky Mountain Poison and Drug Center (which operates a nurse advice line), and Denver Public Health.

### Data Collection

Several Denver Health data sources are being analyzed and evaluated for their utility in syndromic surveillance. Visit-level patient-specific data are available from virtually every source, and chief complaint is recorded by the nurse advice line, emergency department, and emergency medical service computer-aided dispatch. In addition, ICD-9 discharge codes are available from emergency department sites as well as all others (i.e., urgent care, hospital admissions, and community and school-based clinics). The emergency department stores a textual description of a patient's chief complaint and final diagnosis. The computer-aided dispatch system captures a code indicating the nature of the problem. The nurse advice line stores the name of the guideline used to advise a caller, which generally corresponds to the caller's symptoms.

### Design

The system employs ad hoc queries of existing server and mainframe data systems to produce textual reports that are converted to relational databases for analysis. Emergency department data are available on an hourly basis by query, whereas the other data reports are processed nightly. No additional provider input is required, since symptom data are collected at triage (for the emergency department,

computer-aided dispatch, and nurse advice line) and ICD-9-coded data are available for administrative purposes at the end of each encounter.

### Data Analysis

Historic emergency department system data (from approximately 50,000 visits per year) from 1998 to 2000 have been used to test the syndromic surveillance concept. Asthma is used as a model disease because of its high prevalence, seasonal trend, varying severity, and characteristic symptoms (e.g., dyspnea, cough, and shortness of breath), which are similar to those of some illnesses caused by bioterrorist agents (e.g., inhalational anthrax).

Asthma-related utilization data, identified by symptoms (wheezing, shortness of breath, cough) or by diagnosis code (ICD-9 code 493), were collected from all sources to compare utilization trends among the health facilities being accessed.<sup>11</sup> Geographic information systems are also used for reporting and analysis, since patient address is a component of the ad hoc reports.

### Lessons Learned

While development and evaluation of syndromic surveillance continues within Denver Health, a primary goal of the project is the definition of appropriate alert thresholds. Efforts to adequately evaluate and enhance the sensitivity and predictive value positive are essential. Acquisition of new data sources, from non-Denver Health institutions, is being planned. Additional disease modeling analyses, to include seasonal factors, access patterns for health facilities, and inclusion of environmental factors, are under way to better define respiratory symptom-based signals in such surveillance systems.

## Regenstrief Institute/Indiana University

Investigators at the Regenstrief Institute created the Indianapolis Network for Patient Care (INPC) in 1995 with the goal of improving the medical care of patients.<sup>14</sup> The network is an operational community-wide electronic medical record that includes an active surveillance component built around real-time electronic laboratory reporting. The NLM and AHRQ have supported the initial development of the network.

### Scope

The system currently includes data from 11 hospitals in five health systems, the Marion County Health Department, and various physician practices. These hospitals account for over 95 percent of all beds and

emergency department visits in the Indianapolis metropolitan statistical area, which has a population of 1.5 million.

#### Data Collection

The data collected include demographics; laboratory results; and emergency department, inpatient, and outpatient encounter data. The encounter data include chief complaint, coded diagnoses and procedures, immunizations, medications, allergies, electrocardiogram tracings and results, echocardiogram images and results, radiographic images and reports, vital signs and other data, but not all these data elements are available from every participating hospital. The core set of data received from all participating hospitals includes demographics, laboratory data, and chief complaint, coded diagnoses, and coded procedures from emergency department and inpatient encounter data. Results of a pilot study suggest that making these data available to emergency department providers reduces costs and improves care.<sup>15</sup>

The system currently utilizes the real-time laboratory result data for active surveillance of reportable conditions.<sup>16</sup> Under a Memorandum of Understanding with the Indiana State Department of Health, the system compares laboratory data with the Dwyer table (CDC) of reportable results.<sup>17</sup>

#### Design

The INPC receives data from each participant, most as real-time HL7 messages over a secure extranet. The system standardizes the message format and codes and stores the data in the network database in real time. Data for some participants, such as diagnosis and procedure codes and immunization registry data, which are not updated in real time, are sent in a batch file format.

#### Data Analysis

When the system identifies results that indicate a reportable condition, it adds patient demographics and ordering-provider data, such as office telephone numbers and addresses, and previous related results, to the reportable disease databases. The system copies the database to the Marion County Health Department and Indiana State Department of Health each night. In addition, the system sends several public health officers and the investigators an e-mail summary of new cases each morning, which includes a flow sheet showing recent trends. Two general internists with emergency department experience and training in epidemiology review these data daily.

#### Lessons Learned

The INPC follows the National Electronic Disease Surveillance System (NEDSS) architecture and can serve as a laboratory for implementation. Active surveillance provides a sustainable method for monitoring public health. Because of variability in how the HL7 standard is implemented, combining data from multiple health care delivery systems can be difficult.

In addition, few laboratory results are identified with LOINC codes (Logical Observation Identifiers, Names, and Codes (<http://www.regenstrief.org/loinc/index.html>)), so considerable effort is required to map the data to a standard code set. Finally, the results themselves are often unstructured, so text matching is the only method available to identify the value of results.

#### Public Health—Seattle and King County/ University of Washington

The Syndromic Surveillance Information Collection (SSIC)<sup>2</sup> results from collaboration between the Clinical Informatics Research Group at the University of Washington School of Medicine; Jack Ciliberti, MD, medical director of the emergency department at Overlake Hospital Medical Center (Bellevue, Washington); and Public Health—Seattle and King County. This team is working to develop a detection system for regional outbreaks of disease, whether naturally occurring or caused by intentional release of bioterrorism agents.

The SSIC is an automated data collection system that has been in place since March 2001. It is part of a multi-component surveillance system run by Public Health—Seattle and King County, which includes passive and active surveillance of school absenteeism, unexplained deaths (in collaboration with the Medical Examiner's Office), and emergency medical system (ambulance/medic) dispatch data.

#### Scope

This regional system covers King County, Washington, which includes the Seattle metropolitan area, with a total population of 1.7 million. The system receives data from three emergency departments, representing about 120,000 visits annually, and from nine primary care clinics.

#### Data Collection

Data are collected daily via automated transmission from the source information systems. The data include date, time, age, gender, chief complaint/rea-

son for visit, disposition, ICD-9 diagnosis and, in some cases, zip code. In addition, geo-coded emergency medical dispatch data for the city of Seattle are collected in real time. Investigators are currently exploring the collection of both laboratory culture data and poison center call data.

### Design

The SSIC comprises two components, the upload engine and the query engine. The upload engine is a collection of processes that facilitate the secure collection of data from heterogeneous data sources, and their storage. The query engine enables public health experts to manipulate those data and run aberration detection algorithms against it.

The data from the heterogeneous collection sites are sent to the upload engine on a secure production server, where they are filtered into a uniform XML format. These XML files are converted to SQL and stored in a Microsoft SQL database on a highly secure internal server, which communicates only with the production server.

The XML files also trigger a process that converts them to text files, which are then stored using a naming system that signifies the source, date, and time of the data. Then e-mail is sent automatically to principal developers and public health researchers, informing them that new data are available for analysis.

Investigators have built two types of interfaces to the system. The primary interface is an SSL-encrypted channel based on XML-structured data. The second interface, used only in the academic medical center, is based on the clinical e-mail system. Summary reports generated by a clinical information system are sent to the server via clinical e-mail. Monitoring software checks for the arrival of anticipated data sets at pre-selected times and notifies lists of users by e-mail about the presence or absence of those transmissions.

### Data Analysis

The data are made available to Public Health–Seattle and King County for analysis in two ways. First, incoming data sets are normalized to a common format, and an immediate e-mail notification is sent. Second, a query engine is available via a Web-based form, which permits users to request data by inputting range, dates, and source/site. The query triggers a program on the Web server, which then queries the Microsoft SQL database on the internal secure server. The database returns a text file containing the requested data as delimited text sent over an SSL-encrypted channel.

### Lessons Learned

The focus thus far has been on building a heterogeneous, multi-institutional data collection network. Consequently, investigators are slowly learning how to work with hospital management personnel from other institutions and with personnel from outside information technology departments. They have developed a secure, minimally invasive solution for generating and transmitting data sets from a source system; however, some information technology groups prefer to develop their own reporting strategies, transmitting text documents via e-mail, FTP, or other means. The challenge of establishing a common format among various source systems has led investigators to implement translation and conversion capabilities centrally, which will provide more flexibility as the number of sites increases.

### University of Pittsburgh

The Real-time Outbreak and Disease Surveillance (RODS) system is a public health surveillance system that has been deployed since 1999 in western Pennsylvania. It has been developed by the RODS Laboratory of the Center for Biomedical Informatics at the University of Pittsburgh, with funding from NLM, AHRQ, CDC, and DARPA. The legal basis for RODS public health surveillance is established by a set of trilateral Memoranda of Understanding executed between each health system, the health department, and the University of Pittsburgh.

### Scope

In late 2001, RODS was receiving, from ten emergency departments in the region, data about the volume of patients presenting with chief complaints of diarrhea, rash, respiratory illness, and other key symptoms. (RODS collects microbiology data and other data from 17 hospitals, but chief complaints in emergency departments are the current focus of rapid expansion.)

The percentage coverage of regional emergency department visits is as follows: 37 percent of the central urban region (population, 1.3 million), 23 percent of the metropolitan statistical area (population, 2.3 million), and 20 percent of the broader region, encompassing a total of 13 counties (population, 3 million). Interfaces are also under construction for four additional hospitals that have executed Memoranda of Understanding. With the addition of these hospitals, the coverage will increase to 43, 26, and 25 percent for the central urban, metropolitan, and broader areas, respectively. Additional hospitals are reviewing the technical and administrative proposal.

### Data Collection

RODS collects emergency room registration data, microbiology culture results, reports of radiographs, dictations of emergency room clinicians, test orders, and results of laboratory tests, such as cerebrospinal fluid analyses, as follows:

- Item (i): an abstract of data from each emergency department visit, comprising time of visit, patient age, gender, chief complaint, home zip code, work zip code, and a sequential transmission number
- Item (ii): microbiology culture reports that are anonymous and consist of time of culture, patient age, gender, home zip code, work zip code, and sequential transmission number
- Item (iii): orders for stool, throat, and blood cultures that are anonymous and consist of time of order, patient age, gender, home zip code, work zip code, and sequential transmission number
- Item (iv): chest radiograph reports that are anonymous and consist of time of radiograph, radiographic findings identified by natural language processing, patient age, gender, home zip code, work zip code, and sequential transmission number
- Item (v): results of emergency department dictations that are anonymous and consist of time of visit, findings identified by natural language processing, patient age, gender, home zip code, work zip code, and sequential transmission number.
- Item (iv): such other data as agreed on by the parties and the Governing Board that are consistent with the restrictions contained in the agreement

The Memoranda of Understanding have been executed at the option of the health systems for a minimal data set, which includes item (i) only, or for the full set of data.

### Design

Technically, RODS comprises an Oracle database that uses a data model derived from the Public Health Conceptual Data Model and the NEDSS base data model. The database receives new data in real time by means of HL7 messages from other computer systems, such as registration systems and laboratory information systems, over a Secure Shell-protected Internet connection. Reliability of a distributed scheme is dependent on the reliability of the data providers. Mechanisms to restart interfaces automatically and monitor their integrity have produced very high availability over several years. The user interfaces are Web based.

### Data Analysis

RODS provides tools that help detect the presence of a disease outbreak and support the characterization of that outbreak by a public health official. These tools include case definitions, automatic detection algorithms that can be attached to specific data streams, and data analysis tools that support temporal and spatial data analysis and visualization.

Health-related events may be defined at multiple levels—cases, features of cases (e.g., wide mediastinum), and outbreaks. For example, case definitions have been generated for seven prodromes, based on chief complaints. (The seven prodrome groups are rash and botulinic, encephalitic, respiratory, hemorrhagic, diarrheal, and viral groups). Other definitions include patients with first-time positive cultures (especially but not limited to notifiable diseases), patients for whom a blood, sputum, throat, or urine culture was obtained, and positive results from parsing of a chest radiograph impression. When the frequency of one of the seven predefined prodromes exceeds expectations for normal emergency department rates of occurrence, notification takes place by e-mail and pager.

### Lessons Learned

A public health surveillance system must be integrated with the public health investigation and response processes. A lesson learned from RODS is that public health capacity in many parts of the country is overwhelmed. As a result, the RODS Laboratory has been asked to monitor the output of the system themselves and report suspicious events to public health. A group of physicians with public health, emergency department, and infectious disease training has taken on this responsibility.

Critical to the success of public health surveillance is making the tradeoff between the potential benefit of early detection and the potential risk to privacy, in ways that are satisfactory to society. A cornerstone of the RODS approach involves the use of a trusted broker, governed by the community. A trusted broker is a secure computing environment with privacy policies that are governed by the community.

### Discussion

The roundtable met its specific aims to share information, to share lessons learned, and to explore collaboration. This paper is the most visible consequence of the meeting and the interactions among participants over the following week. It addresses the aim of sharing information under System Descriptions.

Table 3 ■

## Comparison of System Characteristics

System	Geography/ Population	Setting/ Data Sources	Data Transmission Standards	Data Update Frequency	Data Collection Technique	Security Protocol	Funding
Bergen County, New Jersey	County; 5 hospitals	ED	Text	Daily	Fax, e-mail, file transfer	—	—
Children's Hospital Boston	Metropolitan; 2 hospitals	ED	XML	Daily	Mining agent and reports	SSL	AHRQ
Denver Public Health	Metropolitan; 600K, 25%	ED, EMS CAD, nurse advice line	ODBC	Daily (ED hourly)	Electronic from ED text and CAD, ad hoc queries		CDC
Dept. of Defense/GEIS	14 countries, 395 installations (307 in USA, 88 other)	Military treatment facilities	Column- delineated	Daily	Data mining	Secure FTP vpn	DARPA
EI Network <sup>23</sup>	21 Pacific Rim countries	Disease alerts	Text, e-mail	2 weeks	Voluntary reporting		CDC
Los Alamos/ U. New Mexico	2 counties, 2 EDs	ED	Http, XML	Voluntary	Reports	SSL	DOE, CDC
Regenstrief/Indiana U.	Metropolitan; 11 hospitals, 5 health systems; 1.5M, 95%	ED, hospital, physician groups	HL7	Real-time	HL7 over secure extranet		NLM AHRQ
Seattle-King County/ U. Washington	1 county, 3 hospital EDs, 9 PCC; 1.7M, 20%	ED, primary care, EMS CAD	XML	Daily	Mining agent and reports	SSL	CDC
U. Pittsburgh	13 counties, 14 hospitals, 10 EDs; 3M, 20%; 1.3M, 37%	ED hospitals	HL7, ODBC	Real-time	HL7, free-text, processing physician case reporting	SSH	DARPA AHRQ NLM CDC

NOTES: The surveillance systems are identified by the organization(s) responsible for their development—Bergen County Department of Health Services, New Jersey; Children's Hospital Boston; Denver Public Health Department; Emerging Infectious Diseases Network, Asia-Pacific Economic Cooperation (EINet-APEC); Regenstrief Institute and Indiana University School of Medicine; Public Health-Seattle & King County and University of Washington; University of Pittsburgh.

ABBREVIATIONS: AHRQ, Agency for Healthcare Research and Quality; CAD, computer-assisted dispatching; CDC, Centers for Disease Control and Prevention; DARPA, Defense Advanced Research Projects Agency; DOE, Department of Energy; ED, emergency department; EMS, emergency medical service; HL7, Health Level 7; NLM, National Library of Medicine; ODBC, Open Database Connectivity; PCC, primary care clinic; SSH, Secure Shell; SSL, Secure Sockets Layer; XML, Extensible Markup Language; vpn, virtual private network.

## Similarities

The systems were developed independently but converged on similar solutions to the problem of early detection utilizing similar types of data and relying on the Internet for connecting institutions. Other interesting similarities between the systems are their stages of development, geographic scope, and the types of data elements they collect (see Tables 2 and 3). These similarities are difficult to attribute; however, they may well relate to the pre-existent availability of relevant information in electronic form and the perception that the population of the country is clustered around metropolitan areas.

All the sites indicated concerns with maintaining security and confidentiality. These concerns appear to be less of a problem for systems operating within

a single health care institution, perhaps because the existing security and confidentiality policies of the institution already address clinical requirements. The systems generally adhere to the principle, espoused by the Health Insurance Portability and Accountability Act (HIPAA) of 1996,<sup>18</sup> of collecting a minimum number of patient identifiers. Although specific methodologies differ, most systems use encryption for the transmission of data. Certain systems, however, in an effort to facilitate data transfer from institutions that are not capable of encryption, do accept automated e-mail of de-identified data.

Several systems use clustering of ICD-9 codes to define disease prodromes of interest in bioterrorism detection. Clustering, instead of using individual codes, is motivated both by the concern that codes are too fine-grained for bioterrorism detection and by

concerns with coding accuracy, inter-coder reliability, and coding variance over time. By clustering codes in prodromal groups, researchers hope to include all codings that might conceivably be applied to a patient presenting with relatively early symptoms of an infectious or toxic syndrome. Clustering schemes have been proposed by USAMRIID<sup>19</sup> and AHRQ<sup>20</sup> but have not yet been universally adopted.

### Differences

The systems have interesting difference. These projects represent a variety of relationships with the public health system. Some projects began inside local, state, or national public health agencies, while others have come from academic medical centers. Regardless of their origin, these projects generally develop in cooperation between organizations providing data, system developers, and public health.

In these systems, data may be acquired at the level of an individual visit (with either primary or multiple diagnoses), of a patient (in either a snapshot or longitudinal view), or of the population aggregate (such as emergency department visits per day for gastrointestinal complaints). Currently, most systems are using the visit or case-report level of detail; however, some systems collect multiple levels of data, such as visit data for certain diagnoses or syndrome clusters combined with emergency department volume data. Some differences are due to data availability: Some systems receive a single diagnosis for a visit, whereas other systems receive multiple diagnoses.

Geographic information is encoded to different levels of granularity by different systems. The spectrum ranges from geo-coding of street address, to zip code, to municipality, and to none at all. Table 3 is a comparison of selected system characteristics.

### Sharing Information and Software

Several types of collaborations among participants were discussed, including sharing of information, code, and data. An e-mail discussion list, established prior to the meeting, remains active. It has 29 members and can be found at <http://bt.cirg.washington.edu>. In addition, the AMIA working group hosting the roundtable has agreed to distribute the addresses of the meeting attendees. Finally, this paper itself is the fruit of collaboration between six groups of people who had not previously worked closely together on surveillance systems.

During the meeting, two of the authors, Wagner and Lober, indicated their willingness to share software

with other sites interested in regional surveillance. The successful implementation of a sentinel surveillance system by a site not involved in developing the software will mark the transition from "research prototype" to "early production." As with any software development, this step would mark an important transition toward wide dissemination of a product.

### Sharing Data

The participants expressed interest in linking their systems; at present, integration is typically limited to the systems providing data. Participants discussed motivations and several models for data integration. First, having test data from another system may be valuable for validating detection algorithms. Second, contemporaneous data from a similar region would enable detection algorithms to compare data against a regional control rather than a historical control. Third, just as it is valuable to understand patterns on a regional basis, rather than patterns specific to a single hospital or clinic, it is important to monitor health status on a national level.

Several models support these types of data integration, which are not mutually exclusive. One model is "peer-to-peer" sharing, based on the exchange of data in a common format between two regional systems. Another model is "centralized" sharing, in which the data are collected by a central agency such as the CDC. NEDSS<sup>21</sup> was discussed as an example of this model. NEDSS is a set of interrelated activities and standards intended to facilitate "... complementary electronic information systems that automatically gather health data from a variety of sources on a real-time basis; facilitate the monitoring of the health of communities; assist in the ongoing analysis of trends and detection of emerging public health problems; and provide information for setting public health policy."<sup>22</sup>

### Conclusion

The systems we examine in this paper, while differing in the details of implementation and detection strategies, share several common characteristics, including common goals and similar data sets. They also share a basis in an approach to early detection of outbreaks that is only partially proven. However, the threat we face is so immediate, and so urgent, that parallel deployment and validation is deservedly a strong theme in the philosophy underlying all these efforts.

As of this writing, 17 persons have had confirmed *Bacillus anthracis* infection, and another 5 cases are suspected. Although the true risk of a widespread

attack with biological weapons is difficult to estimate, the terror produced by the fear of these attacks is very real. Medical and public health informatics have the responsibility to mobilize, much as shipyards and steel mills have done in previous conflicts. We need to set aside traditional concerns with credit and competition, and work together to build the systems that may make our cities, our states, and our nation more secure. Our goal is to learn from one another, improve on one another's work, and do the best job science and informatics can in helping make our society both safer and healthier. We hope the roundtable meeting, and this report, will be first steps in the direction of cooperation.

The authors thank Jeff Duchin from Public Health—Seattle and King County; Marc Paladini from Bergen County Department of Health Services; Mark Oberle, James Gale, and Ann Marie Kimball from University of Washington School of Public Health and Community Medicine; and Patrick O'Carroll, Bill Yasnoff, and John Loonsk from the Centers for Disease Control and Prevention for their discussion and contributions; and David Bliss for his technical insight. They also thank Vasu Brown for sponsoring the roundtable through the AMIA Anesthesia, Critical Care, and Emergency Medicine Work Group, and AMIA staff for helping them organize and publicize this meeting.

#### References ■

1. Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, Kilbourne E. Public Health Informatics: Improving and transforming public health in the information age. *J Public Health Manag Pract.* 2000;6(6):67-75.
2. Duchin JS, Karras BT, Trigg LJ, et al. Syndromic surveillance for bioterrorism using computerized discharge diagnosis databases. *Proc AMIA Annu Symp.* 2001:897.
3. Tsui FC, Wagner MM, Dato V, Chang CCH. Value of ICD-9-coded chief complaints for detection of epidemics. *Proc AMIA Annu Symp.* 2001:711-5.
4. Espino J, Wagner M. The accuracy of ICD-9-coded chief complaints for detection of acute respiratory illness. *Proc AMIA Annu Symp.* 2001:164-8.
5. Zelicoff A, Brillman J, Forslund DW, et al. The rapid syndrome validation project (RSVP). *Proc AMIA Annu Symp.* 2001:771-6.
6. Zeng X, Wagner MM. Modeling the effects of epidemics on routinely collected data. *Proc AMIA Annu Symp.* 2001:781.
7. Malloy WP, Sweeney L. Electronic disease surveillance and reporting: the e-Report System. *Proc AMIA Annu Symp.* 2001:964.
8. Halperin W, Baker EL, Monson RR (eds) *Public Health Surveillance.* New York: Van Nostrand, 1992.
9. Teutsch S, Churchill RE. *Principles and Practice of Public Health Surveillance.* 2nd ed. New York: Oxford University Press, 2000.
10. Centers for Disease Control and Prevention. Guidelines for evaluating surveillance systems. *MMWR Morb Mortal Wkly Rep.* May 6, 1988;37(5-5):1-18.
11. McClung MW, Davidson AJ, Vogt RL, Cantrill SV, Jones RH. Evaluating data sources for syndromic surveillance. Presented at: American Public Health Association 129th Annual Meeting, session 3133; Oct 22, 2001; Atlanta, Georgia.
12. Centers for Disease Control and Prevention. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Morb Mortal Wkly Rep.* Jul 27, 2001;50(RR-13):1-35.
13. Rotz LD, Koo D, O'Carroll PW, Kellogg RB, Lillibridge SR. Bioterrorism preparedness: planning for the future. *J Public Health Manag Pract.* 2000;6(4):45-9.
14. Overhage JM, McDonald CM, Tierney WM. Design and implementation of the Indianapolis Network for Patient Care and Research. *Bull MLA.* 1995;83(1):48-56.
15. Overhage JM, Dexter PR, Perkins SM, et al. A randomized controlled trial of clinical information shared from another institution. *Ann Emerg Med.* 2002;39(1):14-23.
16. Overhage JM, Suico J, McDonald CJ. Electronic laboratory reporting: barriers, solutions and findings. *J Public Health Manag Pract.* 2001;7(6):60-6.
17. McDonald CJ, Overhage JM, Dexter P, Takesue BY, Dwyer DM. A framework for capturing clinical data sets from computerized sources. *Ann Intern Med.* 1997;127:675-82.
18. National Committee on Vital and Health Statistics. Uniform Data Standards for Patient Medical Record Information. Report to Secretary of U.S. Department of Health and Human Services. Health Insurance Portability and Accountability Act (HIPAA) of 1996. Washington, DC: DHHS, 2000.
19. Pavlin J. Personal communication. Department of Defense Global Emerging Infections Surveillance and Response System. Nov 18, 2001.
20. Clinical Classifications Software (ICD-9-CM) Summary and Download Information. Agency for Health Care Policy and Research, Rockville, MD. Available from: <http://www.ahrq.gov/data/hcup/ccs.htm>. Accessed Nov 15, 2001.
21. National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine. *J Public Health Manag Pract.* 2001;7(6):43-50.
22. Centers for Disease Control and Prevention. Supporting Public Health Surveillance through the National Electronic Disease Surveillance System (NEDSS). Oct 2001. CDC Web site. Available from: <http://www.cdc.gov/od/hissb/docs.htm>
23. Kimball AM. APEC-Emerging Infections Network. Oct 2001. Asia-Pacific Economic Cooperation Web site. Available from: <http://www.apec.org/infectious>. Accessed Nov 15, 2001.