Incorporating Social Drivers of Health Information into Health Economics and Outcomes Research: Neighborhood-Level Proxies versus Individual-Level Data

Karl M. Kilgore PhD, Christie Teigland PhD, Matthew McClellan, Zulkarnain Pulungan PhD, Barton Jones MS, Barth Kelly MS

Inovalon Insights, Bowie, MD, United States

Background

- Evidence shows that addressing Social Drivers of Health (SDOH) can improve health outcomes and reduce costs.[1,2]
- But access to data on patients' SDOH remains a barrier.
- Health plans, healthcare providers and life science companies commonly use aggregate proxies to assign SDOH, using census or survey data at the neighborhood level to impute characteristics down to the individuals who reside in that neighborhood.
- Recent advances in data-sharing technologies, such as tokenization of patient identified information (PII), has made it feasible to match patients actual SDOH to information on healthcare diagnosis, utilization, treatments and costs.
- Little is known about the relative advantages and disadvantages of using neighborhood proxies vs. actual patient-level measures of SDOH

Objective

- This study examines systematic differences between SDOH at the patient level vs. the neighborhood level, while holding constant the specific measures use to quantify SDOH at both levels of measurement.

Methods

Data Source

- The Inovalon MORE² closed claims database provided data on patients claims. MORE² is a primary-sourced medical and pharmacy claims database consisting of over 160 health plans and covering all major U.S. payer lines of business including 100% of Medicare Fee-for-Service, Commercial (31% of market), Medicare Advantage (29% of market), and Managed Medicaid (89% of market).
- SDOH data were derived from multiple, comprehensive individual and household databases and the American Community Survey of the US Census aggregated and sourced from Acxiom, Inc.[3]
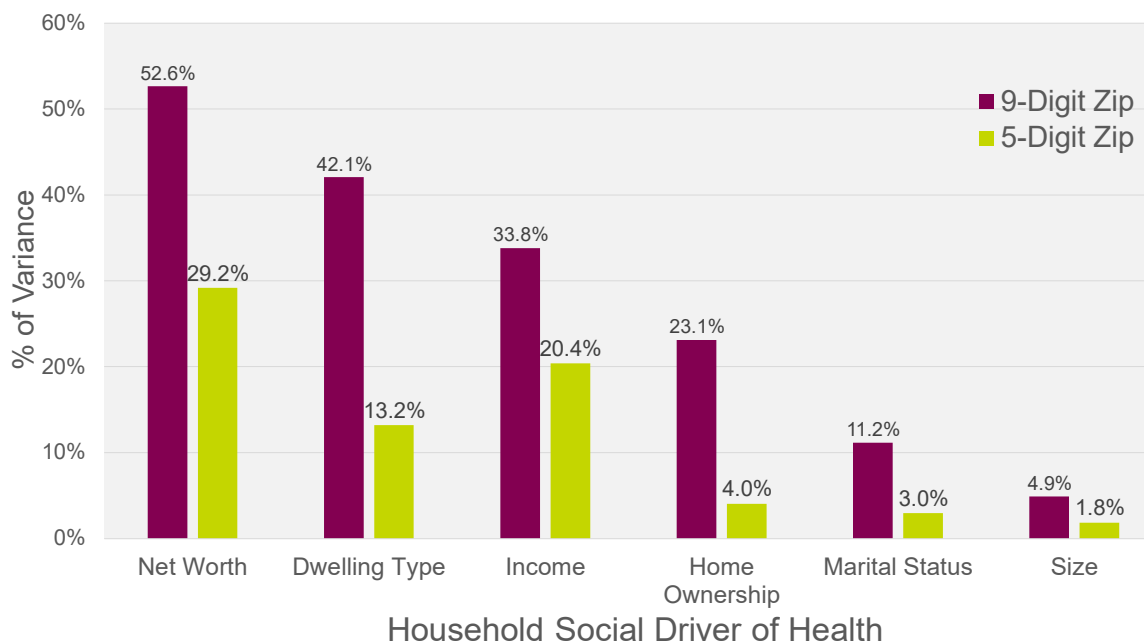
Patients

- A random sample of 100,000 patients in 2019-2021 who had a primary diagnosis of cardiovascular disease was drawn from MORE² and equally stratified by payer type: Medicare FFS, Medicare Advantage, Commercial and Managed Medicaid.
- These patients were matched to a set of Acxiom SDOH measures at 3 levels:
  1. 5-digit ZIP code-level: mean of 1,000 households per neighborhood
  2. 9-digit ZIP code-level: mean of 20 households per neighborhood
  3. Individual SDOH data using tokenized PII.

Analysis

- Each SDOH factor under study was analyzed by simple linear regression analysis where the response variable was the individual-level value and the explanatory variable was the neighborhood-level value.
- 9-digit ZIP and 5-digit ZIP neighborhoods were analyzed separately.
- R^2 statistics, which estimate the proportion of variance in the individual measures which was accounted for by variance in the neighborhood measures, were used as measures of the accuracy of neighborhood level values as proxies for individual level values (higher is better)

Accuracy of Aggregate Proxies: % of Variance in Individual-Level SDOH Measures that is Accounted for by Neighborhood-Level SDOH Measures at 2 Different Neighborhood Sizes

SDOH Factors Examined

- Household Size (number of persons in household)
- Marital Status (married vs. single)
- Home Ownership (own vs. rent)
- Dwelling Type (single- vs. multi-family)
- Household Income
- Household Net worth

Results

- 81% of patients were matched to their individual SDOH characteristics
- 9-digit ZIP R^2 values ranged from .05 (Household Size) to .53 (Net Worth)
- 5-digit ZIP R^2 values were approximately one-half of the corresponding smaller neighborhood values

Conclusions

- The accuracy of aggregate neighborhood characteristics as proxies for individual characteristics varied significantly by SDOH characteristic and size of neighborhood used.
- At best, approximately 50% of variation across individuals was left unexplained by neighborhood-level measures, and in some cases >90% was left unaccounted for.
- Design of effective interventions to address social inequities requires accurate data.
- Recent enhancements in data availability and matching at the individual patient level offer healthcare stakeholders improved information to identify and address health disparities.

References

1. Kilgore KM, McClellan M, Teigland C, Pulungan Z. Using aggregate data to proxy individual-level characteristics in health services research: 9-digit zip code vs. Census Block Group. Value in Health 2018;21:S218-9. Poster presentation PRM50 at the ISPOR 23rd Annual International Meeting. Baltimore MD; 2018.
2. Danos D, Leonardy C, Gilliland A, et al. Increased risk of hepatocellular carcinoma associated with neighborhood concentrated disadvantage. Front Oncol. 2018;8;375.
3. Acxiom Corporation (2022). ZIP+4 InfoBase® Geo Files: demographic, financial and property; Market Indices ACS. Available at: https://www.acxiom.com. Accessed May 1, 2025.