Harnessing Electronic Health Records for Public Health Surveillance

Michael Klompas MD MPH[1], Michael Murphy BA[1], Julie Lankiewicz MPH[1], Jason McVetta[2], Ross Lazarus MBBCh[1], Emma Eggleston MD MPH[1], Patricia Daly MS RN[3], Paul Oppedisano MPH[3], Brianne Beagan MPH[3], Chaim Kirby JD MA[4], Richard Platt MD MSc[1]

[1] *Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA*
[2] *Heliotropic Inc, Los Angeles, CA*
[3] *Massachusetts Department of Public Health, Boston, MA*
[4] *Children's Hospital, Boston, MA*

Abstract

Electronic medical record (EMR) systems are a rich potential source for detailed, timely, and efficient surveillance of large populations. We created the Electronic medical record Support for Public Health (ESP) system to facilitate and demonstrate the potential advantages of harnessing EMRs for public health surveillance. ESP organizes and analyzes EMR data for events of public health interest and transmits electronic case reports or aggregate population summaries to public health agencies as appropriate. It is designed to be compatible with any EMR system and can be customized to different states' messaging requirements. All ESP code is open source and freely available. ESP currently has modules for notifiable disease, influenza-like illness syndrome, and diabetes surveillance.

An intelligent presentation system for ESP called the RiskScape is under development. The RiskScape displays surveillance data in an accessible and intelligible format by automatically mapping results by zip code, stratifying outcomes by demographic and clinical parameters, and enabling users to specify custom queries and stratifications. The goal of RiskScape is to provide public health practitioners with rich, up-to-date views of health measures that facilitate timely identification of health disparities and opportunities for targeted interventions. ESP installations are currently operational in Massachusetts and Ohio, providing live, automated surveillance on over 1 million patients. Additional installations are underway at two more large practices in Massachusetts.

Introduction

The Harvard Center of Excellence in Public Health Informatics has developed an electronic medical record (EMR) based system for comprehensive public health surveillance: the **E**lectronic Medical Record **S**upport for **P**ublic Health (ESP) platform (http://esphealth.org).[1, 2] ESP organizes raw data extracted from EMR systems, maps them to heuristic concepts, analyzes these data for conditions of public health interest, and electronically transmits case-level or

1

population-level data to public health agencies. ESP is designed to be compatible with any EMR system. All source code is available free of charge under a Library General Public License. ESP currently provides notifiable disease reporting for selected infectious diseases, syndromic surveillance for influenza-like illness, and chronic disease surveillance using diabetes mellitus as the demonstration example.

Infectious disease surveillance

ESP was originally designed to identify and electronically report patients with notifiable diseases such as chlamydia, gonorrhea, active tuberculosis and acute viral hepatitis. In contrast to electronic laboratory reports, ESP leverages the full breadth of data present in EMRs to do more than simply report positive test results. For example, ESP algorithms distinguish acute versus chronic infections and active versus latent tuberculosis.[3] ESP algorithms also seek clinical diagnoses that may not trigger positive laboratory tests such as early Lyme disease and culture-negative tuberculosis.[4] Once a case is identified, ESP uses the wealth of data in EMRs to prepare HL7 electronic case reports that include electronically determined symptoms, pregnancy status, and treatments prescribed.

ESP also includes a syndromic surveillance module for influenza-like illness. This module counts patients who fulfill the Centers for Disease Control and Prevention syndromic definition for influenza-like illness. Results are stratified by age and sex and then sent to the Massachusetts Department of Public Health. The health department merges ESP's data into the state -wide sentinel report for integration into the Centers for Disease Control and Prevention National Influenza-Like Illness Surveillance Program for HHS Region 1: (http://www.cdc.gov/flu/weekly/regions2010-2011/hhssenusmap.htm).

Diabetes surveillance

We are currently working on methods to apply the disease detection protocols we pioneered for notifiable disease detection to surveillance for diabetes prevalence, incidence, care, and complications. As with infectious diseases, integration of laboratory data with current and prior diagnoses and prescriptions facilitates more complete and more granular surveillance. For example, we found surveillance for gestational diabetes by assessing oral glucose tolerance test results alone misses a third of cases. One reason is that clinicians may make the diagnosis in unconventional but clinically reasonable ways. For example, the patient may have a history of gestational diabetes from a prior pregnancy and may spontaneously start checking her glucoses on her own and find high results. An algorithm that assesses pregnancy, diagnosis codes, and prescriptions for test strips or lancets captures these extra cases.[5]

Similarly, our algorithm for frank diabetes looks for patients with positive hemoglobin A1Cs, elevated fasting glucose and/or random glucose values, new prescriptions for insulin or oral antiglycemic agents, or recurrent diagnosis codes for diabetes. Including all these criteria increases case capture by 28% compared with assessing diagnosis codes alone (the current de facto standard for electronic population level surveillance) and by 54% compared to using hemoglobin A1Cs alone (an alternative method of population surveillance for diabetes currently mandated in New York City).[6]

We have also developed an algorithm to distinguish between type 1 and type 2 diabetes.[6] This is a major advance since most current population level surveillance tools, such as administrative codes or the Behavioural Risk Factor Surveillance System (BRFSS) do not routinely distinguish between these two very different diseases (some states do add supplemental questions asking patients to self-report diabetes type but this is not a core component of BRFSS). Our algorithm is nested within the population identified by our frank diabetes algorithm and incorporates current and historical ICD9 codes, laboratory tests, and prescriptions. The algorithm's sensitivity and positive predictive value for type 1 diabetes is 100% and 94% respectively.

Benefit to Public Health Practice
ESP provides public health practitioners with more detailed and timely data to identify priority areas for intervention compared to traditional surveillance tools. Public health departments currently rely primarily on voluntary, anonymous telephone surveys such as the Behavioral Risk Factor Surveillance System (BRFSS) to assess chronic disease and health behavior patterns.[7, 8] The BRFSS has significant limitations, including cost, reliance on self-reports, restriction to respondents with telephones, limited language coverage, and capacity for only a limited number of questions. Estimates are not possible for certain populations or geographic areas because of sample size limitations. BRFSS also lacks clinical information such as medications, laboratory tests or vital signs. ESP overcomes many of these limitations by harnessing the wealth of data routinely captured by EMR systems. ESP is able to provide comprehensive surveillance on very large numbers of patients with detailed data on patient demographics (age, sex, race/ethnicity, location), important clinical traits (e.g. body mass index, pregnancy status), patterns of care (e.g. medication prescriptions, use of screening tests, referral to nutrition therapy), health outcomes (hemoglobin A1C, lipid profiles, blood pressure control), and complications (e.g. hypoglycemic episodes, chronic kidney disease, retinopathy, etc.). These data can help health departments identify disparities in health status, care patterns, and outcomes and inform targeted interventions for the most vulnerable members of the population.

Lessons Learned
Effective use of electronic medical record data for public health purposes requires sophisticated understanding of clinical data sources in order to faithfully capture and meaningfully map native data to universal concepts. Surveillance algorithms integrating multiple components of the medical record (diagnosis codes, laboratory test results, and medication prescriptions) are often more sensitive and specific than any of these components alone. Developing sensitive and specific algorithms is painstaking work requiring access to rich clinical data, sophisticated programming staff, engaged clinical staff that can validate electronic cases against manual reviews, and sufficient time and patience to iteratively repeat this process to optimize performance.[9]

Limitations
There are important limitations to surveillance using EMR data. Examples include discrepancies in coding practices between physicians and practices that may affect the performance of case identification algorithms; incomplete data on patients' diagnoses, lab tests, and prescriptions when patients seek care from multiple providers, some of whom may be outside the practices covered by ESP; difficulty determining accurate denominators for incidence and prevalence calculations since some individuals at risk never seek medical care and some individuals seek

3

care from multiple practices and therefore may count in the denominators of both practices but in the numerator of only one, both or neither practice depending on how the specific array of diagnosis codes, labs, and prescriptions accrued at each practice; difficulty keeping concept mapping current in the face of rapidly changing coding nomenclatures; and limited capacity to identify important contextual data and risk factors that are poorly recorded in EMR systems such as incarceration, restaurant work, sick contacts, and recent travel.

Translating Research into Practice

We have developed comprehensive reports for pre-diabetes, gestational diabetes, and frank diabetes that describe demographics (age, sex, race, location), clinical parameters (hemoglobin A1C, body mass index, lipid profile, blood pressure), and indicators of care (prescriptions, nutrition referrals, follow-up testing, changes in glucose control over time). These reports were custom built in collaboration with the Massachusetts Department of Public Health to highlight the parameters of greatest interest and concern to public health practitioners. These reports can also be used by the medical practices themselves to study their patterns of care and identify targets for quality improvement initiatives.

Collaborations

There are four ESP installations at various stages of maturity. The core installation is in Atrius Health, a multispecialty practice with 700 physicians serving over 700,000 patients in 25 sites in Massachusetts. The Atrius Health ESP server resides in the practice's central data processing center. It is populated nightly with text files extracted from Atrius's Epic Care EMR. These contain clinical information on every patient encounter from the preceding 24 hours. A second mature installation is in MetroHealth, an integrated ambulatory and hospital system serving over 350,000 patients in Cleveland, Ohio. It is also populated with text files extracted from MetroHealth's Epic Care EMR. These two installations have together reported over 12,500 notifiable disease case reports to their respective state health departments since inception. Two additional installations of ESP are currently underway. One is in the Northern Berkshires regional health information exchange in North Adams, Massachusetts. This installation is populated by HL7 messages generated by eClinical Works EMRs. It shows ESP's compatibility with different EHR systems and the feasibility of installing ESP in a health information exchange to serve an entire community. The fourth installation is in the Cambridge Health Alliance, an integrated health care system and safety net provider affiliated with the city health department that provides care in hospitals and health centers for the City of Cambridge.

Benefit to Public Health Informatics

All software and protocols developed by the Center are freely available for use by medical practices and public health agencies across the state and nation. ESP can be readily adopted by any medical practice and is fully extensible, allowing users to develop and implement new surveillance targets and reports. ESP is compatible with different electronic medical record systems. It is currently populated by extract-transform-and-load from Epic Care, HL7 messages generated by eClinical Works, and SQL queries from Epic Care's Clarity system.

The Center's work is the platform for a new competitive award from the Office of the National Coordinator for Health Information Technology to build and deploy MDPHnet, a distributed network to support bi-directional communication between the state health department and

practices that have installed ESP. MDPHnet will help integrate surveillance results from distributed ESP installations. It will also facilitate custom queries from health department officials to run in parallel on distributed ESP systems.

Impact on Public Health Practice
The culmination of ESP's data extraction and analysis is intelligent presentation. We aim to make ESP data as intuitive and impactful as possible for users. In collaboration with our Center of Excellence partners at Children's Hospital, we are creating the RiskScape, a web interface to graphically display population level surveillance summaries (Figure 1). The RiskScape is built to be generalizable to any population level surveillance target (e.g., diabetes, asthma, heart disease, influenza like illness), to allow users maximal flexibility to stratify the data in whatever way they wish, and to be as user friendly and visually appealing as possible. The intent of the RiskScape is to highlight geographic regions and population groups that could benefit most from targeted public health interventions. For example, a RiskScape user can specify a report of the rates of postpartum testing for frank diabetes amongst gestational diabetics stratified by zip code, race / ethnicity, and age. A report of this nature could highlight Hispanic women under age 20 in the southern neighborhoods of Boston as having disproportionately lower rates of postpartum testing. This in turn can inform a targeted public health campaign to increase postpartum testing for this well-defined population.
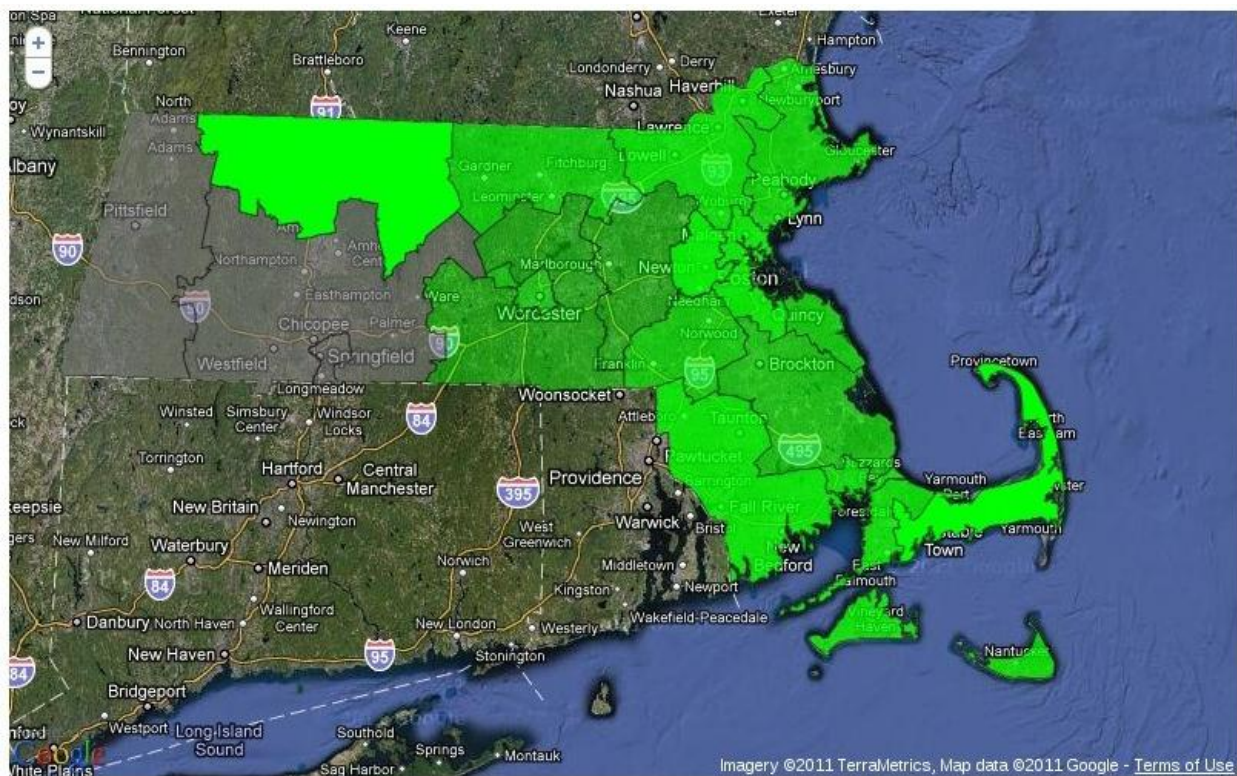

Figure 1. RiskScape Screenshot.

The figure depicts a heat map of the proportion of women with gestational diabetes who receive nutrition counseling from a registered dietician by 3-digit zip code. The figure highlights regions

of the state where disproportionately fewer women are getting counseling from a dietitian. Analyses of this sort can help public health departments develop targeted interventions for the population groups and regions at greatest need.

Conclusions

ESP demonstrates the vast potential of EMRs to change the face of surveillance by improving the accuracy, completeness, efficiency, and granularity of surveillance with relatively little marginal cost for new infections and conditions. The ESP system is customizable and extensible. New surveillance targets such as immunization registries, clinical care monitoring, and drug safety all have the potential to be integrated into ESP. Much unexplored territory remains.

Corresponding author

Michael Klompas
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
133 Brookline Avenue, 6th Floor
Boston, MA 02215
Phone: 617-509-9991
Fax: 617-859-8112
Email: mklompas@partners.org

References

1. Centers for Disease Control and Prevention. 2008. Automated detection and reporting of notifiable diseases using electronic medical records versus passive surveillance--Massachusetts, June 2006-July 2007. MMWR Morb Mortal Wkly Rep. 57(14), 373-76. PubMed
2. Lazarus R, Klompas M, Campion FX, et al. 2009. Electronic Support for Public Health: validated case finding and reporting for notifiable diseases using electronic medical data. J Am Med Inform Assoc. 16(1), 18-24. PubMed http://dx.doi.org/10.1197/jamia.M2848
3. Klompas M, Haney G, Church D, Lazarus R, Hou X, et al. 2008. Automated identification of acute hepatitis B using electronic medical record data to facilitate public health surveillance. PLoS ONE. 3(7), e2626. PubMed http://dx.doi.org/10.1371/journal.pone.0002626
4. Calderwood MS, Platt R, Hou X, 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. PubMed
5. Klompas M, McVetta J, Eggleson E, et al. Automated surveillance and public health reporting for gestational diabetes incidence and care using electronic health record data (abstract). Emerging Health Threats Journal. 2011;4.
6. Klompas M, Eggleson E, McVetta J, et al. Automated detection and classification of diabetes using electronic health records. Paper presented at: CDC Diabetes Translation Conference2011; Minneapolis, MN.

7. Chowdhury P, Balluz L, Town M, et al. 2010. Surveillance of certain health behaviors and conditions among states and selected local areas - Behavioral Risk Factor Surveillance System, United States, 2007. MMWR Surveill Summ. 59(1), 1-220. PubMed

8. Hughes E, Kilmer G, Li Y, et al. 2010. Surveillance for certain health behaviors among states and selected local areas - United States, 2008. MMWR Surveill Summ. 59(10), 1-221. PubMed

9. Klompas M, Bialek SR, Kulldorff M, Vilk Y, Harpaz R. 2011. Herpes zoster and postherpetic neuralgia surveillance using structured electronic data. Mayo Clin Proc. (In press). PubMed http://dx.doi.org/10.4065/mcp.2011.0305