Innovative methods for integrating social determinants of health data with administrative claims to facilitate health equity research

Ivy Tonnu-Mihara, PharmD, MS, Cristian Rojas-Lazic, MA, Michael Mack, MS, Biruk Eshete, MS, Judith J. Stephenson, SM, Michael Grabner, PhD

Carelon Research, Inc., Wilmington, DE, USA

Background

- Social Determinants of Health (SDoH) - defined as the conditions where people are born, live, learn, work, play, worship, and age that are fundamental social and structural factors - affect a wide range of health, functioning, and quality-of-life outcomes and risks.[1] One study estimated that, on average, medical care impacts only 20 percent of health outcomes while SDoH affect as much as 50 percent.[2]
- While published evidence on the impact of SDoH on healthcare delivery and health outcomes is growing, a wide variety of frameworks and definitions are being used, leading to limited reproducibility of estimated effects.
- Administrative claims are an important source of real-world evidence. Linking claims data to sociodemographic and SDoH data in population-based research can improve our methodology and understanding of the impact of SDoH on health outcomes.
- This study presents an innovative way to link individual- and area-level sociodemographic and SDoH indicators to claims data from a large United States (US) managed care database to support health plan operations.

Objectives

- To describe how individual-level race and ethnicity and area-level SDoH indicators are defined and identified for health plan members with administrative claims in the Healthcare Integrated Research Database (HIRD®).
- To compare the distributions of race and ethnicity and SDoH indicators among health plan members with Commercial, Medicare Advantage, and Medicaid health insurance.
- To evaluate the utilization of telehealth services across race and ethnicity and SDoH indicators, stratified by insurance type.

Methods

- The HIRD contains medical and pharmacy claims from a large, national US payer starting in 2006. Overall, the HIRD contains ~64 million commercial, ~ 3 million Medicare Advantage, and ~27 million Medicaid lives with medical and pharmacy enrollment.
- The HIRD includes individual-level race and ethnicity from various sources, including enrollment files, electronic health records (EHRs), self-attestations, and validated algorithmic imputations.
- The HIRD also contains members' nine-digit zip codes and State-County Federal Information Processing Standard (FIPS) codes that enable linkage to US Census block groups and tracts.
- SDoH data from the American Community Survey[3] (at the Census block group level), the Food Access Research Atlas[4] (at the Census tract level), and the National Center for Health Statistics[5] (at the Census tract level) were used to select area-level SDoH measures. Data are reported as missing for members who lack sufficient information for linking. Only a subset of available SDoH measures from these sources were evaluated in this project.
- Members who were currently active (enrolled) in the HIRD with medical and pharmacy coverage as of November 30, 2022, were categorized into cohorts identified by their type of health insurance as Commercial (Comm), Medicare Advantage (MCare), and Medicaid (MCaid).
- Telehealth (TH) service utilization between January 1, 2022, and November 30, 2022, for those with continuous enrollment during the timeframe was determined using a previously published algorithm.[6] The percentage of patients with ≥1 TH claim was cross-tabulated with the race and ethnicity and SDoH measures.
- Descriptions and comparisons were performed among the study cohorts. Standardized mean differences (SMD) were calculated.[7] When appropriate, comparisons to US national data were made.

Results

Table 1. Study cohort demographic characteristics

	Comm	MCare	MCaid
Total member-level sample size	13,560,649	1,031,097	7,876,591
Total number of Census block groups[1]	212,121	74,395	108,051
Total number of Census tracts[2]	71,974	32,299	42,707
Age[3], mean (SD)	36.3 (19.07)	70.6 (10.94)	23.1 (17.93)
Female gender, %	49.9%	56.8%	54.1%
Region of residence[4], %			
Northeast	15.9%	16.1%	10.4%
Midwest	22.5%	52.1%	20.2%
South	36.8%	26.3%	57.7%
West	24.6%	5.6%	10.8%
Urbanicity[5], %			
Urban	57.4%	52.2%	57.2%
Suburban	28.4%	23.8%	28.2%
Rural	14.1%	24.1%	14.6%

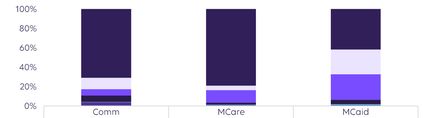
1. Census block groups are geographic blocks uniquely numbered by the US Census Bureau and nested within the Census tracts (https://www.census.gov). In total, the US has 220,333 Census block groups. **2.** Census tracts are small statistical subdivisions of a county or statistically equivalent in the decennial Census (https://www.census.gov). **3.** On 11/30/2022. **4.** Comm and MCaid cohorts had members with missing region of residence. **5.** Urbanicity is determined by linking the members' zip codes to US data from the National Center for Health Statistics (https://www.cdc.gov/nchs/data_access/urban_rural.htm).

- As expected, health plan members' demographic characteristics differed substantially by type of insurance.

Results

Individual-level race and ethnicity

Figure 1. Overall race and ethnicity distribution among members with non-missing data

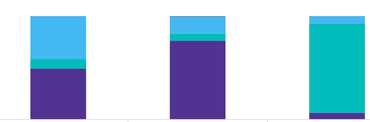


	Comm n = 10,842,013	MCare n = 924,847	MCaid n = 7,571,327
White, NH	70.9%	79.0%	41.8%
Hispanic or Latino of any race	11.8%	4.6%	25.5%
Black or African American, NH	6.6%	13.0%	26.3%
Asian, NH	6.5%	2.1%	4.5%
American Indian or Alaska Native, NH	0.3%	0.4%	0.7%
Native Hawaiian or Other Pacific Islander, NH	0.01%	0.01%	0.3%
Other race, NH	3.8%	0.9%	0.9%

NH = Not Hispanic or Latino

- Non-missing race and ethnicity data were available for ~80%, ~90%, and ~96% of members from the Comm, MCare, and MCaid cohorts, respectively.
- The race and ethnicity distributions differed by insurance type.
  - The largest differences between the Comm and MCare cohorts were between the Hispanic or Latino populations (SMD 0.265) and the Black or African American populations (SMD 0.216).
  - The largest differences between the Comm and MCaid cohorts were between the White (SMD 0.616), the Hispanic or Latino (SMD 0.551), and the Black or African American (SMD 0.358) populations.
- The reported 2020 Census Diversity Index (DI) for the US was 61.1%.[8] For the HIRD health insurance cohorts, the DI was greatest for the MCaid cohort (68.9%), followed by the Comm cohort (47.3%), and lowest for the MCare cohort (35.6%).

Figure 2. Percentage of members identified by race and ethnicity data source within the HIRD

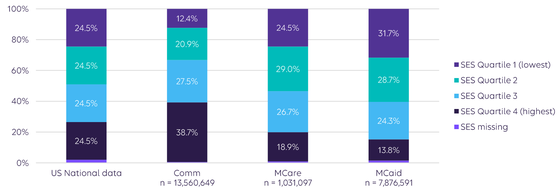


	Comm n = 10,842,013	MCare n = 924,847	MCaid n = 7,571,327
Self-attestations	0.005%	0.4%	0.02%
Algorithmic imputations	41.4%	16.3%	7.5%
Enrollment	9.3%	7.1%	85.9%
Electronic health records	49.2%	76.2%	6.5%

- Race and ethnicity were identified primarily through EHR data for the MCare (~76%) and Comm (~49%) cohorts, while the MCaid cohort had the highest proportion identified through enrollment data (~86%).

Area-level SDoH indicators

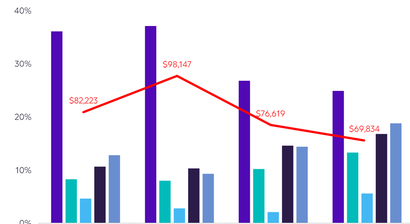
Figure 3. ACS- SES index[2] by quartile



The SES index is a composite measure based on seven SDoH indicators. A score of 4 indicates the member lived in a Census block group in the top 25% of SES and score of 1 indicates the member lived in a Census block group in the bottom 25% of SES using Census block groups from the 2019 ACS as the reference basis for calculation. Members unable to be linked to ACS-SDoH data or where at least one of the seven indicators is missing were categorized as missing. The seven SDoH indicators are 1) Education at least 4 years college, 2) Education less than high school, 3) Standardized median family income (range from 0 to 100), 4) Income below 100% of the federal poverty level, 5) Household with more than 1 person per room (crowding), 6) Unemployment and 7) Standardized median home value (range from 0 to 100).

- The Comm cohort had the highest percentage of members living in the highest SES block group category (SMD 0.209) while the MCaid cohort had the highest percentage of members living in the lowest SES block group category (SMD 0.170).

Figure 4. Proportion of members with selected SDoH indicators in each insurance cohort



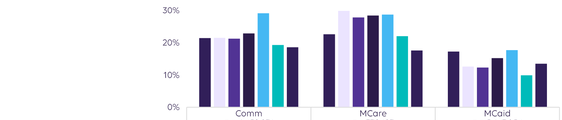
Median Family Income $, 2019	US National data	Comm n = 13,560,649	MCare n = 1,031,097	MCaid n = 7,876,591
Education at least 4 years college	36.1%	37.1%	26.8%	24.9%
Education less than high school	8.3%	8.0%	10.2%	13.3%
English spoken less than well	4.6%	2.8%	2.1%	5.6%
Income below 100% of the federal poverty level	10.7%	10.3%	14.6%	16.8%
Low income and low access (to food)*	12.8%	9.3%	14.4%	18.8%

*Low income and low access (to food) indicator was derived from the US Food Access Research Atlas (FARA). The remaining indicators were derived from the American Community Survey (ACS).

- SDoH indicators in the Comm cohort were close to US national data, while differences were noticeable for the MCare and MCaid cohorts.

Evaluation of TH utilization

Figure 5. Proportion of health plan members with at least 1 TH claim by selected race and ethnicity

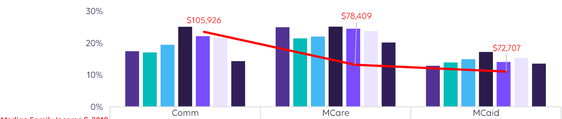


	Comm n = 6,458,931	MCare n = 738,193	MCaid n = 5,647,954
White, NH	21.4%	22.6%	17.3%
Hispanic or Latino of any race	21.5%	29.8%	12.6%
Black or African American, NH	21.2%	27.8%	12.3%
Asian, NH	22.8%	28.4%	15.2%
American Indian or Alaska Native, NH*	29.1%	28.7%	17.7%
Native Hawaiian or Other Pacific Islander, NH*	19.3%	22.0%	9.9%
Other race, NH	18.6%	17.6%	13.5%

Percentages were calculated using total members with at least 1 TH claim as numerators and total members with and without TH services as denominators within each race and ethnicity group and within each insurance type cohort. For example, among all commercially insured White, NH members who were continuously enrolled from January 1, 2022, to November 30, 2022, 21.4% had at least 1 TH claim in 2022.
NH = Not Hispanic or Latino
*Small sample size

- Among the subgroup of members included in the TH evaluation (who had continuous enrollment in 2022), ~87% of Comm, ~93% of MCare and ~83% of MCaid members had race and ethnicity data (data not shown).
- Overall, ~24% of Comm, ~24% of MCare and ~15% of MCaid members utilized TH services in 2022 (data not shown).
- The proportions of members utilizing TH services differed both within and across insurance type cohorts.

Figure 6. Proportion of health plan members with at least 1 TH claim by selected SDoH indicators



Median Family Income $, 2019	Comm n = 7,400,648	MCare n = 793,637	MCaid n = 6,787,724
SES Quartile 1 (lowest)	17.5%	25.0%	12.9%
SES Quartile 2	17.1%	21.5%	13.9%
SES Quartile 3	19.5%	22.1%	14.9%
SES Quartile 4 (highest)	25.1%	25.1%	17.2%
Urban (NCHS)	22.2%	24.5%	14.1%
Suburban (NHCS)	21.7%	23.8%	15.4%
Rural (NCHS)	14.4%	20.2%	13.5%

Percentages were calculated using total members with at least 1 TH claim as numerators and total members with and without TH services as denominators within each SDoH indicator and within each insurance type cohort. For example, among all commercially insured members who lived in SES category 1 block groups and were continuously enrolled from January 1, 2022, to November 30, 2022, 17.5% had at least 1 TH claim in 2022.

- Among Comm and MCaid members, use of TH increased with higher SES status.
- Urbanicity affected TH use in the Comm and MCare cohorts, where rural residents had the lowest rate of TH use.

Limitations

- The proportion of race and ethnicity identified by each available data source differed noticeably by type of insurance. Algorithmic identification remains important for determining race and ethnicity of commercially insured members; a validation against member self-attestations has been conducted in this cohort (publication pending).
- Health plan members identified by administrative claims in the HIRD included in this study were geographically diverse. However, not every US block group was represented, and the proportions reported did not account for the potential impact from missing block groups.
- Sociodemographic and SDoH indicators and health plan membership may change over time. The values provided in this study are cross-sectional using the most recent cohort of currently active members and SDoH data available in the HIRD.
- There are some missing data, primarily for race and ethnicity in the commercially insured member cohort. Additional efforts to collect and standardize these data are needed.

Conclusions

- We successfully integrated SDoH data from a variety of sources with administrative claims data in the HIRD.
- Depending on the specific indicators, the sociodemographic and SDoH indicators differed by type of insurance coverage and were associated with differences in telehealth utilization.
- The integrated data can support a wide array of health equity improvement efforts and other health plan operations.

Funding and disclosures: No funding was received for the conduct of this study. All authors are employees of Carelon Research, which conducts health services research with both internal and external funding, including a variety of private and public entities. MG and JJS are shareholders of Elevance Health.

References

1. U.S. Department of Health and Human Services. Healthy People 2030, Social Determinants of Health. https://health.gov/healthypeople/objectives-and-data/social-determinants-health **2.** Hood CM, Gennuso KP, Swain GR, et al. American Journal of Preventive Medicine. February 2016; 50(2):129- 135. **3.** This project used 2019 5-year estimates from the ACS. For more information on the ACS, please refer to https://www.census.gov/programs-surveys/acs/data.html. **4.** This project used 2019 estimates from the FARA. For more information on the FARA, please refer to https://www.ers.usda.gov/data-products/food-access-research-atlas/about-the-atlas/. **5.** This project used 2013 estimates from the NCHS. For more information on the NCHS urban-rural classification scheme, please refer to https://www.cdc.gov/nchs/data_access/urban_rural.htm. **6.** Avula R, Grabner M, Eshete B, Stanek E, Willey V. Characterizing Telehealth Utilization from Administrative Claims in the US using a Standardized Definition Poster presentation at the ISPOR Annual Meeting, National Harbor, May 2022. **7.** Yang D, Dalton J. SAS Global Forum 2012. Paper 335-2012. https://support.sas.com/resources/papers/proceedings12/335-2012.pdf. [Accessed 12 April 2023] 2. **8.** https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html. [Accessed 11 April 2023]. The DI ranges from 0-100%, and higher numbers represent higher levels of diversity. **9.** Bonito AJ, Bann C, Eicheldinger C, Carpenter L. Creation of New Race-Ethnicity Codes and Socioeconomic Status (SES) Indicators for Medicare Beneficiaries. Final Report, Sub-Task 2. (Prepared by RTI International for the Centers for Medicare and Medicaid Services through an interagency agreement with the Agency for Healthcare Research and Policy, under Contract No. 500-00-0024, Task No. 21) AHRQ Publication No. 08-0029-EF. Rockville, MD, Agency for Healthcare Research and Quality. January 2008.

Poster presented at ISPOR, May 7-10, 2023, Boston, MA, USA.

Ivy Tonnu-Mihara | ivy.tonnu-mihara@carelon.com