Leveraging Location Data in OMOP to Incorporate Area Deprivation Index

Xinzhuo Jiang, Maura Beaton, Jake Gillberg, Andrew Williams, Karthik Natarajan

Background

As the OHDSI community enhances its methods packages to run large scale observational studies, there are new data types that also can be incorporated to better understand diseases, specifically social determinants of health data. One such data is Area Deprivation Index (ADI). ADI is a composite measure of regional socioeconomic variables from US census block group data that have been shown to influence health outcomes such as economic inequality, opportunity structure, income, education and home ownership.[1,2] Based on these factors, a standardized ADI score is calculated to represent the level of socioeconomic deprivation within a given geographic area. The current study aims to assess the feasibility of incorporating ADI scores into a network study that aims to characterize breast cancer patients from two EHR data sites using the location table in the site's OMOP database.

Methods

This feasibility assessment was performed using datasets from Columbia University Irving Medical Center (CUIMC) and Tufts Medical Center (TMC) that have been transformed into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5.3. The CUIMC OMOP database is populated from EHR data from CUIMC and New York-Presbyterian hospital. It consists of 6.69 million unique patient records with data starting from 1989 to the present day. TMC OMOP contains data on approximately 1-million patients who received care beginning in 2006. This database contains EHR data fused with data from TMC's Commission on Cancer accredited tumor registry, its oncology EHR, and death data from the Massachusetts State Registry of Vital Statistics.

Newly diagnosed breast cancer patients were identified using a biopsy-based cohort definition to ensure all patients received their initial diagnosis at either CUIMC or TMC. This cohort consisted of all female patients who had received a breast cancer diagnosis within 90 days of a breast biopsy record. All patients were between 18 and 80 years old at the time of diagnosis and had observations available in their record for at least 90 days prior to the biopsy date. Patients that received a cancer diagnosis at least 30 days prior to the biopsy were excluded.

For each patient, an ADI ranking score was linked to their EHR data using address information in the OMOP location table (Figure 1). At CUIMC, location data was geocoded using ArcGIS Pro 2.8; while TMC patient addresses were geocoded using an open-source geocoding tool developed by OHDSI's Geographic Information System (GIS) working group that leverages PostGIS Tiger Geocoder functionality.[3] Once geocoded, ADI data was downloaded from Neighborhood Atlas, an open-source website developed to make ADI rankings available to researchers,[4] and joined to addresses using census block group FIPS codes, stored in TIGER shapefiles. The process generated an intermediate table containing the mappings between block group FIPS and location id. With the intermediate geocoded location table created, ADI rankings could then be assigned to each patient in a separate output table. Each patient was assigned a county-level ADI score that was determined by taking the median ADI ranking for all patients located within each county.

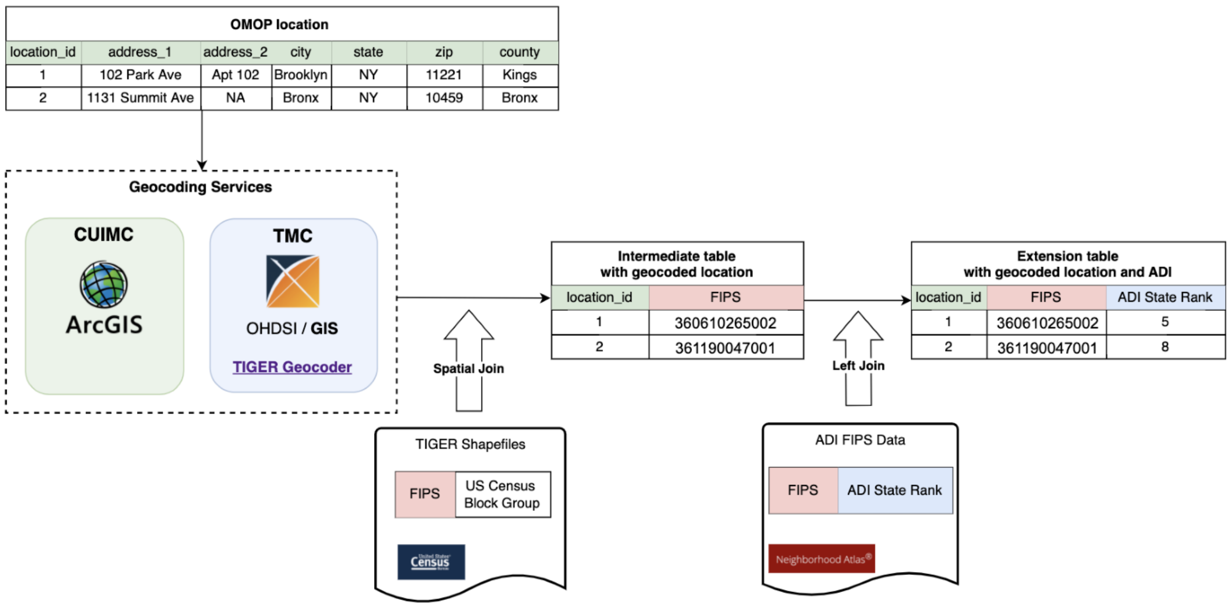

Figure 1. Linkage of ADI to OMOP location table workflow

Results

At CUIMC, 5,845 breast cancer patients were identified, and 392 patients were identified at TMC. Median county-level ADI scores, ranging from 1-10, are shown in Figure 2 and Figure 3. Lower ADI scores are shown in blue and indicate that a county has lower levels of socioeconomic deprivation based on the median ADI for patients living within that county. Higher ADI scores are shown in red and indicate a county has greater disadvantage.

Results show that many of the breast cancer patients at CUIMC and TMC reside in counties with lower ADI scores. However, a substantial portion of CUIMC patients come from Bronx County which has a higher county-level ADI score—and therefore greater socioeconomic deprivation. In addition, many of the patients who travel to CUIMC from upstate New York, and counties in New Jersey and Pennsylvania appear to reside in more disadvantaged counties. Similarly, a small portion of TMC patients reside in Bristol County and western counties in Massachusetts which also have greater socioeconomic deprivation.

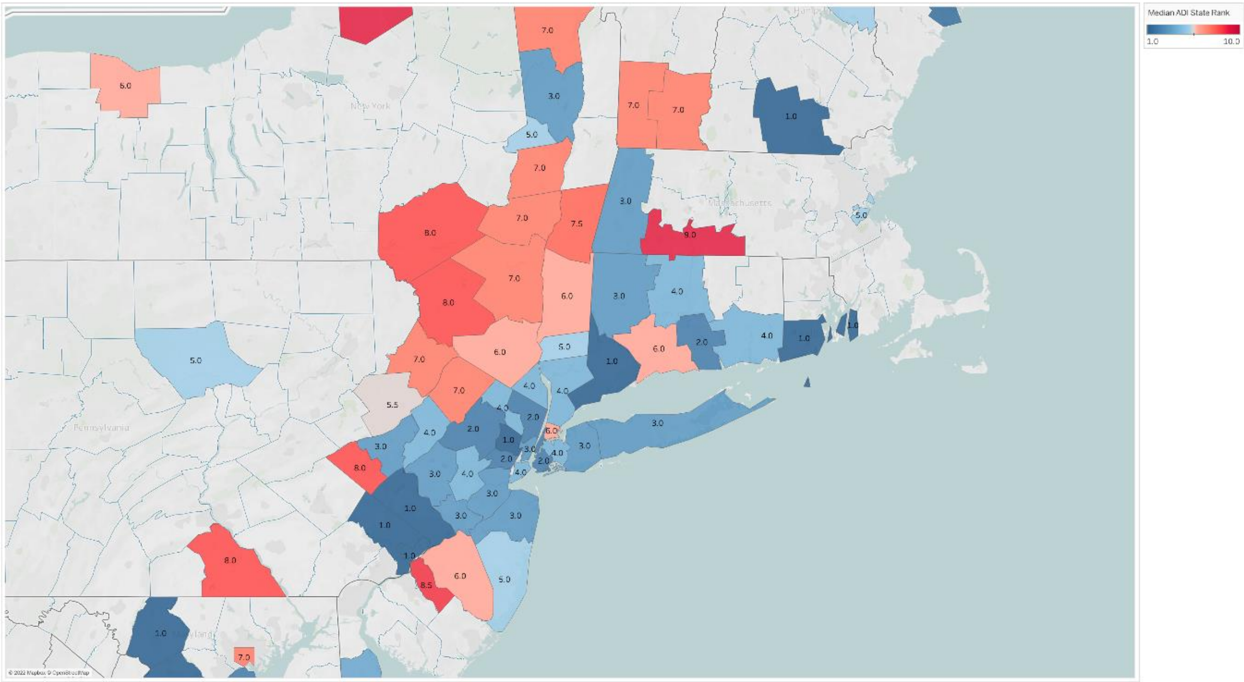

Figure 2. Distribution of CUIMC breast cancer patients, by county-level ADI

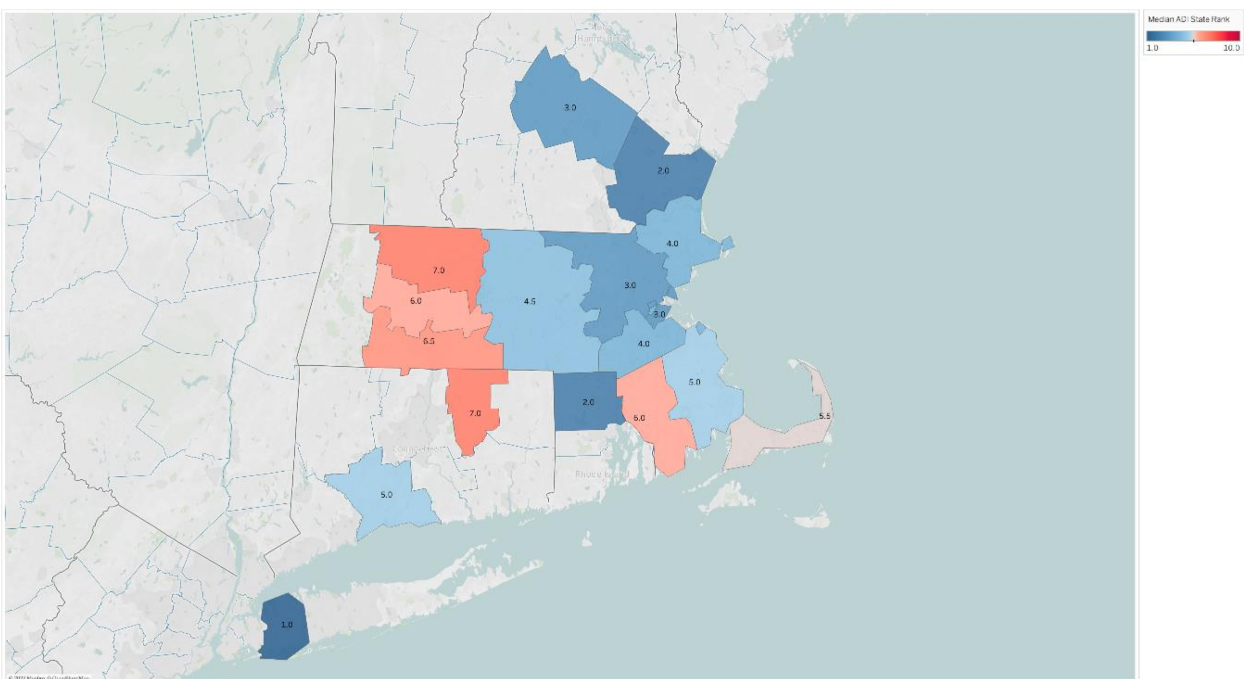

Figure 3. Distribution of TMC breast cancer patients, by county-level ADI

Conclusion

With this study, we have demonstrated that linking ADI scores to OMOP and stratifying cancer cohorts by ADI in a network study is feasible. By successfully running this feasibility study at CUIMC and TMC, we have shown that linking ADI to OMOP using both internal and external data is possible, representing an important first step towards including ADI as a covariate in future OHDSI network studies.

The next steps for this work will be to support other data sites to link ADI to their OMOP instance and run characterization studies using ADI across US sites in the OHDSI network. By following the framework outlined here, data partners can link their OMOP instance not only to ADI rankings, but to any other indices based on US census bureau data such as the Social Vulnerability Index (SVI)[5] or Multidimensional Deprivation Index (MDI)[6]. By supporting more data partners to harmonize their OMOP instance with ADI, SVI or MDI data we will ultimately enable multi-site studies that can investigate the degree to which health outcomes vary by socioeconomic factors.

References

1. Singh GK, Miller BA, Hankey BF, Feuer EJ, Pickle LW. Changing area socioeconomic patterns in U.S. cancer mortality, 1950-1998: Part I--All cancers among men. J Natl Cancer Inst. 2002 Jun 19;94(12):904–15.

2. Singh GK. Area Deprivation and Widening Inequalities in US Mortality, 1969–1998. Am J Public Health. 2003 Jul;93(7):1137–43.

3. OHDSI/GIS [Internet]. Observational Health Data Sciences and Informatics; 2022 [cited 2022 Mar 1]. Available from: https://github.com/OHDSI/GIS

4. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible — The neighborhood atlas. N Engl J Med. 2018 Jun 28;378(26):2456–8.

5. CDC/ATSDR's Social Vulnerability Index (SVI) [Internet]. 2022 [cited 2022 Mar 8]. Available from: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

6. Bureau UC. The Census Multidimensional Deprivation Index: Revised and Updated [Internet]. Census.gov. [cited 2022 Mar 8]. Available from: https://www.census.gov/library/working-papers/2021/demo/SEHSD-WP2021-03.html