Name:	Robert Miller
Affiliation:	University of Southern Maine; Maine Medical Research Institute
Email:	robert.miller@maine.edu
Presentation type (select one):	Poster

OHDSI GIS

Robert T. Miller[1], Andrew E. Williams, PhD[2], Bruce MacLeod, PhD[1,]
Vinton Valentine, PhD[1], Erika Ziller, PhD[3]
[1]University of Southern Maine, Portland, Maine; [2]Maine Medical Research Institute, Portland, Maine; [3]Muskie School of Public Service, Portland, Maine

Abstract

We are developing an approach to integrating Geographic Information Systems (GIS) data into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). We are also developing GIS data visualization tools that will display GIS data layers on top of privacy-preserving maps of patient and facility locations. This work is in the planning stage. To complete it, we will establish an OHDSI GIS workgroup that will propose changes to the CDM and collaborate on tool development. Thus far, we have outlined an approach and identified technical challenges to solve.

Introduction

The OHDSI mission is: "To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care." [1.] The data empowering this vision include clinical, claims, and administrative data on patients, providers, care and outcomes. Attributes of the places where patients live and receive care also affect their health. Distance from care, geographic indicators of the social determinants of health, environmental exposures and food landscapes are among the many types of GIS data that could increase the scope and power of models developed on the OHDSI platform. OHDSI community data holders possess information on populations across the US and around the world. This remarkable richness can be more fully leveraged by taking into account the important drivers of health and health care captured in GIS data.

While many other options exist and are not yet ruled out, R[2] allows for advanced GIS capabilities in an accessible way. R could allow us to push and pull information from the databases, run spatial analysis routines and plot maps on interactive web applications, all within the OHDSI guidelines.

Use cases

Our initial implementation will be driven by specific use cases from a regional research data infrastructure project funded by the Maine Cancer Foundation (Williams, PI). Both analyses are intended to impact cancer care and policy in Maine and may serve as a template for multi-site OHDSI-based studies. The first is focused on characterizing patients with stage 2 muscle-invasive bladder cancer who do not receive recommended chemotherapy prior to

cystectomy. The second is a predictive model of the cardiotoxic effects of anthracyclines in patients with breast cancer.

Obstacles

Two design aspects of the current CDM are limiting GIS functionality:

1) There is an absence of variables to store latitude and longitude data. The addition of two columns to the location table would enable the geolocation of each residence and care facility. Once we have the coordinates for each location, we can easily aggregate these into standardized regions or location-based cohorts.

2) The one-to-one relationship of person to location does not allow a patient to change residences over time. The location_id in the person table is only capable of storing a single residence for an individual. We need to be able to know what a person's residence was for a given date. In another system.[3], this problem is solved by introducing a residency table, which acts as a relational entity between person and location.

There are several situations to consider here:

- A person can change locations over a period of time
- A location can have multiple residents over a period of time
- A person can reside at multiple locations during the same period (e.g. vacation home)

A residency table would resolve the first two issues. The third, if we are concerned about it, could be resolved by adding a "primary residence" Boolean variable to the location table. Note in Figure 1: neither person nor location IDs need to be unique in the residency table.

residency_id	person_id	location_id	start_date	end_date
res1	John_Doe	location1	5/7/1985	-
res2	Jane_Doe	location1	5/7/1985	-
res3	Bob_Smith	location2	8/2/2011	7/29/2012
res4	Bob_Smith	location3	7/30/2012	-

Figure 1. Proposed residency table

Geographic data is often messy

Another hurdle of this project is that location data is often missing or erroneous. To be able to perform GIS analysis we need to have consistent data in the location table. To help solve this issue, as well as push for the OHDSI pursuit of data standardization, we plan to develop tools that do the following:

- Standardize the values of the location variables (e.g. stlouis vs. St. Louis)
- Geocode missing location data (e.g. zip from city and state)
- Geocode coordinate values into existing implementations

Discussion

Once the data is formatted and connected, the next step will be to develop tools that take advantage of the new capacities. Extensive discussion among the workgroup will be needed before decisions are made regarding the specifics of the functionality. For instance, are variables resulting from geospatial analysis calculated dynamically or statically? Do we group locations by cohorts or a new entity altogether?

Conclusion

Introducing GIS into OHDSI can expand analysis capabilities, generate better evidence and help promote better health. An OHDSI workgroup focused on specifying the minimally sufficient changes to the CDM needed to support GIS tool development will help make this possible.

References

1. OHDSI Mission, Vision & Values (2016).(http://www.ohdsi.org/who-we-are/mission-vision-values)
2. The R Project for Statistical Computing (2016).(https://www.r-project.org/)
3. OpenHDS (2016).(https://github.com/OpenHDS)