

SDOH in Health Care Predictive Models

Predictive analytics are used in primary care to direct health care resources to patients with the most need.

Social determinants of health (SDOH) are important features in predictive models for health care utilization.¹⁻⁴

- Poorly measured in administrative claims data⁵

Area-level SDOH can be substitutes for unavailable individual-level indicators.⁶⁻⁷

- Importance of granularity in model performance and utility is unclear

Research Question: Does enhancing the granularity of SDOH predictors from ZIP code tabulation area (ZCTA) to Census Tract strengthen an existing predictive model for avoidable hospitalizations (AH)?⁸

- Model used by primary care practices participating in the Maryland Primary Care Program (MDPCP)

Methods

Population Studied:

Person-month panel data set spanning 35 months

- **465,749 Maryland Medicare fee-for-service beneficiaries** who received primary care at an MDPCP-affiliated practice
- 59.5% female
- 69.8% White; 22.7% Black
- 14.9% dually eligible for Medicaid

Study Design:

Six discrete time survival models estimating risk for an AH – SAS (v 9.4)

- Different combinations of utilization and SDOH risk factors
- **144 claims-based features** indexing demographics, medical history (diagnoses, prescriptions, procedures, utilization)
- **37 SDOH features** from 11 publicly available sources (e.g., American Community Survey) – Census Tract and ZCTA versions

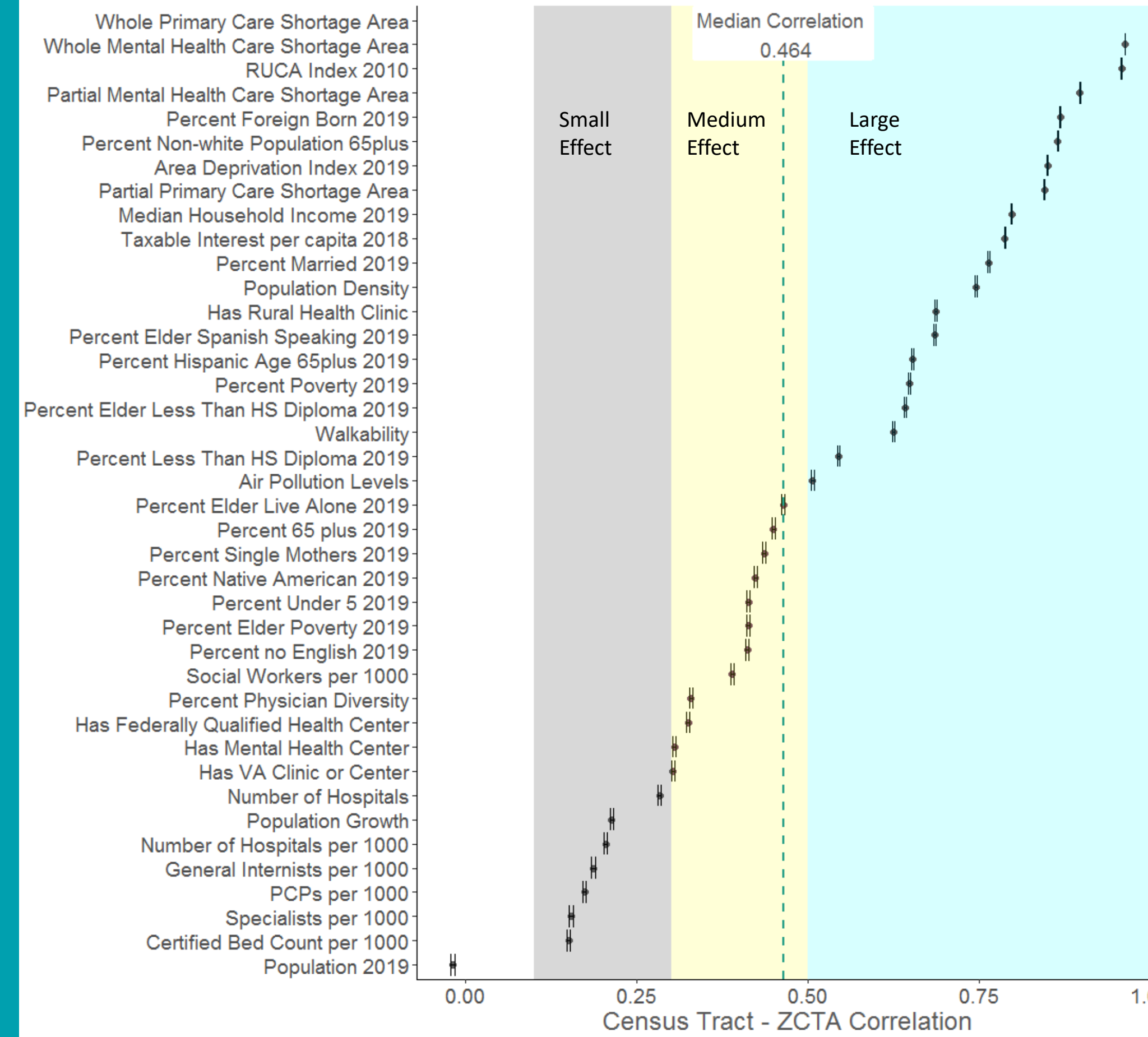
Outcomes:

Calculated agreement between Census Tract and ZCTA indicators

Compared fit, performance, and interpretation across models:

- Akaike Information Criterion (AIC)
- C-statistic
- Gini index
- Percentage of AH events incurred by top 10% riskiest beneficiaries
- Salient SDOH predictors

Variability in Agreement between ZCTA and Census Tract Measures



Conclusions & Policy Implications

Making risk factors more granular did not dramatically improve fit or predictive performance when estimating risk for AH.

Different SDOH variables were retained depending on the granularity level, which does influence model interpretation.

- Because Census Tract measures are likely a better representation of a person's proximal environment, model interpretation may be more meaningful.

Differences in interpretation will be critical if models are used to inform the distribution of resources.

- Especially salient as funds become available to address drivers of health that exist beyond the bounds of traditional health care.

Granularity Had Small Impact on Predictive Performance but Led to Differences in Interpretation

Census Tract-level models fit better (lower AIC) and had stronger predictive ability than models with ZCTA-level or no SDOH risk factors.

- Differences in fit and predictive performance minimal with utilization-based risk factors in model

		Census Tract	ZCTA	No SDOH
Utilization-Based Variables Included	AIC:	460659.70	471279.77	473879.19
	C:	0.8421	0.8378	0.8410
	Gini:	0.6335	0.6323	0.6315
	Top 10%:	51.68%	51.58%	51.61%
No Utilization-Based Variables	AIC:	516673.85	528723.08	528702.73
	C:	0.6864	0.6838	0.679
	Gini:	0.3592	0.3520	0.3320
	Top 10%:	23.82%	23.22%	22.24%

Stepwise variable selection process retained different SDOH depending on geographic level.

	Census Tract	ZCTA
Utilization-Based Variables Included	Median income % > 65 years with < HS diploma Air pollution % Married	Median income % 65 years + with < HS diploma % Physician diversity Primary care shortage area Mental health shortage area Walkability
No Utilization-Based Variables	Median income % 65 years + % 65 years + with < HS diploma % Foreign born % Physician diversity Air pollution Area deprivation index Mental health shortage area % Married % 65 years + non-white Population Taxable interest per capita	Median income % 65 years + % 65 years + with < HS diploma % Foreign born % With < HS diploma % Physician diversity % Poverty % Single mothers Population growth Population density Primary care shortage area

References & Funding

- Gervasi, S.S., Chen, I.Y., Smith-McLallen, A., et al. (2022). The potential for bias in machine learning and opportunities for health insurers to address it. *Health Affairs*, 41(2), 212-218.
- Pickett, K.E., & Wilkinson, R.G. (2015). Income inequality and health: A causal review. *Social Science & Medicine*, 118, 316-326.
- Salgado, M., Madureira, J., Mendes, A.S., Torres, A., Teixeira, J.P., & Oliveira, M.D. (2020). Environmental determinants of population health in urban settings: A systematic review. *BMC Public Health*, 20, 853.
- Abbass, I., Revere, L., Mitchell, J., & Appari, A. (2017). Medication nonadherence: The role of cost, community, and individual factors. *Health Services Research*, 52(4), 1511-1533. doi:10.1111/1475-6773.12547
- Truong, H.P., Luke, A.A., Hammond, G., Wadhwa, R.K., Reidhead, M., & Joynt Maddox, K.E. (2020). Utilization of social determinants of health ICD-10 Z-codes among hospitalized patients in the United States, 2016-2017. *Medical Care*, 58(12), 1037-1043.
- Hanley, G.E., & Morgan, S. (2008). On the validity of using area-based income measures to proxy household income. *BMC Health Services Research*, 8(1), 79.
- Gerominus, A.T., Bound, J., & Neider, L.J. (1996). On the validity of using Census geocode characteristics to proxy individual socioeconomic characteristics. *Journal of the American Statistical Association*, 91(434), 529-537.
- Henderson, M., Han, F., Perman, C., Haft, H., & Stockwell, I. (2021). Predicting avoidable hospital events in Maryland. *Health Services Research*. Online.

This research was funded by the Maryland Department of Health