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

Leigh G. Goetschius,[1] Morgan Henderson,[1,2] Fei Han,[1,3] Dillon Mahmoudi,[4] Chad Perman,[5] Howard Haft[5,*], Ian Stockwell[6,7]

1. The Hilltop Institute at the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA
2. Department of Economics, College of Arts, Humanities, and Social Sciences, UMBC, Baltimore, MD, 21250, USA
3. Department of Computer Science and Electrical Engineering, College of Engineering and Information Technology, UMBC, Baltimore, MD, 21250, USA
4. Department of Geography and Environmental Systems, College of Arts, Humanities, and Social Sciences, UMBC, Baltimore, MD, USA
5. Program Management Office for the Maryland Primary Care Program, Maryland Department of Health, Baltimore, MD, USA
6. Department of Information Systems, College of Engineering and Information Technology, UMBC, Baltimore, MD, 21250, USA
7. Erickson School of Aging Studies, UMBC, Baltimore, MD, 21228, USA

* Independent Consultant

Acknowledgements:

Funding Information: Maryland Department of Health and Mental Hygiene.

Corresponding Author:

Dr. Leigh Goetschius
lgoetschius@hilltop.umbc.edu

DOI: 10.1016/j.socscimed.2023.115943

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.

Keywords: Social determinants of health; information technology in health; Medicare; geographic/spatial factors/small area variations; primary care; health care disparities

Introduction

Predictive analytics and big data are playing a larger role in health care than ever before.[1] 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.[2–4] Social determinants of health (SDOH), such as income, education, and neighborhood conditions, are important to include in these algorithms because research shows they significantly influence health care access, utilization, and overall health.[5–7] 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.[8] 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[9]). As a result, well-formed predictive models often link administrative claims with publicly available, aggregated SDOH data.[10–13]

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.[14] However, there can be substantial

variability of SDOH within ZCTAs,[15] and ZCTAs do not perfectly approximately ZIP codes.[16] More granular, Census Tract-level metrics may provide a more accurate representation of an individual's proximal environment,[13] 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,[15,17] the limited work comparing their ability to predict health outcomes suggests similar predictive performance across different granularity levels.[14,18,19] We are aware of no research to date, however, 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).[4] 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).[4,20] 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.[21] Risk scores and reasons for risk are deployed monthly to more than 475 primary care practices with approximately 2,000 providers (i.e., physicians, nurse practitioners, nurse specialists, physician assistants) across the state.[22] This model has been described in more detail previously.[4,23]

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.[24]

Methods

Sample and Data

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).

Clinical Prediction Model Features

In this clinical prediction model, the AH event outcome was composite of 10 conditions (prevention quality indicators, or PQIs) determined to be potentially preventable with timely, high-quality outpatient care[25] by the Agency for Healthcare Research and Quality, including short- and long-term diabetes complications, hypertension, and asthma, among others.[26] 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.[20]

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-based, and pharmacy use (see Supplemental Methods Table 2 for a complete list of risk factors and previously published work[4] 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.

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.[20] 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).[27,28] 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).[29]

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)[30] 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.

Analytic Strategy

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.

Predictive model for avoidable hospitalization events

We ran six discrete time survival models predicting whether an individual had an avoidable hospitalization or ED visit (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, COPD or asthma, hypertension, heart failure, and bacterial pneumonia.[31] 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.

Comparison of model fit and predictive performance

We compared the six models based on statistical fit in the training data (Akaike Information Criteria – AIC[32]). 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,[33] Gini Index, and cumulative percentage of AH events incurred by beneficiaries with the top 10% of risk scores[34]). 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.[35] 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-squared test.[36] 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 (*table1 v.1.4.1*).

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).

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).

Results

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/Alaskan Native, 1.1% were Hispanic/Latinx, 1.5% were categorized as "Other"[1], 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. Last, 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.

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). Figure 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. Figure 1 color codes each correlation by effect size (small:

[1] "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).

gray; medium: gold; large: teal).[37] 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).

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).

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 eligibility for Medicaid, a proxy for lower income (Figure 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 (Figure 2). The attributed risk was reduced more 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 (Figure 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.

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 whether the extent to which the *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.[4] 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 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.[38,39] 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.[8,40] Not all research has found that neighborhood SDOH explain associations between health disparities and outcomes,[41] however, suggesting that the relationships among 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[13];

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.[42] These supplemental payments are provided for the care of beneficiaries with high clinical risk *and* high neighborhood deprivation (based on the ADI).[42] 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.[14,18,19] 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 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;[43] however, future research is needed that explicitly tests 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[44] and Medicaid claims.[45]

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.[46] 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.[47] 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.

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.[4] 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.[48] This limitation in the data is a

barrier for understanding health disparities based on race or ethnicity; however, efforts are underway by CMS to improve measurement of these variables in the future. Additionally, CCLF claims do not reliably collect individual-level SDOH information,[9] 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[20] 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,[25,26] 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.

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.

References

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

2. Rudolph JL, Zanin NM, Jones RN, et al. Hospitalization in Community-Dwelling Persons with Alzheimer's Disease: Frequency and Causes. *J Am Geriatr Soc*. 2010;58(8):1542-1548. doi:10.1111/j.1532-5415.2010.02924.x

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

4. Henderson M, Han F, Perman C, Haft H, Stockwell I. Predicting Avoidable Hospital Events in Maryland. *Health Services Research*. Published online October 14, 2021. doi:10.1111/1475-6773.13891

5. Pickett KE, Wilkinson RG. Income inequality and health: A causal review. *Social Science & Medicine*. 2015;128:316-326. doi:10.1016/j.socscimed.2014.12.031

6. Salgado M, Madureira J, Mendes AS, Torres A, Teixeira JP, Oliveira MD. Environmental determinants of population health in urban settings. A systematic review. *BMC Public Health*. 2020;20:853. doi:10.1186/s12889-020-08905-0

7. Abbass I, Revere L, Mitchell J, Appari A. Medication Nonadherence: The Role of Cost, Community, and Individual Factors. *Health Serv Res*. 2017;52(4):1511-1533. doi:10.1111/1475-6773.12547

8. Gervasi SS, Chen IY, Smith-McLallen A, et al. The Potential For Bias In Machine Learning And Opportunities For Health Insurers To Address It. *Health Affairs*. 2022;41(2):212-218. doi:10.1377/hlthaff.2021.01287

9. Truong HP, Luke AA, Hammond G, Wadhera RK, Reidhead M, Joynt Maddox KE. Utilization of Social Determinants of Health ICD-10 Z-Codes Among Hospitalized Patients in the United States, 2016–2017. *Medical Care*. 2020;58(12):1037-1043. doi:10.1097/MLR.0000000000001418

10. Hanley GE, Morgan S. On the validity of area-based income measures to proxy household income. *BMC Health Services Research*. 2008;8(1):79. doi:10.1186/1472-6963-8-79

11. Geronimus AT, Bound J, Neidert LJ. On the Validity of Using Census Geocode Characteristics to Proxy Individual Socioeconomic Characteristics. *Journal of the American Statistical Association*. 1996;91(434):529-537. doi:10.1080/01621459.1996.10476918

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

13. Moss JL, Johnson NJ, Yu M, Altekruse SF, Cronin KA. 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 Metr*. 2021;19:1. doi:10.1186/s12963-020-00244-x

14. Berkowitz SA, Traore CY, Singer DE, Atlas SJ. Evaluating Area-Based Socioeconomic Status Indicators for Monitoring Disparities within Health Care Systems: Results from a Primary Care Network. *Health Serv Res*. 2015;50(2):398-417. doi:10.1111/1475-6773.12229

15. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. 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. *American Journal of Epidemiology*. 2002;156(5):471-482. doi:10.1093/aje/kwf068

16. Grubesic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *Int J Health Geogr*. 2006;5(1):58. doi:10.1186/1476-072X-5-58

17. Tian N, Goovaerts P, Zhan FB, Wilson JG. Identification of racial disparities in breast cancer mortality: does scale matter? *Int J Health Geogr*. 2010;9:35. doi:10.1186/1476-072X-9-35

18. Thomas AJ, Eberly LE, Davey Smith G, Neaton JD. ZIP-Code-based versus Tract-based Income Measures as Long-Term Risk-adjusted Mortality Predictors. *American Journal of Epidemiology*. 2006;164(6):586-590. doi:10.1093/aje/kwj234

19. Lovasi GS, Moudon AV, Smith NL, et al. Evaluating Options for Measurement of Neighborhood Socioeconomic Context: Evidence from a Myocardial Infarction Case-Control Study. *Health Place*. 2008;14(3):453-467. doi:10.1016/j.healthplace.2007.09.004

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

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

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

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

24. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885

25. Billings J, Zeitel L, Lukomnik J, Carey TS, Blank AE, Newman L. Impact Of Socioeconomic Status On Hospital Use In New York City. *Health Affairs*. 1993;12(1):162-173. doi:10.1377/hlthaff.12.1.162

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

27. Din A, Wilson R. Crosswalking ZIP Codes to Census Geographies: Geoprocessing the U.S. Department of Housing & Urban Development's ZIP Code Crosswalk Files. *Cityscape*. 2020;22(1). Accessed June 29, 2021. https://www.huduser.gov/portal/periodicals/cityscpe/vol22num1/ch12.pdf

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

29. 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

30. Jordahl K, Bossche JVD, Fleischmann M, et al. *Geopandas*. Zenodo; 2020. doi:10.5281/ZENODO.3946761

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

32. Portet S. A primer on model selection using the Akaike Information Criterion. *Infect Dis Model*. 2020;5:111-128. doi:10.1016/j.idm.2019.12.010

33. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2

34. Llorca J, Delgado-Rodríguez M. Visualising exposure-disease association: the Lorenz curve and the Gini index. *Med Sci Monit*. 2002;8(10):MT193-197.

35. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six Features Of Medicare Coordinated Care Demonstration Programs That Cut Hospital Admissions Of High-Risk Patients. *Health Affairs*. 2012;31(6):1156-1166. doi:10.1377/hlthaff.2012.0393

36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988;44(3):837. doi:10.2307/2531595

37. Cohen J. A Power Primer. *Psychol Bull*. 1992;112(1):155-159. doi:10.1037//0033-2909.112.1.155

38. Wong MS, Steers WN, Hoggatt KJ, Ziaeian B, Washington DL. Relationship of neighborhood social determinants of health on racial/ethnic mortality disparities in US veterans—Mediation and moderating effects. *Health Serv Res*. 2020;55(Suppl 2):851-862. doi:10.1111/1475-6773.13547

39. Skolarus LE, Sharrief A, Gardener H, Jenkins C, Boden-Albala B. Considerations in Addressing Social Determinants of Health to Reduce Race/Ethnic Disparities in Stroke Outcomes in the United States. *Stroke*. 2020;51(11):3433-3439. doi:10.1161/STROKEAHA.120.030426

40. Braveman PA, Arkin E, Proctor D, Kauh T, Holm N. Systemic And Structural Racism: Definitions, Examples, Health Damages, And Approaches To Dismantling. *Health Affairs*. 2022;41(2):171-178. doi:10.1377/hlthaff.2021.01394

41. Piccolo RS, Duncan DT, Pearce N, McKinlay JB. 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*. 2015;130:79-90. doi:10.1016/j.socscimed.2015.01.041

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

43. Braveman PA, Cubbin C, Egerter S, Williams DR, Pamuk E. Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us. *Am J Public Health*. 2010;100(Suppl 1):S186-S196. doi:10.2105/AJPH.2009.166082

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

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

46. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*. 2003;27(3):425. doi:10.2307/30036540

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

48. Bierman AS, Lurie N, Collins KS, Eisenberg JM. Addressing Racial And Ethnic Barriers To Effective Health Care: The Need For Better Data. *Health Affairs*. 2002;21(3):91-102. doi:10.1377/hlthaff.21.3.91

Table 1

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

Risk Factor	Source	Year
Population; Population Growth[1]; Population Density[2]	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[3]	CMS	2021
Hospital Beds/1000 Residents[3]	CMS	2021
Has a VA Clinic or Center	VA	2021
Primary Care Providers/1000 Residents[3]	NPI	2021
Internists/1000 Residents[3]	NPI	2021
Specialists/1000 Residents[3]	NPI	2021
Social Workers/1000 Residents[3]	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

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

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

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

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.

Table 3

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

AH ~ Demographics + Social and Environmental + Individual Utilization					
Comparison	**Estimate**	**Standard Error**	**95% Confidence Interval**	***Chi-Square***	***p***
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					
Comparison	**Estimate**	**Standard Error**	**95% Confidence Interval**	***Chi-Square***	***p***
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)

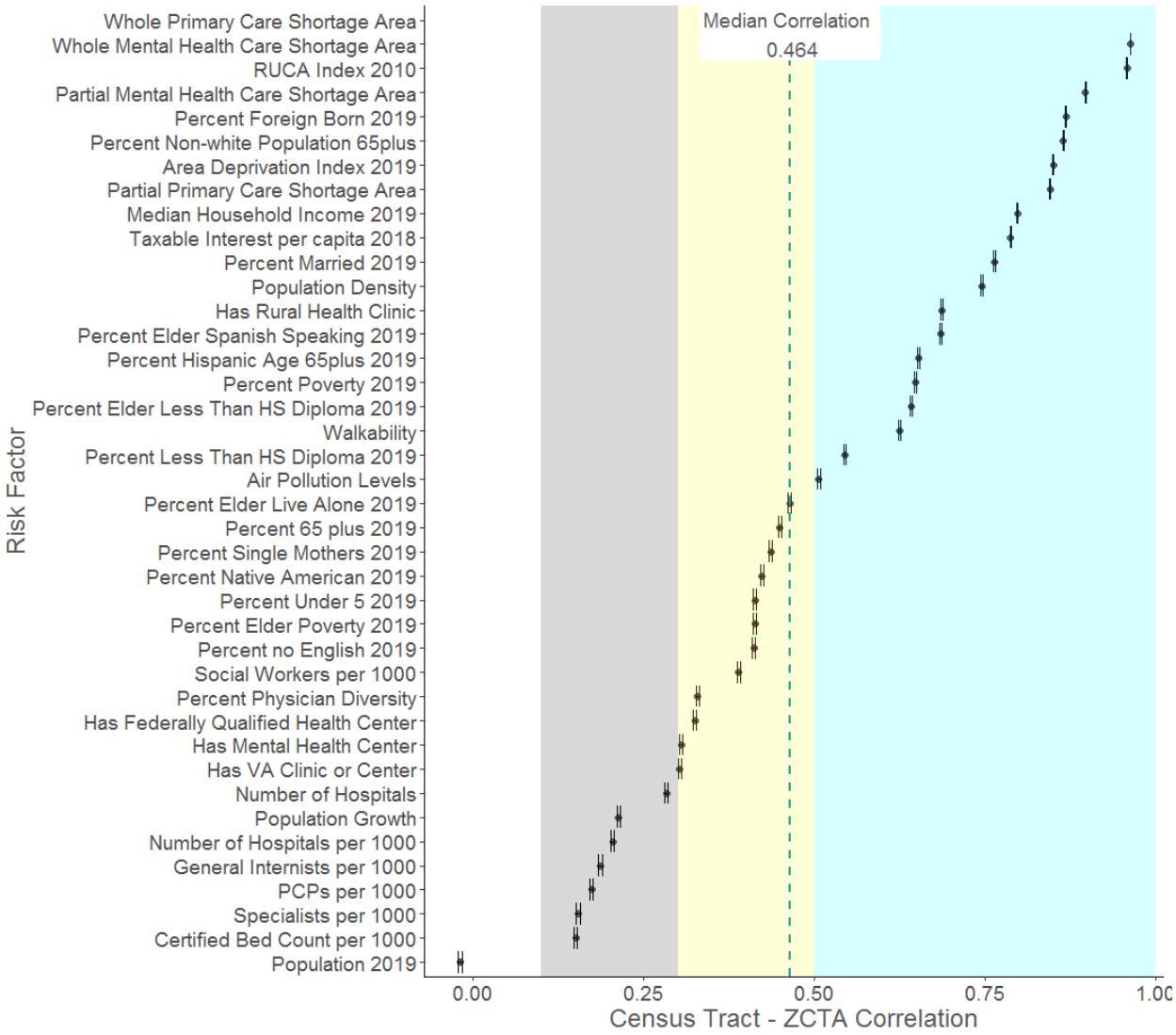

Figure 1: Average correlation and 95% confidence intervals between Census Tract and ZCTA versions of social and environmental risk factors. Correlation estimates are color coded by effect size: estimates in the gray box are small (r=0.1-0.299); estimates in the gold box are medium (r=0.3-0.499); estimates 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.

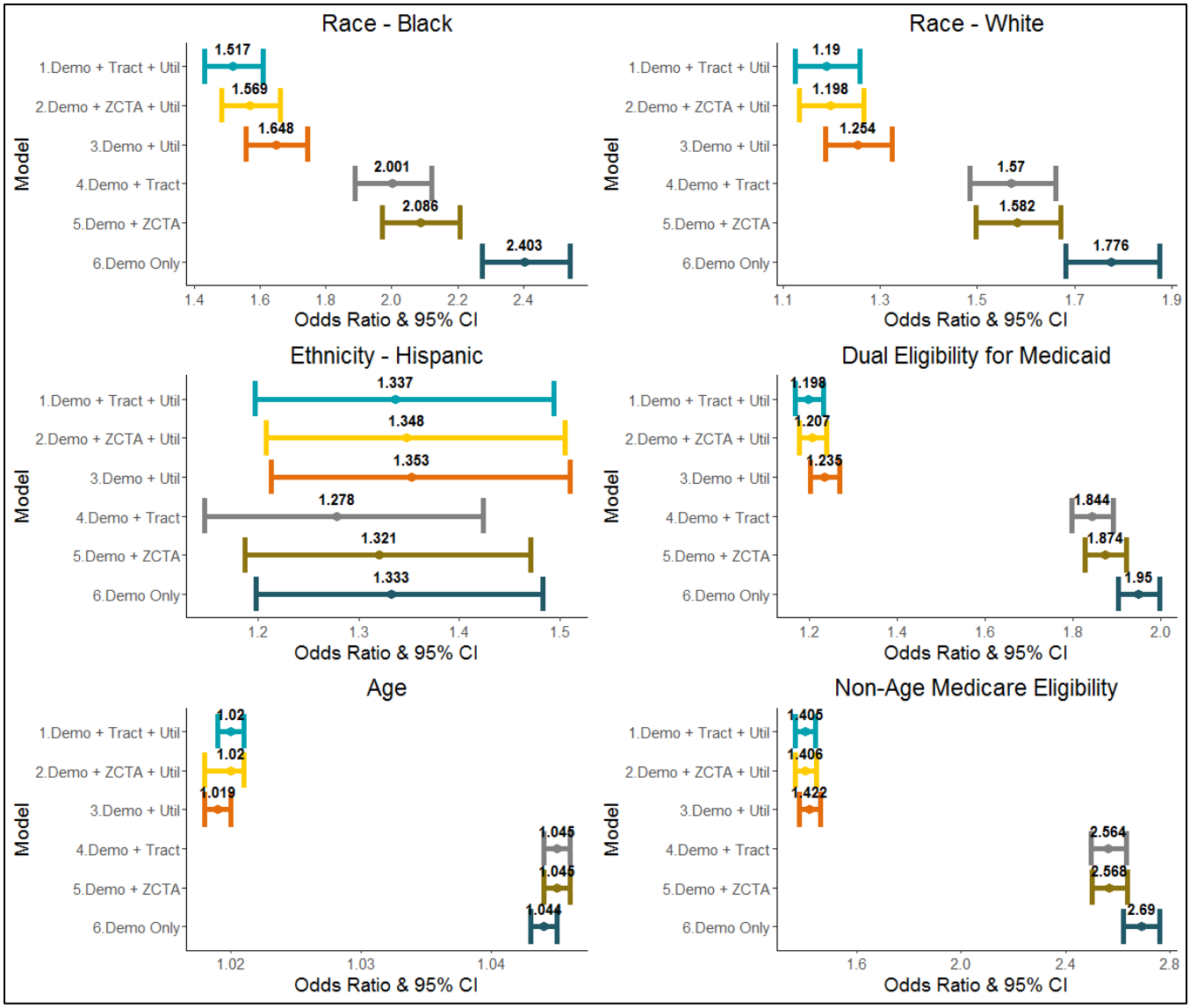

Figure 2: Average odds ratio and 95% confidence intervals for all significant demographic risk factors by model. Including area-level SDOH features, especially Census Tract-level features, reduced the relative risk attributed to the Race and dual-eligibility for Medicaid features.

Supplemental Methods

Utilization and Demographics Risk Factors

Utilization-based risk factors were created from Medicare Part A, B, and D claims (Supplemental Table 2). Medicare part A claims were used to create risk factors that index information on admissions over the past 12 months; nursing home stays over the past 12 months; and certain procedures. Additionally, these claims were used to construct the diagnostic condition flags. These condition flags rely on diagnostic information from hospital, nursing home, physician, and lab claims in conjunction with Chronic Conditions Data Warehouse (CCW) coding specifications to generate beneficiary-level risk factors that represent underlying disease states.[1] Part B claims were used to create risk factors that index utilization of certain services (such as vaccinations, lab tests, or J-code procedures), place of service (for example, urgent care or rural health clinic), and provider specialty (for example, endocrinology or oncology). Part B claims were also used to create an index for beneficiary's primary care utilization and continuity of care. Part D claims were used to create risk factors related to prescription drug use. Demographic risk factors were created from Medicare Beneficiary Demographics files (Supplemental Table 2). To model the impact of the COVID-19 pandemic on healthcare utilization, we also included dummy-variables to flag months within two time periods: "early COVID-19" (from March 2020 – June 2020); "later COVID-19" (from July 2020 – present).

SDOH Risk Factors

All area-level SDOH risk factors came from publicly available data sources (see Supplemental Table 3). The annotated code used to create the census tract and ZCTA versions of each variable is available upon request in SAS and R formats.

Constructed Variables

The majority of the area-level SDOH variables did not need to be constructed (apart from the transformation between geographic units as noted in Table 1 in the main manuscript); however, the variables created from the NPI database, individual-level ACS data, the CMS Provider of Services file, and the VA Facility Listing data required additional steps to create.

Provider Counts. Provider Counts (primary care, general internists, specialists, social workers) per 1000 residents were calculated using data from the National Provider Identifier (NPI) database. Codes for the different provider types were determined using the Health Care Provider Taxonomy Code Set from the National Uniform Claim Committee (Supplemental Table 4).[2] NPI records provide office addresses for each provider. We used the zip code from the reported address to create the ZCTA estimates and geocoded the addresses to get the appropriate census tract for each provider. We then aggregated the providers by type and by geographic location and divided by the population estimates for each geographic area from the 2019 ACS (Table B01003).

Physician Diversity. We used data from the 2019 individual-level ACS data stored in the Integrated Public Use Microdata Series (IPUMS)[3] to estimate the percentage of physicians that are a racial or ethnic minority. Individual-level occupation data were filtered to include only "Physicians and Surgeons" (3060), "Physician Assistants" (3110), and "Nurse practitioners and nurse midwives" (3258). Individuals were coded as a racial or ethnic minority if they were coded

as Black, American Indian or Alaska Native, Hispanic/Latinx, Other, or multiple races within the IPUMs data. We did not classify Asian Americans as a racial or ethnic minority group for this measure based on previous work.[4] Individual-level data was then aggregated to the county-level (most precise geographic level store in the IPUMS data) and we then calculated the percentage of physicians that were a racial or ethnic minority. Census tracts were assigned the value of the county they are within and ZCTAs were assigned their value using the HUD county-zip crosswalk file (https://www.huduser.gov/portal/datasets/usps_crosswalk.html).

CMS Provider of Services. We used data from the 2020 CMS Provider of Services files to construct the binary (0/1) indicators for whether a geographic area has a Community Mental Health Center (PRVDR_CTGRY_CD = 19), Federally Qualified Health Center (PRVDR_CTGRY_CD = 21), Rural Health Clinic (PRVDR_CTGRY_CD = 12), or a For Profit Hospital (GNRL_CNTL_TYPE_CD = 4). We also used this data to calculate the number of hospitals per 1000 residents (PRVDR_CTGRY_CD = 01, divided by the 2019 ACS population estimates) and the number of hospital beds per 1000 residents (CRTFD_BED_CNT). The CMS Provider of Services file contains data on characteristics of hospitals and other types of healthcare facilities, including the name and address of the facility and the type of Medicare services the facility provides. We geocoded each location and aggregated the data by geographic location to get each of the area-level estimates.

VA Facility Listing. We used data from the VA Facility Listing to create the binary indicator (0/1) for whether a geographic area had a VA Medical Center (outpatient clinic, hospital, vet center). We geocoded the address for each location and aggregated the data by geographic location.

Imputation of missing variables

Imputation of missing variables was done for the Percent Physician Diversity per geographic unit variable because, in the ACS public-use microdata (from IPUMS), counties were not identified from 1950 onwards. Therefore, IPUMS assigns county based on other low-level geographic identifiers which is not possible for all counties.[5] This resulted in 10.72% of census tracts and 4.63% of ZCTAs missing percent physician diversity. To avoid large amounts of missing data, percent physician diversity was imputed from a weighted average of physician diversity from the counties in that state.

Taxable interest per capita and area deprivation index (ADI) were missing for 5.53% and 1.57% of ZCTAs in Maryland, respectively. Therefore, data for missing ZCTAs was imputed, when possible, based on a weighted average of those metrics from the other ZCTAs within the same zip code sorting area (first three digits of ZCTA).

Impact of COVID-19 Pandemic on the Model Training

We temporarily transitioned to training the production model on a semi-monthly cadence starting in Summer 2020 to account for altered utilization patterns of health services and continued that increased training frequency until May 2022. We believe that the steps we took to mitigate the effects of the pandemic on the model were successful because we have not observed any systematic changes to model performance during the COVID-19 pandemic relative to the pre-pandemic period.[6]

Model Fit, Discrimination, and Performance Measures

Akaike information criterion (AIC). We evaluated model fit using AIC. AIC quantifies the relative information lost due to prediction error, so lower AIC values are better and AICs can be compared across different versions of models predicting the same outcome in the same sample to determine which model is the best fit for the data.[7]

Concentration Curve and Gini Index. We used a concentration curve to evaluate model calibration.[8] Concentration curves estimate the cumulative share of all outcome events incurred by the riskiest patients, so it is possible to determine the share of all outcome events occurring for individuals above different risk thresholds. To estimate the concentration curve, the patient cohort is ordered from most to least risky (in terms of predicted risk) on the X axis, and the fraction of total outcome events captured by the riskiest patients on the Y axis. We estimate the percent of outcome events, in this case avoidable hospital events, incurred by the top 10% riskiest patients. Concentration curves can be summarized by a Gini coefficient, which is a measure of 0 to 1, that can be interpreted as an index of risk concentration in the population. The higher the Gini index, the more concentrated the risk of the outcome event is in a small proportion of persons.[8] A higher Gini coefficient indicates better model fit.

C-Statistic. The discriminatory power of predictive models can be summarized using the c-statistic, which is a measure of the area under the Receiver Operating Characteristic (ROC) curve.[9] The ROC curve plots the true positive rate against the false positive rate for binary classifiers using successive cutoff thresholds and measures how well the model distinguishes between individuals who did and do not experience the outcome of interest. [10]

Comparison of ROC. The area under ROCs can be statistically compared across correlated versions of a model using a non-parametric approach based on the theory of generalized U-statistics that follows a chi-square distribution.[11] We compared the trapezoidal area under the ROCs for the different models in this paper using the ROCCONTRAST statement within the proc logistic statement in SAS (v.9.4). The ROCCONTRAST statement provides the association statistics, and displays the trapezoidal area under the ROC curve, its standard error, and a confidence interval for each model in the comparison.[12]

Identical SDOH Variable Models

To ensure that the specific SDOH variables selected using stepwise variable selection in the main discrete time survival models did not impact the overall pattern of results for the comparison of model fit and predictive performance, we ran a separate set of models where the SDOH predictors were matched for census tract and ZCTA versions of the models. In this set of models, SDOH risk factors were included if they were selected in either the ZCTA or Census Tract version of the model. We did this separately for the models with utilization-based risk factors and without utilization-based risk factors (4 models total).

Supplemental Results

Variables Excluded from Predictive Models

ZCTA-level Models

The variable indexing the amount of air pollution in a given ZCTA (zcta_air_pollution) was dropped because it was highly correlated with the ZCTA population estimate (zcta_pop19, r=0.888) and the ZCTA walkability estimate (zcta_walkability, r=0.959) in our sample. Percent of population 65 years and older that speak Spanish (zcta_pct_eld_spanish_2019) was excluded because it was highly correlated with the percent of population 65 years and older that is Hispanic/Latinx (zcta_pct_hisp_age65plus_2019, r=0.968). ZCTA-level estimates of the number of specialists (zcta_spec_per_1000) and general internists (zcta_gen_insts_per_1000) per 1000 people were excluded due to high correlation with number of primary care providers per 1000 people (zcta_pcps_per_1000; r=0.970 for both).

Census Tract-level Models

Percent of population 65 years and older that speak Spanish (ct_pct_eld_spanish_2019) was excluded because it was highly correlated with the percent of population 65 years and older that is Hispanic/Latinx (ct_pct_hisp_age65plus_2019, r=0.927). Census tract-level estimates of the number of specialists (ct_spec_per_1000) and general internists (ct_gen_insts_per_1000) per 1000 people were excluded due to high correlation with number of primary care providers per 1000 people (ct_pcps_per_1000; *r*=0.983 and *r*=0.982, respectively). Census tract-level estimates of the number of hospitals (ct_num_hosp) and the number of hospital beds per 1000 (ct_certbed_count_per_1000) were excluded due to their correlation with the number of hospitals per 1000 people (ct_num_hosp_per_1000; *r*=0.867 and *r*=0.855, respectively).

Results from Discrete Time Survival Main Models 1-6

The full results from the final discrete time survival models for each combination of variables (Models 1-6) can be found in Supplemental Tables 6-11.

Identical SDOH Variable Models

The results from the discrete time survival models where the SDOH variables were identical across census tract and ZCTA versions of the model followed the same pattern of results as the main model with free variable selection. Comparison of model fit and performance for the matched models can be found in Supplemental Table 15 and the statistical comparison of the area under the ROCs can be found in Supplemental Table 16. Full results from the discrete time survival models can be found in Supplemental Tables 12-14.

Supplemental References

1. Centers for Medicare and Medicaid Services. Chronic Condition Warehouse (CCW) Condition Algorithms. Published online February 2021.

2. National Uniform Claim Committee. Health Care Provider Taxonomy Code Set. Published July 2021. Accessed July 6, 2021. https://taxonomy.nucc.org/

3. Ruggles S, Flood S, Goeken R, et al. *IPUMS USA: Version 9.0*. IPUMS; 2019. https://doi.org/10.18128/D010.V9.0

4. Laditka JN. Physician supply, physician diversity, and outcomes of primary health care for older persons in the United States. *Health & Place*. 2004;10(3):231-244. doi:10.1016/j.healthplace.2003.09.004

5. County (ICPSR code) Description. IPUMS USA. Accessed January 28, 2022. https://usa.ipums.org/usa-action/variables/COUNTYICP#description_section

6. Henderson M, Han F, Perman C, Haft H, Stockwell I. Predicting Avoidable Hospital Events in Maryland. *Health Services Research*. Published online October 14, 2021. doi:10.1111/1475-6773.13891

7. Portet S. A primer on model selection using the Akaike Information Criterion. *Infect Dis Model*. 2020;5:111-128. doi:10.1016/j.idm.2019.12.010

8. Llorca J, Delgado-Rodríguez M. Visualising exposure-disease association: the Lorenz curve and the Gini index. *Med Sci Monit*. 2002;8(10):MT193-197.

9. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2

10. Mauguen A, Begg CB. Using the Lorenz Curve to Characterize Risk Predictiveness and Etiologic Heterogeneity. *Epidemiology*. 2016;27(4):531-537. doi:10.1097/EDE.0000000000000499

11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988;44(3):837. doi:10.2307/2531595

12. SAS Institute Inc. The LOGISTIC Procedure. In: *SAS/STAT® 15.1 User's Guide*. SAS Institute Inc.; 2018:269.

13. Kim DH, Schneeweiss S. Measuring Frailty Using Claims Data for Pharmacoepidemiologic Studies of Mortality in Older Adults: Evidence and Recommendations. *Pharmacoepidemiol Drug Saf*. 2014;23(9):891-901. doi:10.1002/pds.3674

Supplemental Table 1

Demographic information for included sample

	Overall (N=465,749)
Sex	
Female	276,912 (59.5%)
Male	188,837 (40.5%)
Race-Ethnicity	
Asian	9,766 (2.1%)
Black	105,566 (22.7%)
Hispanic/Latinx	5,257 (1.1%)
American Indian	242 (0.1%)
Unknown	19,962 (4.3%)
White	324,956 (69.8%)
Age (in years)	
Mean (SD)	73.1 (10.5)
Median [Min, Max]	73.0 [20.0, 110]
Non-Age Reason for Medicare Eligibility (i.e., disability)	
No	380,929 (81.8%)
Yes	84,820 (18.2%)
Dually Eligible for Medicaid	
No	396,420 (85.1%)
Yes	69,329 (14.9%)
Months Enrolled in Medicare (35 months max)	
Mean (SD)	33.8 (4.41)
Median [Min, Max]	35.0 [4.00, 35.0]
At Least 12 Months of Claims	
No	4650 (1.0%)
Yes	461295 (99.0%)
Census Tract Data Available	
No	10,207 (2.2%)
Yes	455,542 (97.8%)
ZCTA Data Available	

Supplemental Table 1

Demographic information for included sample

	Overall (N=465,749)
No	1,850 (0.4%)
Yes	463,899 (99.6%)

Supplemental Table 2

List of utilization-based risk factors and their source within Medicare CCLF claims.

Variable	Data Source
Number of emergency department visits within the past 6 months	Part A claims
Indicator for original Medicare eligibility for a non-age related cause	Beneficiary Demographics
Indicator for durable medical equipment (DME) use	Part B DME claims
Indicator for frailty[+]	Part A, B, and DME claims
Indicator for sickle cell anemia	Part A and B claims
Indicator for rivaroxaban use	Part D claims
Indicator for dual eligibility with Medicaid	Beneficiary Demographics
Prior admission length of stay	Part A claims
Number of prior admissions	Part A claims
Indicator for prior nursing home stay	Part A claims
Number of heart-related procedures	Part A claims
Indicator for diabetic foot procedure	Part A claims
Indicator for prior readmission	Part A claims
Indicator for no vaccination (flu or pneumonia)	Part B claims
Number of HbA1c tests	Part B claims
Number of lab tests	Part B claims
Indicator for previous conservative diabetic wound procedure	Part B claims
Number of primary care visits	Part B claims
Indicator for prior surgery	Part B claims
Number of urgent care visits	Part B claims
Number of home health visits	Part B claims
Indicator for endocrinologist visit	Part B claims
Indicator for oncologist visit	Part B claims
Number of specialist visits	Part B claims
Number of outpatient visits	Part B claims
Number of rural clinic visits	Part B claims
Discontinuity of primary care - Proportion	Part B claims
Discontinuity of primary care - Index	Part B claims
Continuity of primary care - Duration	Part B claims
Indicator for no mental health use	Part B claims
Indicator for provider administered drug	Part B claims
Indicator for insulin use	Part D claims
Indicator for no anti-diabetes medication use	Part D claims
Indicator for leukotriene receptor modifier use	Part D claims
Indicator for warfarin use	Part D claims
Indicator for no statin use	Part D claims
Indicator for no losartan use	Part D claims
Indicator for no beta blocker use	Part D claims

Indicator for cilostazol use	Part D claims
Indicator for oral corticosteroid use	Part D claims
Indicator for oral antibiotic use	Part D claims
Number of medications	Part D claims
Total health spending	Part A, B, and D claims
Number of previous avoidable hospitalizations	Part A claims
Indicator for hospice enrollment	Beneficiary Demographics
Age	Beneficiary Demographics
Indicator for acquired hypothyroidism*	Part A and B claims
Indicator for acute myocardial infarction*	Part A and B claims
Indicator for anemia*	Part A and B claims
Indicator for asthma*	Part A and B claims
Indicator for atrial fibrillation*	Part A and B claims
Indicator for benign prostatic hyperplasia*	Part A and B claims
Indicator for cataracts*	Part A and B claims
Indicator for chronic kidney disease*	Part A and B claims
Indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis*	Part A and B claims
Indicator for diabetes*	Part A and B claims
Indicator for glaucoma*	Part A and B claims
Indicator for heart failure*	Part A and B claims
Indicator for hip/pelvic fracture*	Part A and B claims
Indicator for hyperlipidemia*	Part A and B claims
Indicator for hypertension	Part A and B claims
Indicator for ischemic heart disease*	Part A and B claims
Indicator for osteoporosis*	Part A and B claims
Indicator for rheumatoid arthritis/osteoarthritis*	Part A and B claims
Indicator for female/male breast cancer*	Part A and B claims
Indicator for colorectal cancer*	Part A and B claims
Indicator for prostate cancer*	Part A and B claims
Indicator for lung cancer*	Part A and B claims
Indicator for endometrial cancer*	Part A and B claims
Indicator for ADHD, conduct disorders, and hyperkinetic syndrome*	Part A and B claims
Indicator for alcohol use disorders*	Part A and B claims
Indicator for anxiety disorders*	Part A and B claims
Indicator for autism spectrum disorders*	Part A and B claims
Indicator for bipolar disorder*	Part A and B claims
Indicator for cerebral palsy*	Part A and B claims
Indicator for cystic fibrosis and other metabolic developmental disorders*	Part A and B claims
Indicator for drug use disorders*	Part A and B claims
Indicator for epilepsy*	Part A and B claims
Indicator for fibromyalgia, chronic pain and fatigue*	Part A and B claims
Indicator for intellectual disabilities and related conditions*	Part A and B claims
Indicator for learning disabilities*	Part A and B claims

Indicator for leukemias and lymphomas*	Part A and B claims
Indicator for liver disease, cirrhosis and other liver conditions*	Part A and B claims
Indicator for migraine and chronic headache*	Part A and B claims
Indicator for mobility impairments*	Part A and B claims
Indicator for multiple sclerosis and transverse myelitis*	Part A and B claims
Indicator for muscular dystrophy*	Part A and B claims
Indicator for obesity*	Part A and B claims
Indicator for other developmental delays*	Part A and B claims
Indicator for peripheral vascular disease*	Part A and B claims
Indicator for personality disorders*	Part A and B claims
Indicator for post-traumatic stress disorder*	Part A and B claims
Indicator for pressure and chronic ulcers*	Part A and B claims
Indicator for sensory (blindness and visual) impairment*	Part A and B claims
Indicator for sensory (deafness and hearing) impairment*	Part A and B claims
Indicator for spina bifida and other congenital anomalies of the nervous system*	Part A and B claims
Indicator for spinal cord injury*	Part A and B claims
Indicator for tobacco use*	Part A and B claims
Indicator for traumatic brain injury and nonpsychotic mental disorders due to brain damage*	Part A and B claims
Indicator for viral hepatitis*	Part A and B claims
Indicator for depression and depressive disorders*	Part A and B claims
Indicator for Alzheimer's disease and related disorders or senile dementia*	Part A and B claims
Indicator for HIV/AIDS*	Part A and B claims
Indicator for stroke/ischemic transient attack*	Part A and B claims
Indicator for schizophrenia and other psychotic disorders*	Part A and B claims
Indicator for arrhythmia	Part A and B claims
Indicator for albuminuria	Part A and B claims
Indicator for peptic ulcer disease	Part A and B claims
Indicator for cerebrovascular disease	Part A and B claims
Indicator for diabetes with complications	Part A and B claims
Indicator for fluid and electrolyte imbalance	Part A and B claims
Indicator for rheumatoid arthritis/collagen vascular disease	Part A and B claims
Indicator for metastatic cancer	Part A and B claims
Indicator for solid tumor without metastasis	Part A and B claims
Indicator for pulmonary circulatory disorder	Part A and B claims
Indicator for gastroesophageal reflux disease	Part A and B claims
Indicator for gastroparesis	Part A and B claims
Indicator for protein-calorie malnutrition	Part A and B claims
Indicator for sleep apnea	Part A and B claims
Indicator for diabetic ulcer	Part A and B claims
Indicator for urinary tract infection	Part A and B claims
Indicator for sepsis	Part A and B claims
Indicator for neuropathy	Part A and B claims
Indicator for retinopathy	Part A and B claims

Indicator for problems with education and literacy	Part A and B claims
Indicator for problems with employment and unemployment	Part A and B claims
Indicator for occupational exposure to risk factors	Part A and B claims
Indicator for problems with housing and economic conditions	Part A and B claims
Indicator for problems with social environment	Part A and B claims
Indicator for problems with upbringing	Part A and B claims
Indicator for other problems with primary support group	Part A and B claims
Indicator for psychosocial problems	Part A and B claims
Indicator for lifestyle problems	Part A and B claims
Indicator for difficulty with life management	Part A and B claims
Indicator for problems with care provider dependency	Part A and B claims
Indicator for pneumonia	Part A and B claims
Indicator for pancreatitis	Part A and B claims
Indicator for respiratory infection	Part A and B claims
Beneficiary gender	Beneficiary Demographics
Beneficiary race	Beneficiary Demographics
Prior hospitalization discharge status	Part A Claims
Prior hospitalization admission type	Part A Claims

[#] AHRQ's 2018 definition of avoidable hospitalization in defining this outcome

* Operationalized using definitions from CCW warehouse

[+] Clinical definition for frailty is derived from Kim and Schneeweiss, 2014[13].

Supplemental Table 3

List of sources for the publicly available SDOH data.

Data Source	Citation
American Community Survey (ACS)	**Age & Sex** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S0101; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Demographics & Housing Estimates** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table DP05; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Educational Attainment** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S1501; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Fertility** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S1301; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Hispanic & Latino Age** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table B01001l; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Household Characteristics** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S1101; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Household Income** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S1901; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Language Spoken at Home** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S1601; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Marital Status** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S1201; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Poverty in Families** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table S1702; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Selected Social Characteristics** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table DP02; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Total Population**

	U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table B01003; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **White Age** U.S. Census Bureau; American Community Survey, 2019 American Community Survey 5-Year Estimates, Table B01001A; using data.census.gov; <https://data.census.gov/cedsci/>; (30 June 2021). **Individual-level Occupation Data:** Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 11.0 [American Community Survey]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0
Area Health Resources File (AHRF)	Area Health Resources Files (AHRF) 2019-2020. US Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Workforce, Rockville, MD.
Centers for Medicare and Medicaid Services (CMS)	Provider of Services (2020), Centers for Medicare and Medicaid Services, Washington, D.C. https://www.cms.gov/research-statistics-data-systems/provider-services-current-files/2020-pos-file
Environmental Protection Agency (EPA)	United States Environmental Protection Agency. (2020). National Walkability Index. Smart Location Database. https://www.epa.gov/smartgrowth/smart-location-mapping#walkability. Accessed 2021-07-06. Centers for Disease Control and Prevention. National Environmental Public Health Tracking Network. Web. Accessed: 2021-07-05. https://data.cdc.gov/d/qjxm-7fny.
Internal Revenue Service (IRS)	Statistics of Income—2018, Individual Income Tax Returns, Internal Revenue Service, Washington, D.C.
National Provider Identified Database (NPI)	National Provider Identifier records (2021), Centers for Medicare and Medicaid Services, Washington, D.C. https://npiregistry.cms.hhs.gov/
United States Department of Agriculture (USDA)	USDA Economic Research Service. (2020). Rural-Urban Commuting Area Codes. Ag Data Commons. https://data.nal.usda.gov/dataset/rural-urban-commuting-area-codes. Accessed 2021-06-11.
Veteran's Affairs (VA)	Veteran's Affairs Facility Listing. (2021, April 2). https://www.va.gov/directory/guide/rpt_fac_list.cfm
Wisconsin School of Medicine and Public Health (WISC)	Kind AJH, Buckingham W. Making Neighborhood Disadvantage Metrics Accessible: The Neighborhood Atlas. New England Journal of Medicine, 2018. 378: 2456-2458. DOI: 10.1056/NEJMp1802313. PMCID: PMC6051533. University of Wisconsin School of Medicine and Public Health. 2019 Area Deprivation Index v3.0. Downloaded from https://www.neighborhoodatlas.medicine.wisc.edu/ 2021-06-11

Supplemental Table 4

List of healthcare provider taxonomy codes used to identify primary care, internal medicine, specialist providers, as well as, social workers in the NPI data.

Provider Type	Healthcare Provider Taxonomy Codes
Primary Care Providers	208D00000X, 207Q00000X, 207R00000X, 208000000X, 207V00000X
Internal Medicine Providers	07R00000X, 207RA0401X, 207RA0000X, 207RA0002X, 207RA0001X, 207RA0201X, 207RC0000X, 207RI0001X, 207RC0001X, 207RC0200X, 207RE0101X, 207RG0100X, 207RG0300X, 207RH0000X, 207RH0003X, 207RI0008X, 207RH0002X, 207RH0005X, 207RI0200X, 207RI0011X, 207RM1200X,207RX0202X,207RN0300X, 207RB0002X, 207RP1001X, 207RR0500X, 207RS0012X, 207RS0010X
Specialist Providers	**Allergy and Immunology**: 207K00000X **Cardiovascular Disease**: 207RC0000X **Dermatology**: 207N00000X **Internal Medicine**: 207RA0401X, 207RA0000X, 207RA0002X, 207RA0001X, 207RA0201X, 207RC0000X, 207RI0001X, 207RC0001X, 207RC0200X, 207RE0101X, 207RG0100X, 207RG0300X, 207RH0000X, 207RH0003X, 207RI0008X, 207RH0002X, 207RH0005X, 207RI0200X, 207RI0011X, 207RM1200X ,207RX0202X,207RN0300X, 207RB0002X, 207RP1001X, 207RR0500X, 207RS0012X, 207RS0010X **Pediatrics**: 208000000X, 2080A0000X, 2080C0008X, 2080C0008X, 2080P0006X, 2080H0002X,2080T0002X, 2080N0001X, 2080P0008X, 2080B0002X, 2080P0201X, 2080P0202X, 2080P0203X, 2080P0204X, 2080P0205X, 2080P0206X, 2080P0207X, 2080P0208X, 2080P0210X, 2080P0214X, 2080P0216X, 2080T0004X, 2080S0012X, 2080S0010X
Social Workers	104100000X, 1041C0700X, 1041S0200X

Note: Codes for each provider type were taken from the https://taxonomy.nucc.org/

Supplemental Table 5

Correlations between the census tract and ZCTA versions of the social and environmental risk factors in the MDPCP sample.

Risk Factor	Correlation	Lower CI	Upper CI
Whole Mental Health Care Shortage Area	0.963	0.963	0.963
RUCA Index 2010	0.958	0.958	0.958
Partial Mental Health Care Shortage Area	0.897	0.897	0.898
Percent Foreign Born 2019	0.868	0.867	0.869
Percent Non-white Population 65plus	0.864	0.864	0.865
Area Deprivation Index 2019	0.850	0.850	0.851
Partial Primary Care Shortage Area	0.845	0.845	0.846
Median Household Income 2019	0.797	0.796	0.798
Taxable Interest per capita 2018	0.788	0.787	0.789
Percent Married 2019	0.764	0.763	0.765
Population Density	0.745	0.744	0.747
Has Rural Health Clinic	0.686	0.685	0.688
Percent Elder Spanish Speaking 2019	0.685	0.683	0.686
Percent Hispanic Age 65plus 2019	0.653	0.651	0.655
Percent Poverty 2019	0.648	0.647	0.650
Percent Elder Less Than HS Diploma 2019	0.642	0.640	0.644
Walkability	0.625	0.623	0.627
Percent Less Than HS Diploma 2019	0.544	0.542	0.546
Air Pollution Levels	0.507	0.505	0.509
Percent Elder Live Alone 2019	0.464	0.462	0.466
Percent 65 plus 2019	0.449	0.447	0.451
Percent Single Mothers 2019	0.436	0.434	0.438
Percent Native American 2019	0.423	0.421	0.425
Percent Under 5 2019	0.413	0.411	0.416
Percent Elder Poverty 2019	0.413	0.410	0.415
Percent no English 2019	0.411	0.409	0.414
Social Workers per 1000	0.389	0.387	0.392
Percent Physician Diversity	0.329	0.326	0.331
Has Federally Qualified Health Center	0.325	0.322	0.327
Has Mental Health Center	0.304	0.302	0.307
Has VA Clinic or Center	0.302	0.300	0.305
Number of Hospitals	0.283	0.281	0.286
Population Growth	0.213	0.210	0.216
Number of Hospitals per 1000	0.204	0.201	0.207
General Internists per 1000	0.186	0.183	0.189
PCPs per 1000	0.173	0.171	0.176
Specialists per 1000	0.154	0.151	0.156
Certified Bed Count per 1000	0.150	0.147	0.153
Population 2019	-0.020	-0.022	-0.017

Supplemental Table 6

Results from the discrete time survival model predicting AH events by demographic, individual utilization history, and census tract-level social and environmental risk factors (Model 1).

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Intercept	-6.455	0.104	3,827.572			
AH NUMBER	0.356	0.005	5,444.792	1.428	1.414	1.441
Antidiabetes Medication	0.083	0.021	14.934	1.087	1.042	1.134
CCW Albuminuria	0.122	0.022	30.229	1.130	1.082	1.181
CCW Alzheimers Disease	0.085	0.015	32.628	1.089	1.057	1.121
CCW Anemia	0.054	0.012	20.744	1.056	1.031	1.080
CCW Anxiety Disorder	0.121	0.013	83.743	1.128	1.100	1.158
CCW Arrhythmia	0.186	0.013	197.422	1.205	1.174	1.236
CCW Asthma	0.154	0.016	88.967	1.166	1.130	1.204
CCW Atrial Fibrillation	0.091	0.015	34.679	1.095	1.062	1.129
CCW Cataract	-0.134	0.015	81.879	0.874	0.849	0.900
CCW Cerebral Palsy	0.395	0.082	23.443	1.484	1.265	1.740
CCW Chronic Kidney Disease	0.204	0.013	233.485	1.226	1.195	1.259
CCW COPD	0.487	0.013	1,395.597	1.627	1.586	1.669
CCW Depression and Depressive Disorders	0.063	0.013	23.501	1.065	1.038	1.093
CCW Diabetes	0.138	0.018	55.159	1.147	1.107	1.190
CCW Diabetes with Complications	0.134	0.021	40.593	1.144	1.097	1.192
CCW Fluid and Electrolytes Imbalance	0.186	0.014	179.604	1.204	1.172	1.237
CCW Glaucoma	-0.093	0.016	33.565	0.911	0.883	0.940
CCW Heart Failure	0.449	0.013	1,114.790	1.566	1.526	1.608

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
CCW Hyperlipidemia	-0.141	0.012	128.439	0.869	0.848	0.890
CCW Hypertension	0.308	0.016	363.125	1.360	1.318	1.404
CCW Intellectual Disabilities and Related Conditions	0.306	0.051	36.052	1.358	1.229	1.500
CCW Ischemic Heart Disease	0.142	0.012	140.518	1.153	1.126	1.180
CCW Lung Cancer	0.186	0.033	31.818	1.204	1.129	1.285
CCW Neuropathy	0.063	0.016	14.895	1.065	1.032	1.100
CCW Peripheral Vascular Disease	0.087	0.013	47.190	1.091	1.064	1.118
CCW Pneumonia	0.060	0.015	15.645	1.062	1.031	1.094
CCW PTSD	0.132	0.038	12.178	1.142	1.060	1.230
CCW Pressure and Chronic Ulcers	0.144	0.018	64.276	1.155	1.115	1.196
CCW Problems Care Provider Dependency	0.268	0.015	331.187	1.307	1.270	1.346
CCW Pulmonary Circulatory Disorder	0.110	0.016	49.985	1.116	1.082	1.150
CCW Rheumatoid Arthritis/Osteoarthritis	-0.041	0.011	13.527	0.960	0.939	0.981
CCW Respiratory Infection	0.102	0.013	63.803	1.107	1.080	1.135
CCW Retinopathy	0.546	0.055	99.552	1.726	1.551	1.922
CCW Sepsis	-0.141	0.018	58.847	0.868	0.838	0.900
CCW Solid Tumor without Metastasis	-0.060	0.015	15.508	0.941	0.913	0.970
CCW Tobacco Use	0.224	0.015	219.037	1.251	1.215	1.289
CCW Urinary Tract Infection	0.297	0.013	543.411	1.346	1.313	1.380

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Foot Toe Amputation	0.246	0.060	16.619	1.278	1.136	1.438
HbA1c	-0.022	0.005	19.029	0.978	0.968	0.988
Hospice	0.444	0.038	139.728	1.558	1.448	1.677
Insulin	0.236	0.017	190.806	1.267	1.225	1.310
Leukotrine receptor modifier use	0.082	0.021	15.219	1.086	1.042	1.132
Oral Antibiotics	0.116	0.011	106.226	1.122	1.098	1.147
Oral Corticosteroids	0.100	0.018	33.001	1.106	1.068	1.144
Outpatient Visits	0.003	0.000	87.621	1.003	1.002	1.004
PCP Continuity Proportion	0.219	0.026	70.765	1.244	1.183	1.309
Previous Diabetic Wound	0.170	0.027	38.999	1.185	1.124	1.250
Prior Admit Count	-0.056	0.005	105.232	0.945	0.935	0.956
Prior Admit - Length of Stay	-0.005	0.001	13.398	0.995	0.993	0.998
Prior Admit Type - Emergency	0.310	0.019	268.563	1.363	1.313	1.414
Prior Admit Type - Urgent	0.211	0.040	27.797	1.235	1.142	1.335
Prior Discharge - Home	0.188	0.019	102.551	1.207	1.164	1.252
Prior Discharge - Other	0.498	0.075	44.542	1.646	1.422	1.905
Prior Nursing Home Stay	-0.130	0.021	38.642	0.878	0.843	0.915
Prior Surgery	-0.121	0.013	89.361	0.886	0.864	0.908
Provider Administered Drug	0.078	0.012	42.382	1.081	1.056	1.106
Age	0.020	0.001	922.856	1.020	1.019	1.021
Dual Eligible for Medicaid	0.182	0.014	179.014	1.200	1.168	1.232
Race - Black	0.413	0.029	196.601	1.511	1.426	1.600
Race - Hispanic	0.282	0.057	24.935	1.326	1.187	1.482

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Race - White	0.170	0.028	35.772	1.185	1.121	1.253
Medicare Non-Age Eligibility Reason	0.339	0.015	549.702	1.404	1.365	1.444
COVID - Early	-0.587	0.016	1,409.372	0.556	0.539	0.573
COVID - Late	-0.425	0.011	1,457.034	0.654	0.639	0.668
CT Air Pollution	-0.147	0.008	348.763	0.864	0.851	0.877
CT Median Income	0.000	0.000	39.594	1.000	1.000	1.000
CT Percent Less than HS Diploma	0.002	0.000	28.381	1.002	1.001	1.003
CT Percent Married	-0.002	0.001	15.504	0.998	0.996	0.999

Supplemental Table 7

Results from the discrete time survival model predicting AH events by demographic, individual utilization history, and ZCTA-level social and environmental risk factors (Model 2).

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Intercept	-7.940	0.070	12,943.925			
AH NUMBER	0.360	0.005	5,724.024	1.433	1.420	1.447
Antidiabetes Medication	0.078	0.021	13.663	1.082	1.038	1.128
CCW Albuminuria	0.124	0.022	31.622	1.132	1.084	1.182
CCW Alzheimers Disease	0.082	0.015	31.090	1.086	1.055	1.117
CCW Anemia	0.052	0.012	19.787	1.054	1.030	1.078
CCW Anxiety Disorder	0.123	0.013	88.956	1.131	1.102	1.160
CCW Arrhythmia	0.180	0.013	187.576	1.197	1.166	1.228
CCW Asthma	0.155	0.016	91.884	1.167	1.131	1.205
CCW Atrial Fibrillation	0.096	0.015	39.406	1.101	1.068	1.134
CCW Cataract	-0.128	0.015	76.738	0.879	0.854	0.905
CCW Cerebral Palsy	0.421	0.072	34.185	1.523	1.323	1.754
CCW Chronic Kidney Disease	0.210	0.013	251.580	1.233	1.202	1.266
CCW COPD	0.491	0.013	1,448.773	1.634	1.593	1.676
CCW Depression and Depressive Disorders	0.062	0.013	22.961	1.064	1.037	1.091
CCW Diabetes	0.136	0.018	55.210	1.146	1.105	1.188
CCW Diabetes with Complications	0.134	0.021	41.380	1.144	1.098	1.191
CCW Fluid and Electrolytes Imbalance	0.184	0.014	180.055	1.202	1.170	1.235
CCW Glaucoma	-0.096	0.016	36.054	0.909	0.881	0.938
CCW Heart Failure	0.451	0.013	1,149.757	1.569	1.529	1.611

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
CCW Hyperlipidemia	-0.139	0.012	129.326	0.870	0.849	0.891
CCW Hypertension	0.303	0.016	363.204	1.354	1.312	1.397
CCW Intellectual Disabilities and Related Conditions	0.250	0.047	28.466	1.283	1.171	1.407
CCW Ischemic Heart Disease	0.139	0.012	138.049	1.150	1.123	1.177
CCW Lung Cancer	0.193	0.033	34.772	1.213	1.137	1.293
CCW Neuropathy	0.060	0.016	13.403	1.061	1.028	1.096
CCW Peripheral Vascular Disease	0.088	0.013	49.690	1.092	1.066	1.119
CCW Pneumonia	0.060	0.015	15.720	1.062	1.031	1.093
CCW PTSD	0.141	0.037	14.514	1.151	1.071	1.238
CCW Pressure and Chronic Ulcers	0.147	0.018	68.754	1.158	1.119	1.199
CCW Problems Care Provider Dependency	0.268	0.015	336.617	1.307	1.270	1.345
CCW Pulmonary Circulatory Disorder	0.102	0.015	44.047	1.107	1.075	1.141
CCW Rheumatoid Arthritis/Osteoarthritis	-0.039	0.011	12.438	0.962	0.941	0.983
CCW Respiratory Infection	0.101	0.013	64.234	1.107	1.079	1.134
CCW Retinopathy	0.530	0.054	97.178	1.699	1.529	1.888
CCW Sepsis	-0.137	0.018	57.238	0.872	0.841	0.903
CCW Solid Tumor without Metastasis	-0.068	0.015	19.731	0.934	0.907	0.963
CCW Tobacco Use	0.226	0.015	230.088	1.254	1.218	1.292
CCW Urinary Tract Infection	0.298	0.013	559.198	1.347	1.314	1.381

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Foot Toe Amputation	0.213	0.060	12.684	1.237	1.100	1.391
HbA1c	-0.022	0.005	19.181	0.978	0.968	0.988
Hospice	0.442	0.037	140.674	1.555	1.446	1.673
Insulin	0.238	0.017	199.343	1.269	1.228	1.312
Leukotrine receptor modifier use	0.078	0.021	13.938	1.081	1.038	1.126
Oral Antibiotics	0.114	0.011	106.376	1.121	1.097	1.146
Oral Corticosteroids	0.095	0.017	29.970	1.099	1.063	1.137
Outpatient Visits	0.003	0.000	92.692	1.003	1.003	1.004
PCP Continuity Proportion	0.209	0.026	66.348	1.232	1.172	1.296
Previous Diabetic Wound	0.172	0.027	40.775	1.188	1.127	1.252
Prior Admit Count	-0.058	0.005	115.151	0.944	0.934	0.954
Prior Admit - Length of Stay	-0.004	0.001	11.492	0.996	0.993	0.998
Prior Admit Type - Emergency	0.307	0.019	270.270	1.360	1.311	1.411
Prior Admit Type - Urgent	0.241	0.039	37.865	1.273	1.179	1.375
Prior Discharge - Home	0.187	0.018	103.456	1.205	1.163	1.250
Prior Discharge - Other	0.491	0.074	44.769	1.635	1.416	1.888
Prior Nursing Home Stay	-0.130	0.021	39.278	0.879	0.844	0.915
Prior Surgery	-0.123	0.013	94.757	0.884	0.862	0.906
Provider Administered Drug	0.082	0.012	47.921	1.085	1.060	1.111
Age	0.019	0.001	930.482	1.020	1.018	1.021
Dual Eligible for Medicaid	0.188	0.013	196.750	1.207	1.176	1.239

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Race - Black	0.451	0.029	241.436	1.569	1.483	1.661
Race - Hispanic	0.299	0.056	28.477	1.348	1.208	1.505
Race - White	0.181	0.028	40.903	1.198	1.134	1.266
Medicare Non-Age Eligibility Reason	0.341	0.014	569.219	1.406	1.367	1.446
COVID - Early	-0.592	0.016	1,460.607	0.553	0.536	0.570
COVID - Late	-0.423	0.011	1,478.078	0.655	0.641	0.669
ZCTA Median Income	0.000	0.000	47.554	1.000	1.000	1.000
ZCTA Mental Health Care Shortage Area - Partial	-0.043	0.013	10.993	0.958	0.934	0.983
ZCTA Percent 65+ Less than HS Diploma	0.006	0.001	53.382	1.006	1.005	1.008
ZCTA Percent Physician Diversity	0.000	0.000	59.776	1.000	1.000	1.000
ZCTA Primary Care Shortage Area - Whole	0.373	0.108	11.840	1.452	1.174	1.796
ZCTA Walkability	-0.001	0.000	57.796	0.999	0.999	0.999

Supplemental Table 8

Results from the discrete time survival model predicting AH events by demographic and individual utilization history (Model 3).

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Intercept	-8.154	0.063	16,647.942	1.440	1.426	1.453
AH Number	0.364	0.005	5,894.411	1.084	1.040	1.130
Antidiabetes Medication	0.080	0.021	14.377	1.132	1.084	1.182
CCW Albuminuria	0.124	0.022	31.750	1.090	1.059	1.122
CCW Alzheimers Disease	0.086	0.015	33.892	1.052	1.028	1.076
CCW Anemia	0.050	0.012	18.517	1.135	1.107	1.165
CCW Anxiety Disorder	0.127	0.013	94.416	1.197	1.167	1.228
CCW Arrhythmia	0.180	0.013	188.958	1.165	1.129	1.202
CCW Asthma	0.152	0.016	89.746	1.106	1.074	1.140
CCW Atrial Fibrillation	0.101	0.015	43.772	1.086	1.038	1.136
CCW Bipolar Disorder	0.082	0.023	12.863	0.881	0.856	0.907
CCW Cataract	-0.127	0.015	74.959	1.513	1.314	1.742
CCW Cerebral Palsy	0.414	0.072	33.148	1.238	1.207	1.271
CCW Chronic Kidney Disease	0.214	0.013	263.132	1.649	1.608	1.691
CCW COPD	0.500	0.013	1,510.554	1.064	1.037	1.092
CCW Depression and Depressive Disorders	0.062	0.013	22.131	1.159	1.119	1.202
CCW Diabetes	0.148	0.018	65.427	1.148	1.102	1.196
CCW Diabetes with Complications	0.138	0.021	44.012	1.199	1.167	1.231
CCW Fluid and Electrolytes Imbalance	0.182	0.014	176.670	0.907	0.879	0.935
CCW Glaucoma	-0.098	0.016	38.185	1.572	1.532	1.614
CCW Heart Failure	0.452	0.013	1,165.112	0.877	0.856	0.898

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
CCW Hyperlipidemia	-0.131	0.012	115.606	1.370	1.328	1.413
CCW Hypertension	0.315	0.016	395.005	1.263	1.152	1.384
CCW Intellectual Disabilities and Related Conditions	0.233	0.047	24.955	1.150	1.124	1.177
CCW Ischemic Heart Disease	0.140	0.012	139.554	1.219	1.144	1.300
CCW Lung Cancer	0.198	0.033	36.804	1.065	1.031	1.099
CCW Neuropathy	0.062	0.016	14.855	1.097	1.070	1.124
CCW Peripheral Vascular Disease	0.092	0.013	55.004	1.059	1.028	1.091
CCW Pneumonia	0.057	0.015	14.606	1.158	1.076	1.246
CCW PTSD	0.146	0.037	15.400	1.150	1.110	1.190
CCW Pressure and Chronic Ulcers	0.139	0.018	62.321	1.296	1.259	1.333
CCW Problems Care Provider Dependency	0.259	0.015	318.161	1.102	1.069	1.136
CCW Pulmonary Circulatory Disorder	0.097	0.015	39.872	0.965	0.944	0.986
CCW Rheumatoid Arthritis/Osteoarthritis	-0.036	0.011	10.740	1.107	1.080	1.135
CCW Respiratory Infection	0.102	0.013	65.585	1.690	1.521	1.877
CCW Retinopathy	0.525	0.054	95.304	0.871	0.841	0.903
CCW Sepsis	-0.138	0.018	58.076	0.934	0.906	0.962
CCW Solid Tumor without Metastasis	-0.068	0.015	20.151	1.266	1.229	1.303
CCW Tobacco Use	0.236	0.015	250.175	1.348	1.316	1.382
CCW Urinary Tract Infection	0.299	0.013	566.062	1.219	1.085	1.370
Foot Toe Amputation	0.198	0.060	11.068	0.969	0.960	0.979
HbA1c	-0.031	0.005	38.918	1.556	1.447	1.673

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Hospice	0.442	0.037	141.649	1.272	1.231	1.315
Insulin	0.241	0.017	203.997	1.087	1.044	1.132
Leukotrine receptor modifier use	0.084	0.021	16.078	1.083	1.046	1.121
Mental Health Use	0.080	0.018	19.854	1.127	1.103	1.152
Oral Antibiotics	0.120	0.011	117.524	1.093	1.057	1.131
Oral Corticosteroids	0.089	0.017	26.753	1.003	1.002	1.004
Outpatient Visits	0.003	0.000	86.607	1.232	1.172	1.295
PCP Continuity Proportion	0.209	0.026	66.310	1.189	1.128	1.253
Previous Diabetic Wound	0.173	0.027	41.548	0.940	0.931	0.950
Prior Admit Count	-0.061	0.005	132.637	1.329	1.282	1.379
Prior Admit Type - Emergency	0.285	0.019	235.527	1.310	1.214	1.414
Prior Admit Type - Urgent	0.270	0.039	48.217	1.208	1.165	1.251
Prior Discharge - Home	0.189	0.018	106.800	1.556	1.351	1.792
Prior Discharge - Other	0.442	0.072	37.655	0.878	0.843	0.914
Prior Nursing Home Stay	-0.130	0.021	39.935	0.877	0.855	0.899
Prior Surgery	-0.131	0.013	108.124	1.084	1.059	1.109
Provider Administered Drug	0.080	0.012	46.760	1.019	1.018	1.020
Age	0.019	0.001	890.270	1.235	1.203	1.268
Dual Eligible for Medicaid	0.211	0.013	250.064	1.648	1.558	1.743
Race - Black	0.499	0.029	303.756	1.353	1.213	1.510
Race - Hispanic	0.303	0.056	29.267	1.254	1.187	1.325
Race - White	0.226	0.028	65.172	1.422	1.382	1.462
Medicare Non-Age Eligibility Reason	0.352	0.014	607.598	0.552	0.536	0.569
COVID - Early	-0.593	0.016	1,473.173	0.653	0.639	0.667

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
COVID - Late	-0.426	0.011	1,505.110			

Supplemental Table 9

Results from the discrete time survival model predicting AH events by demographic and census tract-level social and environmental risk factors (Model 4).

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Intercept	-7.510	0.133	3,189.564			
Age	0.044	0.001	6,721.521	1.045	1.044	1.046
Dual Eligible for Medicaid	0.612	0.013	2,282.040	1.844	1.798	1.891
Race - Black	0.694	0.030	547.064	2.001	1.888	2.121
Race - Hispanic	0.245	0.055	19.673	1.278	1.147	1.424
Race - White	0.451	0.028	251.686	1.570	1.485	1.660
Medicare Non-Age Eligibility Reason	0.941	0.014	4,840.413	2.564	2.496	2.632
COVID - Early	-0.499	0.015	1,071.958	0.607	0.589	0.625
COVID - Late	-0.372	0.011	1,249.720	0.689	0.675	0.704
CT Air Pollution	-0.136	0.011	159.257	0.873	0.855	0.891
CT ADI	0.002	0.000	23.970	1.002	1.001	1.003
CT Median Income	0.000	0.000	46.364	1.000	1.000	1.000
CT Mental Health Care Shortage Area - Whole	-0.069	0.016	19.494	0.933	0.905	0.962
CT Percent Age 65+	-0.003	0.001	28.185	0.997	0.995	0.998
CT Percent Age 65+ with Less than HS Diploma	0.004	0.001	50.092	1.004	1.003	1.005
CT Percent Foreign Born	-0.002	0.001	13.746	0.998	0.997	0.999
CT Percent Married	-0.003	0.001	11.950	0.997	0.996	0.999
CT Percent Age 65+ Non-White	-0.002	0.000	44.472	0.998	0.998	0.999
CT Percent Physician Diversity	0.002	0.000	30.381	1.002	1.001	1.003
CT 2019 Population	0.000	0.000	10.865	1.000	1.000	1.000

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
CT Taxable Interest Per Capita	0.000	0.000	33.991	1.000	1.000	1.000

Supplemental Table 10

Results from the discrete time survival model predicting AH events by demographic and ZCTA-level social and environmental risk factors (Model 5).

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Intercept	-8.880	0.065	18,853.504			
Age	0.044	0.001	7,005.864	1.045	1.044	1.046
Dual Eligible for Medicaid	0.628	0.013	2,503.003	1.874	1.828	1.921
Race - Black	0.735	0.029	645.218	2.086	1.971	2.208
Race - Hispanic	0.279	0.055	25.835	1.321	1.187	1.471
Race - White	0.459	0.028	264.925	1.582	1.497	1.672
Medicare Non-Age Eligibility Reason	0.943	0.013	4,988.362	2.568	2.502	2.636
COVID - Early	-0.505	0.015	1,117.812	0.604	0.586	0.622
COVID - Late	-0.372	0.010	1,277.787	0.689	0.676	0.704
ZCTA Median Income	0.000	0.000	81.326	1.000	1.000	1.000
ZCTA Percent 65+	-0.007	0.001	44.685	0.993	0.991	0.995
ZCTA Percent Age 65+ with Less than HS Diploma	0.006	0.001	21.264	1.006	1.003	1.008
ZCTA Percent Foreign Born	-0.008	0.001	137.757	0.992	0.991	0.994
ZCTA Percent with Less than HS Diploma	0.014	0.002	58.305	1.014	1.011	1.018
ZCTA Percent Physician Diversity	0.000	0.000	56.818	1.000	1.000	1.000
ZCTA Percent Poverty	-0.006	0.002	16.075	0.994	0.991	0.997
ZCTA Percent Single Mothers	0.001	0.000	12.283	1.001	1.000	1.002
ZCTA Population Growth	0.001	0.000	20.165	1.001	1.001	1.002

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
ZCTA Population Density	0.000	0.000	49.054	1.000	1.000	1.000
ZCTA Primary Care Shortage Area - Whole	0.457	0.107	18.140	1.579	1.280	1.949

Supplemental Table 11

Results from the discrete time survival model predicting AH events by demographic risk factors (Model 6).

Parameter	Estimate	Std. Error	Wald Chi-Square	Odds Ratio	OR Low CI	OR Upp. CI
Intercept	-9.167	0.049	35,135.645			
Age	0.043	0.001	6,895.349	1.044	1.043	1.045
Dual Eligible for Medicaid	0.668	0.012	2,897.248	1.950	1.903	1.998
Race - Black	0.876	0.028	960.408	2.403	2.273	2.540
Race - Hispanic	0.288	0.054	27.876	1.333	1.198	1.483
Race - White	0.574	0.028	429.740	1.776	1.682	1.875
Medicare Non-Age Eligibility Reason	0.990	0.013	5,551.584	2.690	2.621	2.761
COVID - Early	-0.504	0.015	1,118.773	0.604	0.587	0.622
COVID - Late	-0.372	0.010	1,283.127	0.690	0.676	0.704

Supplemental Table 12

Results from the discrete time survival model predicting AH events by demographic, individual utilization history, and social and environmental risk factors using matched predictor variables. Contains the ORs for both census tract and ZCTA versions of the model.

Parameter	CT OR	OR Low CI	OR Upp. CI	ZCTA OR	OR Low CI	OR Upp. CI
Intercept						
AH NUMBER	1.427	1.413	1.440	1.433	1.419	1.446
Antidiabetes Medication	1.086	1.041	1.133	1.081	1.037	1.127
CCW Albuminuria	1.130	1.082	1.180	1.133	1.085	1.183
CCW Alzheimers Disease	1.093	1.061	1.126	1.090	1.059	1.122
CCW Anemia	1.056	1.031	1.080	1.053	1.029	1.078
CCW Anxiety Disorder	1.130	1.101	1.160	1.133	1.105	1.163
CCW Arrhythmia	1.204	1.173	1.236	1.195	1.165	1.226
CCW Asthma	1.165	1.129	1.203	1.165	1.129	1.203
CCW Atrial Fibrillation	1.096	1.063	1.130	1.102	1.069	1.136
CCW Cataract	0.875	0.850	0.901	0.879	0.854	0.905
CCW Cerebral Palsy	1.485	1.266	1.742	1.525	1.324	1.756
CCW Chronic Kidney Disease	1.226	1.194	1.258	1.232	1.200	1.264
CCW COPD	1.625	1.584	1.667	1.635	1.594	1.677
CCW Depression and Depressive Disorders	1.070	1.042	1.099	1.069	1.042	1.097
CCW Diabetes	1.146	1.105	1.188	1.145	1.105	1.187
CCW Diabetes with Complications	1.145	1.099	1.193	1.145	1.099	1.193
CCW Fluid and Electrolytes Imbalance	1.204	1.172	1.237	1.201	1.170	1.234
CCW Glaucoma	0.912	0.884	0.941	0.908	0.880	0.937

Parameter	CT OR	OR Low CI	OR Upp. CI	ZCTA OR	OR Low CI	OR Upp. CI
CCW Heart Failure	1.566	1.525	1.607	1.569	1.529	1.610
CCW Hyperlipidemia	0.869	0.848	0.891	0.871	0.850	0.892
CCW Hypertension	1.359	1.317	1.403	1.355	1.314	1.398
CCW Intellectual Disabilities and Related Conditions	1.357	1.227	1.500	1.282	1.169	1.405
CCW Ischemic Heart Disease	1.152	1.125	1.179	1.149	1.123	1.176
CCW Lung Cancer	1.203	1.128	1.284	1.211	1.136	1.292
CCW Neuropathy	1.065	1.031	1.100	1.061	1.028	1.095
CCW Peripheral Vascular Disease	1.090	1.064	1.118	1.092	1.066	1.120
CCW Pneumonia	1.062	1.030	1.094	1.062	1.031	1.094
CCW PTSD	1.138	1.056	1.227	1.148	1.067	1.235
CCW Pressure and Chronic Ulcers	1.155	1.115	1.197	1.159	1.120	1.200
CCW Problems Care Provider Dependency	1.308	1.271	1.346	1.305	1.268	1.343
CCW Pulmonary Circulatory Disorder	1.115	1.082	1.149	1.107	1.075	1.141
CCW Rheumatoid Arthritis/Osteoarthritis	0.960	0.939	0.981	0.962	0.942	0.983
CCW Respiratory Infection	1.107	1.079	1.135	1.107	1.080	1.135
CCW Retinopathy	1.731	1.555	1.927	1.704	1.533	1.893
CCW Sepsis	0.868	0.837	0.900	0.870	0.840	0.902
CCW Solid Tumor without Metastasis	0.941	0.913	0.970	0.934	0.907	0.963
CCW Tobacco Use	1.247	1.211	1.285	1.251	1.215	1.289

Parameter	CT OR	OR Low CI	OR Upp. CI	ZCTA OR	OR Low CI	OR Upp. CI
CCW Urinary Tract Infection	1.345	1.312	1.379	1.347	1.315	1.381
Foot Toe Amputation	1.278	1.136	1.439	1.237	1.100	1.390
HbA1c	0.977	0.967	0.987	0.978	0.968	0.988
Hospice	1.560	1.450	1.680	1.558	1.448	1.676
Insulin	1.267	1.226	1.311	1.269	1.228	1.312
Leukotrine receptor modifier use	1.085	1.041	1.131	1.081	1.038	1.126
Oral Antibiotics	1.124	1.099	1.149	1.122	1.098	1.146
Oral Corticosteroids	1.106	1.068	1.144	1.099	1.062	1.137
Outpatient Visits	1.003	1.003	1.004	1.003	1.003	1.004
PCP Continuity Proportion	1.249	1.187	1.314	1.236	1.175	1.300
Previous Diabetic Wound	1.183	1.122	1.248	1.187	1.126	1.252
Prior Admit Count	0.946	0.936	0.956	0.945	0.935	0.955
Prior Admit - Length of Stay	0.995	0.993	0.998	0.996	0.993	0.998
Prior Admit Type - Emergency	1.361	1.312	1.413	1.358	1.309	1.409
Prior Admit Type - Other	1.015	0.572	1.800	1.058	0.610	1.837
Prior Admit Type - Urgent	1.234	1.141	1.334	1.269	1.175	1.370
Prior Discharge - Home	1.205	1.162	1.249	1.204	1.161	1.248
Prior Discharge - Other	1.643	1.419	1.901	1.634	1.415	1.887
Prior Nursing Home Stay	0.884	0.848	0.921	0.885	0.850	0.922
Prior Surgery	0.886	0.864	0.909	0.884	0.863	0.906
Provider Administered Drug	1.080	1.055	1.105	1.085	1.060	1.111

Parameter	CT OR	OR Low CI	OR Upp. CI	ZCTA OR	OR Low CI	OR Upp. CI
Age	1.020	1.019	1.021	1.020	1.018	1.021
Dual Eligible for Medicaid	1.198	1.166	1.230	1.204	1.173	1.236
Race - Black	1.505	1.421	1.595	1.571	1.484	1.664
Race - Hispanic	1.319	1.180	1.473	1.331	1.192	1.486
Race - White	1.188	1.123	1.256	1.206	1.141	1.275
Medicare Non-Age Eligibility Reason	1.403	1.364	1.443	1.406	1.367	1.446
COVID - Early	0.555	0.539	0.573	0.553	0.536	0.570
COVID - Late	0.653	0.639	0.667	0.654	0.640	0.668
Air Pollution	0.870	0.852	0.888	0.999	0.999	0.999
Median Income	1.000	1.000	1.000	1.000	1.000	1.000
Percent Less Than HS Diploma	1.001	1.001	1.002	1.005	1.002	1.008
Percent Married	0.998	0.997	1.000	1.000	0.998	1.001
CCW Bipolar Disorder	1.081	1.032	1.132	1.086	1.038	1.136
Mental Health Use	1.057	1.020	1.096	1.065	1.028	1.103
Percent Age 65+ with Less than HS Diploma	1.002	1.001	1.003	1.004	1.001	1.006
Percent Physician Diversity	1.001	1.000	1.001	1.000	1.000	1.000
Mental Health Care Shortage - Partial	0.995	0.969	1.022	0.950	0.926	0.974

Supplemental Table 13

Results from the discrete time survival model predicting AH events by demographic and social and environmental risk factors using matched predictor variables. Contains the ORs for both census tract and ZCTA versions of the model.

Parameter	CT OR	CT OR Low CI	CT OR Upp. CI	ZCTA OR	ZCTA OR Low CI	ZCTA OR Upp. CI
Intercept						
Age	1.045	1.044	1.046	1.045	1.044	1.046
Dual Eligible for Medicaid	1.844	1.798	1.891	1.874	1.828	1.920
Race - Black	1.999	1.886	2.119	2.087	1.971	2.211
Race - White	1.569	1.483	1.658	1.576	1.491	1.666
Race - Hispanic	1.272	1.141	1.417	1.316	1.181	1.465
Medicare Non-Age Eligibility Reason	2.563	2.496	2.632	2.568	2.502	2.637
COVID - Early	0.607	0.589	0.625	0.604	0.586	0.622
COVID - Late	0.689	0.675	0.704	0.689	0.676	0.704
Air Pollution	0.869	0.850	0.888	0.999	0.998	1.000
ADI	1.002	1.001	1.003	1.000	0.998	1.001
Median Income	1.000	1.000	1.000	1.000	1.000	1.000
Mental Health Care Shortage Area - Whole	0.933	0.905	0.962	0.989	0.961	1.017
Percent Age 65+	0.996	0.995	0.998	0.996	0.993	0.998
Percent Age 65+ with Less than HS Diploma	1.004	1.002	1.005	1.006	1.004	1.009
Percent Foreign Born	0.997	0.996	0.999	0.993	0.991	0.994
Percent Married	0.997	0.996	0.999	0.997	0.995	0.999
Percent Age 65+ Non-White	0.998	0.998	0.999	0.999	0.999	1.000
Percent Physician Diversity	1.002	1.001	1.003	1.000	1.000	1.000

Parameter	CT OR	CT OR Low CI	CT OR Upp. CI	ZCTA OR	ZCTA OR Low CI	ZCTA OR Upp. CI
2019 Population	1.000	1.000	1.000	1.000	1.000	1.000
Taxable Interest Per Capita	1.000	1.000	1.000	1.000	1.000	1.000
Percent Less than HS Diploma	1.002	1.001	1.002	1.013	1.010	1.017
Percent Poverty	0.998	0.996	1.000	0.996	0.992	0.999
Percent Single Mothers	1.000	1.000	1.000	1.001	1.000	1.001
Population Growth	1.001	1.000	1.002	1.001	1.000	1.002
Population Density	1.000	1.000	1.000	1.000	1.000	1.000

Supplemental Table 14

Results from the discrete time survival model predicting AH events by demographic and individual utilization history using matched predictor variables.

Parameter	Odds Ratio	OR Low CI	OR Upp. CI
Intercept			
AH NUMBER	1.438	1.425	1.451
Antidiabetes Medication	1.083	1.039	1.129
CCW Albuminuria	1.133	1.085	1.183
CCW Alzheimers Disease	1.091	1.060	1.123
CCW Anemia	1.053	1.029	1.078
CCW Anxiety Disorder	1.135	1.106	1.165
CCW Arrhythmia	1.197	1.167	1.228
CCW Asthma	1.165	1.129	1.202
CCW Atrial Fibrillation	1.107	1.075	1.141
CCW Bipolar Disorder	1.086	1.038	1.136
CCW Cataract	0.881	0.856	0.907
CCW Cerebral Palsy	1.512	1.313	1.741
CCW Chronic Kidney Disease	1.239	1.207	1.271
CCW COPD	1.649	1.608	1.691
CCW Depression and Depressive Disorders	1.064	1.036	1.092
CCW Diabetes	1.159	1.119	1.202
CCW Diabetes with Complications	1.149	1.103	1.196
CCW Fluid and Electrolytes Imbalance	1.202	1.170	1.234
CCW Glaucoma	0.907	0.879	0.935

Parameter	Odds Ratio	OR Low CI	OR Upp. CI
CCW Heart Failure	1.573	1.532	1.614
CCW Hyperlipidemia	0.877	0.856	0.898
CCW Hypertension	1.370	1.329	1.414
CCW Intellectual Disabilities and Related Conditions	1.265	1.154	1.386
CCW Ischemic Heart Disease	1.150	1.123	1.177
CCW Lung Cancer	1.220	1.144	1.300
CCW Neuropathy	1.064	1.031	1.098
CCW Peripheral Vascular Disease	1.096	1.070	1.123
CCW Pneumonia	1.061	1.030	1.093
CCW PTSD	1.157	1.076	1.245
CCW Pressure and Chronic Ulcers	1.151	1.112	1.191
CCW Problems Care Provider Dependency	1.297	1.261	1.335
CCW Pulmonary Circulatory Disorder	1.103	1.070	1.137
CCW Rheumatoid Arthritis/Osteoarthritis	0.965	0.944	0.986
CCW Respiratory Infection	1.107	1.080	1.135
CCW Retinopathy	1.687	1.518	1.874
CCW Sepsis	0.874	0.844	0.906
CCW Solid Tumor without Metastasis	0.934	0.907	0.963
CCW Tobacco Use	1.266	1.229	1.303

Parameter	Odds Ratio	OR Low CI	OR Upp. CI
CCW Urinary Tract Infection	1.348	1.315	1.382
Foot Toe Amputation	1.229	1.094	1.382
HbA1c	0.969	0.959	0.978
Hospice	1.559	1.450	1.677
Insulin	1.273	1.231	1.315
Leukotrine receptor modifier use	1.087	1.043	1.132
Mental Health Use	1.079	1.042	1.118
Oral Antibiotics	1.126	1.102	1.151
Oral Corticosteroids	1.094	1.057	1.132
Outpatient Visits	1.003	1.002	1.004
PCP Continuity Proportion	1.234	1.174	1.298
Previous Diabetic Wound	1.190	1.129	1.255
Prior Admit Count	0.942	0.933	0.952
Prior Admit Type - Emergency	1.335	1.287	1.385
Prior Admit Type - Urgent	1.322	1.225	1.427
Prior Discharge - Home	1.213	1.170	1.257
Prior Discharge - Other	1.629	1.412	1.880
Prior Nursing Home Stay	0.884	0.849	0.921
Prior Surgery	0.878	0.856	0.900
Provider Administered Drug	1.084	1.059	1.109
Age	1.019	1.018	1.020
Dual Eligible for Medicaid	1.235	1.203	1.268
Race - Black	1.647	1.557	1.743

Parameter	Odds Ratio	OR Low CI	OR Upp. CI
Race - Hispanic	1.353	1.213	1.510
Race - White	1.253	1.186	1.324
Medicare Non-Age Eligibility Reason	1.421	1.382	1.461
COVID - Early	0.552	0.536	0.569
COVID - Late	0.653	0.639	0.667
Prior Admit - Length of Stay	0.996	0.994	0.999
Prior Admit Type - Other	1.047	0.603	1.816

Supplemental Table 15

Comparison of model fit and predictive capability using matched predictor variables.

Utilization Risk Factors Included?	Census Tract	ZCTA	No Geographic Predictors
Yes	**Matched - Model 1** **AIC**: 460641.50 **C**: 0.8380 **Gini**: 0.6338 **Top 10%**: 51.62%	**Matched - Model 2** **AIC**: 471278.52 **C**: 0.8379 **Gini**: 0.6326 **Top 10%**: 51.63%	**Matched - Model 3** **AIC**: 473871.97 **C**: 0.8413 **Gini**: 0.6314 **Top 10%**: 51.54%
No	**Matched - Model 4** **AIC**: 516665.00 **C**: 0.6870 **Gini**: 0.3590 **Top 10%**:23.90%	**Matched - Model 5** **AIC**: 528702.73 **C**: 0.6841 **Gini**: 0.3520 **Top 10%**: 23.34%	**Matched - 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.

Supplemental Table 16

Results from the non-parametric tests comparing the trapezoidal area under the ROC from the models using matched predictors across census tract and ZCTA versions of the model.

AH ~ Demographics + Matched Social and Environmental + Individual Utilization					
Comparison	**Estimate**	**S.E.**	**95% CI**	***X²***	***p***
Census Tract – ZCTA *(Model 1 vs. 2)*	0.0004	0.0002	0.00002 – 0.0009	3.85	0.0499
Census Tract – No Geo *(Model 1 vs. 3)*	0.001	0.0003	0.0006-0.0020	15.28	<.0001
ZCTA – No Geo *(Model 2 vs. 3)*	0.0008	0.0002	0.0003-0.001	12.65	0.0004
AH ~ Demographics + Matched Social and Environmental					
Comparison	**Estimate**	**S.E.**	**95% CI**	***X²***	***p***
Census Tract – ZCTA *(Model 4 vs. 5)*	0.003	0.0006	0.001 – 0.004	16.72	<.0001
Census Tract – No Geo *(Model 4 vs. 6)*	0.008	0.001	0.006-0.01	67.49	<.0001
ZCTA – No Geo *(Model 5 vs. 6)*	0.005	0.0001	0.003-0.006	35.64	<.0001