# Chapter 14
# Informatics in Disease Prevention and Epidemiology

**Richard S. Hopkins and J.A. Magnuson**

**Abstract** This chapter provides a description of the components of disease prevention and control programs, and then focuses on information systems designed to support public health surveillance, epidemiologic investigation of cases and outbreaks, and case management. For each such system, we describe sources used to acquire necessary data for use by public health agencies, and the technology used to clean, manage, organize, and display the information. We discuss challenges and successes in sharing information among these various systems, and opportunities presented by emerging technologies.

Systems to support public health surveillance may support traditional passive case-reporting, as enhanced by electronic laboratory reporting and (emerging) direct reporting from electronic health records, and also a wide variety of different surveillance systems. We address syndromic surveillance and other novel approaches including registries for reporting and follow-up of cases of cancer, birth defects, lead poisoning, hepatitis B, etc., and population-based surveys (such as BRFSS or PRAMS).

Systems to support epidemiologic investigation of outbreaks and clusters include generic tools such as Excel, SAS, SPSS, and R, and specialized tool-kits for epidemiologic analysis such as Epi-Info. In addition to supporting outbreak investigation,

R.S. Hopkins, MD, MSPH
Department of Epidemiology,
University of Florida College of Public Health
and Health Professions and College of Medicine,
Gainesville, FL, USA
e-mail: hopkinsrs@comcast.net

J.A. Magnuson, PhD, MD, MSPH (✉)
Department of Medical Informatics and Clinical Epidemiology,
Oregon Health & Science University,
Biomedical Information Communication Center (BICC), 5th Floor,
3181 S.W. Sam Jackson Park Rd., Portland, OR 97239, USA
e-mail: jamagnuson@gmail.com

J.A. Magnuson, P.C. Fu, Jr. (eds.), *Public Health Informatics and Information Systems*,
Health Informatics, DOI 10.1007/978-1-4471-4237-9_14, 

agencies also need systems to collect and manage summary information about outbreaks, investigations, and responses.

Systems to support case management, contact tracing, and case-based disease control interventions are often integrated to some degree with surveillance systems. We focus on opportunities and choices in the design and implementation of these systems.

**Keywords** Case reports • Shared services • Unified systems • Positive predictive value • Syndrome • Incidence • Outbreak • Cluster • Reportable • Notifiable • Registry • Surveillance system

**Learning Objectives**

1. Describe the range of information systems in current use to support public health surveillance, epidemiologic investigations, and disease prevention.
2. Identify opportunities for more effective epidemiology and disease prevention through implementation of emerging technologies.
3. Describe the challenges and opportunities presented by integration of information systems for epidemiology and disease prevention.

## Overview

This chapter provides a description of the components of disease prevention and control programs, and then focuses on information systems designed to support public health surveillance, epidemiologic investigation of cases and outbreaks, and case management. For each such system, we describe sources used to acquire necessary data for use by public health agencies, and the technology used to clean, manage, organize, and display the information. We discuss challenges and successes in sharing information among these various systems, and opportunities presented by emerging technologies.

Systems to support public health surveillance may support traditional passive case-reporting, as enhanced by electronic laboratory reporting and (emerging) direct reporting from electronic health records, and also a wide variety of different surveillance systems. We address syndromic surveillance and other novel approaches including registries for reporting and follow-up of cases of cancer, birth defects, lead poisoning, hepatitis B, etc., and population-based surveys (such as BRFSS or PRAMS).

Systems to support epidemiologic investigation of outbreaks and clusters include generic tools such as Excel, SAS, SPSS, and R, and specialized tool-kits for epidemiologic analysis such as Epi-Info. In addition to supporting outbreak investigation, agencies also need systems to collect and manage summary information about outbreaks, investigations, and responses.

Systems to support case management, contact tracing, and case-based disease control interventions are often integrated to some degree with surveillance systems. We focus on opportunities and choices in the design and implementation of these systems.

## The Main Components of a Disease Prevention Program

Public health programs to prevent disease typically have been designed and implemented one disease at a time. Each disease has its own patterns of distribution in populations, risk factors, and optimal and practical intervention strategies that are effective in controlling, preventing, or even eliminating cases of the disease. For example, an important strategy to prevent measles is vaccination, the main strategy to prevent gonorrhea is antibiotic treatment of case contacts before they become ill themselves, an important strategy to prevent cervical cancer is screening with Pap smears and treatment of preclinical disease, and the main strategy for prevention of neural tube defects is folic acid supplementation of selected foods. Still, each disease prevention program's components are drawn from a relatively short list:

- Planning and evaluation
- Public health surveillance
- Outbreak or cluster recognition and response
- Policy and guidance development
- Clinical services
  - Screening
  - Immunization
  - Prophylaxis
  - Treatment
- Laboratory services
- Case-contact identification and interventions
- Education and training for clinicians
- Public education
- Regulation (for example, of food services, drinking water, child-care centers, hospitals, etc.)
- Administration and financial management

Ideally, program managers choose the most effective combination of these program components to prevent or control the disease or diseases they are charged with addressing. However, as this must be done within the constraints imposed by the available funds, cost-effectiveness is the usual criterion for choosing the preferred combination of program components.

Public health agencies typically are organized both by disease and by function. For example, each disease-specific program usually does not have its own

laboratory, and a single public health clinical facility and its staff may provide varied services such as immunizations for well children, treatment of people with tuberculosis (TB) and their contacts, and Pap smear services. To variable degrees, they may even combine activities in a single patient encounter, for example, testing women for gonorrhea and *Chlamydia trachomatis* infections at the same visit where they get a Pap smear, or offering hepatitis B vaccination during a visit for sexually transmitted diseases (STD) treatment.

As information technology has become more widely used in public health and replaced paper-based systems, it has typically been implemented program area by program area, as resources became available. This has led to the creation of information 'silos.' For example, laboratory information systems usually have developed in isolation from those to support clinical care or public health surveillance.

Information systems to support clinical operations of public health departments (for example, clinical services for STDs, childhood immunizations, HIV/AIDS, TB, or family planning services) have characteristics similar to those of other electronic health record systems in ambulatory care. However, in some health departments, clinical information systems have been separated by disease or clinic.

If one were to design information systems from scratch for a set of disease prevention programs, there would be potential savings and efficiencies from identifying the ways that one program component depends on information from another, or can serve multiple programs, and then designing the system to provide that information seamlessly. One can identify potential efficiencies from two perspectives:

1. *Shared Services*: Information systems can provide the same services for multiple disease programs. For example, electronic reporting of selected laboratory results for surveillance purposes can be implemented only once for any given public health agency, and the same reporting system can receive reportable results related to numerous infectious diseases and acute poisonings, screening tests like Pap smears, and abnormal pathology reports for cancer surveillance.
2. *Unified Systems*: Information systems supporting different program components can be unified, often using a master person index. For example, this would allow clinicians treating people with TB to have ready access to any HIV testing results on their patients, and allow HIV/AIDS clinicians similar access to information about results of tests indicating TB infection.

In reality, it is rare to have an opportunity to design such extensive information systems as a single project. One is dealing with numerous legacy systems that were designed to support program-specific workflows. So a key challenge for the public health informaticist is to help their agency make decisions about where information system 'integration' will yield substantial benefits and where it will not.

For example, if it is desired to know (one time) how many people in the jurisdiction have been reported during a particular time interval with both syphilis and hepatitis B, one could do an *ad hoc* match of information in two independent surveillance information systems. This task might take an analyst a few days or weeks to accomplish – which is almost certainly inexpensive compared to the cost of building a new information system that could do this task almost immediately. For

many purposes, it may be useful and sufficient to be able to display multiple streams of surveillance or programmatic data in the same environment, on the same screen or even in the same chart. In Florida, de-identified reportable disease case information and death certificate information are imported into the ESSENCE analytic environment that was originally designed for syndromic surveillance [1], so that trends for similar conditions by age, sex, and geographic area in the two data streams can be easily compared. On the other hand, if it is desired to have real-time information available to the STD clinic staff about past diagnoses of hepatitis B, or about past receipt of hepatitis B vaccine, then information systems need to be designed to support this kind of look-up; the usual solution is a shared person index between the two systems. Alternatively, a common data repository can be designed in which all information about each person is permanently linked.

As mentioned earlier, there are a number of components common to disease control and prevention programs. In this chapter, we will address information systems designed to support the following:

- Public health surveillance
- Outbreak or cluster recognition and response
- Acquisition of laboratory information
- Case-contact identification and intervention

## *Public Health Surveillance*

CDC defines public health surveillance as "the ongoing, systematic collection, analysis, and interpretation of health data, essential to the planning, implementation, and evaluation of public health practice, closely integrated with the dissemination of these data to those who need to know and linked to prevention and control" [2]. Each word of this definition is carefully chosen, and has implications for the design of surveillance information systems. A one-time data collection activity is not surveillance. Data collection for research purposes is not surveillance. Surveillance data are collected to support public health action, and analyses and recommendations based on these data must be shared with those who provided the data and with others who need to know.

Objectives of surveillance systems differ at the local, state, and federal levels [3]. At the local level, immediate response to individual cases is relatively more important, while at the federal level the analysis of larger-scale patterns is the most important function of surveillance. For state health departments, both uses of surveillance data may be important, depending on the disease and the size of the state.

Public health surveillance systems may be based on data capture from a variety of sources, including case reports, population-based surveys, sentinel providers, electronic health records (including laboratory information management systems for ELR and emergency department records for syndromic surveillance), or administrative data (like hospital or physician claims for reimbursement). For some noninfectious diseases, surveillance is carried out through registries (see below).

Information systems to support reportable disease surveillance contain records representing *case reports* that currently are, for the most part, entered manually into an application by public health staff, based on information received from doctors, infection control practitioners, hospitals, and laboratories. Increasingly, the laboratory information in these records comes from electronic records transmitted by the public health laboratory, hospital laboratories, and commercial laboratories, when there is a positive result meeting certain reporting criteria (like a positive IgM antibody test for hepatitis A). These records typically contain a combination of clinical, laboratory, and epidemiologic information about each case.

In future, increasing proportions of these case reports will be entered directly into a website by the practitioner creating the case report, or be transmitted electronically from the practitioner's electronic health record (EHR) system. Currently almost half the states in the US use the CDC-provided NEDSS Base System (NBS) as their platform for managing case reports. The remainder use either a system developed in-house or one of several commercially-available solutions [4].

In case-based surveillance practice, there is usually a relatively short list of required elements in the initial case report. For some diseases this is the only information received on all cases. For other diseases, usually of more importance and with lower case numbers, an additional data collection form is initiated by the receiving health department, which gathers information as appropriate from the ill person, the treating physician, and health records. The optimum amount of information to collect in the initial case report, as opposed to the disease-specific case report form, is a matter of judgment and may change as technology changes. In a largely manual system, health departments typically desire to minimize barriers to reporting of cases, so the incentive is to keep the initial case report form short. If much of the information desired for the disease-specific case report form can in fact be extracted from an electronic medical record with no additional effort by the person making an electronic case report, then the balance changes. Careful decisions are needed: for which cases of which diseases are follow-up interviews necessary [5]?

Until very recently, virtually all of the case-based surveillance information used at the federal level was collected initially at the local (or sometimes state) level, where it was used in the first instance for local response. As the case report information passes from the local to the state to the federal level, it is subjected to validation and cleaning: cases not meeting the surveillance case definition have been removed from the data submitted to the federal level, missing data have been filled in to the extent possible, and cases have been classified as to whether they are confirmed, probable, or suspected using standard national surveillance case definitions (these case definitions are developed by the Council of State and Territorial Epidemiologists in consultation with CDC) [6].

More recently, advances in technology have allowed case reports, and the information on which they are based, to move almost instantaneously from electronic health record systems, maintained by doctors, hospitals, and laboratories, to public health authorities. There are no technical barriers to these data being available at the federal level essentially as early as they are at the local and state levels. This ready availability of unfiltered clinical information may allow more rapid awareness by

public health officials at all levels of individual cases of high-priority diseases (like botulism or hemorrhagic fevers like Ebola virus infection), and thus lead to more rapid detection and characterization of likely outbreaks.

The simultaneous availability of raw data to multiple agencies at different levels of government also presents certain challenges. The user at the local level will have ready access to information from many sources about local conditions and events, and can use this information to interpret local observations. They will be in a position to understand when an apparent anomaly in their surveillance data is due to an artifact or to local conditions that are not a cause for alarm. They will also know whether a problem is already under investigation. A user at a state or federal level will be able to see patterns over a larger area, and thus may be able to identify multi-jurisdictional outbreaks, patterns, or trends that are not evident at a local level.

The fact that several users may be examining the same raw data at the same time requires that these multiple users be in frequent communication about what they are seeing in their data and which apparent anomalies are already explained or need further investigation. There is a danger that users at a higher level may prematurely disseminate or act on information that, while based on facts, is incomplete or misleading. Similarly, users at a local level may not realize that what they are seeing is part of a larger phenomenon. In the syndromic surveillance domain, the BioSense 2.0 Governance Group [7] has adopted a set of etiquette principles which participating jurisdictions will be required to agree to, that spell out the mutual obligations of analysts at each level of the system (Scott Gordon, Association of State and Territorial Health Officials, 2013, personal communication).

From an information management perspective, an important question is where to put human review of case reports in this information flow. For example, it is becoming technically possible for likely cases of reportable diseases to be recognized automatically in health care electronic record systems. Some of these could be passed on to public health authorities without human review, in the same way that reportable laboratory results are already passed on in Electronic Laboratory Reporting (ELR). For which constellations of findings in the electronic health record would this be appropriate? Should some electronic case reports generated by electronic health record systems be passed to state or even federal public health officials before they are reviewed and validated at the local or state levels? If so, which ones? As always, there is a tension between the speed of information flow and its quality and completeness. There is a need for research to determine which constellations of findings in electronic health records have adequate specificity and sensitivity to warrant automated identification of a person as being likely to have a case of a reportable disease. The acceptable sensitivity and specificity will vary by disease.

In 2001, CDC published the Updated Guidelines for Evaluating Public Health Surveillance Systems [8]. This document identifies a set of key attributes of surveillance systems to be assessed during a surveillance system evaluation, including simplicity, flexibility, data quality, acceptability, sensitivity, predictive value positive, representativeness, timeliness, and stability. These are also useful attributes to consider when designing a surveillance information system [9]. The relative importance of these attributes will vary depending on the condition under

surveillance and the main purposes for surveillance. For example, a surveillance system to detect cases of botulism for immediate public health response puts a high premium on timeliness, and its operators are likely to be willing to accept a modest number of false-positive reports (a lower *positive predictive value*) in order to assure that reports are received very quickly. On the other hand, surveillance to support planning of cancer prevention programs and treatment services is less time-sensitive, given the quite long incubation periods for most cancers, and therefore more concerned with diagnostic accuracy of every case report than with speed of reporting. Timeliness, positive predictive value, and sensitivity of a public health surveillance system are always in tension with each other; increasing two of these always compromises the third.

In systems based on case-reporting from doctors, hospitals, and laboratories, and receipt of electronic health records from these same organizations, records for an individual can in principle be linked with records for that same individual in numerous public health information systems, including those supporting clinical service, immunization registries, case investigation, partner or contact identification, partner or contact notification, and provision of interventions to partners or contacts. Sometimes this will be done best by automated messaging of structured data from one system to another, sometimes by supporting real-time look-up capabilities, and sometimes by development of a master person index to underlie some or all of these applications. One key decision is which application to consider as the hub for this information sharing, for example, the surveillance application itself or a clinical application.

Surveillance systems that are based on sample surveys (such as the Behavioral Risk Factor Surveillance System, BRFSS [10]), on sentinel practices (such as ILI-Net for surveillance of influenza-like illness [11]) or on syndromic surveillance do not have individual patient identifiers, and so intrinsically cannot be linked at the individual level to information systems supporting other disease control program components. Their data are typically managed in systems built on standard statistical software packages, or other independent systems.

Syndromic surveillance systems are based on rapid acquisition of unfiltered, real-time, electronic records without individual identifiers from hospital emergency rooms [12] and urgent care centers, and also, increasingly, from outpatient physicians' offices and from hospital admissions [13]. The primary purpose of these systems is to support detection and characterization of community disease outbreaks, as they are reflected in care received at emergency departments, physicians' offices, or hospitals. Each visit to an emergency department is assigned to a category or *syndrome*, based on words and strings contained in the patient's chief complaint and/or the triage nurse's notes. As the records received by the health department do not have individual identifiers, they cannot be linked to records in other information systems. However, records received by the syndromic surveillance system should contain unique identifiers that could allow the epidemiologist analyzing the data to work back through the sending facility to an identified clinical record. This traceback might become necessary if the person appeared to have a case of a reportable disease or to be part of a significant outbreak. Adding outpatient visits and hospital admissions to the scope of syndromic surveillance is opening up additional

uses for this technology, especially in the areas of real-time non-infectious disease surveillance.

Surveillance for cancers [14], stroke [15], birth defects [16], and some other chronic diseases like amyotrophic lateral sclerosis (ALS) is carried out through registries. *Registries* are usually established by specific legislation, and typically relate to a single topic – for example a registry of records for a disease, or of immunization records. Registries may be restricted to a geographic region.

A distinctive feature of registries is that individual case reports are kept open for long periods of time, up to several or many years, allowing additional information about treatment, hospitalization, and death or other outcomes to be added. Registries thus serve as systems to monitor type, duration, and outcome of treatment for these diseases, in addition to the occurrence of new cases of disease (disease *incidence*). They may also support outreach efforts to patients or their families, as a way to document that appropriate steps have been taken to link patients to needed types and sources of care.

Most cases recorded in state-level cancer registries are acquired from hospital-level registries, using an electronic case report in a standardized format [17]. Some case abstracts are obtained directly by registry personnel or contractors, when hospitals do not have suitable registries of their own. Case reports require extensive review and abstraction of medical records by trained workers. Birth defect registries may also be built by active search for cases in hospital and other medical records, and abstraction of those records to make case reports. They also may be built by electronically linking records from vital statistics (birth and death records), centralized hospital discharge record systems, and clinical service providers for children with birth defects (such as state programs for children with special medical needs) [18]. The latter are much less expensive to develop but cannot be assumed to have captured all cases of the disease under surveillance, or captured them correctly [19].

## *Disease Outbreaks and Clusters*

A disease *outbreak* is defined as a number of cases greater than the number expected during a particular time interval in a geographic area or population. This term usually is used for events due to infectious diseases, and sometimes for those of toxic origin. A similar increase above expected numbers for a non-infectious disease, such as birth defects or cancer, is usually called a *cluster*. Outbreaks and clusters may be due to diseases for which individual cases are reportable (like shigellosis or breast cancer), or diseases for which they are not (like food poisoning due to staphylococcal or *Clostridium perfringens* toxins in most states, SARS when it was new, or multiple sclerosis).

Surveillance systems are designed to facilitate recognition of outbreaks or clusters by frequent examination of the most current information available. The design of the user interface is particularly important. The interface should allow users to: flexibly display line lists, bar charts by date of event (epidemic curves), and maps of location

of cases; flexibly select subsets of cases for display; apply appropriate statistical tests to detect improbable increases in case counts; and display multiple streams of data on the same chart. For example, users may want to display the epidemic curve of an influenza outbreak for several different regions of a state or for several different age groups, or to display counts of positive influenza tests and emergency department visits for influenza-like illness on the same graph with different scales for each.

Syndromic surveillance systems have been leaders in developing and evaluating statistical algorithms for automated detection of anomalies which may, on investigation, turn out to be outbreaks. Such algorithms have less frequently been applied for automated detection of possible outbreaks or clusters in reportable disease data streams.

Most outbreaks and clusters are in fact not recognized by examination of regularly-collected surveillance system data. Instead, they are recognized by private citizens (such as the organizer of a social event, a teacher or school nurse, the manager of a child care center, the manager of a food service facility, an employer, or the ill people themselves) or by practicing doctors, and brought to public health attention via a phone call or e-mail or entry on a web site established for the purpose [20]. Public health workers assess the information and make the decision whether or not to do a formal investigation of the outbreak. One part of such an assessment is to look at available streams of surveillance data and determine whether there is information supporting the occurrence of an outbreak. For example, a report of a possible influenza outbreak in a high school might prompt closer examination of syndromic surveillance data from nearby hospital emergency departments to determine whether there is a more general increase in visits for influenza-like illness. A report of a neighborhood cluster of brain cancers would prompt closer examination of available cancer registry information, which might or might not support an interim conclusion that such a cluster is real and statistically significant.

In order to be accountable for the effectiveness of their work, local and state health departments need to track the occurrence of outbreaks and the public health response to those outbreaks. Since outbreaks can be due to reportable or non-reportable diseases, this cannot be done only by actions such as identifying some cases in the reportable disease data system as being part of an outbreak. Systems to track the occurrence of outbreaks need to document the following:

- time and date the first and last cases occurred
- total (estimated or counted) number of cases
- population group most affected (by age, sex, location)
- setting of the outbreak (school, workplace, restaurant, wedding, etc.)
- suspected or confirmed agent
- most common clinical presentation
- suspected or confirmed source and mode of spread
- methods used to investigate agent, source and mode of spread
- control measures recommended
- control measures implemented
- lessons learned for prevention of future outbreaks and improved investigation and response in future events

This information about outbreaks should be stored for ready retrieval, and to serve as a basis for quality improvement efforts. For quality improvement purposes, it is also helpful to document the content of the summary report written about each outbreak. When the outbreak is due to a reportable disease, individual cases in the reportable disease surveillance information system can be linked to the outbreak, for example by having an outbreak identifier attached to their records.

If preliminary information about outbreaks in a jurisdiction is entered into the outbreak information system in real time, as the investigation is proceeding, and if the outbreak database is readily searchable by all communicable disease investigators in the jurisdiction, then local investigators can use the outbreak database to help them with investigations of new illness or outbreak complaints [21]. For example, if they receive a complaint that illness has occurred in people who consumed a particular food product, they can look in the database and determine whether other recent or current complaints or outbreaks mention the same food product. If they receive a report about a gastroenteritis outbreak in a childcare center, they can determine what agents have been found to be responsible for recent or current similar outbreaks in nearby communities; this can help focus their laboratory testing and initial control strategies.

Some US states have had long-standing systems to document all outbreaks investigated by local or state personnel, but others have not. A major variable in the design of such systems is the state-local division of responsibilities in each state, including the degree of state oversight of 'routine' local outbreak investigations.

The actual investigation of an outbreak or cluster may involve enhanced "active" case-finding, use of case-report forms, group surveys, and formal epidemiologic studies. Active case-finding involves regular solicitation of case reports from doctors, hospitals, and laboratories. Managing the reports of possible, probable, and confirmed cases that are part of the outbreak is an important task. For a reportable disease, the jurisdiction's reportable disease surveillance system may be adequate to manage reported cases. It may be necessary, however, to create a continuously-updated line list of possible cases and their current status, which is outside the scope of the standard reportable disease application.

Outbreak investigation surveys will typically involve interviewing everyone with a possible exposure (like all attendees of a wedding reception), whether they were ill or not. Formal studies may involve interviewing selected non-ill people, for example, as part of a case–control study. The investigation may also involve obtaining and sending to a laboratory a large number of specimens from ill persons, and sometimes from exposed non-ill persons and from environmental sources (food, water, air, soil, etc.). Managing these disparate types of information is a challenge, especially in a large outbreak or one involving multiple jurisdictions. There is currently no one widely-accepted and satisfactory way to manage data in such settings. Each investigation team typically uses the tools it is most familiar with, including some combination of data management tools like MS Excel, MS Access, or EpiInfo [22], and standard statistical packages. Many health departments maintain libraries of standard questionnaires with associated empty data bases, for use during outbreak investigations.

When CDC is involved in a multistate outbreak, the investigation team at the local or state level needs to be able to produce and transmit timely case report and other information in the format desired by CDC. The services of an experienced public health informaticist can be extremely helpful to the investigation team when outbreaks are large and multifocal. An ongoing challenge for CDC and the states is how to make the transition from specialized case reporting during an outbreak of a new disease, such as West Nile Virus encephalitis or SARS, to routine case-based surveillance. If this transition is not well-managed, it is likely to result in the creation of a permanent stand-alone surveillance information system (or silo) for that disease. If the new disease is of national importance, cases should be made nationally notifiable and its surveillance should be incorporated into existing systems.

## *Laboratory Information*

Laboratory information is a critical component of disease surveillance and prevention. Laboratory data form the foundation of many surveillance systems. There are different types of laboratories involved in the public health data stream. Laboratories providing data to public health fall into the general categories of commercial or private industry, hospital or clinical, and public health laboratories.

Public health laboratory information systems (LIS) contain information about test results on specimens submitted for primary diagnosis, for confirmation of a commercial or hospital laboratory's results, for identification of unusual organisms, or for further characterization of organisms into subgroupings (like serotypes) that are of epidemiologic importance. In some states, all clinical laboratories must submit all isolates of certain organisms to the public health laboratory. Many of the results obtained in a public health laboratory turn out to be for diseases that are not reportable and not targets of specific prevention programs. Some of those results may, however, be for cases of non-reportable diseases that are historically rare in the jurisdiction but of great public health importance, or are new or newly-recognized.

The main business of clinical laboratories (located both inside and outside hospitals) is to test specimens for pathogens or groups of pathogens specified by the ordering physician, and return the results to the person who ordered the test. Public health agencies have, since the early 1990s, asked or required such laboratories to also identify results meeting certain criteria (indicating the presence of a case of a reportable disease) and send a copy of the results to the public health agency for public health surveillance. Initially, case reporting by laboratories was accomplished on paper forms, which were mailed or faxed to public health departments. Some laboratories very soon moved to mailing printouts of relevant laboratory results, then to sending diskettes, then to transferring computerized files containing laboratory results by direct modem-to-modem transfer, and eventually to transferring such files via the Internet using standard formats and vocabularies. In some states, public clinics (for example, STD clinics) have used contract laboratories for their testing needs. In this situation, the outside laboratory supplies both positive

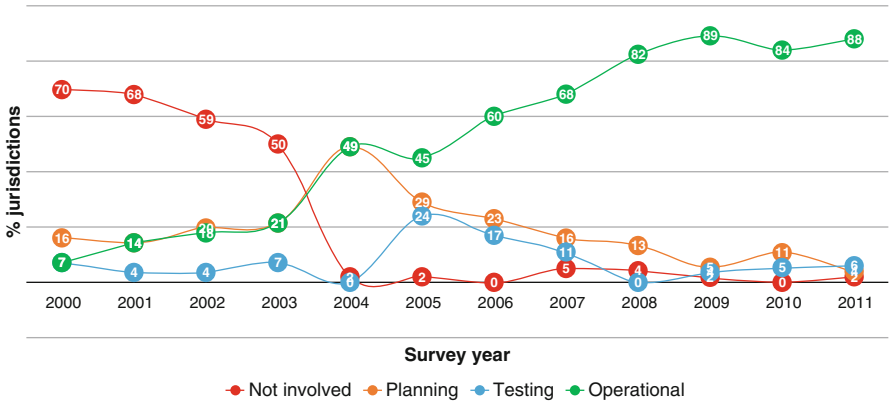


**Fig. 14.1** Trends in ELR adoption; stages of development by percentage of US jurisdictions

and negative results to the public health agency, increasingly by transfer of electronic results in standard formats.

Laboratories provide data on *reportable* conditions to their local or state public health authority. Reportable diseases are determined by each state; clinicians, hospitals, and/or laboratories must report to public health when these conditions are identified. Some reportable conditions are also nationally notifiable. Deidentified cases of these are voluntarily notified by states and territories to CDC, which, in collaboration with the Council of State and Territorial Epidemiologists, maintains a listing of nationally notifiable conditions that includes both infectious (e.g., rabies, TB) and non-infectious (e.g., blood lead, cancer) conditions [23].

The public health partnership with laboratories has led to the very successful and still increasing implementation of electronic laboratory reporting (ELR) in the US. ELR refers to the secure, electronic, standards-based reporting of laboratory data to public health. ELR implementation has been steadily escalating since its inception around the year 2000, replacing previous reporting systems that relied on slower, more labor-intensive paper reporting. The ELR National Working Group conducted annual surveys from 2004 to 2011 [24] which gathered data from all 50 states as well as from several territories and large metropolitan areas. These data were supplemented with data for years 2000–2004, retroactively gathered in the 2010 survey. The tracked growth of ELR (Fig. 14.1) illustrates its rapid rise in the US, from the start of early stage planning to fully operational ELR [25].

The expected benefits of ELR include more rapid reporting of reportable cases to public health departments, allowing faster recognition of priority cases and outbreaks for investigation and response, and thus more effective prevention and control [26]. ELR also is expected to reduce the number of missed cases, as automated systems do not require laboratory staff to actively remember to make case reports, and to improve the item-level completeness and quality of case reports. Although experience shows that the expected improvements in timeliness, sensitivity, completeness, and accuracy are generally being realized [27], timeliness may not be

improved substantially for those diseases where clinicians routinely report based on clinical suspicion without waiting for laboratory confirmation (for example, meningococcal disease) [28]. In addition, laboratories (especially referral laboratories) often do not have access in their own information systems to home addresses for people whose specimens they are testing, and have struggled with providing complete demographic information to public health agencies.

Implementation of an operational ELR system is not a trivial undertaking. Laboratories must configure data into an acceptable message format, most commonly Health Level Seven (HL7®) [29]. Laboratory tests and results should be reported with correlated vocabulary or content codes. Two of the most common code systems used for laboratory tests and their associated results are Logical Observations Identifiers Names and Codes (LOINC®) [30] and Systematized Nomenclature of Medicine (SNOMED CT®) [31]. Neither of these systems is sufficient by itself to encode all the information needed for public health surveillance.

Public health jurisdictions have introduced ELR to their partner laboratories using one or more of the following approaches:

- The "charm" approach – relies on establishing goodwill and collaboration with laboratory partners. While this collegial approach is very appealing, it may be unable to overcome significant barriers such as lack of laboratory funding or resources, and some facilities will supply data only in methods specifically required by law.
- The incentive approach – involves offering either financial or technical assistance to laboratory partners, assisting them in the startup process of ELR. While this approach may be preferred by many laboratories, relatively few jurisdictions have the discretionary funds (or are able to receive federal assistance funds) to implement the approach.
- The enforcement or legislative approach – requires reporting rules or legislation that requires laboratories to participate in ELR. The most successful enforcement approach will include low-cost options for smaller laboratories, such as web data entry, so that they may benefit from an ELR –"lite" implementation [32].

The mainstreaming of ELR systems in the US has pioneered a clear path forward for public health to begin maximizing its presence in the domain of electronic data interchange.

## *Field Investigation Information Systems*

At a local level, case reports for communicable diseases prompt action. Although the specific action varies by disease, the general approach is the same. It starts with an interview of the ill person (or that person's parents or other surrogates) to determine who or what the person was in contact with in ways that facilitate transmission, both to determine a likely source of infection and to identify other people who may be at risk from exposure to this person.

Information systems to support contact tracing, partner notification, and post-exposure prophylaxis (for STDs or TB, for example) contain records about all elicited contacts (exposed persons) for each reported case of the disease in question. These records contain information about each contact, such as whether they were located, whether they received post-exposure prophylaxis, and the results of any additional partner-elicitation interviews or clinical testing that were completed.

Information systems to support surveillance for other reportable diseases also increasingly contain information about what disease-appropriate action was taken in response to each case; such actions may include identification of contacts, education of household members, vaccination or antibiotic prophylaxis of contacts, isolation of the case (including staying home from work or school), or quarantine of exposed people.

STD and TB information systems typically capture full locating information for contacts, and can be used both to support field work and to generate statistics on effectiveness of partner notification activities worker by worker and in the aggregate. Systems for other reportable diseases may capture only the fact that various interventions were done, and the date that these were initiated. Information about the timeliness of initiation of recommended control measures is now required as a performance measure for selected diseases by CDC's Public Health Emergency Preparedness Cooperative Agreement [33].

In the investigation of a case of meningococcal disease, contacts are people who had very close contact with the original person, for example a household member, boyfriend, or regular playmate. Health department staff determines who the close contacts are. Each will then be offered specific antibiotic treatment to prevent illness. For syphilis, contacts are people who have had sex with the original case. Contacts will be examined by a clinician and assessed serologically to see if they are already infected, and offered appropriate prophylactic or curative antibiotic treatment. For measles, contacts may include anyone who spent even a few minutes in the same room as a case. Contacts whose exposure was recent enough, and who are not fully immunized already, will receive a dose of measles-containing vaccine, and all contacts will be asked to self-isolate immediately if they develop symptoms of measles. In investigating a common-source outbreak of legionellosis, histoplasmosis, or anthrax, the local health department may want to locate everyone who had a specified exposure to the apparent source of the infection. These exposed people may need antibiotic prophylaxis or may be advised to seek medical care promptly if they become ill.

Information systems to support this type of work typically have three purposes:

1. Serve as a place for workers to record and look up information about people who are or may be contacts, and to track which contacts have and have not yet received needed interventions.
2. Serve as a source of information for calculating indices of program or worker timeliness and performance, such as the average number of sexual contacts elicited per syphilis patient interviewed, or the percentage of measles contacts who were identified in a timely way and who received post-exposure measles vaccine prophylaxis.
3. Document the workload and effort put in by epidemiology and disease control field staff

It seems logical that the surveillance information system should serve as the basis for a system to support field investigation, and this is often the case. The fact that the recommended interventions vary by disease makes designing a single system more complex. Existing systems that track field worker activities in detail are much more common for STD and TB programs than for others. For general communicable disease fieldwork, it is currently more common that the system simply documents which interventions were done and when, rather than use the application to track specific named contacts or exposed people.

The Public Health Informatics Institute has published a detailed analysis [34] of the typical workflow involved in surveillance, investigation, and intervention for reportable diseases, and the corresponding information system requirements. The work group that PHII convened had representatives of nine different state and local health departments, who were able to identify a large number of processes that were common to all nine jurisdictions, such as case-finding, case investigation, data analysis and visualization, monitoring and reporting, case/contact specific intervention, and others. These common processes can then serve as a basis for designing information systems to support case-reporting, surveillance, and case-based intervention work that are useable in multiple jurisdictions.

## *Interoperability and Integration in Disease Control Information Systems*

Consider existing or planned surveillance systems for multiple diseases and conditions. Broadly, there are three functions in each of these systems – acquiring the raw data, cleaning and managing the data, and making the data available to users. Each of these functions potentially can be integrated, to varying degrees. For example, multiple surveillance systems may benefit from receiving electronic laboratory reports with a result indicating the presence of a case of a reportable disease. Laboratories appreciate having a single set of instructions and a single destination for all their required reports, as this simplifies their work. The laboratories then benefit from the ability of the recipient health department to route the reports internally to the right surveillance information system.

At the other end of the data pathway, users appreciate having a single interface with which to examine data about multiple conditions or diseases, using the same commands and definitions. The users do not have to understand how different surveillance information systems may internally code the same concept in different ways. They also appreciate being able to directly compare information that originally was submitted for the use of different program areas – for example, hepatitis B and gonorrhea in the same chart or table.

In the short to medium term, it is not necessary to build a single integrated data repository or a master person index to achieve these goals, even if that is what one would have designed if one were starting from the beginning. However, if one

wants to be able to see information about the same person that originates and is stored in multiple systems – for example, so that TB clinicians can see HIV data on their patients and vice versa – then an integrated data repository, or a master person index, or a query system that is extremely accurate in finding data on the right person, is needed. Modifying existing systems to be able to carry out these functions is time consuming and expensive, so the business case and requirements need to be especially clear.

**Review Questions**

1. What are some of the methods for surveillance besides case-reporting?
2. How are registries different from other surveillance information systems?
3. What are the advantages and disadvantages of building a master person index across surveillance information systems for multiple diseases?
4. What are the expected benefits of electronic laboratory reporting as a method to enhance surveillance?
5. What are the advantages and disadvantages of building a system to manage information about case contacts as part of the surveillance information system?
6. Who determines for which diseases cases are nationally notifiable?

## References

1. Kite-Powell A, Hamilton J, Wojcik R, Loschen W, Hopkins R. Florida's ESSENCE system: from syndromic surveillance to routine epidemiologic analysis across syndromic and non-syndromic data sources (Abstract). Emerg Health Threats J. 2011;4(Suppl):44–5.
2. Thacker SB, Berkelman RL. History of public health surveillance. In: Halperin W, Baker EL, editors. Public health surveillance. New York: Van Nostrand Reinhold; 1992.
3. Meriwether, RA. Blueprint for a National Public Health Surveillance System for the 21st Century. J Public Health Management and Practice: 1996;2(4):16–23.
4. CDC. Status of state electronic disease surveillance systems – United States, 2007. MMWR Morb Mortal Wkly Rep. 2009;58(29):804–7. Accessed at http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5829a3.htm on 31 Mar 2013.
5. Hopkins RS, Eisenstein L. Prioritizing investigations of reported cases of selected enteric infections. Paper presented at Council of State and Territorial Epidemiologists, Omaha, 6 June 2012.
6. CDC. Nationally notifiable disease surveillance system case definitions. Accessed at http://wwwn.cdc.gov/nndss/script/casedefDefault.aspx on 31 Mar 2013.
7. Association of State and Territorial Health Officers. Biosense 2.0 governance. Accessed at https://sites.google.com/site/biosenseredesign/governance on 31 Mar 2013.
8. CDC Guidelines Working Group. Updated guidelines for evaluating public health surveillance systems, recommendations from the Guidelines Working Group. MMWR Morb Mortal Wkly Rep. 2001;50(RR13):1–35.
9. Hopkins RS. Design and operation of local and state infectious disease surveillance systems. J Public Health Manag Pract. 2005;11(3):184–90.

10. Sosin DM, Hopkins RS. Surveillance. In: Guest C, Ricciardi W, Kawachi I, Lang I, editors. Oxford handbook of public health practice. 3rd ed. Oxford: Oxford University Press; 2013.
11. CDC. Behavioral risk factor surveillance system. http://www.cdc.gov/brfss/about/about_brfss.htm. Accessed 30 Mar 2013.
12. CDC. Overview of influenza surveillance in the United States. http://www.cdc.gov/flu/weekly/overview.htm. Accessed on 30 Mar 2013.
13. International Society for Disease Surveillance Meaningful Use Workgroup. Final recommendation: core processes and EHR requirements for public health syndromic surveillance. Boston, 2011. Accessed at http://www.syndromic.org/storage/ISDSRecommendation_FINAL.pdf on 30 Mar 2013.
14. International Society for Disease Surveillance Meaningful Use Workgroup. Electronic syndromic surveillance using hospital in patient and ambulatory clinical care electronic health record data. Boston, 2012. Accessed at http://www.syndromic.org/storage/ISDS_2012-MUse-Recommendations.pdf on 30 Mar 2012.
15. CDC. National Program of Cancer Registries. http://www.cdc.gov/cancer/npcr/. Accessed on 30 Mar 2013.
16. CDC. Paul Coverdell National Acute Stroke Registry Surveillance – four states, 2005—2007. MMWR Morb Mortal Wkly Rep. 2009;58(SS07):1–23.
17. CDC. Metropolitan Atlanta Congenital Defects Program (MACDP). Accessed at http://www.cdc.gov/ncbddd/birthdefects/macdp.html on 30 Mar 2013.
18. North American Association of Central Cancer Registries, Inc. (NAACCR). Implementation guidelines and recommendations. 2012. Accessed at http://www.naaccr.org/LinkClick.aspx?fileticket=WlmSkHUgKrI%3D&tabid=126&mid=466 on 30 Mar 2013.
19. Florida Department of Health. Report on birth defects in Florida 1998–2007. Accessed at http://www.fbdr.org/pdf/FBDR_report_May2011.pdf on 30 Mar 2013.
20. Salemi JL, Tanner JP, Kennedy S, Block S, Bailey M, Correia JA, Watkins SM, Kirby RS. A comparison of two surveillance strategies for selected birth defects in Florida. Public Health Rep. 2012;127:391–400.
21. Florida Department of Health. Online food and waterborne illness complaint form. Accessed at http://www.doh.state.fl.us/environment/medicine/foodsurveillance/Online_Foodborne_Complaint_Form.html on 13 Mar 2013.
22. CDC. National Biosurveillance plan for human health, version 2.0. Atlanta, 2010. p. 40–4. Accessed at http://www.cdc.gov/osels/pdf/NBSHH_v2.pdf on 30 Mar 2013.
23. CDC. Introducing EpiInfo 7. Accessed at http://wwwn.cdc.gov/epiinfo/ on 30 Mar 2013.
24. CDC. National Notifiable Diseases Surveillance System (NNDSS), CDC. Accessed at http://wwwn.cdc.gov/nndss/ on 30 Mar 2013.
25. National ELR Workgroup. ELR survey summary reports. Available from www.coast2coastinformatics.com. Cited 9 Mar 2013.
26. National ELR Workgroup. 2011 National Electronic Laboratory Reporting (ELR) Snapshot Survey. Available from www.coast2coastinformatics.com. Cited 9 Mar 2013.
27. 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.
28. Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. Am J Public Health. 2008;98(2):344–50.
29. CDC. Potential effects of electronic laboratory reporting on improving timeliness of infectious disease notification – Florida, 2002—2006. MMWR Morb Mortal Wkly Rep. 2008;57(49):1325–8.
30. Health Level Seven (HL7®) homepage. Available at http://www.hl7.org/. Cited 9 Mar 2013.
31. Logical Observation Identifiers Names and Codes (LOINC®). Available at http://www.regenstrief.org/medinformatics/loinc/. Cited 9 Mar 2013.
32. Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT®). Available at http://www.ihtsdo.org/snomed-ct/. Cited 9 Mar 2013.

33. See for example section 64D-3.031(5) of the Florida Administrative Code: notification by laboratories. https://www.flrules.org/.
34. CDC. Public health emergency preparedness cooperative agreement, budget period 1, performance measures specification and implementation guidance, at-a-glance summary. p. 22–4. Accessed at http://www.cdc.gov/phpr/documents/PHEP+BP1+PM+At-a-Glance_v1_1.pdf on 30 Mar 2013.
35. Public Health Informatics Institute. Redesigning public health surveillance in an eHealth world. Decatur, 2012. Accessed at http://www.phii.org/resources/view/1186/Redesigning Public Health Surveillance in an eHealth World on 31 Mar 2013.