Uses of Electronic Health Records for Public Health Surveillance to Advance Public Health

Guthrie S. Birkhead,[1,2] Michael Klompas,[3,4] and Nirav R. Shah[5]

[1]New York State Department of Health, Albany, New York 12237; email: guthrie.birkhead@health.ny.gov

[2]School of Public Health, University at Albany, Rensselaer, New York 12144

[3]Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts 02215; email: mklompas@partners.org

[4]Brigham and Women's Hospital, Boston, Massachusetts 02115

[5]Office of the Chief Operating Officer, Kaiser Foundation Hospitals, Kaiser Permanente, Pasadena, California 91188; email: nirav.r.shah@kp.org

Annu. Rev. Public Health 2015. 36:345–59

First published online as a Review in Advance on January 2, 2015

The *Annual Review of Public Health* is online at publhealth.annualreviews.org

This article's doi: 10.1146/annurev-publhealth-031914-122747

Keywords

public health surveillance, electronic health records, meaningful user

Abstract

Public health surveillance conducted by health departments in the United States has improved in completeness and timeliness owing to electronic laboratory reporting. However, the collection of detailed clinical information about reported cases, which is necessary to confirm the diagnosis, to understand transmission, or to determine disease-related risk factors, is still heavily dependent on manual processes. The increasing prevalence and functionality of electronic health record (EHR) systems in the United States present important opportunities to advance public health surveillance. EHR data have the potential to further increase the breadth, detail, timeliness, and completeness of public health surveillance and thereby provide better data to guide public health interventions. EHRs also provide a unique opportunity to expand the role and vision of current surveillance efforts and to help bridge the gap between public health practice and clinical medicine.

INTRODUCTION

Public health surveillance is a core function of public health practice (41). Surveillance for key diseases and health status indicators is integral to tracking public health interest and to triggering actions by public health agencies to prevent and control disease. State and local health departments in the United States have increasingly moved their surveillance systems toward requiring laboratories to electronically report test results indicative of communicable and some chronic diseases. This practice has resulted in improved completeness and timeliness of reporting (18, 39). However, the collection of detailed clinical information about these reported cases, which is necessary to confirm the diagnosis and provide context to understand disease severity, transmission risk factors, and comorbidities, is still heavily dependent on the surveillance tools of the past century: telephone and fax communication between public health staff and health care providers and manual review of paper medical records by surveillance staff.

The increasing adoption of electronic health records (EHRs) throughout the United States and their increasing functionality present revolutionary opportunities to advance public health surveillance. At the very least, use of EHRs for surveillance should increase further the timeliness and completeness of surveillance. More broadly, EHRs provide a unique opportunity to expand the role and vision of current surveillance efforts. EHRs are rich in a variety of data that can facilitate timely and efficient surveillance on the prevalence of, health care utilization for, treatment patterns for, and outcomes of a host of diseases, including obesity, diabetes, hypertension, and kidney disease. Leveraging these emerging opportunities could vastly improve public health departments' ability to monitor population health, guide public health initiatives, and measure the impact of interventions. In addition, EHRs have the potential to spur closer collaboration and better integration of clinical care and public health practice. Bringing these two worlds closer together, as recommended in a recent Institute of Medicine report, promises to improve and enrich both spheres, leading to better health outcomes for both individuals and populations (25). This article describes the current state of EHR-based surveillance, analyzes the barriers to more effective use of EHRs for surveillance, and lays out a research agenda for the next decade.

PUBLIC HEALTH SURVEILLANCE THEN AND NOW

Public health surveillance is the means by which public health agencies monitor the health status of populations (41). More than a half-century ago, weekly national morbidity statistics were published with the statement "no health department, State or local, can effectively prevent or control disease without knowledge of when, where, and under what conditions cases are occurring" (8, p. 531). However, surveillance is not just gathering health data and generating reports. Langmuir defined surveillance as "the continued watchfulness over the distribution and trends of incidence (of disease) through the systematic collection, consolidation and evaluation of morbidity and mortality reports and other relevant data" as well as the regular dissemination of data to "all who need to know" (31, pp. 182–83). The ultimate purpose of public health surveillance is to drive public health action.

Public health surveillance began with infectious disease surveillance in the late 1800s (41). State and local public health officers mandated infectious disease reporting in order to quarantine patients with smallpox, cholera, typhoid, and tuberculosis (TB). Even now, the legal authority for public health surveillance remains largely at the state and local levels in the United States. Surveillance evolved in the 1930s to include sexually transmitted disease reporting to ensure treatment of cases and contacts. As vaccines were developed, surveillance also evolved to include vaccine-preventable diseases. The Council of State and Territorial Epidemiologists (CSTE)'s

formulation of the list of nationally notifiable diseases beginning in the early 1950s (see **http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/CSTENotifiableConditionListA.pdf**) and development of surveillance case definitions in the 1990s (10) have standardized the way state and local health departments conduct surveillance and improved the comparability of disease rates across states. The CSTE case definitions are usually a combination of laboratory and clinical data that are required to confirm cases for surveillance purposes. Both clinical and laboratory results need to be collected for each reported case. Therefore, in most states, physicians, laboratories, and hospitals are mandated reporters. In the past two decades, public health surveillance has focused on requiring clinical laboratories to electronically report diseases that can be diagnosed via laboratory tests (see below).

Chronic disease surveillance programs were developed after infectious disease programs. The case reporting model was adopted for some conditions such as cancers and lead poisoning, often relying on laboratory results for initial reports and then gathering additional clinical data by having surveillance staff contact health care providers for clinical and risk factor information. Surveillance of very common chronic diseases such as cardiovascular disease, diabetes, and high blood pressure has relied more on population-wide administrative databases often collected for other purposes such as paying insurance claims or certifying deaths. There are insufficient resources to have public health surveillance staff collect clinical and risk factor information on individual cases of these chronic diseases; information about the epidemiology of these diseases therefore is gleaned from diagnosis codes contained in these databases. These tools allow only limited insight into the causative risk factors for chronic diseases. Population-based risk factor surveillance is therefore often conducted by random-digit-dialed telephone surveys. The best example is the Behavioral Risk Factor Surveillance System (BRFSS), which collects self-reported risk factors and diagnoses on a random sample of the population (11). The BRFSS provides useful estimates on the prevalence of tobacco use, hypertension, and diabetes. However, these data can lack timeliness, are limited by being self-reported, and do not allow analysis of disease patterns in small geographic areas.

The terrorist attacks of 2001 and the need to better prepare for these and other mass health events such as influenza pandemics led to the development of so-called "syndromic" surveillance systems (40). These are near real-time reporting systems usually based on hospital emergency department (ED) data. Because of the need for rapid reporting to detect a terrorist attack, for example, reports may be sent before a formal diagnosis has been made and coded. Therefore, syndromic surveillance systems are often based on patients' chief complaints as reported to triage nurses. Analysis entails doing a text search of the chief complaint field for terms of interest. Little else other than basic patient demographics is collected. Detailed clinical and laboratory information are not available in these systems currently. The data are crude but can be helpful in tracking broad population events such as influenza during the influenza season (14).

LIMITATIONS OF CURRENT SYSTEMS

Each of the surveillance systems noted above faces significant challenges in one or more of the parameters by which surveillance systems are evaluated (21). These challenges include the completeness and timeliness of reporting, the accuracy and specificity of the coding of the data, the availability of risk factor information to guide preventive interventions, and the cost of data collection. Reporting of notifiable diseases by physicians has always been viewed as incomplete (37, 39). Many health care providers lack adequate knowledge of reporting requirements and are encumbered by the additional workload required to file a report. Collecting detailed clinical information from health care providers is currently still paper- or telephone-based and is labor intensive. This

practice results in delays in obtaining critical information for diseases of immediate public health importance and to a lack of data on the most common causes of morbidity and mortality in the population other than self-reported survey information and limited death certificate coding.

Beginning in the mid-1990s, the timeliness and completeness of public health surveillance were significantly enhanced with implementation of electronic reporting of clinical laboratory test results for notifiable diseases (37, 39). However, laboratory data do not contain clinical information about patients. Gathering this information still requires significant surveillance and health care provider staff time, resulting in delays and gaps in the availability of surveillance data and, ultimately, delays in real-time public health action. In addition, laboratory reporting is not useful for surveillance of conditions that do not require a laboratory test result to confirm the diagnosis, for example early Lyme disease, toxic shock syndrome, hypertension, and asthma.

Limitations of surveillance systems relying on vital records and hospital discharge data include the small number of diagnostic codes that accompany each record, failure to use specific codes, the possibility of coding bias (e.g., to maximize reimbursement or avoid penalties), lack of risk factor information, and embargoes on the availability of certain data, sometimes on the order of years. Even when surveillance staff review medical records, key information about risk factors such as smoking, blood pressure, diet, exercise, and occupation may not be recorded. Population surveys are further limited by the self-reported nature of the information, by a lack of information for small geographic areas, and sometimes by substantial delays in data availability.

Most efforts to improve public health surveillance over the past two decades have focused on electronic laboratory surveillance. Clinical laboratories increasingly use electronic systems and standardized coding to record, store, and report laboratory results. There are relatively few laboratories, and they developed electronic systems earlier than did other health care settings. These factors have allowed health departments to focus their surveillance improvement efforts on electronic laboratory reporting over the past decade. In contrast, the health care system is a very large, broad, and diverse enterprise that has been slow to move to EHRs with universally agreed on standards. Health departments have not had the resources to engage with the health care system to collect surveillance data from EHRs (34).

EHRs: A NEW TOOL

The use of EHRs has blossomed in the past decade. Their increasing penetration into routine medical practice holds great promise to improve the quality of health care, reduce medical errors, and facilitate more coordinated care (38). The adoption and implementation of EHRs have been spurred by federal government support under the American Recovery and Reinvestment Act (ARRA), which included nearly $30 billion in incentives for hospitals and physician practices to adopt certified EHRs through the Health Information Technology for Economic and Clinical Health (HITECH) Act (3). In the years since ARRA was enacted, the number of hospitals and physicians using EHRs has grown. At the time the law was drafted, only 12% of acute care hospitals had a basic EHR system. By 2012, that number had grown to 44% (38). Among office-based physicians, the percentage of doctors using EHRs rose from 48% to 72% (24), despite the cost and effort required to install and maintain EHR systems (this is the primary barrier to their introduction and use).

Like paper files, EHRs record and store the patient's health data, possibly including, but not limited to, demographic information; problem list and active and past diagnoses; laboratory test orders and results; current prescriptions; radiological images and reports; hospitalization information; consultant reports; immunizations; pathology reports; social history; allergies; health screening study results; and physician, nurse, social worker, and physical therapy notes. In

addition to having this information at the health care provider's fingertips in a searchable format on any securely connected computer, EHRs can include additional functionality such as access to clinical and public health guidelines, reminders about routine screenings or disease reporting responsibilities, and graphical display of trends in key parameters such as blood glucose for diabetic patients or blood pressure measurements in hypertensive patients. Some EHR systems can generate practice-level statistics. Despite this promising array of possible features, there has been a lack of standardization, and many EHRs have been developed with many different designs, formats ("look-and-feel"), and functionality. For example, a recent survey of office-based physicians found that only 60.9% of EHRs could easily generate a list of patients by diagnosis, only 48.2% could easily track a patient referral to completion, and only 51.4% could easily generate a report on quality measures. It was difficult or impossible to generate such information in 11.8%, 20.5%, and 18.1% of the respondents' EHR systems, respectively (24). The meaningful use program, described below, has been an important effort to bring standardization and interoperability to EHRs.

One purpose of EHRs is to enable providers to share patient information so that care can be delivered seamlessly across different settings and separate encounters. This practice helps avoid duplicate tests, prevents drug-drug interactions, and enhances patient care. EHRs can also enable patients to access their records remotely and to use that information to better manage their health and health care. The Institute of Medicine has recently recommended that certain social and behavioral domains—commonly known as the social determinants of health—be incorporated in a patient's health records as well, given their critical impact on health. These include factors such as housing, social isolation, and food insecurity (26).

To ensure that EHRs reach their potential, networks are being developed to link EHRs so that health care providers can share information needed for care and patients can access their own records electronically. Such health information exchange systems, referred to in some states as regional health information organizations (RHIOs), also make it possible for public health workers to access EHRs to collect legally mandated disease reports.

Use of EHRs for Public Health Surveillance

The rapid spread of EHRs in clinical medicine presents an opportunity to take surveillance conducted by public health agencies forward, beyond the gains realized by the promotion of electronic clinical laboratory reporting systems in the 1990s. EHRs could support surveillance and make it more efficient in several ways. First, EHRs or RHIO networks could be designed to use standard computer algorithms to identify cases that meet surveillance case definitions and automatically report them to public health agencies. This practice could supplement traditional, notifiable disease reports triggered by laboratory results and could also be used to facilitate surveillance for common chronic diseases such as cardiovascular disease, asthma, and hypertension because these are defined not by laboratory testing but by clinical criteria often found in the EHR. The volume of such reports could be mitigated by developing automated systems to collect the data. For diseases first reported to the health department by an electronic laboratory test result, EHRs could be programmed to automatically send additional data on clinical parameters and risk factors to provide context for the reportable test. Alternatively, providing public health surveillance staff with EHR access could enable much more rapid collection of clinical data elements and risk factors. For patients seeing multiple physicians, information from multiple EHRs could be gathered and combined into a single surveillance report. RHIOs or linked EHRs could also be queried electronically to gather aggregate, deidentified information on conditions of public health interest, replacing current syndromic and survey-based surveillance systems. Finally, electronic

connections between health departments and health care providers could be bidirectional and be used to provide health care providers with information such as public health updates on active disease outbreaks, diagnosis and treatment recommendations, and/or guidance on how to manage conditions of public health importance.

Meaningful Use

Significant resources are being devoted to implementing EHRs in medical settings across the United States. Investments are also needed to develop the capability of EHRs to be used for public health surveillance and for health departments to be able to receive and process these surveillance reports. One avenue to support EHR use for public health surveillance has been the Centers for Medicare and Medicaid Services (CMS) Meaningful Use (MU) program (4). The MU program is part of the HITECH Act passed by Congress in 2009. MU provides incentive payments totaling nearly $30 billion over 10 years through Medicare and Medicaid for eligible health care providers and hospitals that adopt EHRs with certain capabilities. One of these capabilities is the capacity to exchange health data with public health departments (4, 12). The MU program is being rolled out in three stages. MU Stage I, initiated in 2010, contained three population health/public health objectives, which included the provider's ability to submit electronic data to the public health department immunization registry, a laboratory reporting system (hospitals only), or a syndromic surveillance system (12). The Stage I measure was the successful submission of one test message to one of these surveillance systems. MU Stage 2 regulations were published in 2012 and added two additional public health objectives: reporting to the state cancer registry and reporting to another specialized public health registry (12, 36). Health department participation is necessary for providers and hospitals to receive MU incentive payments by receiving and certifying their surveillance report data and allowing for ongoing data submission. Public health agencies now face the challenge and opportunity of finding ways to integrate and use these new reports within their workflows.

The MU program has suffered from periodic delays and the scaling back of its objectives. For example, notifiable disease data were to have been included in MU Stage 2, but this has now been dropped. Stage 2 was originally scheduled to be implemented by 2014 but is now delayed. MU Stage 3 was to have been implemented by 2015 but has also been delayed until at least 2017 (22). Nevertheless, the MU program has benefited public health surveillance efforts by promoting increased adoption and standardization of electronic data in the health care system. For example, the MU program is expanding the number of hospitals that are reporting laboratory and syndromic surveillance using current national electronic messaging standards. Recommendations for syndromic surveillance in the ambulatory care setting were issued by the International Society for Disease Surveillance (ISDS) in 2012, but messaging standards have not yet been developed (25).

Implementation of the MU program has been challenging for health departments because no dedicated funding has been provided as part of the program (34). As a result, health departments' ability to receive, process, and use electronic health data is still limited. A recent survey of states indicated a lack of resources to implement MU (2). In addition, if electronic reporting fulfills its promise of increasing the completeness of surveillance reporting, health departments risk being overwhelmed by additional surveillance reports. Coping with the increased flow of data from automated surveillance systems can require additional staff and investment in electronic systems necessary to effectively manage the data. Data standards for public health surveillance are also needed, as described further in this article, so that EHRs do not need to be customized to report to each public health jurisdiction (16, 30, 35).

The Affordable Care Act

The Affordable Care Act (ACA) for the first time provides dedicated funding to improve the nation's public health infrastructure in the form of the Prevention and Public Health Fund (PPHF) (14). Although the PPHF does not earmark funds specifically to help develop EHR reporting to public health surveillance systems, it does provide support for specific surveillance systems such as those for reporting health care–associated infections, which in the future may draw reports directly from EHR systems. In addition, the PPHF supports the training of the public health workforce in areas such as informatics, which are necessary skills that will help the Centers for Disease Control and Prevention (CDC) and state public health departments to develop public health surveillance systems that draw on EHR data. The ACA does not directly support the development of EHRs in health care settings.

THE ELECTRONIC MEDICAL RECORD SUPPORT FOR PUBLIC HEALTH (ESPnet) SURVEILLANCE PLATFORM

The literature on the use of EHRs for public health surveillance is limited. The largest body of work describes the experience of the Electronic Medical Record Support for Public Health (ESPnet) surveillance platform developed by the Harvard Center of Excellence in Public Health Informatics and the Massachusetts Department of Public Health with funding from the CDC (29, 32). The system is in use in four large ambulatory care practices in Massachusetts and Ohio, serving almost two million people in conjunction with their respective state health departments. ESPnet is configured as a separate server independent of the host EHR, allowing compatibility with different EHR systems and avoiding an additional computing burden on the host EHR system. The host EHR is programmed to send structured data from every patient encounter to ESPnet either in real time or nightly. These data include demographics, diagnosis codes, laboratory orders and results, prescriptions, vaccinations, and social history. ESPnet then applies algorithms to identify conditions of public health interest. Results are transmitted to the state health department either as individual case reports for notifiable diseases or as population-level aggregate summaries for chronic diseases and influenza-like illness (ILI). Case reports are formatted as HL 7 messages (see **http://www.hl7.org/**), which is the international standard for electronic transmission of health data. ESPnet source code and algorithms are available for download for free from esphealth.org (20).

ESPnet's algorithms utilize physician diagnostic coding, laboratory test orders and results, vital signs, and prescriptions to identify conditions of public health interest. For example, the combination of a diagnostic code for TB, laboratory tests for TB, and prescription of two or more anti-TB medications triggers an automatic surveillance report to the health department (see below). ESPnet currently has modules for notifiable diseases (including active TB, chlamydia, gonorrhea, syphilis, Lyme disease, pelvic inflammatory disease, pertussis, and acute hepatitis A, B, and C), chronic diseases (asthma, type 1 diabetes, type 2 diabetes, gestational diabetes, pre-diabetes, obesity, hyperlipidemia, and smoking), ILI, and vaccine adverse effects. The algorithm for each disease must be calibrated to maximize sensitivity and predictive value positive to be of greatest use to the health department without overloading them with false-positive reports. ESPnet relieves the providers of the responsibility to report to the health department and makes available to the health department detailed clinical information with the initial report, potentially streamlining surveillance investigations. Individual ESP installations can also be linked using a distributed network model to allow public health departments to query multiple providers at a time in order to obtain data on conditions that have not been specifically coded into ESPnet for

routine reporting. These results can be automatically aggregated by ESPnet to get a population-level picture of public health that transcends the population mix of any one medical practice (43).

Two of the biggest challenges ESPnet faces are the effort required for new installations and the need to update detection algorithms whenever new tests are introduced, new coding systems are implemented, and treatments change. Algorithm development and validation are labor-intensive tasks that require time, expertise, interest, and resources. Standardization of EHR systems and the development of standard surveillance algorithms by the CDC and groups like CSTE could lessen these challenges in the future. Other limitations include data gaps if patients receive care from multiple providers who are not all connected to ESPnet. Finally, health departments need the appropriate electronic infrastructure to receive reports from EHRs. Massachusetts has addressed this challenge by adapting its existing electronic laboratory report receiver (2, 34).

A second approach to public health surveillance with EHRs is to use continuity of care documents (CCDs) (17). CCDs are a standard clinical summary that the 700+ federally certified EHRs must be able to produce (see **http://oncchpl.force.com/ehrcert**). The CCD format is acceptable for data transfer to meet the MU requirements and uses the HL 7 architecture familiar to both EHR vendors and health department surveillance programs. CCDs are generated by the EHR and contain information on patients' demographics, diagnoses, laboratory results, and medications. With appropriate selection criteria, such as ESPnet's case-detection algorithms, CCDs could be generated for reportable diseases and transmitted to the health department. The advantage of this approach is that the CCD is standardized across all EHRs, so surveillance could be rapidly scaled. This standardization would also allow health departments to develop their systems to receive surveillance reports based on national standards. However, the development and maintenance of algorithms to select cases to report to the health department would be an issue, as it is with ESPnet. No citations using CCDs for surveillance could be found, but this approach has promise for the future because it is linked to the MU initiative and could be standardized across EHRs.

EHRs FOR COMMUNICABLE DISEASE SURVEILLANCE

Tuberculosis

TB is a disease of great public health importance that warrants a rapid public health response to each case. Surveillance needs to be highly sensitive to avoid missing a case of active TB. In achieving a high sensitivity, false-positive reports can be tolerated but should be minimized to avoid unnecessary effort on the part of surveillance staff. Thus the number of reported cases with active TB (positive predictive value, or PPV) should be maximized. Using the ESP system for reporting of TB cases illustrates the process, the challenges, and the promise of using EHRs for public health surveillance (6).

Algorithm development began with an analysis of EHR data from a 14-month period of time from a single large practice. Investigators created algorithms from various ICD-9 codes for TB, drugs for TB treatment, and TB laboratory tests to try to detect all active TB cases. The sensitivity and PPV for each candidate algorithm were assessed. The algorithm, which had 100% sensitivity and had the highest PPV, was a prescription for pyrazinamide (used to treat active TB) or an ICD-9 code for TB with a laboratory order for acid-fast bacilli (AFB) testing for TB diagnosis, or an ICD-9 code for TB diagnosis and a prescription for at least two anti-TB drugs other than pyrazinamide within 60 days. This algorithm detected 6 of 6 patients in the practice with active TB (sensitivity 100%, PPV 64%). In the next phase, this algorithm was validated against EHR

data from 1990 through May 2006. The algorithm again had high sensitivity and a PPV for physician-suspected active TB of 84%, but the PPV for confirmed active TB was only 47%. In the final prospective phase, the algorithm was implemented into ESPnet for live, active surveillance. Between August 2007 and January 2009, ESPnet reported 7 cases with sensitivity of 100% (6/6) and PPV for confirmed disease of 86% (6/7). Two cases were reported to the health department more rapidly than the standard provider (hospital) and laboratory reporting systems by 12 and 36 days, respectively.

The experience with EHR-based TB surveillance illustrates a number of important points. First, a significant amount of work needs to be done upfront to define the automated algorithms by which the computer search of EHR records will be done for the disease of interest. Second, it illustrates the advantages of basing surveillance on more than laboratory results. TB laboratory results may take days to weeks to be completed, and a significant portion of TB cases may be culture negative, at least initially, owing to disease occurring at extrapulmonary sites, the difficulty of culturing TB, and interference with culturing due to empiric treatment for TB. For this reason, health departments ask providers to report TB as soon as they suspect disease rather than waiting for culture confirmation, but providers may not be aware of this. This case also highlights the trade-off between sensitivity and PPV, which must be made when selecting the algorithm to trigger public health surveillance reporting. Use of EHRs for surveillance could result in so-called surveillance bias, which is defined as overreporting of suspected cases or a low PPV, resulting from reporting of cases that do not meet the case definition (15). The tolerance for overdiagnosis and surveillance bias needs to be calibrated for each disease.

Acute Viral Hepatitis

A second area of notifiable communicable disease surveillance where EHR-based surveillance has been developed is viral hepatitis (1). Hepatitis A and hepatitis B are two forms of viral hepatitis with very similar symptoms of liver inflammation. However, the modes of transmission, and thus the public health interventions in response to reported cases, differ. Hepatitis A is enterically transmitted, and the public health reaction is to provide immunoglobulin or vaccine-based postexposure prophylaxis to household contacts and to identify any persons who might have been exposed to food prepared by the index case. Information on employment as a food worker is an important piece of surveillance information for each case of hepatitis A. Hepatitis B is a sexually transmitted or blood-borne virus, so public health measures focus on contacts exposed by those routes. Hepatitis B can have a chronic carrier state, which may lead to confusion about whether a newly diagnosed case is acute, the result of recent transmission, or chronic, the result of transmission in the past. Distinguishing acute from chronic disease is important in order to prevent further transmissions and to identify which segments of the population are currently at high risk for new infections. Elevated levels of immunoglobulin M (IgM) are a laboratory indicator of any acute infection. Blood testing for hepatitis A antigen and hepatitis B core antigen IgM is useful to confirm cases of acute illness from either virus. However, there may be false-positive hepatitis A IgM tests because of other infections or vaccination, and health care providers may not always order the appropriate IgM tests in patients with possible hepatitis B.

Algorithms to identify hepatitis A and B in EHRs may be based on the national surveillance case definitions for these diseases promulgated by CSTE (10). These include both clinical and laboratory criteria: an illness with discrete onset and jaundice, or elevated serum aminotransferase and positive IgM test. ESPnet has algorithms for acute hepatitis A and B that mirror CSTE criteria (9, 27). These algorithms were compared to laboratory surveillance alone and to ICD-9 diagnosis codes. For hepatitis A, the laboratory reporting of hepatitis A IgM, physician ICD coding and

EHR algorithm all had very high sensitivity. Only two cases identified by IgM testing alone were missed by EHR reporting. However, EHR reporting had a 100% PPV compared to 71% for IgM reporting, where there were a number of false-positive tests, and 48% for ICD-9 coding. Therefore, in terms of efficiency of surveillance, EHR reporting was superior in that less effort was required to follow up on false-positive cases. In addition, EHR reports included liver function test results, which are needed to apply the surveillance case definition, thereby saving surveillance staff the need to contact the health care provider to obtain this information.

For acute hepatitis B surveillance, EHR reporting had far higher sensitivity and PPV than hepatitis B core antigen IgM testing or ICD-9 coding (1). IgM testing detected only 9% of confirmed cases, an indication that physicians were not ordering this test. There were very few false-positive EHR reports of acute hepatitis B. Therefore, EHR surveillance was superior to alternative methods.

Sexually Transmitted Diseases

Methods similar to those used for hepatitis A and B have been used to automate surveillance reporting for chlamydia, gonorrhea, and syphilis from EHRs (11). Sensitivity and PPV were very high for all three conditions. Compared with passive provider reporting, EHR reporting increased the number of chlamydia reports by 39% and gonorrhea by 53%. Pelvic inflammatory disease (PID), which is not a reportable disease in many jurisdictions, was also evaluated. PID is a condition for which there is no specific laboratory test, so electronic laboratory reporting is not a suitable method for surveillance. Twenty cases of PID were reported versus none by passive reporting (9).

Influenza/Influenza-Like-Illness Surveillance

Syndromic surveillance is a more efficient method than is case reporting for tracking very common conditions such as influenza or for rapid identification of potential clusters of disease. Because individual public health action is not taken in response to case reports, personal identifiers need not be collected. Influenza surveillance has historically relied on sentinel physician practices that manually report the number of cases meeting a definition of ILI that they see each week (13). In recent years, these reports have been supplemented by electronic reporting of chief complaints from EDs and some ambulatory care settings. To identify cases, a text search is performed of the chief complaint field in electronic ED systems using standard search terms (e.g., “fever,” “flu”). The reports contain basic demographic information but generally do not contain any clinical, vaccination, or laboratory information. EHR-based systems can provide these data and relieve the burden on health care providers to report via the sentinel surveillance system.

In one study, data from a network of community health centers sharing a common EHR in New York were evaluated for surveillance of ILI and other syndromes and were compared with hospital-ED syndromic data (23). Cases were identified both by queries of structured EHR data such as ICD-9 code, temperature, and respiratory rate, and by natural language processing of narrative data. EHR-based data were found to correlate well with hospital ED–based syndromic reporting; the data from structured queries were slightly better correlated. The same data were useful during the 2009 H1N1 influenza pandemic to determine the geographic distribution of influenza cases and to direct public health resources such as vaccines (7). In another study, EHR-based syndromic surveillance using the ESPnet system to detect ILI was highly correlated with traditional sentinel surveillance data (44).

EHRs FOR CHRONIC DISEASE SURVEILLANCE

The medical literature contains fewer studies of public health agencies using EHRs for chronic disease surveillance than it does for communicable diseases. EHR data have been used to estimate the burden of selected diseases in the population receiving health care. As with syndromic surveillance, these conditions are much more common than the traditional reportable communicable diseases, and public health action is not taken in response to individual cases; therefore, identifiable information need not be collected in the surveillance process. Both prevalence and incidence of chronic disease are important surveillance indicators to track. More detailed clinical, laboratory, and treatment data are needed to determine risk factors and to fully assess the burden of disease. EHR-based surveillance for chronic diseases is limited to populations receiving care.

EHR data have been used to estimate the prevalence of asthma in the population receiving care in a health care provider network with more than 375,000 patients in Wisconsin (42). Reports were based exclusively on physician ICD-9 coding. The distribution of asthma cases in the EHR data was comparable for age, sex, and race ethnicity to data from the state's BRFSS survey. The limitations of the study were the exclusive reliance on physician ICD-9 coding and limited coverage of the whole population.

EHR data have also been used to assess the validity of smoking prevalence estimates in England (5). Smoking rates in a random sample drawn from the national Clinical Research Datalink (CPRD), a comprehensive, population-wide health care database, showed close correlation with those from a representative national health survey. Rates of former smokers were lower in the EHR data, possibly because physicians did not record this information systematically, as the survey did.

Clinical algorithms for types I and 2 diabetes have been developed in the ESPnet system using ICD-9 codes for diabetes, an elevated fasting glucose or hemoglobin A_{1c}, prescription for an oral hypoglycemic agent other than metformin, or a prescription for insulin not during pregnancy (28). The optimized algorithm for type 1 diabetes had a sensitivity of 97% and a PPV of 88%.

Eye health is related to diabetes; blindness is not an uncommon complication. Data from three different data systems, Kaiser Permanente, CMS (Medicare), and the Veterans Health Administration, were combined to detect eye disorders (19). The advantages of using EHRs for surveillance of eye disease were the real time nature of the data, the automated collection, and the ability to study incidence and prevalence over time. Disadvantages were a lack of full documentation of eye diseases in EHRs and a lack of interoperability of different EHR systems, which limited the ability to combine data.

Finally, EHRs provide an opportunity to conduct better surveillance for preventive health screenings such as colonoscopies and mammograms, which now are tracked only by population surveys such as BRFSS (7).

THE PROMISE AND CHALLENGES OF EHRs FOR PUBLIC HEALTH SURVEILLANCE

The EHR revolution provides great promise for improving public health surveillance, but this promise is still unrealized. Use of EHRs for surveillance has many potential benefits. EHRs provide an opportunity to streamline and improve current surveillance practices. Their use could greatly improve the reporting of nonlaboratory diagnosed diseases as well as the collection of treatment and risk factor data, so long as these data are entered in the EHR. EHRs could be built with integrated, disease-specific surveillance modules with prompts for disease-specific

questions—symptoms, treatment, exposures, and contacts—to assure that the necessary data are collected. EHRs could be used to greatly enhance surveillance for chronic diseases. They have the potential to expand the purview of routine surveillance to include obesity, asthma, diabetes, heart disease, and cancer at little marginal cost per condition. They can also provide rich data on health care utilization, treatment patterns, and outcomes that are currently very difficult to assess with routine surveillance methods. Finally, EHRs provide multiple opportunities for greater integration of public health and clinical health by enhancing information sharing in both directions.

Using EHRs for public health surveillance has many challenges. At the moment, many providers have still not implemented an EHR. To further develop EHRs for public health surveillance, standardization of EHRs for public health content, technical design, and communication protocols must continue. There is also a great need for systematic and unified case definitions of all diseases under surveillance based on data routinely coded in all EHRs. For the information to be consistent across the population, surveillance algorithms should be based on standardized EHR data elements and should report a standardized set of data about each case. EHRs need to be linked into networks both for clinical benefit to the patient and for aggregation of data for public health surveillance and population health measurement. The ongoing maintenance of the algorithms to identify cases of each disease in each EHR is labor and resource intensive. MU incentive funding under the Affordable Care Act is intended to support the development of EHR-based surveillance, but resources are needed for public health departments, too. MU must continue to drive the standardization of both EHRs and public health surveillance needs. Unfortunately, MU is currently delayed and does not provide resources to build the public health infrastructure necessary to receive surveillance reports from EHRs. All this will require public health jurisdictions to work together and to overcome historical local preferences for how surveillance is conducted so that we can establish national standards.

The published research on EHR-based public health surveillance is very limited. Much more research is needed to document the potential benefits and limitations of EHRS. Some questions for future research include the following: For communicable diseases, does surveillance using EHRs result in higher sensitivity and timeliness of completed investigations with an acceptable PPV compared with traditional electronic laboratory-based surveillance? Can cost savings for the surveillance system be demonstrated? Are public health control measures for communicable diseases more effective, e.g., prophylaxis administered and secondary cases averted or identified more quickly? For chronic disease surveillance, is the population coverage with EHRs sufficient to yield representative, population-wide data? How do EHR-based data, which include only people in care, compare with population-wide survey data? Are EHR data sufficiently detailed in terms of risk factor information (e.g., current smoker, past smoker) to yield population-based estimates comparable with those of current surveillance? What is the capacity of public health agencies to receive, analyze, and disseminate findings from surveillance reports from EHRs? Do public health alerts transmitted through EHRs change clinical practice or improve public health outcomes? Can useful information about social determinants of health be gleaned from EHRs?

Researchers also have an opportunity to document the improved integration of clinical services and public health programs that results from the increased interaction between clinicians and public health professionals, which is stimulated by EHR-based surveillance, as called for in a recent Institute of Medicine report (25). Such interaction could lead to collaborations in other areas, such as the development of automated public health decision support tools accessible through the EHR. Finally, new forms of public health surveillance using EHRs have the potential to monitor population health in as-yet-unimagined ways. Ideally, public health and clinical medicine will work together to develop meaningful surveillance measures that are useful at both the clinical population level and the total population health level. We are limited only by our imaginations.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank Christopher Davis, Geraldine Johnson, and Winnie Yu for invaluable assistance in researching the literature and drafting the manuscript.

LITERATURE CITED

1. Allen-Dicker J, Klompas M. 2012. Comparison of electronic laboratory reports, administrative claims, and electronic health record data for acute viral hepatitis surveillance. *J. Public Health Manag. Pract.* 18(3):209–14
2. ASTHO (Assoc. State Territ. Health Off.). 2011. *Meaningful use readiness survey*. ASTHO, Arlington, VA. **http://www.astho.org/programs/e-health/meaningful-use-readiness-survey/**
3. Blumenthal D. 2010. Launching HITECH. *N. Engl. J. Med.* 362(5):382–85
4. Blumenthal D, Tavenner M. 2010. The 'meaningful use' regulation for electronic health records in U.S. hospitals. *N. Engl. J. Med.* 363(6):501–4
5. Booth HP, Prevost AT, Gulliford MC. 2013. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiol. Drug Saf.* 22(12):1357–61
6. Calderwood MS, Platt R, Hou X, Malenfant J, Haney G, et al. 2010. Real-time surveillance for tuberculosis using electronic health record data from an ambulatory practice in eastern Massachusetts. *Public Health Rep.* 125(6):843–50
7. Calman N, Hauser D, Lurio J, Wu WY, Pichardo M. 2012. Strengthening public health and primary care collaboration through electronic health records. *Am. J. Public Health* 102(11):e13–18
8. CDC (Cent. Dis. Control Prev.). 1996. Historical perspectives notifiable disease surveillance and notifiable disease statistics—United States, June 1946 and June 1996. *MMWR* 45(25):530–36
9. CDC (Cent. Dis. Control Prev.). 2008. Automated detection and reporting of notifiable diseases using electronic medical records versus passive surveillance—Massachusetts, June 2006–July 2007. *MMWR* 57(14):373–76
10. CDC (Cent. Dis. Control Prev.). 2013. *Case definitions for notifiable disease surveillance*. Updated Dec. 10. CDC, Atlanta. **http://wwwn.cdc.gov/nndss/script/casedefDefault.aspx**
11. CDC (Cent. Dis. Control Prev.). 2014. *Behavioral Risk Factor Surveillance System*. Updated Sept. 29. CDC, Atlanta. **http://www.cdc.gov/brfss/**
12. CDC (Cent. Dis. Control Prev.). 2014. *Meaningful use*. Updated Sept. 16. CDC, Atlanta. **http://www.cdc.gov/ehrmeaningfuluse/**
13. CDC (Cent. Dis. Control Prev.). 2014. *Overview of influenza surveillance*. Updated Oct. 6. CDC, Atlanta. **http://www.cdc.gov/flu/weekly/overview.htm**
14. CDC (Cent. Dis. Control Prev.). 2014. *Prevention and Public Health Fund*. Updated Feb. 11. CDC, Atlanta. **http://www.cdc.gov/fmo/topic/PPHF/**
15. Chiolero A, Santschi V, Paccaud F. 2013. Public health surveillance with electronic medical records: at risk of surveillance bias and overdiagnosis. *Eur. J. Public Health* 23(3):350–51
16. Chute CG, Koo D. 2002. Public health, data standards, and vocabulary: crucial infrastructure for reliable public health surveillance. *J. Public Health Manag. Pract.* 8(3):11–17
17. D'Amore JD, Sittig DF, Ness RB. 2012. How the continuity of care document can advance medical research and public health. *Am. J. Public Health* 102(5):e1–4
18. Effler P, Ching-Lee M, Bogard A, Ieong MC, Nekomoto T, Jernigan D. 1999. Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods. *JAMA* 282(19):1845–50

19. Elliott AF, Davidson A, Lum F, Chiang MF, Saaddine JB, et al. 2012. Use of electronic health records and administrative data for public health surveillance of eye health and vision-related conditions in the United States. *Am. J. Ophthalmol.* 154(6 Suppl.):S63–70
20. GNU Operat. Syst. 1991. *GNU Library General Public License*. Version 2.0, June. Free Software Found., Boston. **http://www.gnu.org/copyleft/library.html**
21. German RR, Lee LM, Horan JM, Milstein RC, Pertowski CA, et al. 2001. Updated guidelines for evaluating public health surveillance systems. Recommendations from the Guidelines Working Group. *MMWR* 50(RR-13):1–35
22. HITPC Meaningful Use Workgr. 2014. *Meaningful Use stage 3 recommendations*. Updated March 11. Off. Natl. Coord. Health Inf. Technol., Washington, DC. **http://www.healthit.gov/FACAS/sites/faca/files/hitpc_muwg_stage3_recs_2014_03_11.pdf**
23. Hripcsak G, Soulakis ND, Li L, Morrison FP, Lai AM, et al. 2009. Syndromic surveillance using ambulatory electronic health records. *J. Am. Med. Inf. Assoc.* 16:354–61
24. Hsiao CJ, Hing E. 2012. *Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2012*. NCHS Data Brief. Updated Dec. 6, CDC, Atlanta. **http://www.cdc.gov/nchs/data/databriefs/db111.htm**
25. IOM (Inst. Med.). 2012. *Primary Care and Public Health: Exploring Integration to Improve Population Health*. Washington, DC: IOM, Natl. Acad.
26. IOM (Inst. Med.). 2014. *Capturing Social and Behavioral Domains on Electronic Health Records: Summary of Selected Domains*. Washington, DC: Natl. Acad. Sci. **http://www.iom.edu/~/media/Files/Report%20Files/2014/EHR-Phase-1/EHRdomains.pdf**
27. ISDS (Int. Soc. Dis. Surveill.). 2012. *Electronic Syndromic Surveillance Using Hospital Inpatient and Ambulatory Clinical Care Electronic Health Record Data. Recommendations from the ISDS Meaningful Use Workgroup*. Brighton, MA: ISDS. **http://www.syndromic.org/storage/ISDS_2012-MUse-Recommendations.pdf**
28. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. 2013. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care* 36(4):914–21
29. Klompas M, McVetta J, Lazarus R, Eggleston E, Haney G, et al. 2012. Integrating clinical practice and public health surveillance using electronic medical record systems. *Am. J. Public Health* 102(S3):S325–32
30. Kukafka R, Ancker JS, Chan C, Chelico J, Khan S, et al. 2007. Redesigning electronic health record systems to support public health. *J. Biomed. Inf.* 40(4):398–409
31. Langmuir AD. 1963. The surveillance of communicable diseases of national importance. *N. Engl. J. Med.* 268:182–92
32. Lazarus R, Klompas M, Campion FX, McNabb SJ, Hou X, et al. 2009. Electronic support for public health: validated case finding and reporting for notifiable diseases using electronic medical data. *J. Am. Med. Inf. Assoc.* 16(1):18–24
33. Lee LM, Teutsch SM, Thacker SB, St. Louis ME, eds. 2010. *Principles and Practice of Public Health Surveillance*. New York: Oxford Univ. Press. 3rd ed.
34. Lenert L, Sundwall DN. 2012. Public health surveillance and meaningful use regulations: a crisis of opportunity. *Am. J. Public Health* 102(3):e1–7
35. Loonsk JW, McGarvey SR, Conn LA, Johnson J. 2006. The Public Health Information Network (PHIN) preparedness initiative. *J. Am. Med. Inf. Assoc.* 13:1–4
36. OPHSS (Off. Public Health Sci. Serv.). 2014. *Summary of public health objectives in stage 2 Meaningful Use - ONC and CMS final rules*. Updated April 1. OPHSS, Atlanta. **http://www.cdc.gov/ehrmeaningfuluse/Docs/Summary%20of%20PH%20Objectives%20in%20Stage%202%20MU%20ONC%20and%20CMS%20Final%20Rules_04_01_2014.pdf**
37. Overhage JM, Grannis S, McDonald DJ. 2008. A comparison of the completeness and timeliness of automated laboratory reporting and spontaneous reporting of notifiable conditions. *Am. J. Public Health* 98(2):344–50
38. RWJF (Robert Wood Johnson Found.). 2013. *Health Information Technology in the United States: Better Information Systems for Better Care, 2013*. Princeton, NJ: RWJF. **http://www.rwjf.org/en/research-publications/find-rwjf-research/2013/07/health-information-technology-in-the-united-states-2013.html**

39. Silk BJ, Berkelman RL. 2005. A review of strategies for enhancing the completeness of notifiable disease reporting. *J. Public Health Manag. Pract.* 11(3):191–200
40. Sosin DM, Hopkins RS. 2010. Public health surveillance for preparedness and emergency response: biosurveillance for human health. See Ref. 33, pp. 306–20
41. Thacker SB. 2010. Historical development. See Ref. 33, pp. 1–17
42. Tomasallo CD, Hanrahan LP, Tandias A, Chang TS, Cowan KJ, Guilbert TW. 2014. Estimating Wisconsin asthma prevalence using clinical electronic health records and public health data. *Am. J. Public Health* 104(1):e65–73
43. Vogel J, Brown JS, Platt R, Land T, Klompas M. 2014. MDPHnet: secure, distributed sharing of electronic health record data for public health surveillance, evaluation, and planning. *Am. J. Public Health* 104:2265–70
44. Yih CW, et al. 2014. Automated influenza-like illness reporting—an efficient adjunct to traditional sentinel surveillance. *Public Health Rep.* 129:55–63