Social Science & Medicine 326 (2023) 115943

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Assessing performance of ZCTA-level and Census Tract-level social and environmental risk factors in a model predicting hospital events

Leigh G. Goetschius [a,*], Morgan Henderson [a,b], Fei Han [a,c], Dillon Mahmoudi [d], Chad Perman [e], Howard Haft [e,1], Ian Stockwell [f,g]

[a] The Hilltop Institute at the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA
[b] Department of Economics, College of Arts, Humanities, and Social Sciences, UMBC, Baltimore, MD, 21250, USA
[c] Department of Computer Science and Electrical Engineering, College of Engineering and Information Technology, UMBC, Baltimore, MD, 21250, USA
[d] Department of Geography and Environmental Systems, College of Arts, Humanities, and Social Sciences, UMBC, Baltimore, MD, USA
[e] Program Management Office for the Maryland Primary Care Program, Maryland Department of Health, Baltimore, MD, USA
[f] Department of Information Systems, College of Engineering and Information Technology, UMBC, Baltimore, MD, 21250, USA
[g] Erickson School of Aging Studies, UMBC, Baltimore, MD, 21228, USA

ARTICLE INFO

Handling Editor: Susan J. Elliott

Keywords:
Social determinants of health
Information technology in health
Medicare
Geographic/spatial factors/small area variations
Primary care
Health care disparities

ABSTRACT

Predictive analytics are used in primary care to efficiently direct health care resources to high-risk patients to prevent unnecessary health care utilization and improve health. Social determinants of health (SDOH) are important features in these models, but they are poorly measured in administrative claims data. Area-level SDOH can be proxies for unavailable individual-level indicators, but the extent to which the granularity of risk factors impacts predictive models is unclear. We examined whether increasing the granularity of area-based SDOH features from ZIP code tabulation area (ZCTA) to Census Tract strengthened an existing clinical prediction model for avoidable hospitalizations (AH events) in Maryland Medicare fee-for-service beneficiaries. We created a person-month dataset for 465,749 beneficiaries (59.4% female; 69.8% White; 22.7% Black) with 144 features indexing medical history and demographics using Medicare claims (September 2018 through July 2021). Claims data were linked with 37 SDOH features associated with AH events from 11 publicly-available sources (e.g., American Community Survey) based on the beneficiaries' ZCTA and Census Tract of residence. Individual AH risk was estimated using six discrete time survival models with different combinations of demographic, condition/utilization, and SDOH features. Each model used stepwise variable selection to retain only meaningful predictors. We compared model fit, predictive performance, and interpretation across models. Results showed that increasing the granularity of area-based risk factors did not dramatically improve model fit or predictive performance. However, it did affect model interpretation by altering which SDOH features were retained during variable selection. Further, the inclusion of SDOH at either granularity level meaningfully reduced the risk that was attributed to demographic predictors (e.g., race, dual-eligibility for Medicaid). Differences in interpretation are critical given that this model is used by primary care staff to inform the allocation of care management resources, including those available to address drivers of health beyond the bounds of traditional health care.

1. Introduction

Predictive analytics and big data are playing a larger role in health care than ever before (Shilo et al., 2020). Algorithms designed to estimate a patient's risk for health care utilization, such as avoidable hospitalizations and emergency department visits (hereafter referred to as AH events), are increasingly being used by primary care teams to statistically triage patients and efficiently target care resources. The goal of these efforts is to prevent or delay patients from experiencing these unnecessary, costly services which increase their risk for future cognitive and functional decline (Rudolph et al., 2010; Godard-Sebillotte et al., 2019; Henderson et al., 2021). Social determinants of health (SDOH), such as income, education, and neighborhood conditions, are important to include in these algorithms because research shows they

* Corresponding author.
E-mail address: lgoetschius@hilltop.umbc.edu (L.G. Goetschius).
[1] Independent Consultant.

https://doi.org/10.1016/j.socscimed.2023.115943
Received 28 July 2022; Received in revised form 3 April 2023; Accepted 30 April 2023
Available online 2 May 2023
0277-9536/

significantly influence health care access, utilization, and overall health (Pickett and Wilkinson, 2015; Salgado et al., 2020; Abbass et al., 2017). Additionally, including SDOH may decrease bias in predictive models for outcomes related to health and health care by reducing reliance on patient demographics and prior utilization, which might reflect disparities in access to care (Gervasi et al., 2022). However, individual-level information about many SDOH, such as income, education, and proximity to health care resources, is typically missing from, or poorly captured in, administrative health data (e.g., sparse use of SDOH Z codes in health records (Truong et al., 2020)). As a result, well-formed predictive models often link administrative claims with publicly available, aggregated SDOH data (Hanley and Morgan, 2008; Geronimus et al., 1996; Zhang et al., 2020; Moss et al., 2021).

Linking individual-level administrative claims with publicly available, aggregated SDOH datasets presents challenges depending on the geographic identifiers included in the administrative data and the desired level of geographic granularity for the area-based SDOH risk factors. Typically, the most straightforward way to link individual-level and spatial-based data on a large, sustainable scale is to merge Census ZIP Code Tabulation (ZCTA) level geographies with beneficiary addresses stored in administrative data (Berkowitz et al., 2015). However, there can be substantial variability of SDOH within ZCTAs (Krieger et al., 2002), and ZCTAs do not perfectly approximately ZIP codes (Grubesic and Matisziw, 2006). More granular, Census Tract-level metrics may provide a more accurate representation of an individual's proximal environment (Moss et al., 2021), but with additional (and potentially non-trivial) development cost of geocoding patient addresses. Although previous research shows varying levels of concordance between ZCTA and Census Tract-level variables (Krieger et al., 2002; Tian et al., 2010), the limited work comparing their ability to predict health outcomes suggests similar predictive performance across different granularity levels (Berkowitz et al., 2015; Thomas et al., 2006; Lovasi et al., 2008). However, we are aware of no research to date, that has evaluated the effect of SDOH geographic granularity on predictive performance in the context of risk factors indexing individual medical history. Additionally, there is a paucity of work that moves beyond predictive performance to understand whether the geographic unit for SDOH variables affects the *interpretation* of model results.

This study evaluated the impact of SDOH geographic granularity on predictive performance and model interpretation in the context of an existing clinical prediction model for AH events in Maryland Medicare fee-for-service (FFS) beneficiaries, which is deployed as a tool for practices affiliated with the Maryland Primary Care Program (MDPCP) (Henderson et al., 2021). In this model, risk for AH events is estimated monthly from a targeted set of predictors indexing utilization-based medical history (i.e., diagnoses, prescriptions, procedure history, prior utilization) and demographics from Medicare claims, as well as area-based SDOH risk factors from publicly available sources (e.g., American Community Survey, Neighborhood Atlas, and others) (Henderson et al., 2021; Pelser et al., 2019). Actionable reasons for risk accompany individual risk scores, which correspond to the top risk factors contributing to a patient's risk for an AH event. This rank-based, risk-stratification model is used by primary care providers and care-management teams to identify patients at high risk for AH events within their practice panels, so they can focus their limited time and care-coordination resources on the patients who are most likely to benefit (Obermeyer et al., 2019). Risk scores and reasons for risk are deployed monthly to more than 475 primary care practices with approximately 2000 providers (i.e., physicians, nurse practitioners, nurse specialists, physician assistants) across the state (Schrader et al., 2021). This model has been described in more detail previously (Henderson et al., 2020, 2021).

We enhanced the granularity of the SDOH risk factors from ZCTA to Census Tract as part of regular improvements to the production model in October 2021. Prior to deploying the updated model, we sought to determine whether use of more granular Census Tract-level SDOH, versus more aggregated ZCTA-level measures, strengthened the model's predictive performance for AH events in the Maryland Medicare population. First, we compared the association between our Census Tract and ZCTA-level risk factors to verify that the different levels of granularity captured different information in our sample. Then, we compared model fit and predictive performance across versions of the predictive model with ZCTA-level risk factors, Census Tract-level risk factors, and no area-based risk factors. We did this comparison for the full model (i.e., utilization, demographic, and SDOH risk factors) and a reduced model that does not include individual utilization risk factors (six models total). We tested whether the granularity of area-based SDOH variables affects model performance with and without utilization-based risk factors, because they are not created for beneficiaries with fewer than 12 months of claims history. Additionally, the effect of geographic granularity in the reduced model informs predictive models that use area-based SDOH risk factors without access to extensive utilization data. We hypothesized that the inclusion of *any* area-based risk factors would improve model fit and performance, and that the addition of Census Tract-level risk factors would strengthen the model more than including ZCTA-level risk factors. Last, we examined whether the granularity of the SDOH risk factors influenced model interpretation, given that it is intended to inform the distribution of care-management resources. This study followed the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) guidelines (Benchimol et al., 2015).

2. Methods

2.1. Sample and data

2.1.1. Sample

This study used Medicare Claim and Claim Line Feed (CCLF) data for approximately 465,000 community-dwelling Medicare FFS beneficiaries in Maryland who were attributed to an MDPCP-enrolled primary care practice in Quarter 3 (July–September) of 2021. Medicare FFS beneficiaries currently living in long-term care facilities or nursing homes are not attributed to MDPCP primary care practices and thus are not included in this sample. From this cohort, we created a person-month panel dataset with risk factors that spanned 35 months (September 2018–July 2021).

2.1.2. Clinical prediction model features

In this clinical prediction model, the AH event outcome was a composite of 10 conditions (prevention quality indicators, or PQIs) determined to be potentially preventable with timely, high-quality outpatient care (Billings et al., 1993) by the Agency for Healthcare Research and Quality, including short- and long-term diabetes complications, hypertension, and asthma, among others (AHRQ Quality IndicatorsTM Prevention Quality, 2022). Risk for incurring an AH event was estimated using 182 risk factors indexing utilization-based medical history (i.e., diagnoses, pharmacy utilization, procedure history, prior utilization), demographic information, and SDOH. These features were selected for inclusion in the pool of risk factors based on their association with avoidable hospital events or ambulatory care sensitive conditions in the literature and stakeholder feedback (Pelser et al., 2019).

2.1.3. Utilization and demographic data

Data used for the utilization risk factors come from Medicare CCLF data. We created a person-month panel dataset that uses Part A (i.e., facility), Part B (i.e., professional), and Part D (i.e., pharmacy) claims across 35 months (September 2018–July 2021) to characterize individual procedural, diagnostic, utilization, and pharmaceutical history (see Supplemental Methods Table 2 for a complete list of risk factors and previously published work (Henderson et al., 2021) for more detail). The demographic risk factors, including sex, age, race-ethnicity, and dual-eligibility for Medicaid were created using the beneficiaries' demographic data from CCLF.

2.1.4. SDOH data

Our model development process includes a rigorous, literature-based feature selection methodology. The SDOH risk factors used in the model predicting AH events were identified based on a previously shown association with AH events in the literature (Pelser et al., 2019). All variables were created from publicly available data sources (Table 1) using a spatial joining process, described below. We created two versions of the environmental risk factors: Census Tract-level and ZCTA-level. Most risk factors were available for both Census Tracts and ZCTAs. For risk factors that were only available at the ZCTA level (1 risk factor) or the Census Tract (or other census polygon) level (7 risk factors), we used the U.S. Department of Housing and Urban Development (HUD) United States Postal Service ZIP Code Crosswalk files to transform the variables to the appropriate geographic unit (see Table 1 and the Supplemental Methods for more detail). (Din and Wilson, 2020; Office of Policy Development and Research, 2021) As of the 2010 U.S. Census, there were 73,057 Census Tracts in the United States (1,406 in Maryland), and 32,989 ZCTAs (468 in Maryland). (US Census Bureau, 2010).

Table 1
Risk factors indexing social and environmental determinants of health and their source.

Risk Factor	Source	Year
Population; Population Growth[a]; Population Density[b]	ACS (B01003)	2019
Percent Age 0–4; Percent Age 65+	ACS (S0101)	2019
Percent Married	ACS (S1201)	2019
Percent Single Mothers	ACS (S1301)	2019
Median Household Income	ACS (S1901)	2019
Percent in Poverty	ACS (S1702)	2019
Percent Less than High School Diploma	ACS (S1501)	2019
Percent Native American	ACS (DP05)	2019
Percent Non-English Speakers	ACS (S1601)	2019
Percent Foreign Born	ACS (DP02)	2019
Percent Age 65+ Live Alone	ACS (S1101)	2019
Percent Age 65+ Non-White	ACS (B01001A)	2019
Percent Age 65+ Latinx	ACS (B01001L)	2019
Percent Age 65+ in Poverty	ACS (S1702)	2019
Percent Age 65+ Less than High School Diploma	ACS (S1501)	2019
Rural Urban Index	USDA	2010
Area Deprivation Index	WISC	2019
Taxable Interest	IRS	2018
Has a Mental Health Center	CMS	2021
Has a Federally Qualified Health Center	CMS	2021
Has a Rural Health Clinic	CMS	2021
Has a For Profit Hospital	CMS	2021
Number of Hospitals	CMS	2021
Hospitals/1000 Residents[c]	CMS	2021
Hospital Beds/1000 Residents[c]	CMS	2021
Has a VA Clinic or Center	VA	2021
Primary Care Providers/1000 Residents[c]	NPI	2021
Internists/1000 Residents[c]	NPI	2021
Specialists/1000 Residents[c]	NPI	2021
Social Workers/1000 Residents[c]	NPI	2021
Partial Primary Care Shortage Area	AHRF	2020
Whole Primary Care Shortage Area	AHRF	2020
Partial Mental Health Care Shortage Area	AHRF	2020
Whole Mental Health Shortage Area	AHRF	2020
Percent Physician Diversity (racial or ethnic minority, excluding Asian Americans)	ACS Individual-Level Data	2019
Air Pollution (average daily PM2.5 concentration)	EPA	2011–2015
Walkability	EPA	2020

ACS = American Community Survey, 5-year estimates, data table number in, AHRF = Area Health Resources Files, CMS = Centers for Medicare & Medicaid Services, EPA = Environmental Protection Agency, IRS = Internal Revenue Service, NPI = National Provider Identifier database, USDA = United States Department of Agriculture, VA = Veterans Affairs, WISC = Wisconsin School of Medicine and Public Health.

[a] Due to data availability, growth for Census Tracts is from 2013 to 2019 and from 2011 to 2019 for ZCTAs.

[b] Density calculated using land area (square miles) according to the 2019 Census Gazetteer records.

[c] Calcuated using the 2019 population estimates from ACS.

2.1.5. Geocoding process

We used an automated, two-step geocoding procedure to identify an individual's unique Census Tract. First, we used Microsoft® Azure Maps' "Get Search Address" feature to transform individuals' home addresses from the CCLF data into geographical coordinates (i.e., latitude, longitude). Then, we mapped the coordinates to a Census Tract using the GeoPandas (v0.8.1) (Jordahl et al., 2020) python package. When an individual's unique Census Tract was identified, we linked the environmental risk factors from their Census Tract and five-digit ZCTA of residence to their individual utilization risk factors. ZCTAs were assigned based on the ZIP code of the beneficiary's address.

2.2. Analytic strategy

2.2.1. Association between ZCTA and Census Tract environmental risk factors

First, we examined the association between the Census Tract and ZCTA-level social and environmental risk factors at the beneficiary level using Pearson's correlation. Given the sample size, the p-values for these correlations were not interpreted; however, effect sizes were considered. Next, prior to model building, we examined the associations between all Census Tract-level social and environmental variables and all ZCTA-level variables.

2.2.2. Predictive model for avoidable hospitalization events

We ran six discrete time survival models predicting whether an individual had an AH event in the following month (0/1) from different combinations of utilization and geographic risk factors: Model 1 included demographic, Census Tract-level, and utilization risk factors; Model 2 included demographic, ZCTA-level, and utilization risk factors; Model 3 included demographic and utilization risk factors, but no geographic risk factors; Model 4 included demographic and Census Tract-level risk factors; Model 5 included demographic and ZCTA-level risk factors; Model 6 included only demographic risk factors. AH events were defined using 2020 technical definitions for prevention quality indicator (PQI) measures from the Agency for Healthcare Research and Quality (AHRQ) and include diagnoses for diabetes complications, chronic obstructive pulmonary disease (COPD) or asthma, hypertension, heart failure, and bacterial pneumonia (Prevention Quality Indicators Technical Specifications, 2020). Each regression model was trained on 80% of the person-month data (randomly sampled at the person level) and used a stepwise variable selection process so that only risk factors that significantly improved model fit were retained in the final model. We then applied the coefficients from the training model to the risk factors in the remaining 20% of the data (i.e., the testing data) and evaluated its predictive performance.

2.2.3. Comparison of model fit and predictive performance

We compared the six models based on statistical fit in the training data (Akaike Information Criteria, or AIC (Portet, 2020)). Additionally, we evaluated each model's predictive discrimination, that is, the ability to discriminate between beneficiaries who did and did not experience AH events, in the testing data (C-statistic (Steyerberg et al., 2010), Gini Index, and cumulative percentage of AH events incurred by beneficiaries with the top 10% of risk scores (Llorca and Delgado-Rodríguez, 2002)). We were interested in the predictive capability for the top 10% of risk scores, because research suggests that care-management efforts targeting high-risk patients are the most effective at reducing health care utilization and costs (Brown et al., 2012). Last, we used a nonparametric approach to compare the receiver operating curves (ROCs) using the testing data to determine which models, if any, were significantly more effective at discriminating between beneficiaries who did and did not experience AH events using a chi-square test (DeLong et al., 1988). We used a Bonferroni-correction to adjust significance threshold based on comparisons ($p = 0.05/6 = 0.008$). More details about these metrics are included in the Supplemental Methods. All statistical analyses were done in SAS (v.9.4). R (v.4.1.0) was used to make plots (*tidyverse v.1.3.1*) and tables (Table 1 v.1.4.1).

2.2.4. Comparison of model interpretation

We compared the risk factors retained after the stepwise variable selection process for each model to determine whether the granularity of the risk factors affected reasons for risk and model interpretation. Additionally, we compared the odds ratios for the demographic risk factors across the final models to determine whether the granularity of risk factors and inclusion of utilization-based risk factors changed the relative risk attributed to demographic variables, such as race-ethnicity and dual-eligibility status (a proxy for low income).

2.2.5. Supplementary analyses

Because each discrete time survival model used a stepwise variable selection method to determine the final variables included in the model, the SDOH variables included in the Census Tract and ZCTA models differed. Although we find the difference in SDOH variable selection across models to be meaningful, we ran an additional set of models in which the SDOH variables were identical to ensure that variable selection did not impact the overall pattern of results for the predictive performance comparisons (see Supplemental Methods for more detail).

3. Results

3.1. Sample characteristics

The cohort comprised 465,749 Medicare beneficiaries with a total of 16, 962, 894 person-months. Beneficiaries were on average 73.1 years old, 59.5% were female, 14.9% were dually eligible for Medicaid, and 18.2% were eligible for Medicare for a reason other than age (i.e., disability). Approximately 70% were Non-Hispanic White, 22.7% were Black or African American, 2.1% were Asian, 0.1% were American Indian or Alaska Native, 1.1% were Hispanic/Latinx, 1.5% were categorized as “Other”,[2] and the race-ethnicity for the remaining 2.7% was unknown. Beneficiaries were excluded from predictive models that included utilization risk factors (models 3–5) if they did not have at least 12 months of Medicare claims (1%). Beneficiaries without a valid Census Tract (2.2%) or ZCTA (0.4%) were excluded in predictive models using the Census Tract or ZCTA versions of the SDOH risk factors, respectively. Lastly, 196 (<0.1%) beneficiaries were excluded from the cohort because they could not be assigned a valid Census Tract or ZCTA. See Supplemental Methods Table 1 for complete cohort characteristics.

3.2. Association between ZCTA and Census Tract environmental risk factors

The average correlation between the Census Tract and ZCTA versions of the SDOH risk factors for the beneficiaries in this sample was $\mu_{correlation} = 0.529$ ($SD = 0.269$), meaning they shared approximately 28% of their variance ($R^2 = 0.2798$). This variability underscores that the level of geographic granularity can impact an individual’s area-level estimate for a given risk factor which can have meaningful implications when it comes to model interpretation. The risk factor with the lowest Census Tract-ZCTA correlation was 2019 population ($r = -0.021$), and the risk factor with the highest correlation was an indicator for whether the whole county in which the region is located lies within a mental health care shortage area ($r = 0.963$). Fig. 1 depicts the correlation and the 95% confidence interval (CI) for each social and environmental risk factor. All variables were correlated at $p < 0.0001$; however, given the large sample size, the p-values should be interpreted with caution. Fig. 1 color codes each correlation by effect size (small: gray; medium: gold; large: teal) (Cohen, 1992). The correlations between all Census Tract- and ZCTA-level environmental variables are included in Supplemental Methods Table 5. To reduce the risk of multicollinearity, we excluded risk factors that were correlated at greater than $r = 0.8$ with another risk factor at the same granularity level (Census Tract: $N = 5$; ZCTA: $N = 4$; see Supplemental Results for specific variables).

3.3. Comparison of model fit and predictive performance

We estimated the six discrete time survival models predicting whether a beneficiary would incur an AH event the following month from different sets of predictors in the training sample. All models converged normally, and the included risk factors for each had p-values of <0.00012 (see Supplemental Results for full details about the final models). We recorded the AIC for each model to compare model fit (Table 2; note that lower AIC values mean better fit). Model 1 (demographic, Census Tract-level, and utilization risk factors) was the best fit for the data. The models that included individual utilization risk factors (models 1–3) fit the data better than the models that did not (models 4–6). Among the models that did not include individual utilization history, model 4 (demographic and Census Tract-level risk factors) fit the data better than the ZCTA-level model (model 5) and the model with only demographic predictors (model 6). This pattern of results held when the list SDOH risk factors were identical across the ZCTA and Census-Tract versions of the model rather than varying due to variable selection (see Supplemental Results for more detail).

The coefficients from each discrete time survival model estimated in the training data were then applied to the risk factors in the testing data to evaluate the predictive performance of each model in the 20% of the sample that was held in reserve (Table 2). Similar to our observation of model fit in the training data, model 1 had the best predictive performance measured, using both the C-statistic and Gini index. Additionally, beneficiaries with the top 10% of risk scores from model 1 accounted for the highest percentage of AH events (51.68%, see Table 2 for all models). As noted above, models that included individual utilization history (models 1–3) performed better than models that did not (models 2–4).

In general, models with the Census Tract-level risk factors outperformed the ZCTA-level models; however, the differences were slight, particularly when individual utilization history was included. When we statistically compared predictive discrimination across models, the models with the Census Tract-level social and environmental risk factors outperformed both the models that did not include area-based risk factors and those with ZCTA-level risk factors (Table 3). The difference in performance was relatively small between Census Tract and ZCTA-level models; however, and in models that included individual utilization risk factors, the difference was marginal ($p = 0.0314$) and was non-significant when using a significance threshold that was adjusted for multiple comparisons ($p < 0.008$; Table 3).

3.4. Comparison of model interpretation

Interestingly, the automated, stepwise selection process retained the same demographic and utilization risk factors, regardless of the granularity of the area-based risk factors; however, different social and environmental risk factors were retained, depending on whether they were at the ZCTA or Census Tract-level (Table 2). Additionally, the inclusion of area-based social and environmental risk factors, at either level of granularity, meaningfully reduced the relative risk attributed to demographic predictors, including the indicators for race and dual-

[2] “Other” includes persons identifying as two or more races and Native Hawaiian, Other Pacific Islander, or any other racial-ethnic group (e.g., Middle Eastern, North African).

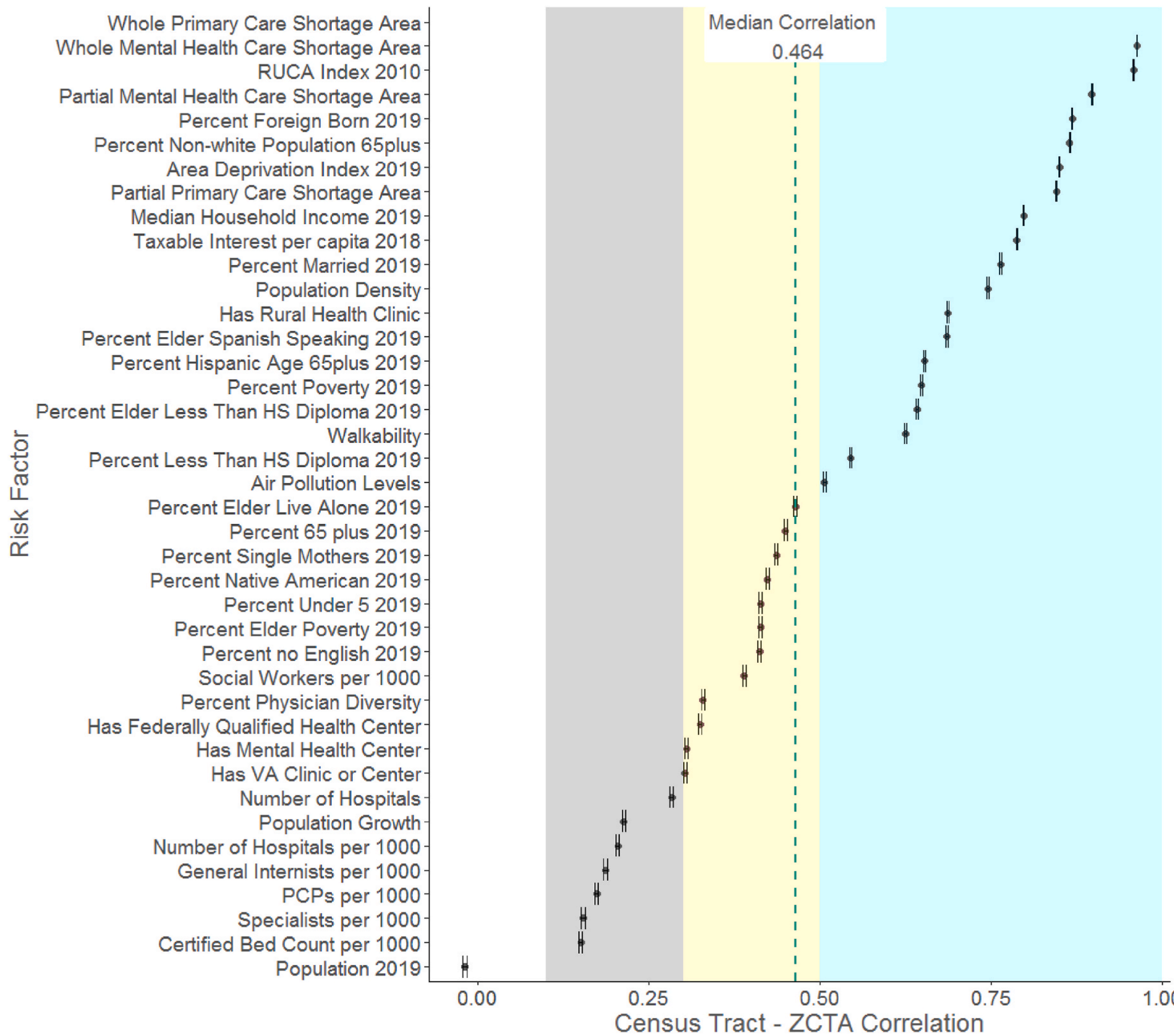

Fig. 1. Average correlation and 95% confidence intervals between Census Tract and ZCTA versions of social and environmental risk factors. Correlation estimates are color coded by effect size: estimates in the gray box are small ($r = 0.1–0.299$); estimates in the gold box are medium ($r = 0.3–0.499$); estimates in the teal box are large ($r \geq 0.5$). Results show that the agreement between Census Tract and ZCTA versions of risk factors for beneficiaries in this sample varies considerably suggesting that, for many features, Census Tract and ZCTA measures would differ for individuals. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

eligibility for Medicaid, a proxy for lower income (Fig. 2). The attributed risk for the predictor indexing Medicare eligibility for a reason other than age was reduced when SDOH variables were added, but only when individual utilization variables were not included in the model. Attributed risk was not meaningfully different for the indicators for Hispanic/Latinx ethnicity or age when area-based SDOH were included (Fig. 2). The attributed risk was further reduced when Census Tract-level risk factors were added compared with ZCTA-level factors; however, as described above, this difference was less pronounced when individual utilization history was accounted for (Fig. 2). Although race and dual-eligibility for Medicaid were still identified as a statistically significant features in each model, area-level SDOH, especially Census Tract-level SDOH, redistributed some of the variance attributed to them to more actionable risk factors that race and dual-eligibility are likely proxies for.

4. Discussion

Area-level risk factors indexing SDOH are important features of preventive predictive models using administrative claims to estimate a patient's risk for health outcomes. However, it is not clear how the extent of *granularity* of the risk factors affects the predictive performance and utility of those models, particularly within the context of a larger pool of factors indexing individual-level medical history. We examined whether the use of Census Tract-level SDOH, rather than ZCTA-level, strengthened an existing clinical prediction model estimating the risk for AH events for Medicare FFS beneficiaries in Maryland. Although there were varying degrees of overlap between ZCTA and Census Tract versions of SDOH, we found that increasing the granularity of the area-based risk factors did not dramatically improve the model's fit or predictive performance. In fact, when risk factors characterizing individual procedural, diagnostic, utilization, and pharmaceutical history were included, the difference in the model's predictive discrimination and calibration attributable to geographic level was negligible. Further, predictive performance for the enhanced, Census Tract-level model was similar to that of the original production model (Henderson et al., 2021). Geographic granularity made the most meaningful difference in the interpretation of the model and its potential impact on care-management decisions.

Modeling risk from SDOH, particularly using more granular estimates, reduced the relative risk attributed to race and, to a lesser extent, dual-eligibility and disability status, and redistributed it to more actionable reasons for risk that can be targeted using proactive care-management efforts. Interestingly, there was no reduction in relative

Table 2
Comparison of model fit and predictive capability.

Utilization Risk Factors Included?	Census Tract	ZCTA	No Geographic Predictors
Yes	**Model 1** **AIC**: 460659.70 **C**: 0.8421 **Gini**: 0.6335 **Top 10%**: 51.68% **Selected Risk Factors**: • Median income • % > 65 years with less than high school diploma • Air pollution • % Married	**Model 2** **AIC**: 471279.77 **C**: 0.8378 **Gini**: 0.6323 **Top 10%**: 51.58% **Selected Risk Factors**: • Median income • % 65 years + with less than high school diploma • % Physician diversity • Primary care shortage area (whole) • Mental health shortage area (partial) • Walkability	**Model 3** **AIC**: 473879.19 **C**: 0.8410 **Gini**: 0.6315 **Top 10%**: 51.61%
No	**Model 4** **AIC**: 516673.85 **C**: 0.6864 **Gini**: 0.3592 **Top 10%**:23.82% **Selected Risk Factors**: • Median income • % 65years + • % 65 years + with less than high school diploma • % Foreign born • % Physician diversity • Air pollution • Area deprivation index • Mental health shortage area (whole) • % Married • % 65 years + non-white • Population • Taxable interest per capita	**Model 5** **AIC**: 528723.08 **C**: 0.6838 **Gini**: 0.3520 **Top 10%**: 23.22% **Selected Risk Factors**: • Median income • % 65years + • % 65 years + with less than high school diploma • % Foreign born • % With less than high school diploma • % Physician diversity • % Poverty • % Single mothers • Population growth • Population density • Primary care shortage area (whole)	**Model 6** **AIC**: 528702.73 **C**: 0.679 **Gini**: 0.3320 **Top 10%**: 22.24%

Note: AIC is based on model fit in the training data. The C-statistic, Gini coefficient, and top 10% predictive statistics are derived from applying the model coefficients from the training data in the testing data. Significant risk factors are not included for models 3 and 6 because they do not include social and environmental risk factors.

risk attributed to Hispanic/Latinx ethnicity or age with the inclusion of SDOH risk factors. Modeling risk associated with individual medical history also markedly reduced the relative risk attributed to race, dual-eligibility status, age, and enrollment in Medicare for a reason other than age. These findings align with previous research that suggests neighborhood SDOH, such as residential segregation, mediate the association between race-ethnicity and health outcomes (Wong et al., 2020; Skolarus et al., 2020). Additionally, race-ethnicity is often a proxy for disparities in SDOH, such as access to quality health care, environmental exposures (e.g., pollutants), and neighborhood disadvantage, which can be tied to systemic racism (Gervasi et al., 2022; Braveman et al., 2022). However not all research has found that neighborhood SDOH explain associations between health disparities and outcomes (Piccolo et al., 2015), suggesting that the relationships among

Table 3
Results from the non-parametric tests comparing the trapezoidal area under the ROC curves.

Comparison	Estimate	Standard Error	95% Confidence Interval	*Chi-Square*	*p*
AH ~ Demographics + Social and Environmental + Individual Utilization					
Census Tract – ZCTA *(Model 1 vs. 2)*	0.0005	0.0002	0.00004–0.001	4.63	0.0314*
Census Tract – No Geo *(Model 1 vs. 3)*	0.001	0.0003	0.001–0.002	15.65	<.0001
ZCTA – No Geo *(Model 2 vs. 3)*	0.001	0.0002	0.0003–0.001	11.42	0.0007
AH ~ Demographics + Social and Environmental					
Census Tract – ZCTA *(Model 4 vs. 5)*	0.003	0.001	0.002–0.005	25.22	<.0001
Census Tract – No Geo *(Model 4 vs. 6)*	0.008	0.001	0.006–0.0103	67.49	<.0001
ZCTA – No Geo *(Model 5 vs. 6)*	0.005	0.001	0.003–0.006	35.64	<.0001

*Non-significant when using a Bonferroni-correction to adjust for six comparisons ($p < 0.008$).

disparities, SDOH, and health outcomes are complex and likely shaped by study population, location, and choice of outcome. More research is needed to understand these complex associations and identify which of the SDOH in our pool of risk factors would be effective targets for primary care intervention.

Different SDOH variables were salient predictors of AH events, depending on the granularity level, which may influence the application of model output by primary care providers who use the reasons for risk to guide their allocation of care resources. Evidence suggests that Census Tract-level estimates may more accurately approximate individual-level SDOH (Moss et al., 2021); therefore, Census Tract-level reasons for risk may be more informative of a patient's unmet social needs. Differences in SDOH reasons for risk are critical in light of the recent initiation of supplemental care-management fees to primary care providers designed to advance health equity, such as the MDPCP's new Health Equity Advancement Resource and Transformation (HEART) payments (Maryland Primary Care Program, 2021). These supplemental payments are provided for the care of beneficiaries with high clinical risk *and* high neighborhood deprivation (based on the Area Deprivation Index, or ADI) (Maryland Primary Care Program, 2021). SDOH reasons for risk in the predictive model output can help determine where such payments could be best directed. Thus, it is critical that they provide the most precise estimates of an individual patient's social and environmental risk factors.

Our findings documenting the marginal improvement of model performance using Census Tract-level SDOH risk factors relative to less granular ZCTA-level SDOH risk factors are consistent with previous research (Berkowitz et al., 2015; Thomas et al., 2006; Lovasi et al., 2008). However, this study expanded the evidence base informing the use of area-based SDOH in predictive models in two ways. First, we showed that, when predicting risk for AH events, the choice of geographic granularity had a smaller effect when individual, utilization-based risk factors were included in the model, than in models without utilization-based risk factors. Findings from the ROC comparison analyses showed that predictive performance was only marginally improved in the model using Census Tract-level SDOH. Further, when

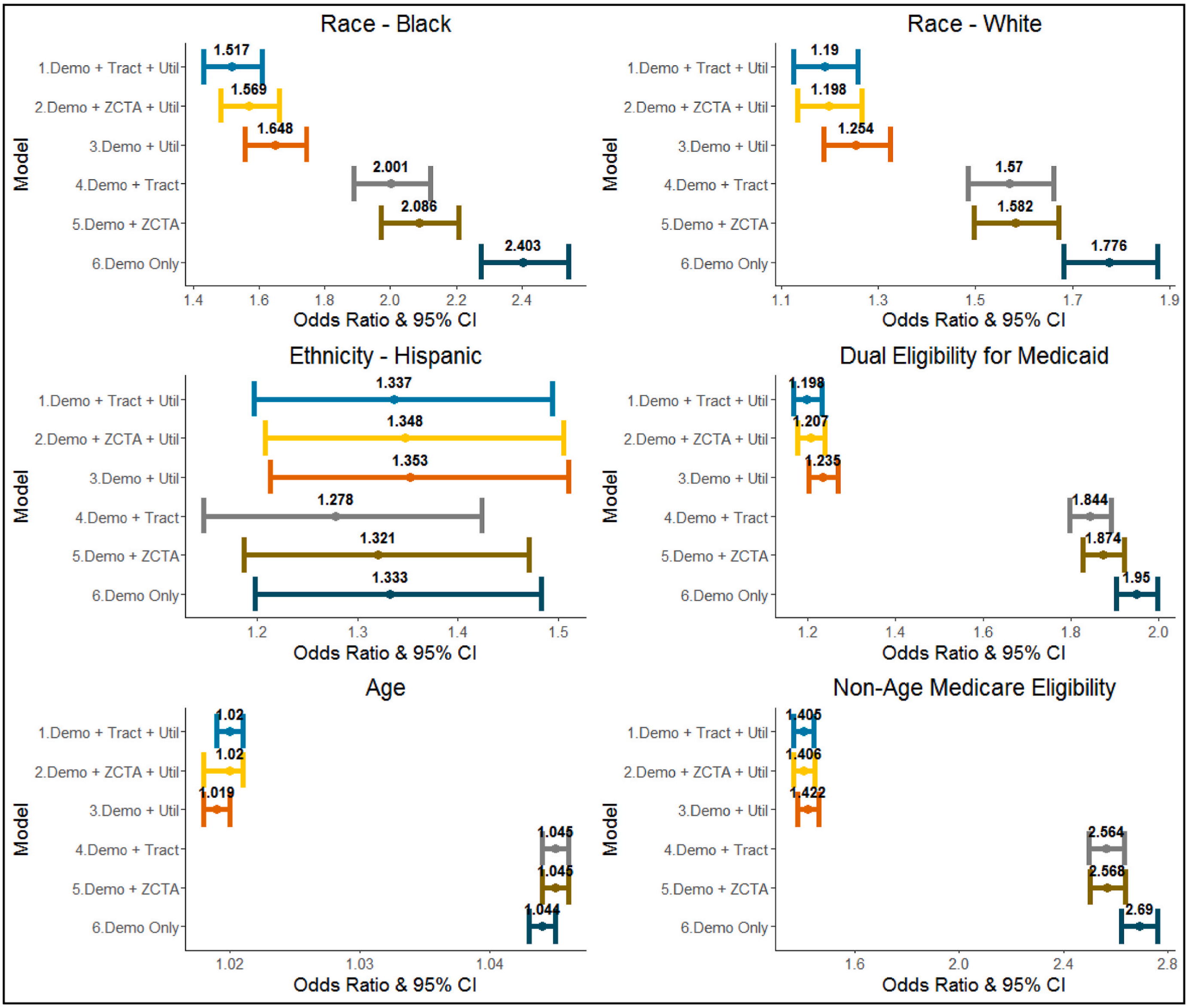

Fig. 2. Average odds ratio and 95% confidence intervals for all significant demographic risk factors by model. Including area-level SDOH features, especially Census Tract-level features, reduced the relative risk attributed to the Race and dual-eligibility for Medicaid features.

utilization-based risk factors were included, predictive performance was effectively the same, even when area-based SDOH were not included. This reduced effect may be because risk factors indexing an individual's utilization-based medical history are more proximal predictors for our utilization-based outcome than area-level SDOH. Further, differences in utilization and diagnostic history may also be indexing similar disparities in access to health care as the area-level SDOH (Braveman et al., 2010); however, future research is needed to explicitly test that hypothesis. Although we do not advocate for excluding area-level SDOH from predictive models, particularly given our findings related to model interpretation, these results suggest that less granular ZCTA measures may be appropriate if the model is solely intended for risk prediction, even when utilization-based risk factors are not available. This may be particularly relevant for predictive models built using data sources in which only an individual's ZIP code or county of residence is available, such as the nationwide datasets with Medicare (LDS Master Beneficiary Summary File, 2017) and Medicaid claims (Centers for Medicaid and CHIP Services, 2020).

Second, although the granularity of area-based SDOH risk factors did not dramatically impact affect predictive power for AH events, increased granularity appeared to have a meaningful impact on risk factor coefficient *interpretation*. That is, in models with more granular area-level risk factors, less relative risk was attributed to certain demographic risk factors; this suggests that, in models with less granular area-level risk factors, these demographic risk factors may be capturing both individual and environmental risk. Thus, the improvement in model interpretation may justify the added development costs of linking administrative claims with Census Tract-level, area-based SDOH for the production model. We developed an automated method to geocode beneficiary addresses, which makes it a feasible component of our production pipeline, and the benefits outweigh any added burden. However, findings from the present study suggest that taking the additional time and resources to regularly geocode beneficiary addresses may not be necessary if the primary objective is limited to estimating risk scores.

It is important to note that the optimal level of geographic granularity of area-based risk factors for clinical prediction models, such as this one, is also influenced by how the model output is intended to be used, as well as the needs and resources of the intended users (Venkatesh et al., 2003). For example, individual risk scores and reasons for risk from this model are intended to guide the direction of care management resources by primary care clinic staff and inform the discussion of interventions to address an individual's specific needs. Therefore, more granular features (i.e., Census Tract) that can act as proxies for individual-level SDOH may be more useful. However, less granular, community-level features from the ZCTA or even county level may be more appropriate in predictive models that focus on the impact of the neighborhood environment on patient outcomes or that are used for different purposes (e.g., to identify targets for community-level interventions or policies). Additionally, it may be important to consider using broader geographic areas when indexing the availability or accessibility of resources, such as health care professionals or facilities, where potential service areas are larger than a single Census Tract or ZCTA (Wang and Luo, 2005). Ultimately, there is no one-size-fits-all approach to using area-based SDOH features in predictive analytics; however, this research adds to the growing literature underscoring the importance of modeling their influence on health outcomes.

4.1. Limitations

Medicare CCLF data are well-suited for modeling a patient's risk for AH events. However, there are limitations to this data source. First, CCLF claims are not updated in real time; there is approximately a 40-day lag between the most recent claims and the release of the scores. This limitation is unavoidable, but theoretically has a minor impact on the model, because it indexes a minimum of 12 months of claims (Henderson et al., 2021). Second, Medicare claims do not contain clinical information, such as lab results, or vital statistics, such as blood pressure and weight, which could potentially increase the predictive capability of the model. Third, reporting of the race-ethnicity information for the Medicare beneficiary demographics file is voluntary and currently combines race and ethnicity into a single variable (Bierman et al., 2002). This limitation in the data is a barrier for understanding health disparities based on race or ethnicity; however, efforts are underway by Centers for Medicare & Medicaid Services (CMS) to improve measurement of these variables in the future. Additionally, CCLF claims do not reliably collect individual-level SDOH information (Truong et al., 2020), so we cannot determine whether the SDOH predictors that are salient at the Census Tract or ZCTA-level provide a more accurate estimate of an individual patient's needs or proximal environment.

In addition to limitations of the CCLF data, this study did not include all potentially relevant SDOH indicators. We focused on SDOH that have been previously associated with AH events in the literature (Pelser et al., 2019) that could be created at the ZCTA and Census Tract levels using publicly available data. However, risk factors indexing SDOH such as safe housing conditions or access to healthy food (e.g., food deserts) may also be meaningful and would be important to examine in future research. Additionally, it is tenable that the SDOH included in the present study may be differentially associated with the individual PQIs that make up the AH event composite. However, given that all PQIs are considered to be preventable through timely, quality primary care (Billings et al., 1993; AHRQ Quality Indicators™ Prevention Quality, 2022), we believe it is also relevant to understand predictors of the composite outcome. Further, focusing on a single composite rather than multiple, individual PQIs may also make it easier to incorporate the risk scores and reasons for risk into primary care workflows, thus making the tool more useful. Last, although the data used in this study comprise almost 500,000 Medicare FFS beneficiaries, our findings may not generalize to other populations – for example, individuals on Medicaid or who are commercially insured – or other outcomes. It is possible that, in certain circumstances, more granular environmental data *would* significantly improve model performance; however, for this to occur, these environmental risk factors would need to capture variation in the outcome, which is not accounted for by individual-level predictors.

5. Conclusions

Risk factors indexing area-level SDOH, such as household income and health care access, strengthen predictive models for AH events. Enhancing the granularity of SDOH predictors from ZCTA to Census Tract did not dramatically improve the model's predictive performance. However, doing so may meaningfully affect model interpretation by changing which SDOH are selected as potential reasons for risk and the relative risk attributed to demographic variables (e.g., race, dual-eligibility status). Further, Census Tract-level SDOH describe a smaller area than ZCTA versions and therefore may provide a more accurate estimate of risk and protective factors within an individual's proximal environment. Differences in interpretation are critical because predictive models such as this one are increasingly being used to inform the distribution of resources, especially as funds become available to address the drivers of health that exist beyond the bounds of traditional health care.

Data availability

We cannot share the raw claims data because they are protected health information under HIPAA that cannot be de-identified but will share the model structure, coefficients, and analysis code.

Acknowledgements

Funding Information: Maryland Department of Health and Mental Hygiene.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2023.115943.

References

Abbass, I., Revere, L., Mitchell, J., Appari, A., 2017. Medication nonadherence: the role of cost, community, and individual factors. Health Serv. Res. 52 (4), 1511–1533. https://doi.org/10.1111/1475-6773.12547.

AHRQ Quality Indicators™ Prevention Quality Indicators. Agency for Healthcare Research and Quality; 2022.

Benchimol, E.I., Smeeth, L., Guttmann, A., et al., 2015. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement. PLoS Med. 12 (10), e1001885 https://doi.org/10.1371/journal.pmed.1001885.

Berkowitz, S.A., Traore, C.Y., Singer, D.E., Atlas, S.J., 2015. Evaluating area-based socioeconomic status indicators for monitoring disparities within health care systems: results from a primary care network. Health Serv. Res. 50 (2), 398–417. https://doi.org/10.1111/1475-6773.12229.

Bierman, A.S., Lurie, N., Collins, K.S., Eisenberg, J.M., 2002. Addressing racial and ethnic barriers to effective health care: the need for better data. Health Aff. 21 (3), 91–102. https://doi.org/10.1377/hlthaff.21.3.91.

Billings, J., Zeitel, L., Lukomnik, J., Carey, T.S., Blank, A.E., Newman, L., 1993. Impact of socioeconomic status on hospital use in New York city. Health Aff. 12 (1), 162–173. https://doi.org/10.1377/hlthaff.12.1.162.

Braveman, P.A., Cubbin, C., Egerter, S., Williams, D.R., Pamuk, E., 2010. Socioeconomic disparities in health in the United States: what the patterns tell us. Am. J. Publ. Health 100 (Suppl. 1), S186–S196. https://doi.org/10.2105/AJPH.2009.166082.

Braveman, P.A., Arkin, E., Proctor, D., Kauh, T., Holm, N., 2022. Systemic and structural racism: definitions, examples, health damages, and approaches to dismantling. Health Aff. 41 (2), 171–178. https://doi.org/10.1377/hlthaff.2021.01394.

Brown, R.S., Peikes, D., Peterson, G., Schore, J., Razafindrakoto, C.M., 2012. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Aff. 31 (6), 1156–1166. https://doi.org/10.1377/hlthaff.2012.0393.

Centers for Medicaid and CHIP Services, 2020. Transformed Medicaid Statistical Information System (T-MSIS) Data Dictionary Version: V2.4.0. Center for Medicare and Medicaid Services.

Cohen, J.A., 1992. Power primer. Psychol. Bull. 112 (1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44 (3), 837. https://doi.org/10.2307/2531595.

Din, A., Wilson, R., 2020. Crosswalking ZIP codes to census geographies: geoprocessing the U.S. Department of housing & urban development's ZIP code Crosswalk files.

Cityscape 22 (1). https://www.huduser.gov/portal/periodicals/cityscpe/vol22num1/ch12.pdf. (Accessed 29 June 2021).

Geronimus, A.T., Bound, J., Neidert, L.J., 1996. On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. J. Am. Stat. Assoc. 91 (434), 529–537. https://doi.org/10.1080/01621459.1996.10476918.

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 Aff. 41 (2), 212–218. https://doi.org/10.1377/hlthaff.2021.01287.

Godard-Sebillotte, C., Le Berre, M., Schuster, T., Trottier, M., Vedel, I., 2019. Impact of health service interventions on acute hospital use in community-dwelling persons with dementia: a systematic literature review and meta-analysis. PLoS One 14 (6). https://doi.org/10.1371/journal.pone.0218426.

Grubesic, T.H., Matisziw, T.C., 2006. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. Int. J. Health Geogr. 5 (1), 58. https://doi.org/10.1186/1476-072X-5-58.

Hanley, G.E., Morgan, S., 2008. On the validity of area-based income measures to proxy household income. BMC Health Serv. Res. 8 (1), 79. https://doi.org/10.1186/1472-6963-8-79.

Henderson, M., Han, F., Stockwell, I., 2020. Maryland Primary Care Program (MDPCP) Pre-AH Risk Score Specifications and Codebook (Version 3). The Hilltop Institute at UMBC. https://www.hilltopinstitute.org/wp-content/uploads/publications/MDPCPPre-AHRiskScoreSpecsAndCodebook-V3-June2020.pdf. (Accessed 9 June 2021).

Henderson M, Han F, Perman C, Haft H, Stockwell I. Predicting avoidable hospital events in Maryland. Health Serv. Res.. Published online October 14, 2021. doi:10.1111/1475-6773.13891.

Jordahl, K., Bossche, J.V.D., Fleischmann, M., et al., 2020. Geopandas. Zenodo. https://doi.org/10.5281/ZENODO.3946761.

Krieger, N., Chen, J.T., Waterman, P.D., Soobader, M.J., Subramanian, S.V., Carson, R., 2002. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the public health disparities geocoding project. Am. J. Epidemiol. 156 (5), 471–482. https://doi.org/10.1093/aje/kwf068.

LDS Master Beneficiary Summary File (MBSF), Chronic Conditions Data Warehouse (CCW) Data Dictionary, 2017. Centers for Medicare and Medicaid Services.

Llorca, J., Delgado-Rodríguez, M., 2002. Visualising exposure-disease association: the Lorenz curve and the Gini index. Med. Sci. Mon. Int. Med. J. Exp. Clin. Res. 8 (10), MT193–197.

Lovasi, G.S., Moudon, A.V., Smith, N.L., et al., 2008. Evaluating options for measurement of neighborhood socioeconomic context: evidence from a myocardial infarction case-control study. Health Place 14 (3), 453–467. https://doi.org/10.1016/j.healthplace.2007.09.004.

Maryland Primary Care Program, 2021. Maryland Department of Health. COVID-19 Updates from the Program Management Office. Published online November 10. https://health.maryland.gov/mdpcp/Documents/MDPCP%20COVID-19%20Updates%20from%20the%20PMO.pdf. (Accessed 7 January 2022).

Moss, J.L., Johnson, N.J., Yu, M., Altekruse, S.F., Cronin, K.A., 2021. Comparisons of individual- and area-level socioeconomic status as proxies for individual-level measures: evidence from the Mortality Disparities in American Communities study. Popul. Health Metrics 19, 1. https://doi.org/10.1186/s12963-020-00244-x.

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Science 366 (6464), 447–453. https://doi.org/10.1126/science.aax2342.

Office of Policy Development and Research, 2021. HUD USPS ZIP Code Crosswalk Files. HUD User. https://www.huduser.gov/portal/datasets/usps_crosswalk.html#codebook. (Accessed 16 November 2021).

Pelser, C., Henderson, M., Stockwell, I., 2019. Risk Factors for Potentially Avoidable Hospital Admissions and Emergency Department Visits: A Literature Review. The Hilltop Institute at UMBC. https://www.hilltopinstitute.org/wp-content/uploads/publications/MDPCPAvoidableHospitalizationLiteratureReview-April2019.pdf.

Piccolo, R.S., Duncan, D.T., Pearce, N., McKinlay, J.B., 2015. The role of neighborhood characteristics in racial/ethnic disparities in type 2 diabetes: results from the Boston Area Community Health (BACH) Survey. Soc. Sci. Med. 130, 79–90. https://doi.org/10.1016/j.socscimed.2015.01.041.

Pickett, K.E., Wilkinson, R.G., 2015. Income inequality and health: a causal review. Soc. Sci. Med. 128, 316–326. https://doi.org/10.1016/j.socscimed.2014.12.031.

Portet, S., 2020. A primer on model selection using the Akaike Information Criterion. Infect Dis Model 5, 111–128. https://doi.org/10.1016/j.idm.2019.12.010.

Prevention Quality Indicators Technical Specifications, 2020. Version V2020 (ICD 10-CM/PCS). Agency for Healthcare Research and Quality (AHRQ), Department of Health and Human Services, p. 2. https://qualityindicators.ahrq.gov/Archive/PQI_TechSpec_ICD10_v2020.aspx.

Rudolph, J.L., Zanin, N.M., Jones, R.N., et al., 2010. Hospitalization in community-dwelling persons with alzheimer's disease: frequency and causes. J. Am. Geriatr. Soc. 58 (8), 1542–1548. https://doi.org/10.1111/j.1532-5415.2010.02924.x.

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 Publ. Health 20, 853. https://doi.org/10.1186/s12889-020-08905-0.

Schrader, D., Haft, H., Perman, C., et al., 2021. Maryland Primary Care Program (MDPCP) 2020 Annual Report. Maryland Department of Health. https://health.maryland.gov/mdpcp/Documents/MDPCP_2020_Annual_Report.pdf. (Accessed 4 February 2022).

Shilo, S., Rossman, H., Segal, E., 2020. Axes of a revolution: challenges and promises of big data in healthcare. Nat. Med. 26 (1), 29–38. https://doi.org/10.1038/s41591-019-0727-5.

Skolarus, L.E., Sharrief, A., Gardener, H., Jenkins, C., Boden-Albala, B., 2020. Considerations in addressing social determinants of health to reduce race/ethnic disparities in stroke outcomes in the United States. Stroke 51 (11), 3433–3439. https://doi.org/10.1161/STROKEAHA.120.030426.

Steyerberg, E.W., Vickers, A.J., Cook, N.R., et al., 2010. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 21 (1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2.

Thomas, A.J., Eberly, L.E., Davey Smith, G., Neaton, J.D., 2006. ZIP-Code-based versus tract-based income measures as long-term risk-adjusted mortality predictors. Am. J. Epidemiol. 164 (6), 586–590. https://doi.org/10.1093/aje/kwj234.

Tian, N., Goovaerts, P., Zhan, F.B., Wilson, J.G., 2010. Identification of racial disparities in breast cancer mortality: does scale matter? Int. J. Health Geogr. 9, 35. https://doi.org/10.1186/1476-072X-9-35.

Truong, H.P., Luke, A.A., Hammond, G., Wadhera, 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. Med. Care 58 (12), 1037–1043. https://doi.org/10.1097/MLR.0000000000001418.

US Census Bureau, 2010. Census Tallies. Census.gov. Published October 8, 2021. Accessed March 2, 2022. https://www.census.gov/geographies/reference-files/time-series/geo/tallies.html.

Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: toward a unified view. MIS Q. 27 (3), 425. https://doi.org/10.2307/30036540.

Wang, F., Luo, W., 2005. Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. Health Place 11 (2), 131–146. https://doi.org/10.1016/j.healthplace.2004.02.003.

Wong, M.S., Steers, W.N., Hoggatt, K.J., Ziaeian, B., Washington, D.L., 2020. Relationship of neighborhood social determinants of health on racial/ethnic mortality disparities in US veterans—mediation and moderating effects. Health Serv. Res. 55 (Suppl. 2), 851–862. https://doi.org/10.1111/1475-6773.13547.

Zhang, Y., Zhang, Y., Sholle, E., et al., 2020. Assessing the impact of social determinants of health on predictive models for potentially avoidable 30-day readmission or death. PLoS One 15 (6), e0235064. https://doi.org/10.1371/journal.pone.0235064.