RICHARD M. FAIRBANKS
SCHOOL OF PUBLIC HEALTH
INDIANA UNIVERSITY
IUPUI

RWJF Product Cover Page

RWJF ID#: 71271
File Name: 71271GPmeeting_5.pdf
Presenters Name: Brian E Dixon
Affiliation: IU Richard M Fairbanks School of Public Health, Regenstrief Institute, Department of Veterans Affairs
Title: Measuring Population Health Using Electronic Health Records: Exploring Biases and Representativeness in a Community Health Information Exchange
Organization and Meeting: MEDINFO 15th World Congress on Health and Biomedical Informatics
Date and Location: August 19-23; San Paulo, Brazil

Measuring Population Health Using Electronic Health Records: Exploring Biases and Representativeness in a Community Health Information Exchange

Brian E. Dixon[a,b,c], P. Joseph Gibson[a,d], Karen Frederickson Comer[e], Marc Rosenman[b,f]

[a] *Indiana University Fairbanks School of Public Health, Department of Epidemiology, Indianapolis, IN, USA*
[b] *Regenstrief Institute, Center for Biomedical Informatics, Indianapolis, IN, USA*
[c] *Veterans Health Administration, Health Services Research and Development Service, Indianapolis, IN, USA*
[d] *Marion County Public Health Department, Indianapolis, IN, USA*
[e] *Indiana University-Purdue University Indianapolis, Polis Center, Indianapolis, IN, USA*
[f] *Indiana University School of Medicine, Department of Pediatrics, Indianapolis, IN, USA*

Abstract

Assessment is a core function of public health. Comprehensive clinical data may enhance community health assessment by providing up-to-date, representative data for use in public health programs and policies, especially when combined with community-level data relevant to social determinants. In this study we examine routinely collected and geospatially-enhanced EHR data to assess population health at various levels of geographic granularity available from a regional health information exchange. We present preliminary findings and discuss important biases in EHR data. Future work is needed to develop methods for correcting for those biases to support routine epidemiology work of public health.

Keywords:

Electronic Health Records; Health Information Exchange; Geographic Information Systems; Community Health Planning; Health Services Needs and Demand

Introduction

Public health authorities monitor population health to identify burden of disease, manage health assets, establish policy, and evaluate interventions. This assessment usually relies on a limited set of information available through surveys, vital records, and paper-based disease reporting. Electronic health record (EHR) systems may provide more timely data for a larger portion of a population. Yet there exist a number of challenges to routine use of EHR data, including linking them to community data about social determinants of health. In this study, we sought to develop and evaluate neighborhood-level indicators of population health using EHR data integrated with a community information system (CIS).

Methods

The Indiana Network for Patient Care (INPC), a large health information exchange with over 5 billion clinical observations from EHR systems, was geospatially enhanced and combined with social determinant data from SAVI, a community information system serving the same geographical region (1). We then assessed the prevalence of diseases of public health interest (e.g., diabetes mellitus type 2, depression, chlamydia infection) and calculated several HEDIS-like clinical quality indicators (e.g., number of eligible patients screened for chlamydia, number of diabetics receiving annual HbA1c testing) using the linked records. Using various statistical methods we assessed the reliability and representativeness of these data to measure population health at various levels of geographic granularity (e.g., zip code, census tract, census block group, health planning area, neighborhood, city-county council district).

Results

Rates of diabetes ranged from 1.5% to 16.07% with an average of 8.9% of the population among neighborhoods spread across a large metropolitan area. When examined at the census tract level, diabetes rates ranged from 1.5% to 12.83% with an average of 8.9% of the population for a given area.

During the project we identified three biases in using EHR data. First, EHR data only represent the portion of the population which seeks health care services. Second, linked EHR data are biased based on the algorithm used to match patient records. The HIE uses a very conservative probabilistic technique, resulting in duplicate records for the same person if they provide incomplete demographic data in some healthcare encounters. Third, the HIE proportionally contains more data from low income providers given its relationships with the safety net provider and local health department. Many private practices in the suburbs are not yet part of the HIE. We are exploring ways to adjust rates and correct these biases so they do not overestimate burden of disease and poor care quality in inner-city neighborhoods due mainly to over-representativeness of low-income populations.

Our efforts remain a work in progress. Next we seek to compare census tract, neighborhood, and other geographic area measures with data from a recent population survey. We also seek to present our results to public health authorities for feedback as we expand to other diseases and health indicators.

Conclusion

EHR systems capture data about more people than do population surveys, but they have biases that affect their estimates of population health indicators such as disease prevalence or preventative screening rates. Future work is needed to develop methods for correcting for those biases to support routine epidemiology work of public health.

Acknowledgements

This project was supported by the Robert Wood Johnson Foundation (No. 71271) through its National Coordinating Center for Public Health Services and Systems Research.

References

[1] Comer KF, Grannis S, Dixon BE, Bodenhamer DJ, Wiehe SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. Public Health Rep. 2011;126 Suppl 3:54-61.

Address for correspondence

Brian E. Dixon, MPA, PhD
410 W. 10th St, Suite 2000
Indianapolis, IN, 46202, USA
bedixon@regenstrief.org

Measuring Population Health Using EHRs: Exploring Biases and Representativeness in a Community Health Information Exchange

Brian E. Dixon, P. Joseph Gibson,
Karen Frederickson Comer, Marc Rosenman

Background

Community Health Assessment

- is a core function of public health;
- enables health departments to measure disease incidence, target resources, and evaluate public health policies and programs; and
- requires comprehensive, representative information.

Traditional methods have relied on the following data sources

- agency data sets, such as birth certificates;
- population, behavioral surveys; and
- paper-based disease reporting.

Electronic health record (EHR) and health information exchange (HIE) systems present an opportunity to improve the content, timeliness, and comprehensiveness of community health assessments.

Research Objective

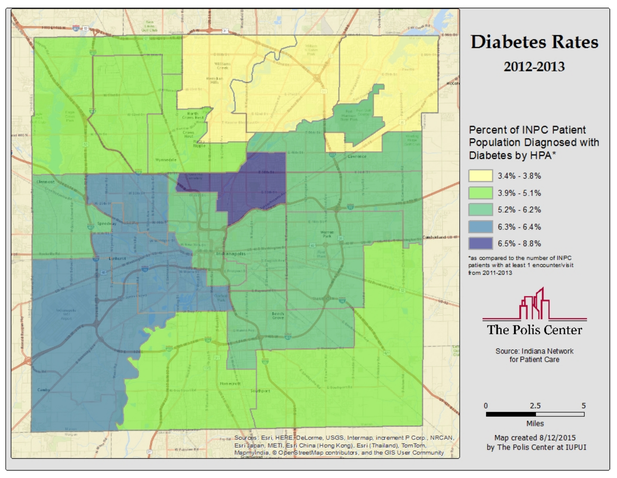
We seek to develop and evaluate neighborhood-level indicators of population health using EHR data integrated with a community information system (CIS).

Methods and Materials

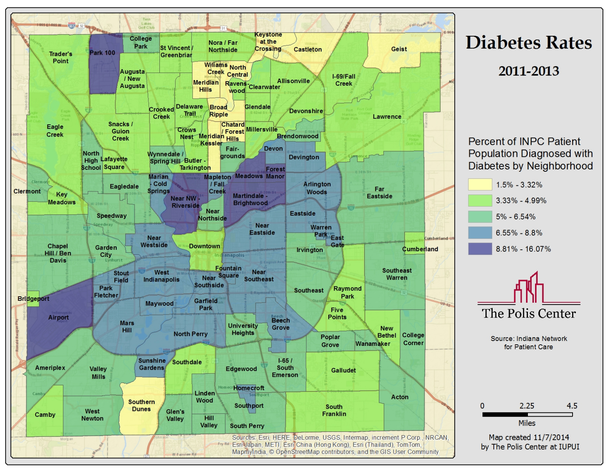
Records from the Indiana Network for Patient Care (INPC), a large health information exchange with over 5 billion clinical observations, were geocoded and combined with geographic social determinant data from SAVI, a CIS serving Indiana.

Using the linked data, we assessed

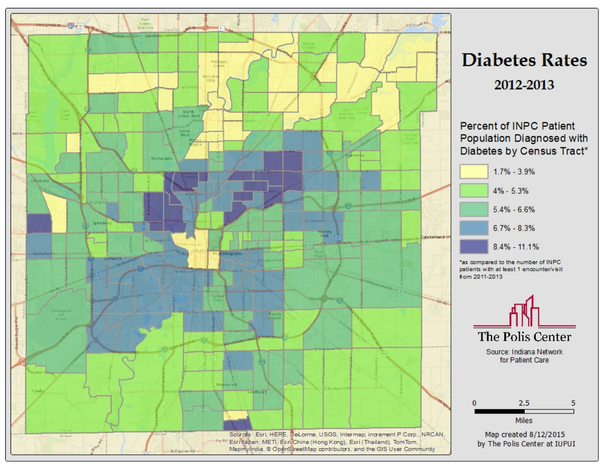
- Prevalence of diseases (e.g., diabetes mellitus type 2); and
- Calculated several HEDIS-like clinical quality indicators (e.g., number of diabetics receiving annual HbA1c testing).



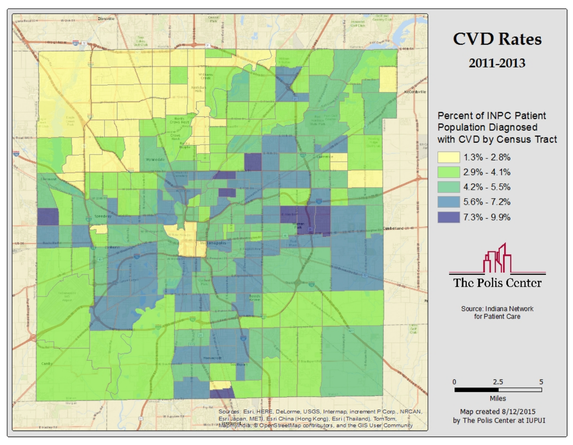
Percent of HIE population diagnosed with diabetes by health district



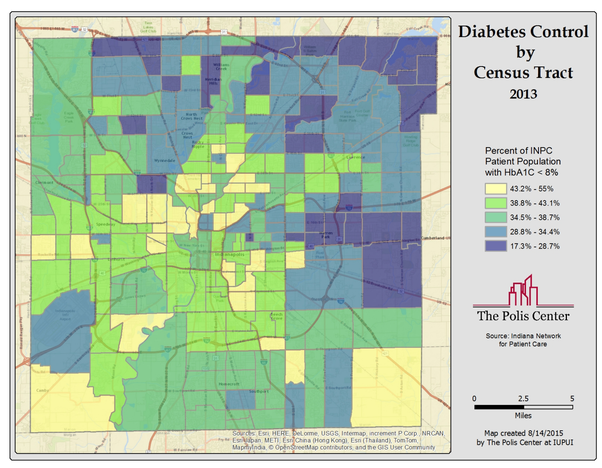
Percent of HIE population diagnosed with diabetes by neighborhood



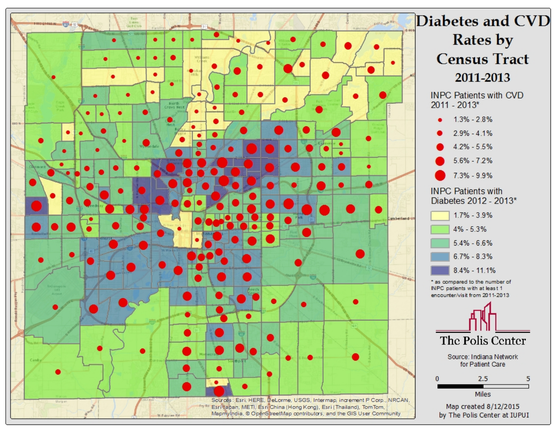
Percent of HIE population diagnosed with diabetes by census tract



Percent of HIE population diagnosed with cardiovascular disease (CVD) by census tract



Percent of patients with diabetes "under control" (HbA1c < 8%) by census tract



HIE population rates of diabetes and cardiovascular disease (CVD) together by census tract

Potential Sources of Bias

EHR data appear to have three potential biases:

1. EHR data represent the portion of the population which obtains health care services;
2. Linked EHR data may be biased based on the algorithm used to match patient records; and
3. The completeness of the data in the Indiana HIE varies by hospital/clinical system; for example, proportionally more data are available from low income patients given inclusion of safety net providers and local health department clinics.

Future Directions

We continue to explore ways to adjust rates and correct these biases so they do not overestimate burden of disease and poor care quality in inner-city neighborhoods due mainly to over-representativeness of low-income populations.

We are exploring adjustments based on:

- race;
- income;
- age, especially Medicare population;
- insurance, especially Medicaid and Medicare; and
- education.

Acknowledgements

PEDAL (Population EHR Data For Assessment At The Local Level) is funded by the Robert Wood Johnson Foundation (RWJF) Public Health Services and Systems Research (PHSSR) portfolio, which is managed by the National Coordinating Center for PHSSR.