Using Aggregate Data to Proxy Individual-Level Socioeconomic Characteristics in Research on Medication Adherence: 9-Digit ZIP Code vs. Census Block Group

Presented at AMCP Nexus 2018, Oct 22-25, 2018, Orlando, FL
Karl M. Kilgore, PhD; Christie Teigland, PhD; Zulkarnain Pulungan, PhD

Avalere Health | An Inovalon Company, 1350 Connecticut Avenue NW, Suite 900, Washington, DC 20036
T 202.207.1300 | F 202.467.4455 | avalere.com

Visit us online at www.avalere.com
@AvalereHealth | Scan QR code to view poster as a pdf.

Background

- Socioeconomic characteristics of patients, or Social Determinants of Health (SDH), are receiving increased emphasis in research on healthcare processes and outcomes, including medication adherence (MA)
- The use of aggregate proxy data, where SDH characteristics of residential areas are imputed to the individual and used to risk adjust health outcomes, is common in research on social risk factors
- Extant studies on biases introduced by aggregate proxies yield inconsistent results, and may suffer from confounds due to differences between aggregate vs. individual data collection methods[1,2,3]
- This study compares two models relating individual patient MA outcomes to a set of similarly defined SDH proxy variables calculated from two different levels of aggregation: 9-digit ZIP Code (ZIP9) vs. Census Block Group (CBG)

The Sides of the Aggregate Proxy Debate

Aggregate Proxies **Under-Estimate** True Effects	Aggregate Proxies **Over-Estimate** True Effects	The Impact of Aggregate Proxies is **Unpredictable**
Because publicly available data are aggregated over large geographic areas which contain disparate sub-populations, any true differences associated with social risk factors are "washed out"	Because the aggregate variable is a characteristic of the geographic area and is measuring a broader construct, e.g., higher-income individuals are more likely to be employed and own their own home	Because the surveys from which proxies are calculated utilize different methodologies, not only from each other, but also from the source of individual-level variables

Objective

To compare proxies for individual level SDH drawn from two different neighborhood sizes, CBG vs. ZIP9, in modeling healthcare outcomes, specifically 3 standard MA measures

Methods

Study Population and Data Sources

- 1.7 million Medicare Advantage beneficiaries, continuously enrolled in 2015, extracted from a national claims database
- Based on their address, beneficiaries were matched to household SDH variables from:
  - A commercial market research database drawn from multiple, comprehensive individual and household databases (e.g. public records, buying activity, financial behavior) and aggregated at the 9-digit ZIP code level (> 30 million discrete observations)[4]
  - The American Community Survey, aggregated at the CBG level (approximately 220,000 areas)
- Data represented all 50 states and the District of Columbia, and were drawn from 30 Medicare Advantage Organizations (MAOs)

Variables

- Explanatory (predictor) variables: SDH variables common to both aggregate proxy data sources were income, education level, marital status, home ownership, and Hispanic ethnicity
- Response (outcome) variables: Individual-level binary indicators of whether or not the member qualified for the numerator of 3 PQA MA measures, also used in CMS Star Ratings for Part D Health plans: Medication Adherence for Hypertension (MA-H), Diabetes Medications (MA-D), and Cholesterol (MA-C)[5]

Statistical Analysis

- Generalized linear mixed model regression analyses were used to model the relationships between the response variables and the predictor variables (main effects only)
- A random effect for Medicare Advantage organization was included to account for the non-independence of data from members of the same health plan
- The 3 binary response variables were analyzed with separate models, using a logistic link function
- For each response variable, the set of ZIP9 and CBG predictors were analyzed separately, and their parameters estimates, standard errors and significance levels were compared

Results

- Full sample: N=1,709,146; 57% female; age<65: 17%, age≥80: 20%
- MA-C sample: N=709,455; MA-C ≥ 80%: 50%
- MA-D sample: N=242,952; MA-D ≥ 80%: 59%
- MA-H sample: N=671,998; MA-H ≥ 80%: 55%

Figure 1. Comparison of Predictor Variable Distributions and Correlation between ZIP9 and CBG

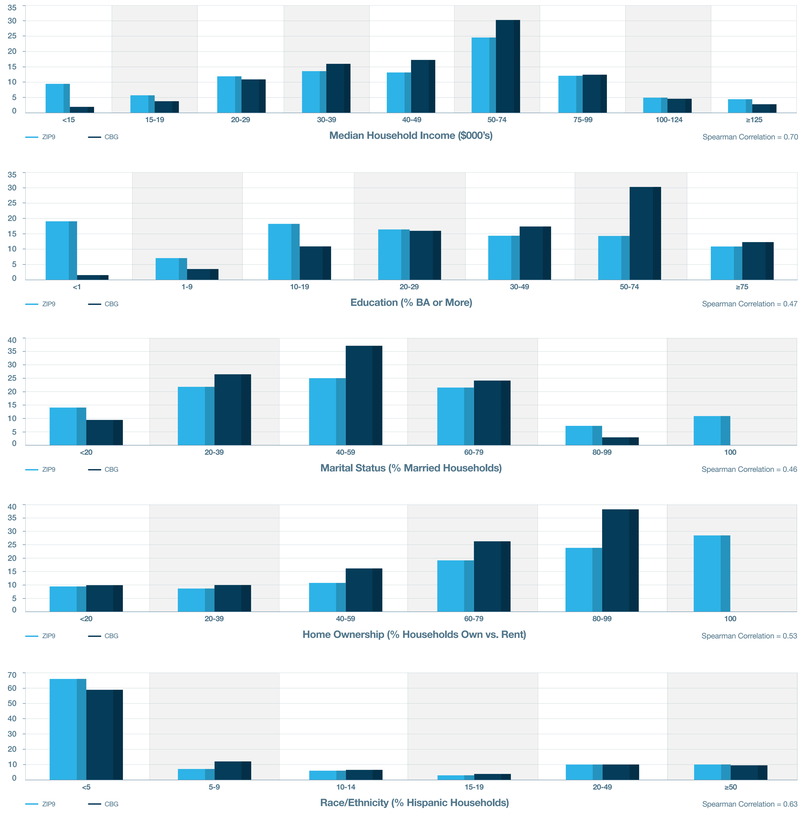

Table 1. Comparison of Modeled Parameter Estimates, Standard Errors (SE) and P-values: MA-C
(Highlighted rows indicate a significant effect for ZIP9 but not CBG)

		ZIP9			CBG		
	Level	Parameter	SE	p-value	Parameter	SE	p-value
Median Household Income	<15*						
	15-19	**0.027**	**0.013**	**0.031**	**0.001**	**0.021**	**0.979**
	20-29	0.019	0.011	0.073	-0.029	0.019	0.123
	30-39	0.041	0.011	0.000	-0.012	0.019	0.533
	40-49	**0.064**	**0.011**	**<.0001**	**-0.017**	**0.020**	**0.385**
	50-74	**0.053**	**0.011**	**<.0001**	**-0.026**	**0.021**	**0.213**
	75-99	**0.059**	**0.013**	**<.0001**	**-0.037**	**0.023**	**0.108**
	100-124	**0.067**	**0.016**	**<.0001**	**-0.033**	**0.026**	**0.209**
	≥125	0.026	0.017	0.123	-0.050	0.031	0.109
Education (% BA or More)	<1						
	1-9	-0.020	0.011	0.081	0.016	0.026	0.548
	10-19	0.019	0.008	0.022	0.067	0.026	0.010
	20-29	0.030	0.009	0.001	0.080	0.027	0.003
	30-49	0.034	0.009	0.000	0.103	0.027	0.000
	50-74	0.042	0.009	<.0001	0.118	0.028	<.0001
	≥75	0.064	0.010	<.0001	0.119	0.038	0.002
Marital Status (% Married Households)	≤20						
	20-39	0.042	0.009	<.0001	0.061	0.010	<.0001
	40-59	0.089	0.010	<.0001	0.117	0.012	<.0001
	60-79	0.105	0.011	<.0001	0.129	0.014	<.0001
	80-99	0.111	0.014	<.0001	0.102	0.022	<.0001
	100	**0.116**	**0.012**	**<.0001**	**n/a**	**n/a**	**n/a**
Home Ownership (% Households Own vs. Rent)	≤20						
	20-39	0.008	0.011	0.505	-0.039	0.012	0.001
	40-59	**-0.025**	**0.012**	**0.031**	**-0.011**	**0.012**	**0.391**
	60-79	0.007	0.011	0.565	0.002	0.013	0.900
	80-99	0.018	0.012	0.134	0.021	0.014	0.141
	100	**0.035**	**0.012**	**0.004**	**n/a**	**n/a**	**n/a**
Race/Ethnicity (% Hispanic)	<5						
	5-9	**0.024**	**0.011**	**0.023**	**0.006**	**0.008**	**0.456**
	10-14	0.005	0.011	0.675	0.009	0.011	0.436
	15-19	0.010	0.016	0.539	0.025	0.014	0.075
	20-49	**-0.021**	**0.009**	**0.027**	**-0.014**	**0.010**	**0.159**
	≥50	-0.085	0.010	<.0001	-0.062	0.012	<.0001

*The first level of each variable was the reference level

Table 2. Comparison of Modeled Parameter Estimates, Standard Errors (SE) and P-values: MA-H
(Highlighted rows indicate a significant effect for ZIP9 but not CBG)

		ZIP9			CBG		
	Level	Parameter	SE	p-value	Parameter	SE	p-value
Median Household Income	<15*						
	15-19	0.003	0.013	0.844	-0.041	0.021	0.055
	20-29	0.005	0.011	0.649	-0.061	0.019	0.002
	30-39	0.034	0.011	0.003	-0.051	0.020	0.010
	40-49	0.043	0.012	0.000	-0.048	0.020	0.019
	50-74	0.061	0.011	<.0001	-0.054	0.021	0.011
	75-99	0.075	0.013	<.0001	-0.064	0.023	0.006
	100-124	0.079	0.017	<.0001	-0.060	0.027	0.028
	≥125	0.076	0.018	<.0001	-0.066	0.033	0.048
Education (% BA or More)	<1						
	1-9	0.005	0.012	0.692	-0.014	0.027	0.588
	10-19	**0.018**	**0.009**	**0.037**	**0.047**	**0.027**	**0.077**
	20-29	0.047	0.009	<.0001	0.076	0.027	0.005
	30-49	0.042	0.010	<.0001	0.125	0.027	<.0001
	50-74	0.052	0.010	<.0001	0.146	0.029	<.0001
	≥75	0.078	0.011	<.0001	0.136	0.040	0.001
Marital Status (% Married Households)	≤20						
	20-39	0.037	0.009	<.0001	0.059	0.011	<.0001
	40-59	0.062	0.010	<.0001	0.113	0.012	<.0001
	60-79	0.072	0.011	<.0001	0.112	0.014	<.0001
	80-99	0.072	0.014	<.0001	0.088	0.023	0.000
	100	**0.061**	**0.013**	**<.0001**	**n/a**	**n/a**	**n/a**
Home Ownership (% Households Own vs. Rent)	≤20						
	20-39	0.026	0.012	0.029	-0.035	0.012	0.004
	40-59	-0.006	0.012	0.587	-0.013	0.013	0.310
	60-79	**0.031**	**0.012**	**0.010**	**0.000**	**0.014**	**0.985**
	80-99	**0.053**	**0.013**	**<.0001**	**0.022**	**0.015**	**0.127**
	100	**0.066**	**0.012**	**<.0001**	**n/a**	**n/a**	**n/a**
Race/Ethnicity (% Hispanic)	<5						
	5-9	0.003	0.011	0.786	0.013	0.009	0.137
	10-14	0.032	0.012	0.007	0.044	0.012	0.000
	15-19	0.025	0.017	0.132	0.040	0.015	0.006
	20-49	0.045	0.010	<.0001	0.051	0.011	<.0001
	≥50	0.133	0.010	<.0001	0.173	0.012	<.0001

*The first level of each variable was the reference level

Key Findings

- SDH proxy variables aggregated at two different neighborhood sizes (ZIP9 vs. CBG), and re-coded to have identical levels, were highly correlated
  - However, ZIP9 distributions tended to have slightly higher frequencies in the tails of the distribution
- SDH variables were significantly predictive of MA for all 3 measures, regardless of the level of aggregation
  - However, predictors at the ZIP9 level yielded statistically significant parameter estimates nearly twice as often as CBG level predictors
- Results for the MA-D are not shown due to space constraints, but are very similar to MA-C and have been presented previously[6]

Discussion

- This study systematically compared two different aggregation levels of SDH proxy variables, while holding constant the patient sample, the conceptual definition and coding systems of the predictor variables, as well as the outcome measures modeled
- The results suggest that finer granularity groupings (9-digit ZIP code) of SDH can provide a more precise measurement of the sociodemographic differences which exist within the population and their association with health outcomes

Conclusion

- SDH are highly associated with medication adherence
- Averaging population risk factors across large groups of individuals and assigning to individuals may mask highly relevant associations of those factors with MA measures, or other outcomes being measured
- This greater variation found using more granular neighborhood data may be meaningful and improve the predictability of multivariate models such as risk adjusted quality measures or models predicting patient cost or utilization

References

1. Geronimus AT, Bound J, Neidert LJ. On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. J Am Stat Assoc 1996; 91: 529-537
2. Soobader M, LeGrew FB, Hadden W, Maury B. Using aggregate geographic data to proxy individual socioeconomic status: does size matter? Am J Public Health 2001; 91: 632-636
3. Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, Posner S. Socioeconomic status in health research: one size does not fit all. JAMA 2005; 294: 2879-2888
4. Acxiom Corporation (2015). ZIP+4 InfoBase® Geo Files: Demographic, Financial and Property. Sept 2015 release. www.acxiom.com
5. Centers for Medicare and Medicaid Services. Medicare 2015 Part C & D Star Ratings Technical Notes. CMS, 2015
6. Kilgore KM, McClellan M, Teigland C, Pulungan Z. Using Aggregate Data to Proxy Individual-Level Characteristics in Health Services Research: 9-Digit Zipcode vs. Census Block Group. Value in Health 2018; 21: S218-9. (Poster presentation PRM50 at the ISPOR 23rd Annual International Meeting, Baltimore MD, May, 2018.)