October 2016

Using Electronic Health Record Data to Improve Community Health Assessment

Brian E. Dixon
Indiana University Richard M. Fairbanks School of Public Health, bedixon@regenstrief.org

Jian "Frank" Zou
Worcester Polytechnic Institute, jzou@wpi.edu

Karen F. Comer
Polis Center at IUPUI, kfrederi@iupui.edu

Marc Rosenman
Regenstrief Institute, mrosenma@iu.edu

Jennifer L. Craig
Regenstrief Institute, craigjen@regenstrief.org

P Gibson
Marion County Public Health Department, jgibson@marionhealth.org

Follow this and additional works at: https://uknowledge.uky.edu/frontiersinphssr

🎨 Part of the Epidemiology Commons, Health Information Technology Commons, and the Health Services Research Commons

Using Electronic Health Record Data to Improve Community Health Assessment

Abstract

Background: Community health assessments assist health departments in identifying health needs as well as disparities, and they enable linking of needs with available interventions. Electronic health record (EHR) systems possess growing volumes of clinical and administrative data, making them a valuable source of data for ongoing community health assessment.

Purpose: To produce population health indicators using data from EHR systems that could be combined and visually displayed alongside social determinants data, and to provide data sets at geographic levels smaller than a county.

Methods: Data from multiple EHR systems used by major health systems covering >90% of the population in a metropolitan urban area were extracted and linked using a health information exchange (HIE) network for individuals who had at least two clinical encounters within the HIE network over a 3-year period. Population health indicators of highest interest to public health stakeholders were calculated and visualized at varying levels of geographic granularity.

Results: Ten population health indicators were calculated, visualized, and shared with public health partners. Indicators ranged from the prevalence of a disease to the proportion of individuals with poor maintenance of their chronic condition. Calculating rates at the census-tract level or larger (e.g., average population size > 4000 people) is preferable to smaller geographic units of analysis.

Implications: Extraction and linking of EHR system data are feasible for public health via an HIE network. While indicators can be derived, biases exist in the data that require more study. Further, HIE networks do not yet possess data for all conditions and measures desired by local public health stakeholders. The data that can be extracted, however, can be combined with public datasets on social determinants

Keywords

Community health assessment, electronic health records, health information exchange, public health informatics, population health indicator, geospatial analysis

INTRODUCTION

Community health assessment enables health departments to identify disparities and high-risk groups and to target interventions appropriately. Community health assessment is also a core requirement for public health accreditation. Ideally, public health decision-makers have up-to-date, comprehensive, representative information about the population they serve, yet the existing, commonly used data sources for community assessment fall short of ideal.[1] Electronic health record (EHR) systems possess growing volumes of clinical and administrative data that describe the safety, quality, and efficiency of care delivered by providers, hospitals, and health systems. However, it remains unclear whether and how the data captured by EHR systems can enable improved assessment of population health.

Whereas existing metrics and prior studies have focused on state and county indicators, this study sought to develop community health indicators at geographic levels smaller than a county. At each geographic level, the team sought to extract and integrate data from multiple EHR systems to create indicators based not on population surveys but recorded health status and outcomes documented by care providers. The goal was to produce data sets of population health indicators that could be combined and visually displayed alongside social determinants data. The study further sought to provide data sets at geographic levels smaller than a county to enable comprehensive community health assessment.

METHODS

Data from multiple EHR systems within the five major health systems serving the residents of Marion County, Indiana, were extracted from the Indiana Network for Patient Care (INPC). The INPC is one of the oldest and most comprehensive health information exchange (HIE) networks in the U.S.,[2] which not only captures data as they are generated by healthcare delivery processes but also integrates clinical data for individual patients to create a longitudinal health record. Founded in 1996 by the Regenstrief Institute and managed by the Indiana Health Information Exchange since 2004, the INPC receives data from over 100 hospitals, several physician practices, laboratories, and other healthcare facilities representing over 12 million unique patients and 6 billion pieces of clinical data. The INPC routinely supports clinical and public health research[3] as well as population health surveillance.[4]

The current study included data on living individuals who resided in Marion County and had at least two clinical encounters within the health systems participating in INPC during a 3-year period beginning January 1, 2011 and ending December 31, 2013. The INPC represents 80% of all Indiana citizens, but it contains data on >90% of citizens in Marion County where it was founded. Representativeness is based on the 2010 decennial census data from the U.S. Census Bureau.

Using a participatory process involving surveys and one-on-one discussions with public health stakeholders across Indiana, several population health indicators were identified that were of high interest, typically not available to health departments at levels below the county, and which relied on data routinely captured in EHR systems and reported to the INPC. The indicators included items measuring prevalence (e.g., prevalence of diabetes, prevalence of asthma) as well as chronic disease management (e.g., control of HbA1c, lipid control); they were based largely on HEDIS (Healthcare Effectiveness Data and Information Set) measures that are routinely reported by health

systems to payers and employers. Other measures were items, like the prevalence of diabetes or depression, not available to health departments in geographic area smaller than a county.

The denominator for prevalence measures represented the number of living individuals who received care from INPC providers during the 3-year time period. Health management denominators represented the number of individuals diagnosed with the disease as documented in the INPC. Numerators were defined using a combination of ICD9-CM clinical diagnostic codes as well as laboratory results and medication history data available from the longitudinal records in INPC.

Indicators were calculated at the following geographic levels: Marion County local health department planning area (average population 40,000–50,000); ZIP code (average population 8000); geopolitical neighborhood (average population 6000); census tract (average population 4000); and census block group (average population 1500). Once calculated, indicator data were mapped using ArcGIS (ESRI, Redlands CA) to visualize the data for local health department personnel. The study received ethics approval by the Institutional Review Board at Indiana University (Study No. 1310502217).

RESULTS

Using the INPC, ten different population health indicators were generated, using data integrated from multiple EHR systems (Table 1). Indicators ranged from the proportion of individuals with depression or a sexually transmitted infection to the proportion of patients with cardiovascular disease who had poor lipid control. The master person index[5] and other foundational technologies for integrating data across unique individuals made this possible with little effort on the part of the study team or our local public health partners.

Figure 1 is a map of Marion County, Indiana, with the prevalence of diabetes by census tract. Figure 2 is a map of Marion County, Indiana, with the rates of diabetes control as measured by glycosylated hemoglobin (HbA1c) levels below 8.0% by those with a confirmed diagnosis of diabetes by census tract. Individuals at the Marion County Public Health Department, when viewing the maps and data depicted in Figures 1 and 2, stated that the availability of population data at geospatial levels smaller than a county is more insightful than examining the countywide rate of diabetes. Granularity was perceived as useful, because health department staff believe it may be able to target diabetes education or prevention efforts in the areas shaded in dark blue where there are either high rates of diabetes (Figure 1) or low rates of diabetes management (Figure 2) within a census tract.

Figures 1, 2, and 3 are presented as Additional Files (attached to the published paper) due to their complexity and to ensure that readers will be able to appreciate the detail.

During this project, it was noted that calculating rates at the census tract level or larger (e.g., average populations >4000 people) is preferable. Rates at the census block level were more challenging to interpret as many blocks contained a count of zero or one. This makes comparing rates difficult, and it would be hard for health departments to release small numbered cells given concerns for privacy and confidentiality of health data.

Table 1. Complete list of population health indicators calculated during the study

Indicator	Description
Prevalence of diabetes	Rates of individuals living with diabetes within a certain geographic area
Prevalence of asthma	Rates of individuals living with asthma within a certain geographic area
Prevalence of chronic obstructive pulmonary disease	Rates of individuals living with COPD within a certain geographic area
Prevalence of depression	Rates of individuals diagnosed with major depression or currently taking medications associated with depression within a certain geographic area
Prevalence of cardiovascular disease	Rates of individuals diagnosed with CVD or who received a procedure stemming from CVD (e.g., coronary artery bypass grafting)
Prevalence of chlamydia, gonorrhea, or syphilis	Rates of individuals with a positive, laboratory confirmed diagnosis of an STD
HbA1c testing for patients with diabetes	Proportion of individuals living with diabetes who received a HbA1c test within the defined period.
HbA1c controlled at <8% for patients with diabetes	Proportion of individuals living with diabetes whose most recent HbA1c test was <8%.
LDL-C screening for patients with cardiovascular conditions	Proportion of individuals with CVD who received an LDL-C test within the defined period.
LDL-C levels < 100 mg/dL for patients with cardiovascular conditions	Proportion of individuals with CVD whose most recent LDL-C levels <100 mg/dL

COPD, chronic obstructive pulmonary disease; CVD, cardiovascular disease; HbA1c, hemoglobin A1c; LDL, low-density lipoproteins; STD, sexually transmitted disease

While novel, population health indicators drawn from HIE networks and/or EHR systems are not perfect, and the data needs of public health departments are larger than what any one data source can provide. For example, while the granular rates of diabetes could help target finite resources at the health department, the overall rate of diabetes as captured in the INPC was only half that as measured by the 2013 community health survey conducted by the Marion County Public Health Department. The data captured by EHR systems are biased toward individuals who have access to or seek care from healthcare providers. Further, the presence of chronic disease is not fully documented by providers in EHR systems.[6,7] In addition, patients receive care from a wider array of facilities than just those affiliated with large health systems, such as grocery store pharmacies and urgent care clinics. All of these issues represent challenges to representativeness in EHR data that must be considered when interpreting population health measures from HIE networks like the INPC.

IMPLICATIONS

Electronic health record systems include more people with more frequent contacts than do existing, standard public health data sources like surveys and death certificates. Therefore, data from EHR systems could allow more frequent measurements of disease incidence and outcomes as well as more precise identification of disparities. Further, EHR-based population health indicators could be combined with education levels, socioeconomic data, and other social determinants of health to better understand disease rates and their underlying causes. For example, Figure 3 plots the proportion of individuals with controlled diabetes (darker areas indicate greater control) in relation to the proportion of the population living in poverty. In addition, access to near real-time

population health indicators may enable health departments to conduct more complete and timely community health assessments, enabling them to be used also for evaluation of public health programs.

Yet data from EHR systems also have limitations that need to be understood to fully interpret their meaning when influencing population health decisions. For example, data captured by EHR systems largely represents "sick care" as opposed to health, as many people do not utilize healthcare facilities every year. Further, some health care is provided outside of the INPC network, including but not limited to urgent care clinics independent of large health systems and public health clinics that do not report data to the HIE. EHR systems can also fail to document diagnoses[7] and/or properly maintain problem lists[8] for patients with chronic illnesses.

Efforts to integrate EHR system data are nascent in many parts of the U.S. More work must be done in research and practice to integrate EHR and public health systems into datasets that can overcome current limitations. Public health practitioners could make HIE network data, drawn from multiple EHR systems, more representative of their communities by encouraging healthcare providers to participate and use HIEs, perhaps by making it easy for providers to use HIEs for public health reporting. Public health practitioners might also explore using HIEs to gather health care quality measures required by meaningful use, such as smoking status or body mass index. Researchers could study how biases in EHR-based community health assessment measures might be quantified and mitigated, or characterize EHR-based measures that best reflect measures from community-based surveys. If the biases can be attenuated, the volume and timeliness of EHR data may take public health a quantum leap forward in its ability to establish continuous improvements in community health.

SUMMARY BOX

What is already known about this topic? Prior work showed that health systems are rapidly adopting electronic health record (EHR) systems and that integration of data can create population health datasets. Further, it is well known that public health departments generally have access to only infectious disease data in EHR systems based on existing policies that require reporting of such data to public health authorities.

What is added by this report? Longitudinal health records using HIE networks for patients with chronic illness represent a novel data source to assist health departments in better understanding the prevalence and management of infectious and chronic disease at local levels.

What are the implications for public health practice, policy, and research? While constructing population health indicators at local levels using EHR systems is feasible, the capacity to do this does not exist everywhere yet. Furthermore, there are inherent biases in EHR data that must be understood and accounted for when applying EHR-based population indicators to public health planning and decision-making processes.

REFERENCES

1. Wetta RE, Dong F, LaClair B, Pezzino G, Orr SA. Factors affecting the progress of community health assessment and improvement activities in Kansas. J Public Health Manag Pract 2015;21(4):E1–9. PMID: 24717555; DOI:10.1097/PHH.0000000000000086.

2. Overhage JM. The Indiana Health Information Exchange. In: Dixon BE, editor. Health Information Exchange: Navigating and Managing a Network of Health Information Systems. 1 ed. Waltham MA: Academic Press, 2016;267–79. Print Book ISBN: 9780128031353.

3. Dixon BE, Whipple EC, Lajiness JM, Murray MD. Utilizing an integrated infrastructure for outcomes research: a systematic review. Health Info Libr J 2016;33(1):7–32.PMID: 26639793; DOI: 10.1111/hir.12127

4. Dixon BE, Grannis SJ, Revere D. Measuring the impact of a health information exchange intervention on provider-based notifiable disease reporting using mixed methods: a study protocol. BMC Med Inform Decis Mak 2013;13(1):121.PMID: 24171799; PMCID: PMC3819468; DOI: 10.1186/1472-6947-13-121.

5. McFarlane TD, Dixon BE, Grannis SJ. Client Registries: Identifying and Linking Patients. In: Dixon BE, editor. Health Information Exchange: Navigating and Managing a Network of Health Information Systems. 1 ed. Waltham MA: Academic Press, 2016;163–82. Print Book ISBN: 9780128031353.

6. Dixon BE, Rosenman M, Xia Y, Grannis SJ. A vision for the systematic monitoring and improvement of the quality of electronic health data. Studies in health technology and informatics 2013;192:884–8. PMID: 23920685.

7. Liaw ST, Chen HY, Maneze D, Taggart J, Dennis S, Vagholkar S, et al. Health reform: is routinely collected electronic information fit for purpose? Emergency medicine Australasia: EMA. 2012;24(1):57–63. PMID: 22313561; DOI: 10.1111/j.1742-6723.2011.01486.x

8. Wright A, McCoy AB, Hickman TT, Hilaire DS, Borbolla D, Bowes WA, 3rd, et al. Problem list completeness in electronic health records: A multi-site study and assessment of success factors. Int J Med Inform 2015;84(10):784–90. PMID: 26228650; PMCID: PMC4549158 [Available on 2016-10-01]; DOI: 10.1016/j.ijmedinf.2015.06.011.