PUBLISHED BY THE INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Advances in Disease Surveillance

VOLUME 3
NUMBER 3
DATE 2007

ARTICLES

*E*lectronic Medical Record *S*upport for *P*ublic Health (ESP): Automated Detection and Reporting of Statutory Notifiable Diseases to Public Health Authorities

Michael Klompas[1,2], Ross Lazarus[2], James Daniel[3], Gillian A. Haney[3], Francis X. Campion[4], Benjamin A. Kruskal[4], Xuanlin Hou[2], Alfred DeMaria[3], and Richard Platt[1,2]

[1] Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA.
[2] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.
[3] Massachusetts Department of Public Health, Boston, MA.
[4] Harvard Vanguard Medical Associates, Boston, MA.

Received for publication August 10, 2006; accepted for publication March 27, 2007.

Clinician-initiated reporting of notifiable conditions is often delayed, incomplete, and lacking in detail. We report on the deployment of Electronic medical record Support for Public health (ESP), a system we have created to screen electronic medical record (EMR) systems automatically for evidence of reportable diseases, to transmit disease reports securely to health authorities, and to respond to queries from health departments for clinical details about laboratory detected cases. ESP uses a software that constructs and analyzes a temporary database that is regularly populated with comprehensive codified encounter data from a medical practice's EMR system. The ESP database resides within the host medical practice's firewall, configured on either a central workstation to service large multisite, multi-physician practices or as a software module running alongside a small practice's EMR system on a personal computer. The encounter data sent to ESP include patient demographics, diagnostic codes, laboratory test results, vital signs, and medication prescriptions. ESP regularly analyzes its database for evidence of notifiable diseases. When a case is found, the server initiates a secure Health Level 7 message to the health department. The server is also able to respond to queries from the health department for demographic data, treatment information, and pregnancy status on cases independently reported by electronic laboratory systems. ESP is designed to be compatible with any EMR system with export capability: it facilitates translation of proprietary local codes into standardized nomenclatures, shifts the analytical burden of disease identification from the host EMR system to the ESP database, and is built from open-source software. The system is currently being piloted in Harvard Vanguard Medical Associates, a multi-physician practice serving 350,000 patients in Eastern Massachusetts. Disease detection algorithms are proving to be robust and accurate when tested on historical data. In summary, ESP is a secure, unobtrusive, flexible, and potentially portable method for bidirectional communication between EMR systems and health departments. It is currently being used to automate the reporting of notifiable conditions but has promise to support additional public health objectives in the future.

Public Health Practice; Epidemiologic Measurements; Disease Notification; Medical Records Systems; Computerized

Correspondence to Michael Klompas, Harvard Medical School, Department of Ambulatory Care and Prevention, 133 Brookline Avenue, Boston, MA 02215 (e-mail: mklompas@partners.org)

Abbreviations: EMR, electronic medical record; ESP, Electronic medical record Support for Public health; HL7, Health Level 7; HVMA, Harvard Vanguard Medical Associates; ICD9, International Classification of Diseases, Ninth Revision; LOINC, logical observation identifiers names and codes; PHIN-MS, Public Health Information Network Messaging System; PID, pelvic inflammatory disease.

No health department, State or local, can effectively prevent or control disease without knowledge of when, where, and under what conditions cases are occurring.

Introductory statement printed each week in *Public Health Reports* from 1913 through 1951

INTRODUCTION

For over 100 years, public health authorities have actively tracked the incidence and distribution of communicable diseases as a first step toward understanding and preventing their spread (1). The adoption of direct electronic reporting by clinical laboratories has increased the breadth and efficiency of reporting for many diseases (2, 3). Laboratory reporting systems, however, have not obviated the need for a clinician's participation in reporting. Some reportable diseases are only established by a clinical impression, such as pelvic inflammatory disease (PID) or culture-negative tuberculosis. Other diseases require clinician's interpretation of laboratory results such as Lyme disease and acute hepatitis C. In addition, electronic laboratory reporting systems do not provide detailed demographic data on patients or pertinent clinical details such as symptoms, treatment rendered, and pregnancy status. Clinician reporting, however, is dependent upon clinician initiative and is still largely done manually by mail, fax, or telephone. Some jurisdictions have instituted web-based reporting systems but these continue to depend upon clinician initiative to report (4). Unfortunately, clinician-initiated health reporting suffers from incomplete capture of incident infections, incomplete descriptions of case data, delay between detection and processing of reports, and substantial administrative cost for manual processing of reports (5, 6, 7). In light of these limitations of current reporting, we have created an electronic system to automate the detection and reporting of notifiable conditions by leveraging the information coded into electronic medical record (EMR) systems. The name of the system is Electronic medical record Support for Public health (ESP). It supports secure, bidirectional communication between EMR systems and health departments. It detects and reports notifiable conditions identified from a clinician's medical records. It can also report specific patient level clinical information to health authorities in response to health department queries about patients identified independently by electronic laboratory reporting. The system has been developed using the EpicCare EMR system (8) at Harvard Vanguard Medical Associates (HVMA), a multipractice physician group serving 350,000 patients in Eastern Massachusetts. It has been designed, however, to be potentially compatible with any EMR system that has export capability.

Conceptual framework

ESP has been designed to embody the following principles:

1. Automatic—shifts the initiative for reporting from healthcare providers to electronic systems.
2. Unobtrusive—invisible to clinicians during routine clinical care; transfers analytical workload away from the host EMR server so as not to interfere with clinical computing.
3. Secure—employs stringent measures to protect sensitive clinical data.
4. Universal—designed to be potentially compatible with almost any electronic medical record system.
5. Flexible—easily accommodates rule modifications to detect new conditions or improve the detection of existing conditions.
6. Provider controlled—clinical data is stored on the provider's premises and cases can be reviewed for approval prior to transmission.

METHODS

Architecture

ESP consists of a database and analytical software placed within a medical practice. The database is regularly populated with specific data elements extracted from each encounter recorded in the practice's EMR system. ESP analyzes the database nightly for evidence of notifiable conditions. When notifiable conditions are identified, ESP formats a Health Level 7 (HL7) message and initiates a secure, encrypted electronic message to the state health department. ESP is also able to respond to HL7 messages from the health department containing queries for clinical details about cases reported directly by laboratories. Practices with multiple locations that share an integrated EMR system can be accommodated by a single central ESP server. Our initial deployment, for example, uses a single server located in a central data center that serves 16 different clinical sites. ESP can also be configured for small practices as an independent software module that can run on a practice's microcomputer alongside its EMR system.

The model of an independent database and analytical software system, housed on a practice's premises and secured by the practice's staff, was selected for the following reasons: 1) to avoid interference with clinical operations by offloading analytical and processing burden from the host EMR system; 2) to permit flexibility to modify ESP case detection algorithms frequently without interfering with the host EMR system; 3) to secure confidential patient

information by retaining all data behind the practice's electronic firewall until a message is sent to the health department; and 4) to allow portability of ESP to a wide array of EMR systems with different internal structures, proprietary codes, and analytical capabilities.

The ESP database is regularly populated with data extracts of encounter data sent by the host EMR system via FTP protocol. The source EMR system formats the clinical data into a defined sequence of delimited text fields containing patient demographics, diagnostic codes, laboratory orders and results, medication prescriptions, vital signs, and pregnancy status. Clinician notes and other free-text entries are not currently included. ESP has a built-in code remapping tool to convert local codes into standardized nomenclatures that can be analyzed by ESP. Laboratory test codes are translated into logical object identifier names and codes (LOINC), result names are translated into systematized nomenclature of medicine (SNOMED) codes, diagnoses are translated into International Classification of Diseases, Ninth Revision, (ICD9) codes, and prescribed medication codes are translated into national drug code (NDC) numbers. The ESP code remapping tables are maintained and managed by authorized practice staff using a local ESP web application.

When a case is identified, ESP queries the database for additional clinical information of importance to the health department including pregnancy status and relevant prescriptions. The system also assesses recent ICD9 codes and vital signs to determine whether the patient was symptomatic (e.g., recorded temperature for presence of fever; ICD9 codes for urethritis or vaginal leukorrhea).

ESP preferentially sends an immediate case report to the health department. If a clinician wishes to review cases prior to transmission, however, ESP can queue cases for manual approval before they are sent. Cases queued for confirmation are reviewed by authorized personnel using a secure, web-based case management system within the practice's firewall. The case management system presents the user with a list of patients with suspected reportable diseases. When the reviewer selects a patient, ESP displays a summary of recent encounters and their types and dates, ICD9 diagnoses, lab test results, medication prescriptions, and ordering clinicians. The particular data terms that led to ESP flagging the case for reporting are highlighted. The reviewing clinician can choose to reject the case (false positive), to authorize transmission (true positive), or to place the case on hold while further information is being gathered. The web interface also has links to contact information for each ordering clinician and to the CDC (Centers for Disease Control and Prevention) website with criteria for notifiable disease identification (9). A demonstration version of the ESP case management interface using fictional patient data can be viewed at http://esphealth.org.

When a case is ready for transmission, ESP generates a standardized HL7 message. The initial deployment sends encrypted HL7 messages to the Massachusetts Department of Public Health server using a simple object access protocol web service, but the ESP messaging sub-system is also able to send messages using the Centers for Disease Control's PHIN-MS (Public Health Information Network Messaging System) secure messaging protocol. Any authorized health authority capable of receiving HL7 messages can potentially receive ESP messages. ESP also supports queries initiated by authorized health departments to provide clinical context for positive lab tests they have received independently from their electronic lab reporting system. Patient matching is done using test accession numbers, patient names, and test dates. As with conditions first identified by ESP, the system assesses for pregnancy status, patient symptoms, and relevant prescriptions. Response messages are transmitted back to the health department in HL7 format following the same protocol as ESP initiated cases, including optional case review by a practice designee.

The ESP database is purged of all patient encounter data after 90 days other than the data relevant to confirmed cases. Data from these cases are retained to permit the generation of reports that detail the system's notification activity.

The ESP server runs custom software built from the following open-source applications: MySQL RDBMS (MySQL AB, Uppsala, Sweden), CDC PHIN-MS (Centers for Disease Control and Prevention, Atlanta, GA), and the Python language (Python Software Foundation, Ipswich, MA). It can run upon both Linux (Red Hat, Raleigh, NC) and Windows (Microsoft Corporation, Redmond, WA) operating systems.

Security

Confidential patient data are protected by keeping it under the physical and logical control of the practice until the point of message transmission. ESP does require outbound access to the public internet in order to communicate with the health department server but it is protected from external access behind an internet firewall. Messages are preferentially transmitted using PHIN-MS over 128-bit encrypted communication channels, but the messaging protocol can be tailored to suit the requirements of different health departments. Public key encryption infrastructure certificates are required of both sending and receiving machines before patient data are transmitted.

Case identification logic

Cases are identified by analyzing diagnostic codes, laboratory tests and results, and medication prescriptions. The case definitions are based on those published by the Centers for Disease Control, but can be customized for local users (9). They run the gamut from simple to complex. A positive DNA probe for *Chlamydia trachomatis* from a urethral swab, for example, is sufficient to establish a case of chlamydia. Other conditions require more sophisticated analysis of multiple laboratory tests. Acute hepatitis C, for example, is diagnosed when the patient has a concurrent positive hepatitis C enzyme-linked immunosorbent assay (ELISA), confirmatory recombinant immunoblot assay, negative IgG and/or IgM for hepatitis A, negative core and surface antigens of hepatitis B, and a serum alanine aminotransferase level seven times above the upper limit of normal for the assay. The system can also assess for suggestive changes in laboratory tests over time. Acute hepatitis C

can also be diagnosed by serial negative followed by a positive hepatitis C ELISA, or by the combination of a negative ELISA paired with a positive hepatitis C RNA PCR assay. Conditions that are diagnosed on clinical grounds alone are sought by looking for suggestive combinations of ICD9 codes, laboratory test orders, and medication prescriptions. Lyme disease, for example, can be diagnosed from the combination of 1) an ICD9 code 088.81 (erythema chronicum migrans) or positive Lyme serology, and 2) prescription for at least 14 days of doxycycline or other suggestive antibiotic.

Compatability

ESP has been designed to enable broad adoption by minimizing technical and processing demands on source EMR systems, by embracing standardized nomenclatures, by permitting custom case identification algorithms, and by reliance on low cost software components. The ESP input data files extracted from source EMR systems consist only of unformatted text delimited into our specified field sequence. The EMR system can export encounter data using its own internal codes as the ESP data load tool includes user-configurable translation tables to map proprietary local codes into standardized vocabularies. Local users are responsible for building, maintaining, and confirming the accuracy of their proprietary to standard code mapping tables. Users need only translate the small subset of laboratory codes pertinent to each condition into LOINC terms understood by ESP thereby limiting the burden of code translation. The ESP remapping table facilitates this process by soliciting the local user for their proprietary equivalent of each LOINC code used by ESP. In addition, users need to specify the range of possible positive result designations for each code such as activation of an abnormal result flag or specific free-text such as "detected" or "positive" or "present." Free-text result fields are parsed by ESP for these terms using a simple natural language processor. In addition, ESP constantly screens all incoming labs for the emergence of new codes that might be consistent with notifiable diseases by screening the text accompanying each lab (regardless of result) for truncated keywords such as "chlam*" for chlamydia or "hep*" for hepatitis. When a novel code containing one of these terms is identified by ESP, an alert is generated querying the user whether the new code ought to be added to the ESP code map or henceforth ignored. We successfully used the code remapping web application for the initial deployment of ESP at HVMA where we needed to translate CPT (current procedural terminology) and component combinations into LOINC codes. We have yet to test the application at a second implementation site, however, hence the true extent of the system's portability has yet to be fully assessed. Source code for ESP can be downloaded under a lesser general public license from http://esphealth.org.

RESULTS

Validation

Case identification algorithms were validated by applying each algorithm to a five-year span of historical data from HVMA. The total number of patients identified by each algorithm was compared with the Massachusetts Department of Public Health's historical counts for HVMA patients manually reported to have the disease of interest. A subset of 50 randomly chosen charts of patients who met the rule's conditions was then reviewed manually to assess the rule's accuracy. These 50 patients were then matched to individuals in the health department records using limited identifiers (gender, date of birth, HVMA office site, and test date) in order to assess congruity between patients identified electronically and those reported manually. For example, between 2000 and 2004, HMVA reported 1,629 cases of chlamydia to the health department. Application of the ESP chlamydia detection algorithm to electronic records from this period identified 1,927 episodes. Review of 50 randomly selected charts confirmed that all cases were true positives (100% positive predictive value). We were able to match 90% of these 50 cases to individuals in health department records, a percentage that reassuringly mirrors the ratio of manually reported cases to electronically identified cases from our five-year cohort.

We also assessed the accuracy of case identification for cases of PID. ESP reports a case of PID when a patient is given an ICD9 code for PID and has a positive test for gonorrhea or chlamydia within 30 days. This case definition is a strict interpretation of the PID case definition published by the CDC (10, 11). It was designed under guidance from the Massachusetts Department of Public Health to maximize the specificity at a cost of sensitivity on the rationale that 1) privacy concerns preclude reporting false positive results; and 2) current reporting of PID is extremely sparse and therefore any increase in reporting (even if incomplete) will be an improvement on current practice. Using this case identification algorithm, we identified 74 cases of PID at HVMA during the years 2000–2004. During this same period, HVMA manually reported only 1 case. Chart review of 12 patients identified by this algorithm revealed one possible false positive case (a patient who had pelvic pain and concurrent culture confirmed *Herpes simplex* type 2 as well as *Chlamydia trachomatis*).

Current status

ESP has been implemented at HVMA, a large multisite, multi-specialty group practice with over 350,000 patients in Eastern Massachuestts. Messages are sent to the Massachusetts Department of Public Health. The system currently detects and reports chlamydia, gonorrhea, and PID. Next we will add reporting for pertussis, acute hepatitis C, and Lyme disease, and then, for the remaining 75 conditions that are reportable in Massachusetts. The State health department anticipates initiating queries to ESP triggered by their Electronic Lab Reporting system in late 2007. We have also formed a partnership with the Massachusetts eHealth Collaborative to pilot ESP in the community-wide EMR system being deployed in North Adams, Massachusetts.

DISCUSSION

Potential future applications

The ESP model of a time-limited database, refreshed daily with comprehensive clinical data and capable of bi-directional communication with health departments, has

potential to serve many additional public health functions beyond identification and reporting of notifiable diseases. These include syndromic surveillance [including the distributed data model used by the National Bioterrorism Syndromic Surveillance Demonstration Program (12, 13)], vaccine registries, clinical decision support for the management of notifiable conditions, auditing of mandated screening programs such as lead assays in children, assessment for spatial clusters of environmentally linked diseases such as asthma, and population level surveillance for non-notifiable diseases. We anticipate also building modules to generate reports for clinicians' or practices' use that tally the number of notifiable conditions detected and reported for a given period relative to the denominator of their choice (e.g., total ambulatory visits, total number of patients tested for the condition, total number of patients with an ICD9 code suggestive of the presenting clinical syndrome, etc) to assist with internal quality control and education.

Conclusion

We have created a system to identify and report individual patients with notifiable conditions automatically using routine data collected by EMR systems. ESP also supports queries initiated by health departments for clinical data to guide management of positive laboratory reports received independently from electronic laboratory reporting systems. ESP uses industry standard nomenclature. The system has been designed to be secure, unobtrusive to clinicians, and potentially compatible with most EMR systems. The system also has the potential to support additional public health objectives including syndromic surveillance, vaccine registries, and auditing of mandated screening programs. It is currently being deployed and actively tested in Massachusetts.

ACKNOWLEDGMENTS

The development of ESP is supported by a grant from the Centers for Disease Control (PH000238D).

REFERENCES

1. Centers for Disease Control and Prevention (CDC). Notifiable disease surveillance and notifiable disease statistics—United States, June 1946 and June 1996. MMWR Morb Mortal Wkly Rep 1996;45:530–6.
2. Effler P, Ching-Lee M, Bogard A, Ieong MC, 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.
3. Panackal AA, M'ikanatha NM, Tsui FC, et al. Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. Emerg Infect Dis 2002;8(7): 685–91.
4. Centers for Disease Control and Prevention (CDC). Progress in improving state and local disease surveillance—United States, 2000–2005. MMWR Morb Mortal Wkly Rep 2005;54: 822–5.
5. Thacker SB, Choi K, Brachman PS. The surveillance of infectious diseases. JAMA 1983;249:1181–85.
6. Thacker SB, Berkelman RL. Public health surveillance in the United States. Epidemiol Rev 1988;10:164–90.
7. Standaert SM, Lefkowitz LB Jr, Horan JM, Hutcheson RH, Schaffner W. The reporting of communicable diseases: a controlled study of *Neisseria meningitidis* and *Haemophilus influenzae* infections. Clin Infect Dis 1995;20: 30–6.
8. EpicCare Ambulatory. Available at http://www.epicsystems.com/Software/EnterpriseClinical.php#ambulatory. Accessed December 3, 2006.
9. Centers for Disease Control and Prevention (CDC). Summary of notifiable diseases—United States, 2003. MMWR Morb Mortal Wkly Rep 2005;52:1–85.
10. Centers for Disease Control and Prevention (CDC). Pelvic inflammatory disease. 1996 case definition. Available at http://www.cdc.gov/epo/dphsi/casedef/pelviccurrent.htm. Accessed February 20, 2007.
11. Centers for Disease Control and Prevention (CDC). Sexually transmitted diseases treatment guidelines, 2006. MMWR Recomm Rep 2006;55:1–94.
12. Yih WK, Caldwell B, Harmon R, et al. National Bioterrorism Syndromic Surveillance Demonstration Program. MMWR Morb Mortal Wkly Rep 2004;53(Suppl):43–9.
13. Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. BMC Public Health 2006; 6:235.