In Brief

Although electronic health records (EHRs) were designed as individual-level documentation tools for patient-provider clinical interactions, some have advocated for their use in population-level health surveillance efforts. Limited research exists on this utility of EHRs, but some investigators have used them successfully to estimate the prevalence of chronic disease events and heart disease risk factors. This article reviews such studies and describes EHR use in the Heart of New Ulm project, a large 10-year intervention designed to reduce the incidence of myocardial infarctions and improve heart disease risk factors in a rural Minnesota community.

Methods of Using Electronic Health Records for Population-Level Surveillance of Coronary Heart Disease Risk in the Heart of New Ulm Project

Jeffrey J. VanWormer, MS

Ever since the call for universal adoption of electronic health records (EHRs) by the American health care system, EHRs are being rapidly adopted by hospitals and entire medical care systems to streamline patient care, increase the efficiency of documentation and monitoring of health outcomes, and reduce medical errors.1,2 EHRs hold much promise to simplify at least some aspects of clinical care and are often touted as a major cost-savings technology. The actual rate at which EHRs have been adopted around the country, however, varies greatly by region and health care system size. However, <20% of health care provider systems currently report using an EHR,3 with a somewhat lower rate among smaller clinics and in rural areas.

EHRs are mainly designed to be individual-level, patient-provider encounter-oriented technology. They capture and store patient health characteristics and provider actions. However, the scale and sustainability by which EHRs collect this sort of routine data also present the potential for major epidemiological research. In some health care systems, EHRs may provide an optimal technology for surveillance of at least some health conditions at the population level.2

This offers a major advantage for epidemiological researchers who have difficulty collecting community-wide data in areas beyond the reach of existing surveillance systems. For example, the National Health and Nutrition Examination Survey4 provides estimates of blood lipid levels at the national level but is not adequately powered to characterize blood lipid levels at much smaller, local levels such as a county, zip code, or city. Similarly, the Behavioral Risk Factor Surveillance System,5 a very large national survey, provides national and state-level estimates on behavioral risk factors such as smoking, but it does not permit reliable estimates of such risk factors at smaller geographical levels.

Given the major expense and expertise required for large, traditional epidemiological surveillance options and the fact that EHRs are becoming more common, there is great potential to use a new paradigm for population-level surveillance. This is particularly true at more localized community levels where interventions are most likely to occur and therefore need to be monitored for effectiveness.

EHRs can be queried and essentially contain naturally collected clinical data on large segments of the communities they serve. Models of EHR-based surveillance of a given population’s health risks, however, are scarce in the scientific literature. This article highlights methods of EHR-based surveillance of coronary heart disease (CHD) risk factors and describes a model of population-level EHR-based surveillance of CHD risk factors that is being used in the Heart of New Ulm (HONU) project, an ongoing 10-year study designed to reduce the rate of myocardial infarctions (MIs) in a rural area of southwestern Minnesota. The article
also outlines special considerations for other population-level health surveillance projects.

**Previous Research**

**CHD events**

At the population level, surveillance of CHD events is sometimes performed using aggregated administrative data. For example, in some states, hospitals report de-identified CHD events data (e.g., International Classification of Diseases, Ninth Revision [ICD-9], codes 410–414) on an annual basis to the state health department. The state health department then collates and analyzes these data to monitor CHD trends at the state or regional level. These limited datasets are often inadequate for detailed epidemiological investigations because they rarely contain individual-level information beyond CHD event hospital discharge code, age, sex, and resident address.

In a similar manner, publicly available data from county records can be examined to gauge CHD mortality. These data also provide only very limited demographic information and, because no individual identifiers are present on both hospital and death records, cannot be reliably linked to calculate an accurate incident CHD rate.

A recent study utilized an approach in which the investigators summed the annual hospital discharged acute coronary syndrome cases with out-of-hospital CHD death cases (divided by the estimated person-years of follow-up using census estimates). Such a composite index estimates the overall CHD event rate based on the assumption that all hospital discharged cases did not die and all out-of-hospital CHD death cases did not reach the hospital. Although it obviously produces an overestimate of the true CHD attack rate because of double counting many cases, it may be a reasonably sensitive method to detect gross changes in CHD rates associated with policy interventions.

A more pointed example relying on EHR data comes from the Marshfield Epidemiologic Study Area (MESA). The MESA draws from EHRs from the Marshfield Clinic’s health care delivery system, which captures > 95% of all hospitalizations and deaths that occur among residents of 14 zip codes in north central Wisconsin. Specifically, the incidence rate of acute MIs was monitored throughout 7 years. MESA investigators used a validation algorithm whereby, with the exception of death certificate MIs, all discharge-coded MIs must appear twice in the EHR before an incident case is counted. Person-time for the denominator of incident rate calculations comes from a combination of census data with a correction for prevalent MI cases, deaths, and changes of address (in or out of the MESA area). This method of EHR-based incident MI surveillance has shown nearly 100% sensitivity and 92% specificity in MESA validity studies.

Other studies of estimating CHD event rates over time have also drawn selectively from medical records data. However, these studies tended to be large, ongoing cohort investigations of specific areas in which extensive case validation procedures were used to clinically verify incident CHD. Such procedures set the gold standard for validity of CHD surveillance but may be impractical for typical surveillance purposes because of their high costs and required technical expertise.

**CHD risk factors**

In addition to CHD events, some EHR-based population-level surveillance has occurred for CHD risk factors. Surveillance of CHD risk factors at the population level is very important because it addresses the modifiable components of CHD events and therefore can more precisely inform the design and monitoring of interventions for prevention.

Precedents in this area of EHR surveillance are less common in the scientific literature. The most comprehensive EHR-based population surveillance of CHD risk factors was performed by Persell et al. and was based on earlier methods piloted by Van Wyk. They examined the records of 23,111 patients aged 20–79 years who were seen at least twice during a 2-year period at a Chicago-area academic internal medicine practice. Age, sex, systolic blood pressure, total cholesterol, HDL cholesterol, use of antihypertensive medications, use of lipid-lowering medications, use of antithrombotic medications, and cigarette smoking data were collected. In addition, Framingham risk scores (FRSs), an estimate of the 10-year probability of experiencing a CHD event, were also calculated from these data where possible.

Results indicated that 77% of this patient population had complete risk factor data available to determine their FRS. Lipid values were the least available metric, with 23% missing their total cholesterol measurements. The overall FRS was similar to national estimates, with ~ 15% of the population indicated as high risk (i.e., having a > 20% estimated 10-year probability of a CHD event). A random audit of 100 charts comparing EHR-derived (automated) CHD risk assessment versus physician-derived (manual) CHD risk assessment found a kappa statistic of 0.91, indicating a high level of agreement between the EHR and manual methods.

Solberg et al. used EHRs to assess the prevalence of tobacco use—a strong behavioral risk factor for CHD—among 204,800 patients seen at a large health care delivery organization in the Minneapolis metropolitan area. Data on tobacco use and demographics were pulled from EHRs, and 69% of all eligible patients had complete data available for analyses. Current tobacco use prevalence was found to be 19%, which is in line with other state-level estimates. In addition, further analyses found that current tobacco use was significantly associated with male sex, having Medicaid insurance, having English as a preferred language, and being an American Indian.

**Putting It All Together: The HONU Project**

**Project overview**

The HONU project is a systems-level, population health–oriented initiative designed to reduce the CHD event rate in the 56073 zip code (i.e., the New Ulm, Minn., area; see conceptual impetus from Boucher et al.). New Ulm is a rural area in southwestern Minnesota and was selected in an attempt by the Allina health care system to focus and utilize resources to conduct broader population health management centered on heart disease prevention and using advanced EHR informatics tools.

The HONU project is delivering a suite of programs and services to the target population that is designed to enhance the existing heart disease prevention activities within the com-
munity. Design of this intervention package was informed by previous community-level heart disease prevention programs, including the Minnesota Heart Health Program. The intervention package of the HONU project targets modifiable CHD risk factors, most notably tobacco use, blood pressure, blood lipids, obesity, stress, physical activity, and fruit and vegetable consumption.

Intervention components are rolled out across three spheres of influence, including health care, worksites, and the community. Health care–oriented interventions include EHR-based clinical decision support tools (e.g., information flow sheets), quarterly continuing medical education seminars, and a telephonic coaching program for high-risk patients. Worksite activities include employer environment and culture audits, wellness plan development consulting, and employee-focused quarterly lifestyle challenge campaigns and biannual CHD risk factor screenings. Community-oriented interventions include food service provider environmental assessments (with follow-up consulting from a community registered dietitian on menu or shelf modifications), neighborhood lifestyle challenge programs led by lay community health leaders, and biannual CHD risk factor screenings.

All project activities are paired with a strong health-related social marketing component (e.g., newspaper articles, radio interviews, televised documentaries, grocery bag and utility bill inserts, and Internet social networking). This effort is designed to promote awareness of risk factors, healthy lifestyle benefits, and project-related programs and services to create strong interest in health improvement efforts and reinforcement among community residents.

EHR-based surveillance of outcomes

The long-term objective of the HONU project is to reduce the CHD event rate among residents of the 56073 zip code (aged 40–79 years) during 10 years. Medium-term objectives are to improve the proportion of 56073 zip code residents with controlled modifiable risk factors during 5 years. These risk factors include high LDL cholesterol, high blood pressure, high blood glucose, obesity, smoking, and medication underutilization (for anti-thrombotic, anti-dyslipidemia, or anti-hypertension medications).

Informed by the previous research outlined above, EHR data are used on an ongoing basis to monitor each of these metrics in the HONU project. For the primary outcome of interest—MIs—an EHR-based surveillance system has been set up to track all ICD-9 code 410 hospital discharges (primary and secondary) from Allina facilities. In addition, death certificates, obituary notices, and ambulance records are continuously monitored to track CHD mortality (specifically ICD-10 code I22–I25.x) among area residents. These data are made available in quarterly intervals and are presented to study investigators in aggregated, de-identified form.

A manual process is set up whereby each hospital discharge and death certificate MI case is reviewed by a designated reviewer at the New Ulm Medical Center to determine whether a given MI case is an incident or recurrent event and to ensure that cases that have both a hospital discharge and death certificate indication of MI are not double counted if they occur within 30 days of each another. In this way, deaths, incident MIs, and recurrent MIs can be disentangled to provide the unique number of MI cases that occurs among zip code 56073 residents per Allina health care system records.

For the secondary outcomes—CHD risk factors—the most recently collected data from EHRs are pulled at annual intervals to examine descriptive statistics for each metric, similar to the processes conducted by Persell et al. For example, the most recent LDL cholesterol value collected during 2009 would be used for each individual, and then aggregate descriptive statistics (e.g., mean, standard deviation) would be calculated to represent the 2009 characterization of LDL cholesterol for the entire community. Obviously, in any given 1-year timeframe, not all adults in the target population will have their LDL measured because this surveillance system relies on naturally occurring office visits, with the occasional supplementation of community heart health screenings that are offered by the project every few years. Recent analyses of Allina EHR data have indicated that a 1-year surveillance period for total cholesterol, blood pressure, and tobacco use offers a statistically indistinguishable parameter estimate relative to longer timeframes (with larger samples), even out as far as a 5-year time period.

Discussion

Evidence is evolving on methods of using EHRs as a population health surveillance tool, particularly with regard to monitoring CHD events and, to a lesser degree, CHD risk factors. In addition, EHR-derived estimates of global CHD risk appear to have a high degree of agreement with manual physician-derived estimates of global CHD risk. This is important because of the potential of EHRs to rapidly assess the CHD risk of large segments of the population, something that is not commonly done during standard office visits.

However, EHRs are clearly not yet a common tool used to monitor the health of populations. This is likely because of several reasons such as the limited availability of EHR data in many locations (although this is changing), the limited access to aggregated EHR data in a given health care system, and the fact that few hospital systems are responsible (or reimbursed) for managing the health of their entire community. Given the recent efforts at the national level, particularly as they relate to the recently passed health care reform bill, these factors may soon change, as health plans and health care provider systems are held accountable (and possibly rewarded) for more effective prevention and control of chronic diseases in the communities they serve, rather than only in the patients that present for treatment.

Special considerations

Using EHRs to conduct population-level surveillance of CHD risk factors could be a major breakthrough in the U.S. public health infrastructure, in which tracking large-scale regional trends is complicated by a health care system that tends to operate in “silos.” For example, it is nearly impossible to calculate a reliable national estimate of the incidence rate of MIs because that would require centralized data from all medical centers at which individuals present with MIs. Because these institutions are a mix of public and private organizations across different states and with different nonstandardized tracking mechanisms, access to that level of data would be extremely problematic. Indeed, national esti-
mates of CHD incidence rates are typically extrapolated from the few large cohort studies in which this information is routinely collected.\textsuperscript{12,22}\n
There are also ethical considerations involved in using EHR data for population health surveillance. This is essentially a retrospective or quality improvement type of evaluation whereby obtaining classic informed consent a priori on such a large scale would render the approach impractical. Patients in Minnesota have the option to opt out of health services research using their health records during an office visit at the facility collecting the information. In the HONU project, efforts are made only to use limited analytical datasets for precisely what is needed to answer the research questions of interest. In addition, personally identifiable information is stripped for analytical review boards at different facilities may hold more or less conservative philosophies on the most appropriate use of EHR data for population health surveillance.

A third concern relates to whether data from EHRs are representative of the population they are purported to represent or whether they present bias. In the HONU project, the EHR data used for analyses are purported to represent adult residents aged 40–79 years in the 56073 zip code. Of course, it is likely impossible to know the precise number of individuals and their health status at any given time in a given large population, but relative to U.S. census data, it appears that ~92% of all individuals in the target population have at least one CHD risk factor data point present in the Allina EHR during the preceding 5 years. Therefore, the EHR offers a reasonable population health surveillance tool for the geographical area of interest in the HONU project.

This may not be the case in other areas. For example, large metropolitan areas are typically served by multiple health care provider systems and health plans. Data from just one of these systems may not adequately represent a given suburban city where the populace is divided among several health care provider systems (particularly if the divisions are differential based on age, sex, race, or economics) and individuals frequently switch among them. In such cases, EHR data from multiple sources would likely need to be combined to avoid a biased representation of the target population.

Conclusions

EHRs are fast becoming a ubiquitous tool in the American health care system. Although EHRs were designed as a documentation tool for individual patient-provider encounters, their potential to serve as a tool for managing the health of entire communities is just being realized. Evidence from the scientific literature is quite limited in this area so far, but some health care organizations have demonstrated the successful use of EHRs to characterize crucial health endpoints such as quantifying CHD incidence or identifying type 2 diabetes cases.

It seems timely to now expand that logic to begin programs designed for population-level surveillance of CHD, diabetes, and other chronic disease risk factors, which are arguably the more important health characteristics to assess and intervene on if prevention efforts are to be successful. For example, Hivert et al.\textsuperscript{23} were recently able to mine 12 primary care clinics’ EHR data to identify metabolic syndrome cases and track their use of clinical resources (e.g., hospital admissions and outpatient visits). Findings indicated that patients with metabolic syndrome tend to be more expensive to treat in the short term and therefore are a justifiable target for investment in aggressive preventive interventions through clinical outreach initiatives.

The HONU project has designed and is using a basic platform (i.e., population health dashboard) to track the health metrics of one Minnesota zip code as part of a broader effort to intervene on CHD risk factors and reduce the rate of CHD events in that area. If proven successful, it may be appropriate to translate and transfer this model to other areas within the Allina health care system and beyond.

References

16Solberg LI, Flottemesch TJ, Foldes SS, Molitor BA, Walker PF, Crain AL: Tobacco-use prevalence in special populations taking


20Johnson PJ: “Comparison of Electronic Health Record and Community Screening Data.” Presentation at the Minnesota Health Services Research Conference, Minneapolis, Minn., March 2010


Jeffrey J. VanWormer, MS, is the director and principal investigator of the Heart of New Ulm Project at the Minneapolis Heart Institute Foundation in Minnesota.