SOA Research INSTITUTE

Diversity, Equity and Inclusion

Health Care Cost Trends

Interactions of Health Care Rating Factors with Race, Ethnicity and Socioeconomic Factors

July | 2022

Interactions of Health Care Rating Factors with Race/Ethnicity and Socioeconomic Factors

AUTHORS Lydia Tolman, FSA, MAAA

Jacquelyn Young, ASA, MAAA

Jelena Milovanovic, PhD, AIAA, ACIA

Hongjuan Zhou, PhD, ASA

John Zicarelli, PhD, FCAS, CFA

Lydia Gabric

Hayley Osterkorn

Joe Simpson

SPONSOR Diversity, Equity and Inclusion Research Advisory Council

CONTENTS

Interactions of Health Care Rating Factors with Race/Ethnicity and Socioeconomic Factors

Executive Summary

The health insurance industry has historically used several factors to develop premium rates. Many of these, e.g., gender rating, have been largely eliminated from health insurance rating practices by regulations like the Patient Protection and Affordable Care Act (ACA) because they were determined as inequitable. However, several other factors still remain as differentiators in premium rates charged to individuals and groups. The aim of this research was to determine if, in fact, there could be unintended factors, like race, ethnicity or socioeconomic status, that are inherently included in some of these remaining allowable factors.

Our research focused on three allowable rating factors—tobacco usage, geographic area or industry—may be impacting insurance premiums across certain socioeconomic or racial/ethnic groups. Our research did not examine whether these rating factors accurately reflect the differences in claims costs for each group. Certain geographic areas may cost more than other areas because of provider (e.g., hospital system) contracting, utilization patterns within urban versus rural populations, etc. People using tobacco generally may have higher health costs than a similar person who does not use tobacco. Certain industries may be riskier and prone to higher health costs than other industries. We did not have access to the underlying claims costs (actual or projected) and cannot conjecture whether rating factors are indicative of actual costs. However, if rating factors are indicative of claims costs, our results indicate the costs of health care in the region studied could be greater in the areas where some socioeconomic and demographic groups are more likely to live. The following is a summary of key findings for each factor within the limited scope of our analysis.

We studied tobacco surcharges and area factors for ACA-compliant individual and small group health insurance. Because of the complexity of manually pulling together rating manuals for all states and the need to mesh data with publicly available data for proportions of race and ethnicity by rating area, we focused our analysis on a state with a large population in a large number of counties, where each county represents a unique rating area, and where numerous insurers are active in each rating area. Of the states that meet the criteria, we focused on Florida because it has the largest population. Rating factors for larger populations are less likely to be influenced by outlier events or situations than are those for smaller populations.

Tobacco surcharges in the individual and small group ACA markets in Florida may be capturing unintended factors, such as socioeconomic status or other comorbid conditions, and therefore may be resulting in higher rating factors for certain populations in an unequitable fashion. Smokers are more likely to have lower incomes, have a less formal education and have worse access to healthy foods than their non-smoking peers. Thus, in addition to the direct health care costs of smoking, it is possible that the load calculated could include the cost of conditions that are correlated with smoking. On top of this, premium subsidies allowed for in the ACA cannot be used to cover the tobacco surcharge. So, smokers can have higher out of pocket premium expenses than the maximum allowable 50% surcharge would imply at first glance.

Based on the geographic distribution of current Florida residents, the Hispanic/Latino population is more likely to see a higher premium rate, where the Non-Hispanic/Latino white population is more likely to see a lower premium

rate. Our analysis also showed that in Florida, Non-Hispanic/Latino whites are generally more likely to smoke than are Hispanic/Latinos.

This study did not consider federal reimbursement, uninsured rates, practice patterns or other external variables that could drive this correlation. Race/ethnicity data was also available for Non-Hispanic Black/African American and Asian/American populations. However, Non-Hispanic/Latino Black/African American and Asian/American populations in Florida were not as strongly correlated with ACA area factors in the individual or small group markets.

We were unable to study industry factors in the large group market at a sufficiently granular level to draw any conclusions in our study of industry factors for two reasons. First, many insurance companies do not disclose detailed factors publicly because of competitive pressures. We focused on Colorado large group filings, where limited detail is made available. Secondly, matching the somewhat outdated 1987 4-digit Standard Industrial Classification (SIC) code in use by the health insurance industry today to publicly available demographic data sets, which rely on other industry classification systems, necessitates a reduction in the granularity of industry definitions which can be studied.

Introduction and Scope of Project

The health insurance industry has always used a wide variety of factors to develop rates for individuals seeking commercial coverage. Before the passage of the Patient Protection and Affordable Care Act (ACA), these included factors based on gender and health status. Women in their childbearing years were often charged more than men of the same age in states that had not implemented gender neutral rating rules. Pregnancy and other preexisting conditions meant significantly higher rates or outright declination of coverage for an individual with such conditions. Although women in their childbearing years on average do cost more than men of the same age, legislators disallowed gender rating in the ACA. Preexisting condition exclusions were also removed in ACA rules and regulations.

While the ACA disallows rating based on gender and health status in the individual and small group market, this paper explores health rating factors in use today. We studied whether tobacco, geographic and industry factors might lead to higher rates charged for specific subsets of race/ethnicity or socioeconomic status within the individual, small group and large group markets.

Take smoking, for example. Research has consistently shown that "smoking prevalence is generally higher among underprivileged groups," particularly those with a low socioeconomic status.[1] Recent statistics from the Centers for Disease Control (CDC) indicated that smokers are more likely to have a less formal education and have lower income than their non-smoking counterparts. LGTBQ+ people, certain racial/ethnic groups, people with disabilities and people with mental illnesses are also more likely to use tobacco than the general population.[2] On the other hand, highly educated, wealthier populations are more likely to drink alcohol.[3] Tobacco usage is clearly linked with deteriorations in morbidity and mortality, and rating for tobacco use is considered appropriate in health insurance rating. However, excessive alcohol use is also linked to worsened morbidity and mortality, but alcohol usage is not a rating factor for health insurance. Thus, the question that arises is whether the use of conventional rating factors is leading to higher rates charged for certain racial/ethnic or underprivileged groups.

This research paper is specifically focused on tobacco and geographic factors from the ACA individual and small group markets and industry factors from the large group market. The aim of the research is to determine if there could be "ghost" rating factors, like race/ethnicity or socioeconomic status, that are inherently included in allowable rating factors and to encourage more nuanced thinking from health actuaries and health insurers when developing their rating factors. This paper does not consider whether the rating factors examined (tobacco, area and industry) accurately reflect the underlying claims expenses used to develop those factors. The researchers did not have access to the underlying claims costs (actual or projected) and cannot conjecture whether rating factors are indicative of actual costs. However, if rating factors are indicative of cost, our results indicate the health care delivery system studied could be impacting socioeconomic and demographic groups disparately. The intent of this paper is to advocate neither for nor against the inclusion of tobacco, area or industry rating factors in health insurance. Nor do we advocate for or against proposing the inclusion of alternative rating factors.

[1] Hiscock, R., Bauld, L., Amos, A., Fidler, J. A., & Munafo, M. (2012). Socioeconomic status and smoking: a review. Annals of the New York a Academy of Sciences, 1248(Addiction Reviews), 107–123. https://doi.org/10.1111/j.1749-6632.2011.06202.x

[2] Centers for Disease Control and Prevention. (2021, April 7). Burden of Cigarette Use in the U.S. Centers for Disease Control and Prevention. https://www.cdc.gov/tobacco/campaign/tips/resources/data/cigarette-smoking-in-united-states.html

[3] Jones, J. M. (2021, March 23). Drinking Highest Among Educated, Upper-Income Americans. Gallup.com. https://news.gallup.com/poll/184358/drinking-highest-among-educated-upper-income-americans.aspx

Background

Before delving into the analysis, it is important to step back and review how premiums are developed in the health insurance industry, what rating factors are, and what goes into determining those factors.

Premiums are the amount a group or an individual pay to an insurer every month to fund health care expenses and gain access to the insurer's network of physicians, hospitals, pharmacies and so on. Premiums are made up of several components, including expected average claims costs, the expenses to administer the plan (e.g., expenses to pay claims), taxes and fees and a net risk margin or profit.

Claims make up the bulk of the cost of insurance premiums. The ACA mandated that claims (and costs to improve quality of care) must make up 80% of premiums for individuals and small groups and 85% of premiums for large groups. Thus, estimating expected claims costs is one of the key challenges that insurers face. If an insurer overestimates claims, they may be uncompetitive and also have to pay refunds to their members. On the other hand, if an insurer underestimates claims, they may not have enough money to pay claims or to administer the plan.

Several characteristics of an individual or group can drive large differences in expected claims costs. Some of these characteristics are not allowed to be considered by insurance carriers when differentiating premiums. For individuals and small groups, the ACA only allows for five characteristics or rating factors, to be used to differentiate premiums between members (or between groups):

- Age: Medical and pharmaceutical costs tend to increase as we age.
- Location, also known as geographic area: Where a person lives has a big effect on claims costs. Several differences between locations can drive these differences.
    - Cost differences between health care providers. For example, if a very rural area of a state has one hospital, that hospital may be able to demand higher rates for their services from an insurer than a more urban area with multiple competing hospitals.
    - Cost of living. Certain areas are more expensive for physicians and hospitals to operate in because of higher necessary salaries for medical staff or higher property/rental costs for physical locations.
    - Utilization of medical services. Consumers access care differently depending on their location. Sometimes prevailing physician practice can vary across areas; or the presence of a specialty hospital like a cardiac center may drive more interventions because of additional capacity and resources.
- Tobacco use. Tobacco users tend to have more medical costs than non-tobacco users.
- Individual versus family enrollment: Expected claims costs are greater when multiple people are covered.
- Plan category: Expect claims costs are greater when the insured pays lower cost share, e.g., an insurer will have greater claims costs for a plan with a $500 deductible than for one with a $1,000 deductible.

For large groups, which are typically groups defined as having 50 or more employees, insurers may consider the group's previous total claims experience, the group's industry (a doctor's office may be more expensive than a hair salon to insure, for example) and some of the other factors mentioned above.

With all of this in mind, we want to reiterate that this paper does not consider whether the rating factors examined (tobacco, area and industry) accurately reflect the underlying claims expenses used to develop those factors. The researchers did not have access to the underlying claims costs (actual or projected) and cannot conjecture whether rating factors are indicative of actual costs. Examining whether rating factors reflect actual costs is a topic for further research. However, if rating factors are indicative of actual claims cost, our results indicate the health care delivery system studied could be impacting socioeconomic and demographic groups disparately.

Data and Methodology

DATA SOURCES

Publicly available data sets were used to segment the population into common rating factors (tobacco, geographic, industry) along with other demographic data, e.g., race/ethnicity and socioeconomic status. Publicly available data used included:

- The 2021 County Health Rankings (CHR)
- The American Community Survey (ACS) 5-year estimates—Public Use Microdata Sample 2019
- 2021 rating manuals available through state filing websites. Specifically:
    - Florida 2021 ACA Individual and Small Group filings for several carriers
    - Colorado 2021 Large Group filings for several carriers

CHR data was used for our tobacco and area factor analyses. The CHR data is a nationwide dataset offering demographic data by county. This dataset was used to acquire the demographic data such as median household income, home ownership status, high school and college statistics, smoking statistics, as well as race/ethnicity. CHR data was used rather than ACS data since ACS data did not have statistics available for the smallest 26 of the 67 counties in Florida. Unfortunately, CHR data is not available at the person or household level, so statistics were calculated at the county level.

ACS data was used for our industry factor analysis. The ACS data is a nationwide dataset offering demographic data by county. This dataset was used to acquire demographic data including household income and race/ethnicity. This set was used instead of CHR data because it included industry factor statistics that more easily aligned with the SIC codes typically used by the health insurance industry.

Area rating factors were pulled from the 2021 ACA Individual and Small Group filings in the state of Florida for this analysis. Florida's rating areas were determined to be the most useful because each county has its own unique geographic rating area. This allowed all county-level statistics to be associated directly with the rating factor specific to that county. There were a few states that split geographic rating areas this way, but Florida had the most counties as well as the largest population.

Tobacco surcharges were pulled from the 2021 ACA Individual filings in the state of Florida for consistency with the area factor analysis.

We studied industry factors in the state of Colorado. Florida was not an optimal state for the study of industry factors because of the large number of carriers offering large group products, the quantity of non-standardized filing memorandums that would need to be studied and the lack of detailed industry factors included in the filing documents. Industry factors were instead compiled from 2021 large group filings in the state of Colorado. There were fewer carriers, which allowed for less manual manipulation of text documents. More importantly, Colorado requires the full disclosure of rating factors for most carriers. Since industry factors do not typically vary depending on the geographic location, we determined these factors were appropriate.

DATA FIELD DEFINITIONS

Data field definitions for fields pulled from the CHR dataset are as follows:[4]

- Adult Smoking: Percentage of adults who are current smokers (age-adjusted).
- Proportion of Non-Hispanic/Latino white: Percentage of population that is Non-Hispanic white
- Proportion of Non-Hispanic/Latino Black/African American: Percentage of population that is non-Hispanic/Latino Black or African American
- Proportion of Hispanic/Latino: Percentage of population that is Hispanic/Latino
- Proportion of Asian/American: Percentage of population that is Asian
- Proportion of Other: Percentage of population that is not Non-Hispanic/Latino white, Non-Hispanic/Latino Black/African American, Hispanic/Latino or Asian/American
- Food Environment Index:[5] Index of factors that contribute to a healthy food environment, from 0 (worst) to 10 (best). The factor equally weights two indicators of the food environment:
    - Limited access to healthy foods: Percentage of population that is low income and does not live close to a grocery store
    - Food insecurity: Percentage of the population that did not have access to a reliable source of food during the past year
- High School Completion Proportion: Percentage of adults ages 25 and over with a high school diploma or equivalent
- Some College Proportion: Percentage of adults ages 25–44 with some post-secondary education
- Median Household Income: The income where half of households in a county earn more and half of households earn less
- Home Ownership Status Proportion: Percentage of occupied housing units that are owned

Data field definitions for fields pulled from the ACS dataset are as follows:[6]

- NAICSP: North American Industry Classification System (NAICS) recode for 2018 and later based on 2017 NAICS codes
- HINCP:
    - Household income for the past 12 months
    - Ranges were created using the census MDAT tool
- RAC1P: Recoded detailed race/ethnicity code

DATA MANIPULATION

Further manipulation of the rate manuals was required to obtain tobacco, area and industry factors by the demographic cuts chosen from the ACS and CHR data sets.

TOBACCO SURCHARGE

In order to calculate the Individual tobacco surcharges, the 2021 rates with the additional charge for tobacco were divided by comparable rates without the tobacco charge. This gives the implied tobacco load by carrier. For carriers

[4] Explore health rankings: Rankings Data & Documentation. County Health Rankings & Roadmaps. (n.d.). Retrieved January 7, 2022, from https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation

[5] Food environment index. County Health Rankings & Roadmaps. (n.d.). Retrieved January 7, 2022, from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/health-behaviors/diet-exercise/food-environment-index

[6] United States Census Bureau. (n.d.). Data Definitions for American Community Survey (ACS) 5-year estimates – Public Use Microdata Sample 2019. MDAT Table Creator. Retrieved January 7, 2022, from https://data.census.gov/mdat/#/search?ds=ACSPUMS5Y2019

that do not rate for tobacco usage, this was set to 1.0. Since the tobacco load is allowed to vary by age, the tobacco loads were calculated for all rating areas by age. Enrollment splits in the data were available for age ranges 21–34, 35–54, and 55+, so the median tobacco load was calculated for each plan and rating area combination within these age ranges. Then, the available enrollment splits were used to weight the average tobacco loading for each age range and rating area combination. The tobacco factors used by individual carriers do not vary by county in Florida. Thus, when we explore the proportion of smokers by county, it can be used in a way to study relationships between tobacco usage and socioeconomic variables.

AREA FACTORS

Area factors were compiled from both Individual and Small Group Florida rate filings. Given the wide variance of average area factors across carriers, rating factors were normalized relative to one rating area. We observed that the area with the closest factor to the carrier's average, with credible enrollment levels, was rating area 10 across both the individual and small group market. To normalize, revised geographic factors were calculated using the originally filed area factor, divided by the area factor in rating area 10 for that carrier. For this reason, our analysis excluded insurers that did not offer products in rating area 10. The companies excluded were Health First Commercial Plans, Inc., Florida Health Care Plan, Inc., and Capital Health Plan, Inc. As a result, our analysis included 96% of the enrolled Individual ACA enrollees and 94% of enrolled Small Group ACA enrollees in the state of Florida in 2019. The final area factor for each county used the weighted average enrolled across all carriers. Note that the weights were based on statewide enrollment levels since more granular enrollment counts are not publicly available.

INDUSTRY FACTORS

Colorado Large Group industry factors (SIC codes) were compiled from six carriers, including UnitedHealthcare, Aetna, Cigna, Humana, Kaiser, Anthem BlueCross BlueShield of Colorado, under the assumption that industry factors should not vary considerably across states as previously discussed. In order to map the industries from these filings to those included in the ACS dataset, the filed SIC code ranges were mapped to ACS NAICS 2-Digit industry code categories. Details on the mapping are provided in Appendix A. Under each carrier, the median factor was calculated in each category.

METHODS OF ANALYSIS

After manipulating the data into a more useful conglomerate dataset that could be associated with the CHR socioeconomic dataset, the following analyses were conducted.

HIERARCHICAL CLUSTERING

Hierarchical clustering was performed using the average linkage methods for which the mean inter-cluster dissimilarity is computed. The Euclidean distance was used to measure dissimilarity—where all pairwise dissimilarities between the observations in cluster A and the observations in cluster B are used to record the average of these dissimilarities, as the distance between cluster A and cluster B. The hierarchical clustering results in a tree-based representation of the counties, known as a dendrogram. We cut the dendrogram to form the desired clusters. The clustering results for Florida counties can be visualized in a heat map, such as Figure 2.

BUBBLE PLOTS

Bubble plots were applied to visualize the relationship between demographic and rating variables. Bubble sizes corresponded to the population of the rating area. In some cases, we used various colors of bubbles to identify counties with similar characteristics, which used hierarchical clustering for grouping. This allowed for deeper

exploration into variable relationships. As an example, Figure 3 is a bubble plot that uses the hierarchical clustering of median household income as a third variable of exploration.

CORRELATION ANALYSIS

We calculated correlations between all available variables (both rating and demographic) to help determine which relationships to explore further. There were two correlation methods considered:

- Pearson Correlation measures the linear correlation between the two variables
- Spearman Correlation is a nonparametric measure of the rank correlation between the two variables. It equals the Pearson correlation between the rank values of these two variables.

Pearson correlation assesses linear relationships well, while the Spearman correlation works well with monotonic relationships (whether linear or not) without assuming knowledge of the variables' distributions. A positive correlation coefficient corresponds to an increasing monotonic relationship between the two variables and vice versa. We found consistent results from both correlation measures.

To evaluate the significance of correlations and to account for the effect of county size at the same time, we apply the population-weighted simple linear regressions and conduct hypothesis tests. The p-values of F test for these regressions were used to measure significance. We considered any p-value lower than 0.05 as statistically significant for the given demographic variable/rating variable combination. A positive estimated regression coefficient for the demographic variable was interpreted as a positive correlation between the rating variable and the demographic variable and vice versa. Data was centered to make the intercept of the regression as close to 0 as possible to make the output easier to interpret directionally.

Review of Results

TOBACCO SURCHARGE

Before the implementation of the ACA, health insurers rated or declined coverage for individuals for several factors that could be linked to a person's lifestyle, e.g., smoking, alcohol or other drug abuse or obesity. Post ACA implementation, preexisting health conditions are no longer allowable rating factors with the exception of tobacco use. Smokers can still be charged up to 50% more than non-smokers under the ACA.

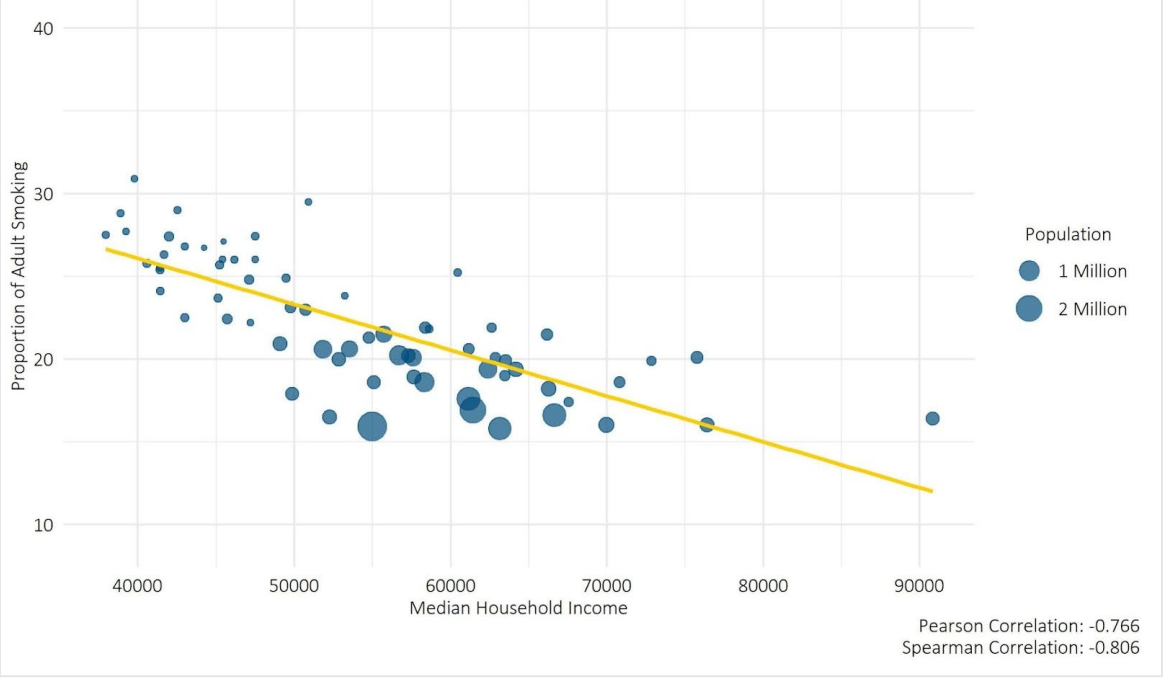
The below summarizes our analysis of the interaction between the smoker load and several demographic and socioeconomic variables, including household income, race/ethnicity, high school completion and food environment index. Conclusions and commentary related to the tobacco surcharge are included at the end of this section. Complete statistically significant (<0.05 p-value) linear regression results have been included in Appendix B but generally align with the conclusions developed below based on correlation (Pearson, r, and Spearman, rho) and bubble plots. Remember that correlation does not imply causation. Thus, a correlation between smoker load and a demographic or socioeconomic variable is not meant to imply that the variable was used to develop (or "cause") the smoker load.

CORRELATION OF SMOKING AND HOUSEHOLD INCOME

As discussed in the introduction to this paper, smokers are more likely to be socioeconomically disadvantaged as compared to their non-smoking peers. The data in our research affirms this as shown in Figure 1. To orient you to the figure, the proportion of adult smokers is on the y-axis and median household income is on the x-axis. The dark blue bubbles represent Florida counties. Thus, each bubble plots a unique Florida county against smoking proportion and household income. The size of the bubble corresponding to the population size of the county. The yellow line represents a simple linear regression trend line. As you can see in Figure 1, counties with higher proportions of adult smoking are correlated with lower household income levels.

Figure 1

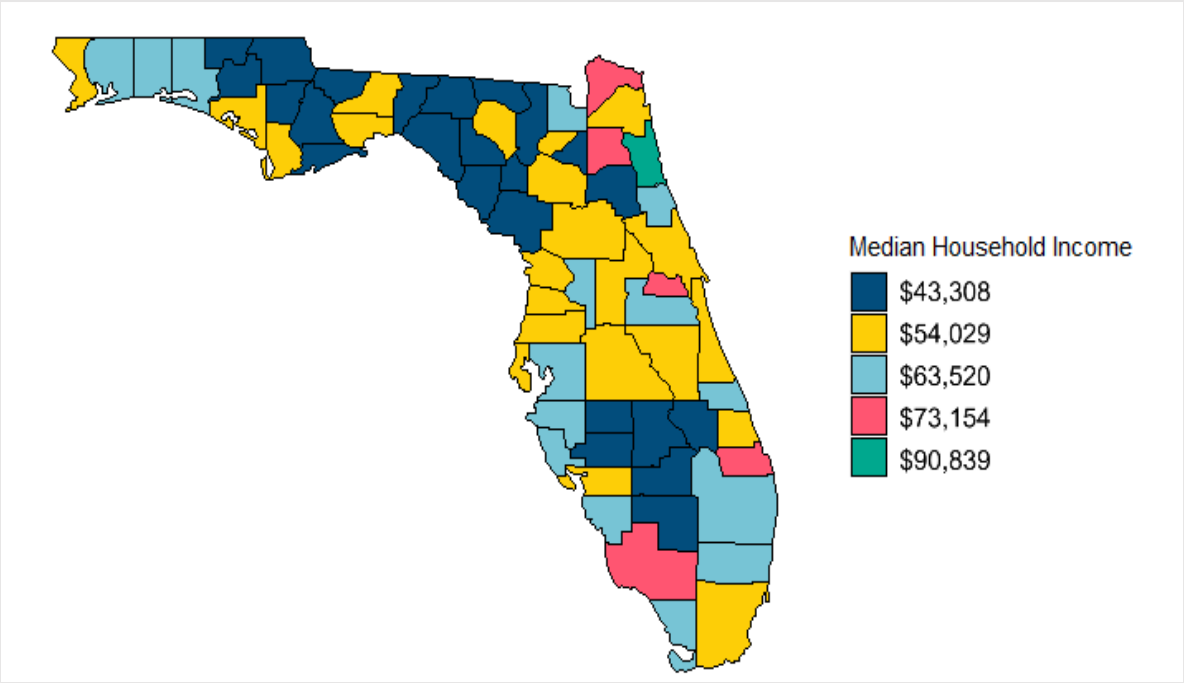
PREVALENCE OF ADULT SMOKING COMPARED TO MEDIAN HOUSEHOLD INCOME IN FLORIDA



Adult smoking is strongly negatively correlated with median household income (−0.766 Pearson correlation and −0.806 Spearman correlation). Variables are negatively correlated if an increase in one variable is associated with a decrease in the other variable. In this case, the proportion of adults who smoke decreases as median household income increases.

The results demonstrate that tobacco loads clearly affect lower income populations much more substantially than higher income populations. This correlation with higher rates of smoking and lower household income holds for all racial/ethnic groups as well. To demonstrate this visually, first, counties were grouped into five clusters with the following average median household income. Figure 2 shows clusters of counties using the color associated with the cluster's median household income.

Figure 2
MEDIAN HOUSEHOLD INCOME CLUSTERS IN FLORIDA



Bubble charts were created plotting the proportion of adult smoking on the y-axis and the proportion of a given racial or ethnic group on the x-axis. Data points represent a unique Florida county with the size of the data point, or bubble, commensurate to population size of the county. The color of the bubbles corresponds to the median household income shown in Figure 2. The resulting charts are shown in Figures 3 through 6. Racial/ethnic groups studied include Hispanic/Latino, Asian American, Non-Hispanic/Latino Black/African American and Non-Hispanic white. The focus of this paper is on statistically significant results. In cases where no statistically significant relationships were observed, we excluded or only briefly mentioned results. This is not meant to imply that the results are not important; rather, that there was insufficient statistical evidence to draw conclusions for this paper.

The correlation between smoking proportion and median household income in Figure 1 holds for all the following figures. Specifically, the lowest median household income is associated with the highest rates of smoking, the second lowest median household income is associated with slightly lower rates of smoking and so on.

Figure 3 plots counties as bubbles with proportion of Hispanic/Latino shown on the x-axis and proportion of adult smoking on the y-axis. Counties with lower proportions of smoking show as lower on the y-axis. Counties with smaller proportions of Hispanic/Latino are shown further on the left of the x-axis.

The following items are of note in this chart:

- Hispanic/Latino adults are less likely to smoke than adults of other races/ethnicities. Higher proportions of adult smoking are negatively correlated with higher proportions of Hispanic/Latino populations. The correlation statistics are shown in Table 1 (Pearson, r, −0.408 and Spearman, rho, −0.513). Thus Hispanic/Latino populations may be less likely to experience smoking loads than other races/ethnicities.
- An interesting pattern emerges when looking at the clustering. Regardless of the proportion of Hispanic/Latino population, the counties with the lowest median household income counties have the highest proportion of adult smoking and are therefore more likely to be charged a smoking load.

Figure 3

PREVALENCE OF ADULT SMOKING AND PROPORTION OF HISPANICS/LATINOS IN FLORIDA

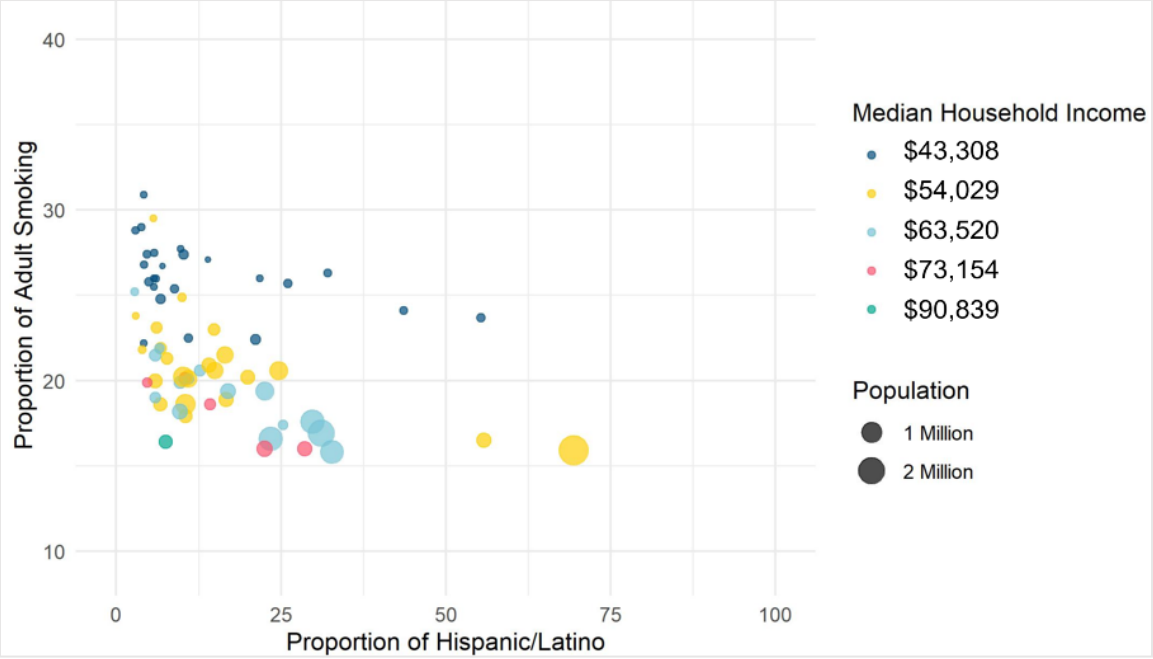


Figure 4 plots counties as bubbles with proportion of Asian American shown on the x-axis and proportion of adult smoking on the y-axis. Counties with lower proportions of smoking show as lower on the y-axis. Counties with smaller proportions of Asian American are shown further on the left of the x-axis.

The following items are of note in this chart:

- Higher proportions of adult smoking are negatively correlated with higher proportions of Asian American populations. The correlation statistics are shown in Table 1 (Pearson, r, −0.730 and Spearman, rho, −0.827). Thus, Asian American populations may be less likely to experience smoking loads as compared to other races/ethnicities.
- Like the Hispanic/Latino chart, an interesting pattern emerges when looking at the clustering. Regardless of the proportion of Asian American population, the lowest median household income counties have the highest proportion of adult smoking and are therefore more likely to be charged a smoking load.

Figure 4

PREVALENCE OF ADULT SMOKING AND PROPORTION OF ASIAN AMERICANS IN FLORIDA

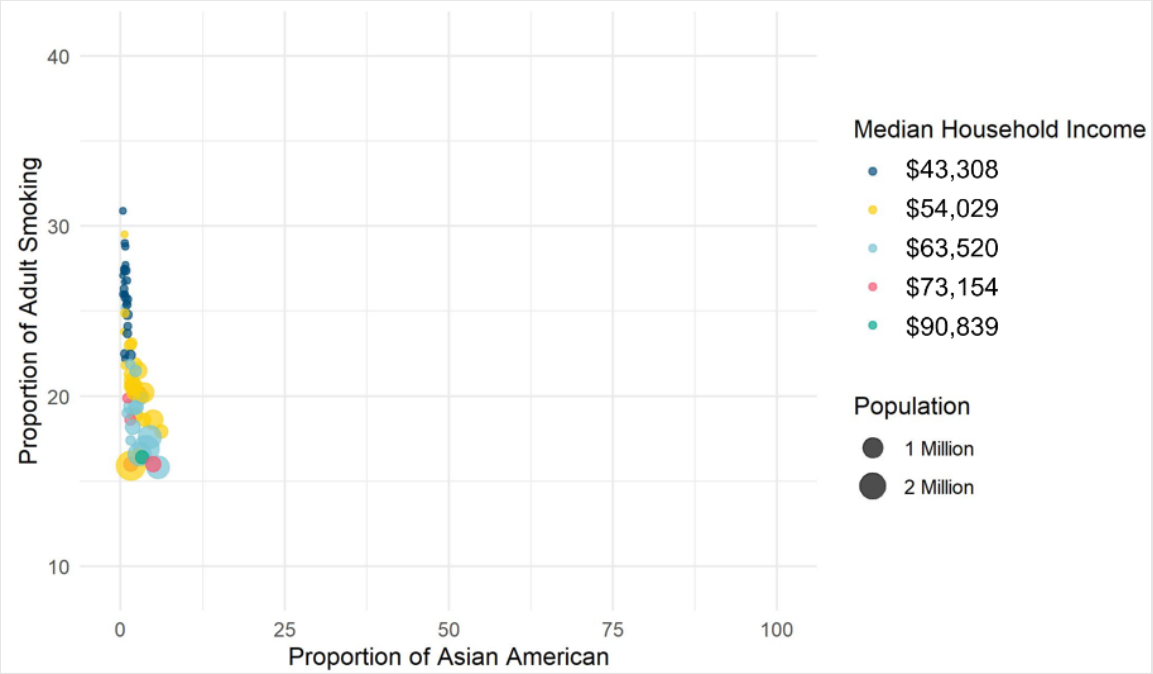


Figure 5 plots counties as bubbles with proportion of Non-Hispanic/Latino Black/African American shown on the x-axis and proportion of adult smoking on the y-axis. Counties with lower proportions of smoking show as lower on the y-axis. Counties with smaller proportions of African American are shown further on the left of the x-axis.

The following items are of note in this chart:

- There is no statistically significant correlation between proportion of adult smoking and proportion of Non-Hispanic/Latino Black/African American. The correlation statistics are shown in Table 1 (Pearson, r, 0.116 and Spearman, rho, 0.160).
- Like the Hispanic/Latino chart, an interesting pattern emerges when looking at the clustering. Regardless of the proportion of African American population, the lowest median household income counties have the highest proportion of adult smoking and are therefore more likely to be charged a smoking load.

Figure 5

PREVALENCE OF ADULT SMOKING AND PROPORTION OF NON-HISPANIC/LATINO BLACK/AFRICAN AMERICANS IN FLORIDA

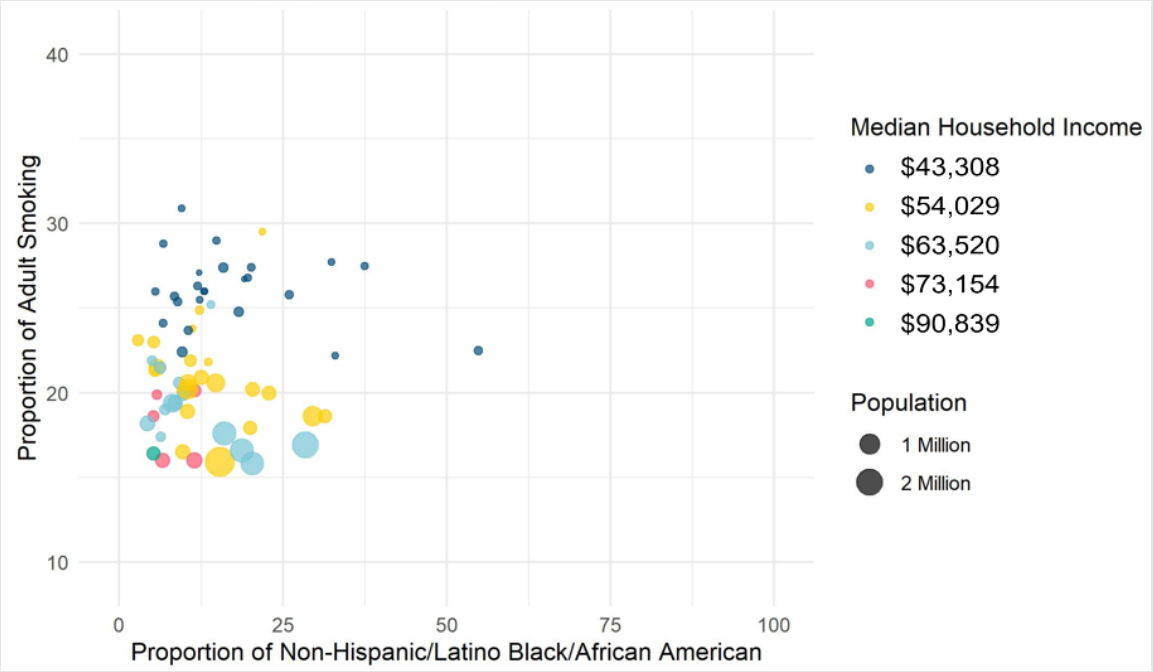


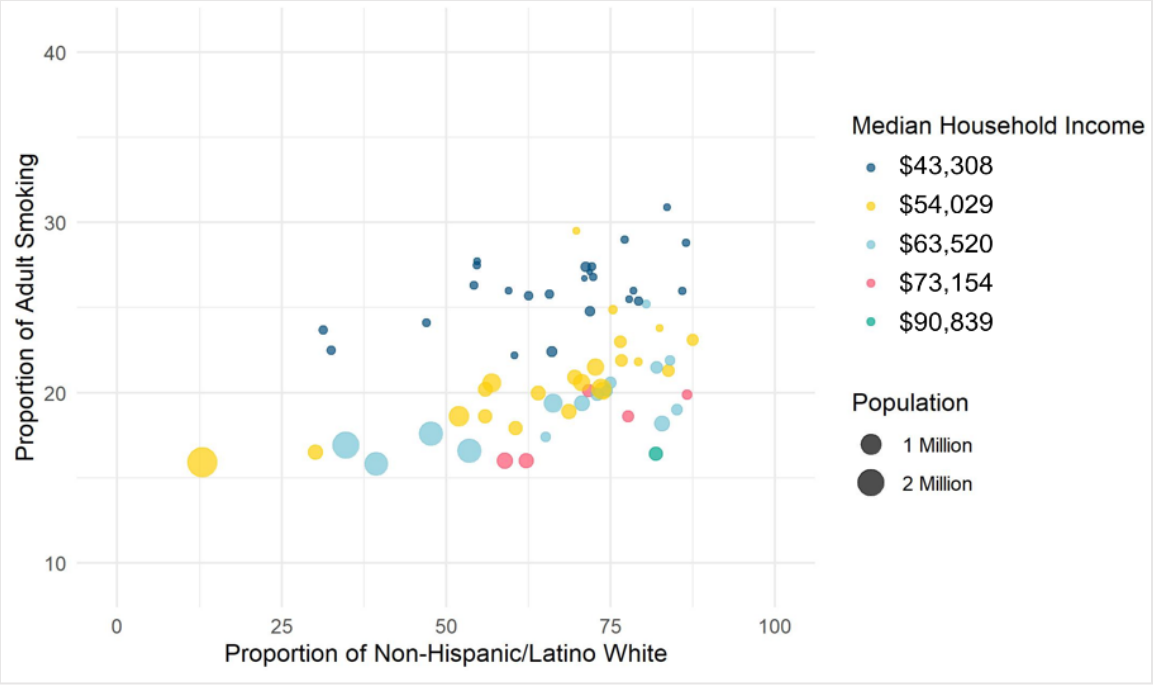
Figure 6 plots counties as bubbles with proportion of Non-Hispanic/Latino white shown on the x-axis and proportion of adult smoking on the y-axis. Counties with lower proportions of smoking show as lower on the y-axis. Counties with smaller proportions of Non-Hispanic/Latino white are shown further on the left of the x-axis.

The following items are of note in this chart:

- There is a weak correlation between proportion of adult smoking and proportion of Non-Hispanic/Latino white. The correlation statistics are shown in Table 1 (Pearson, r, 0.330 and Spearman, rho, 0.274). It would be tempting to conclude that lower income Non-Hispanic/Latino whites may be more affected by the tobacco load surcharge than other racial groups within the same income level. However, the F-test on the regression has a p-value greater than 0.05, which indicates the results are not statistically significant. If data had been available at the person or household level, further analysis may have indicated statistical significance.
- Like the Hispanic/Latino chart, a pattern emerges when looking at the clustering. Regardless of the proportion of Non-Hispanic/Latino white population, the lowest median household income counties have the highest proportion of adult smoking and are therefore more likely to be charged a smoking load.

Figure 6

PREVALENCE IN ADULT SMOKING AND PROPORTION OF NON-HISPANIC/LATINO WHITES IN FLORIDA



CORRELATION OF SMOKING AND RACE/ETHNICITY IN FLORIDA

We observed a significant negative correlation between the percentage of Asian American or Hispanic/Latino race/ethnicity in a county and the percentage of adult smokers. Interestingly, the strongest positive correlation between proportion of adult smoking and proportion of a given race/ethnicity is for the non-Hispanic/Latino white population as can be seen in Table 1 and Figure 6, but this correlation is still relatively weak. It would be tempting but without sufficient evidence to conclude that lower income Non-Hispanic/Latino whites may be more affected by the tobacco load surcharge than other racial groups within the same income level. This is because of lack of data at the person or household level.

Table 1

CORRELATION OF PERCENTAGE OF ADULT SMOKING AND PERCENTAGE OF EACH RACE/ETHNICITY

Race/Ethnicity	Pearson Correlation, r	Spearman Correlation, rho
Hispanic/Latino	−0.408	−0.513
Asian American	−0.730	−0.827
Non-Hispanic/Latino Black/African American	0.116	0.160
Non-Hispanic/Latino White	0.330	0.274

CORRELATION OF SMOKING AND OTHER SOCIO-ECONOMIC VARIABLES

Unsurprisingly, median household income has a strong positive Spearman correlation with other data fields typically associated with higher incomes, i.e., High School Completion, Some College and Food Environment Index is included in this group as well. Thus, the correlation between smoking and these data fields (High School Completion, Some

College and Food Environment Index) is also particularly strong. Note that as previously stated in the Data and Methodology section, the Food Environment Index is a scaled index (0 is worst, 10 is best) and equally weights the percentage of the population that is low income and does not live close to a grocery store with the percentage of the population that did not have access to a reliable source of food in the last year. Results are shown in Tables 2 and 3 as well as Figures 7–9.

Table 2
CORRELATION OF MEDIAN HOUSEHOLD INCOME AND OTHER SOCIOECONOMIC INDICATORS

Socioeconomic Indicator	Pearson Correlation, r	Spearman Correlation, rho
High School Completion	0.740	0.806
Some College	0.720	0.744
Food Environment Index	0.675	0.719

Table 3
CORRELATION OF ADULT SMOKING AND OTHER SOCIOECONOMIC INDICATORS

Socioeconomic Indicator	Pearson Correlation, r	Spearman Correlation, rho
High School Completion	−0.731	−0.757
Some College	−0.817	−0.826
Food Environment Index	−0.648	−0.688

Figure 7 plots counties as bubbles with proportion of the population completing high school shown on the x-axis and proportion of adult smoking on the y-axis. Counties with lower proportions of smoking show as lower on the y-axis. Counties with lower proportions of high school completion are shown further on the left of the x-axis.

The following items are of note in this chart:

- There is a significant positive correlation between median household income and high school completion proportion. The correlations are shown in Table 2 (Pearson, r, 0.740 and Spearman, rho, 0.806).
- There is a significant negative correlation between proportion of adult smoking and high school completion proportion. The correlations are shown in Table 3 (Pearson, r, −0.731 and Spearman, rho, −0.757).
- Hence, those who have less formal education are more likely to have lower incomes and thus, are more likely to be charged a tobacco surcharge.

Figure 7

PREVALENCE OF ADULT SMOKING AND PROPORTION OF HIGH SCHOOL COMPLETION IN FLORIDA

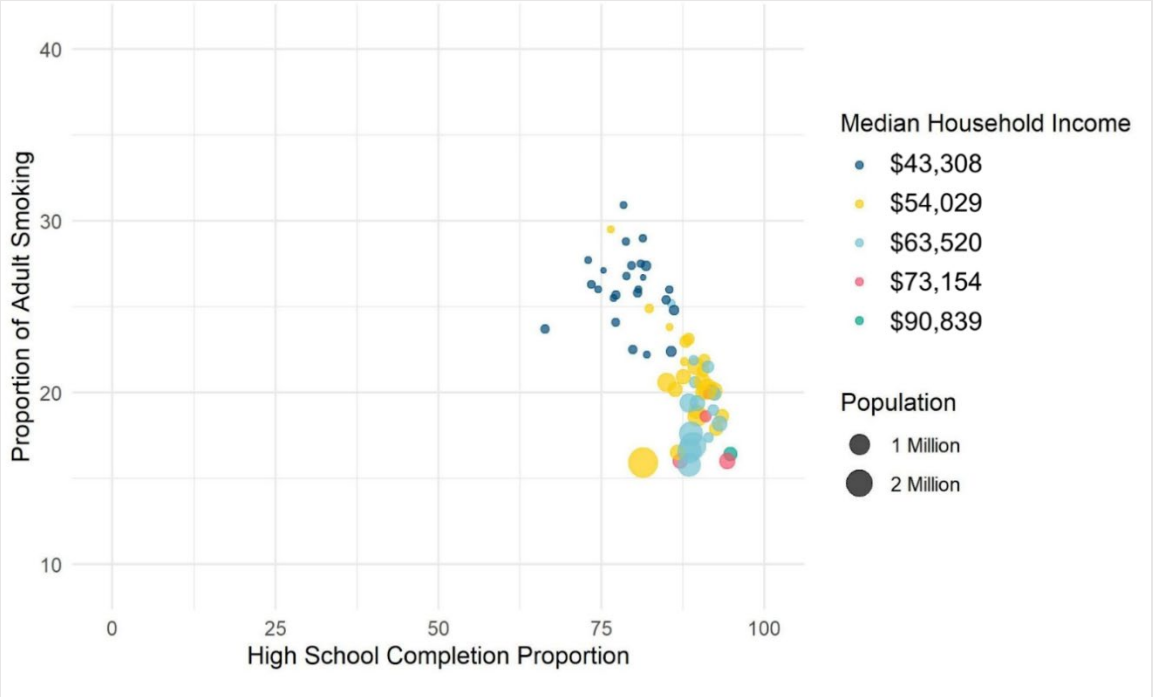


Figure 8 plots counties as bubbles with proportion of some college shown on the x-axis and proportion of adult smoking on the y-axis. Counties with lower proportions of smoking show as lower on the y-axis. Counties with lower proportions of some college are shown further on the left of the x-axis.

The following items are of note in this chart:

- There is a significant positive correlation between median household income and having completed some college education. The correlations are shown in Table 2 (Pearson, r, 0.720 and Spearman, rho, 0.744).
- There is a significant negative correlation between proportion of adult smoking and having completed some college education. The correlations are shown in Table 3 (Pearson, r, −0.817 and Spearman, rho, −0.826).

A similar conclusion to the prior figure can be drawn. Those who have less formal education and have lower incomes are more likely to be charged a tobacco surcharge.

Figure 8

PREVALENCE OF ADULT SMOKING AND PROPORTION OF SOME COLLEGE EDUCATION IN FLORIDA

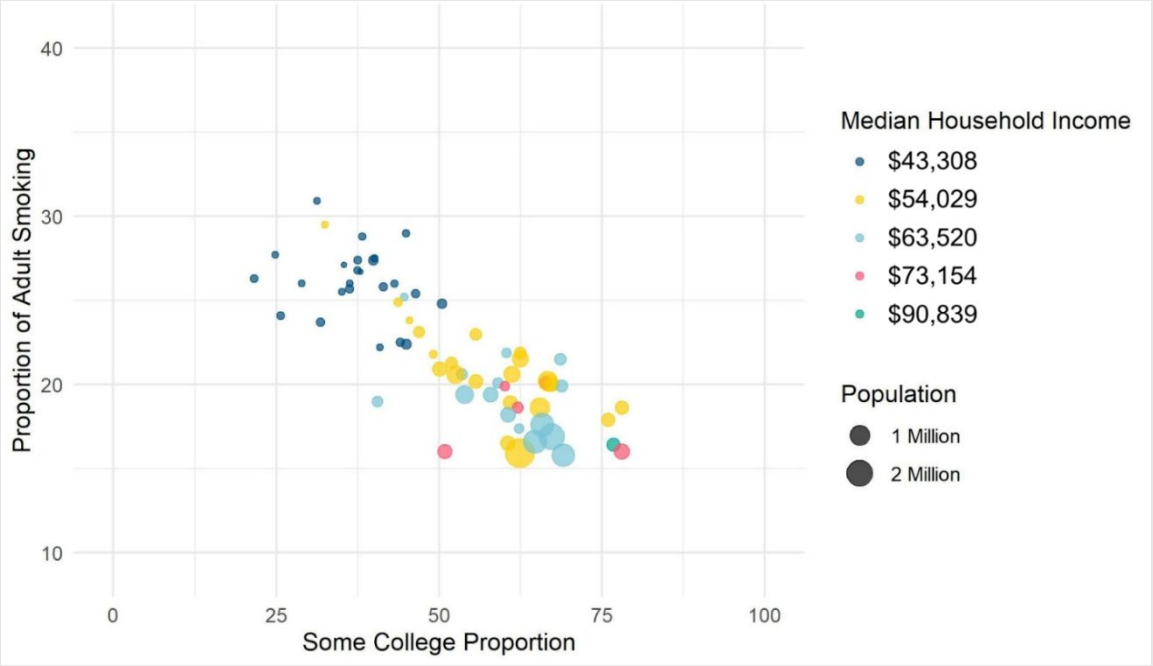


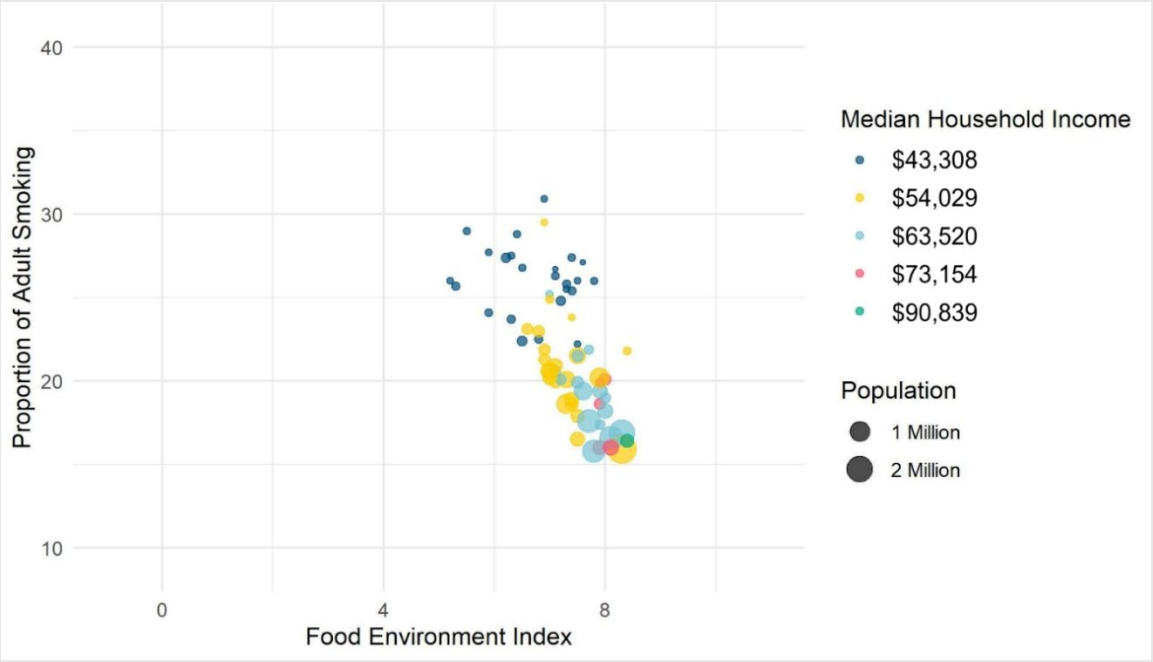
Figure 9 plots counties as bubbles with food environment index shown on the x-axis and proportion of adult smoking on the y-axis. Counties with lower proportions of smoking show as lower on the y-axis. Counties with lower food environment (less access to healthy food) indices are shown further on the left of the x-axis.

The following items are of note in this chart:

- There is a significant positive correlation between median household income and food environment index. The correlations are shown in Table 2 (Pearson, r, 0.675 and Spearman, rho, 0.719).
- There is a significant negative correlation between proportion of adult smoking and food environment index. The correlations are shown in Table 3 (Pearson, r, −0.648 and Spearman, rho, −0.688).

Those with worse food insecurity are more likely to smoke and, therefore, be charged a smoker surcharge.

Figure 9
PREVALENCE OF ADULT SMOKING AND FOOD ENVIRONMENT INDEX



COMMENTARY AND CONCLUSIONS REGARDING THE USE OF A TOBACCO SURCHARGE

Our analysis concludes that smokers are more likely to not only have lower incomes than their non-smoking peers, but they are also more likely to have a less formal education and have worse access to healthy foods. This held across all racial/ethnic groups. Persons in counties with higher proportions of Asian American or Hispanic/Latino populations were less likely to smoke and were therefore less likely to be impacted by a smoking load.

Additionally, the researchers found this commentary included in the description of the Food Environment Index from the County Health Ranking's website notable: "There is strong evidence that food deserts are correlated with high prevalence of overweight, obesity and premature death as supermarkets traditionally provide healthier options than convenience stores or smaller grocery stores. Additionally, those with low income may face barriers to accessing a consistent source of healthy food. Lacking consistent access to food is related to negative health outcomes such as weight gain, premature mortality, asthma and activity limitations, as well as increased health care costs."[7] Thus, it is possible that the load calculated could include the cost of conditions that are correlated with smoking rather than caused by smoking. However, that analysis is outside the scope of this research.

The actual ranges of tobacco surcharges by age in Florida are shown in Table 4. For a 21–34-year-old, the median surcharge is 12%. A few insurers have decided no tobacco load is necessary in some, but not all, geographic areas in Florida, while others are charging up to 40%. Although tobacco rates on average are generally less than the 50%

[7] Food environment index. County Health Rankings & Roadmaps. (n.d.). Retrieved January 7, 2022, from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/health-behaviors/diet-exercise/food-environment-index

federally mandated maximum allowable surcharge and some states have limited the maximum surcharge of 50% to lower amounts, including 1% in Vermont, most states have not limited the maximum.

Table 4
TOBACCO SURCHARGE RANGES IN FLORIDA BY AGE

Age Range	Min	25th	Median	Mean	75th	Max
21–34	0%	10%	12%	12%	15%	20%
35–54	0%	15%	18%	17%	20%	30%
55+	0%	15%	20%	19%	20%	40%

On top of this, because the premium subsidies allowed for in the ACA cannot be used to cover the tobacco surcharge, smokers, who are already more likely to be an economically disadvantaged population, can have significantly higher net rates than non-smokers. To illustrate this, we pulled rates (after subsidization) for sample individuals in Florida in 2022 from the latest plan preview tool from Healthcare.gov.[8] We considered three hypothetical individuals:

- Scenario 1: 64-year-old making $20,000 per year
- Scenario 2: 50-year-old making $27,500 per year
- Scenario 3: 35-year-old making $35,000 per year

Tables 5 through 7 build up the net premiums for a non-smoker and smoker for each of these scenarios, respectively. First, the monthly non-smoker rate is provided. Then, the smoking load is applied (0% for non-smokers) to get the monthly premium rate prior to premium subsidies. A monthly subsidy reduces this to the final monthly net premium cost to the member. The totals are then annualized. The additional premium smoking members experience because of the smoking load are then expressed in dollars and as a percentage.

Table 5
SCENARIO 1: 64-YEAR-OLD MAKING $20,000 PER YEAR

Scenario 1	Non-Smoker	Smoker
Age	**64**	**64**
Income	**$20,000**	**$20,000**
Monthly Non-Smoker Premium Rate	$1,049	$1,049
Smoking Load	0%	20%
Monthly Premium Rate (without Subsidy)	$1,049	$1,259
Monthly Subsidy	($1,049)	($1,049)
Monthly Premium Net Subsidy (Cost to Member)	**$0**	**$210**
Annual Premium Cost to Member	**$0**	**$2,517**
Annual Additional Premium for Smoking ($)		**$2,517**
Annual Additional Premium for Smoking (%)		**N/A**

[8] https://www.healthcare.gov/see-plans/#/

Table 6

SCENARIO 2: 50-YEAR-OLD MAKING $27,500 PER YEAR

Scenario 2	Non-Smoker	Smoker
Age	50	50
Income	$27,500	$27,500
Monthly Non-Smoker Premium Rate	$624	$624
Smoking Load	0%	20%
Monthly Premium Rate (without Subsidy)	$624	$749
Monthly Subsidy	($568)	($568)
Monthly Premium Net Subsidy (Cost to Member)	**$56**	**$181**
Annual Premium Cost to Member	**$678**	**$2,177**
Annual Additional Premium for Smoking ($)		**$1,499**
Annual Additional Premium for Smoking (%)		**221%**

Table 7

SCENARIO 2: 35-YEAR-OLD MAKING $35,000 PER YEAR

Scenario 3	Non-Smoker	Smoker
Age	35	35
Income	$35,000	$35,000
Monthly Non-Smoker Premium Rate	$427	$427
Smoking Load	0%	15%
Monthly Premium Rate (without Subsidy)	$427	$491
Monthly Subsidy	($287)	($287)
Monthly Premium Net Subsidy (Cost to Member)	**$140**	**$204**
Annual Premium Cost to Member	**$1,683**	**$2,452**
Annual Additional Premium for Smoking ($)		**$769**
Annual Additional Premium for Smoking (%)		**46%**

As these tables show, in some cases, smokers have to pay a premium while their non-smoking counterparts pay nothing out-of-pocket. In other cases, the same plan can be over 200% more expensive once a tobacco surcharge is added. This furthermore comes with little impact to smoking cessation levels. As noted in a 2016 Health Affairs article, "tobacco surcharges conflicted with a major goal of the ACA—increased financial protection—without increasing smoking cessation."[9]

9 Friedman, A. S., AB, K., CM, K., T, L., Al., E., K, C., SC, S., KG, V., SD, H., JR, H., BD, S., J, G., AC, L., U, G., BL, C., TJ, L., Manz, K. C., & Pesko, M. F. (2016, July 1). Evidence suggests that the ACA's tobacco surcharges reduced insurance take-up and did not increase smoking cessation: Health Affairs Journal. Health Affairs. Retrieved January 5, 2022, from https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2015.1540

Tobacco usage is clearly linked with increased health care costs. But, even if the tobacco factors are indicative of cost, our results indicate the health care delivery system studied could be impacting socioeconomic groups disparately. However, current tobacco surcharges may be capturing unintended factors, such as socioeconomic status or other comorbid conditions and if so, could be affecting socioeconomically disadvantaged populations in an inequitable fashion. Acquiring the detailed data and conducing the analysis necessary to investigate this question is a topic for further research.

AREA FACTORS

Beyond tobacco loads, one of the other allowable rating factors under the Affordable Care Act is based on the member or group's place of residence or rating area. The geographic rating factor is implemented to capture differences in health care unit cost and mix of services by geographic location and can vary between individual and small group markets. The state of Florida is comprised of 67 different rating areas—one for each county in the state. Each can potentially be assigned a unique rating factor, depending on the geographic differences as justified by the carrier. Our analysis was based on the average normalized area rating factor across carriers offering coverage in each these counties, relative to race/ethnicity proportions within that county. Correlations were also analyzed across area rating factors and other socioeconomic characteristics such as household income and the Food Environment Index. However, the most statistically significant correlation with area factors was race/ethnicity.

AREA FACTOR CORRELATION WITH HISPANIC/LATINO VERSUS NON-HISPANIC/LATINO WHITE POPULATIONS

The relationship between geographic area factors and race/ethnicity was further pursued after correlation analyses were run using the combination of tobacco, area and industry rating factors versus the available racial/ethnic and socioeconomic variables. Two of the strongest relationships were area factor versus proportion of Hispanic/Latino and area factor versus proportion of Non-Hispanic/Latino whites. This was true for both individual and small group ACA lines of business. These relationships were further studied and visualized through the use of bubble charts.

Figure 10 is the first of a set of bubble plots to follow. Each bubble on the graph represents one county or rating area of Florida. The y-axis for each of these graphs represents the average normalized area factor for a particular county. A higher area factor represents a higher average premium for ACA enrollees, all else being equal. In Figure 10 the x-axis represents the portion of Hispanic/Latino population within that county. The plots to follow are similar but for other races/ethnicities. As an example, the largest bubble plotted in the first graph represents Miami-Dade County. It has the largest population in the state of Florida, which is why it is the largest bubble on the graph. Based on its positioning, approximately 73% of the population is Hispanic/Latino and the average normalized area factor is 1.1. The county map in Figure 11 provides further visualization, created by the U.S. Census Bureau, showing the geographic location with the densest populations for each race/ethnicity.

The statistical significance of the relationships between these variables can be further verified through linear regression analyses, which is represented by the yellow lines on the bubble plots. A positive line slope (the line is higher on the right than on the left) implies that the two variables have a positive relationship. Therefore, the larger the proportion of the race/ethnicity, the higher the area factor. Inversely, a negative slope (the line is lower on the right than on the left) implies a negative relationship, and the larger the proportion of the race/ethnicity, the lower the area factor. Figures 10 and 11 represent county information within the individual market.

The regression line in Figure 10 is representative of the relationship between area factor and Hispanic/Latino proportions within Florida counties. Each bubble represents an individual county, and its size corresponds to the population size. The positive slope and positive correlation (Pearson, r, 0.296 and Spearman, rho, 0.224), implies that the greater a county's proportion of Hispanic/Latino population, the greater its area factor. In other words, Hispanics/Latinos in the Florida ACA individual market are more likely to have a great-than-average area factor.

Figure 10

INDIVIDUAL ACA MARKET AREA FACTORS AND PROPORTION OF HISPANICS/LATINOS BY FLORIDA COUNTY

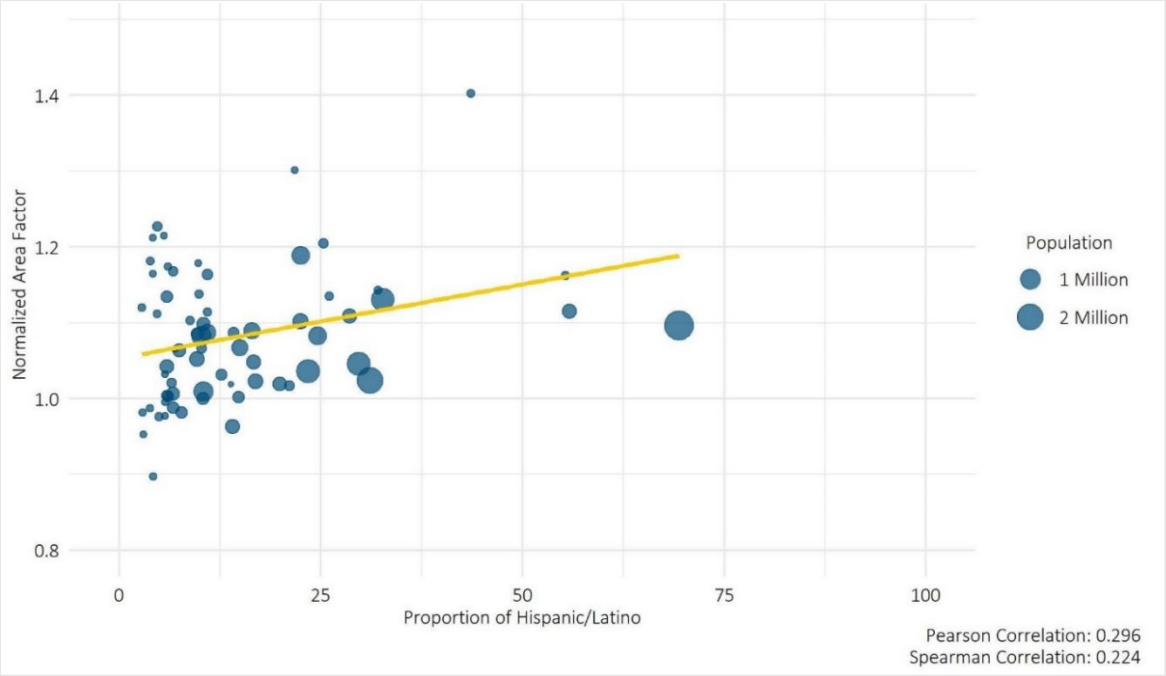
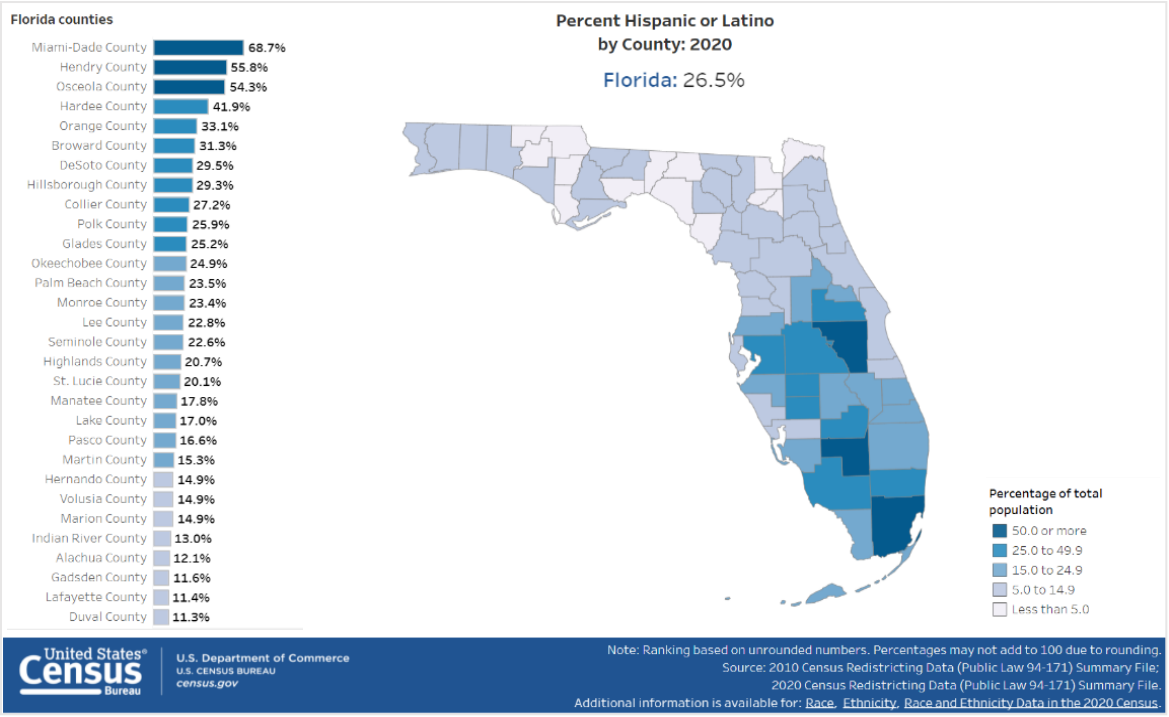


Figure 11

PERCENT HISPANIC OR LATINO BY COUNTY IN FLORIDA



Source: U.S. Census Bureau, 2010 Census Redistricting Data (Public Law 94–171) Summary File; 2020 Census Redistricting Data (Public Law 94–171) Summary File.

Beyond the regression takeaways, we had a few general observations. Note that Miami-Dade county is the county with the highest proportion of people that identify as Hispanic/Latino. This is the largest bubble and the farthest to the right in Figure 10. Although the cost of care in this county is generally higher than other areas of Florida, this county does not have the highest rating area factor, contrary to the regression line. There were also concerns that variances based on county size may drive rating factors. For example, we may expect that smaller counties with less access to health care resources would have higher cost care and subsequently a higher area rating factor as discussed in the Background section. However, this is not observed in Figure 10. The small bubbles, or smaller counties, cover both the lows and highs of the factor range on the y-axis. Similarly, larger bubbles, or larger counties, can be found on the low and high end of the spectrum. Therefore, county size does not appear to be significantly correlated to area factors.

Similar to Figure 10, Figure 12 is representative of the relationship between area factor and Non-Hispanic/Latino white proportions within Florida counties. However, the regression line has a negative slope, and the relationship has a negative correlation (Pearson, r, –0.205 and Spearman, rho, –0.195). The negative slope implies that the Non-Hispanic/Latino white population is more likely to see lower area factors in the Florida ACA individual market, which may translate to greater health care premiums. Again, the county map in Figure 13 provides further visualization, created by the U.S. Census Bureau, showing the geographic location with the densest populations of the Non-Hispanic/Latino white race/ethnicity.

Figure 12

INDIVIDUAL ACA MARKET AREA FACTORS AND PROPORTION OF NON-HISPANIC/LATINO WHITES BY FLORIDA COUNTY

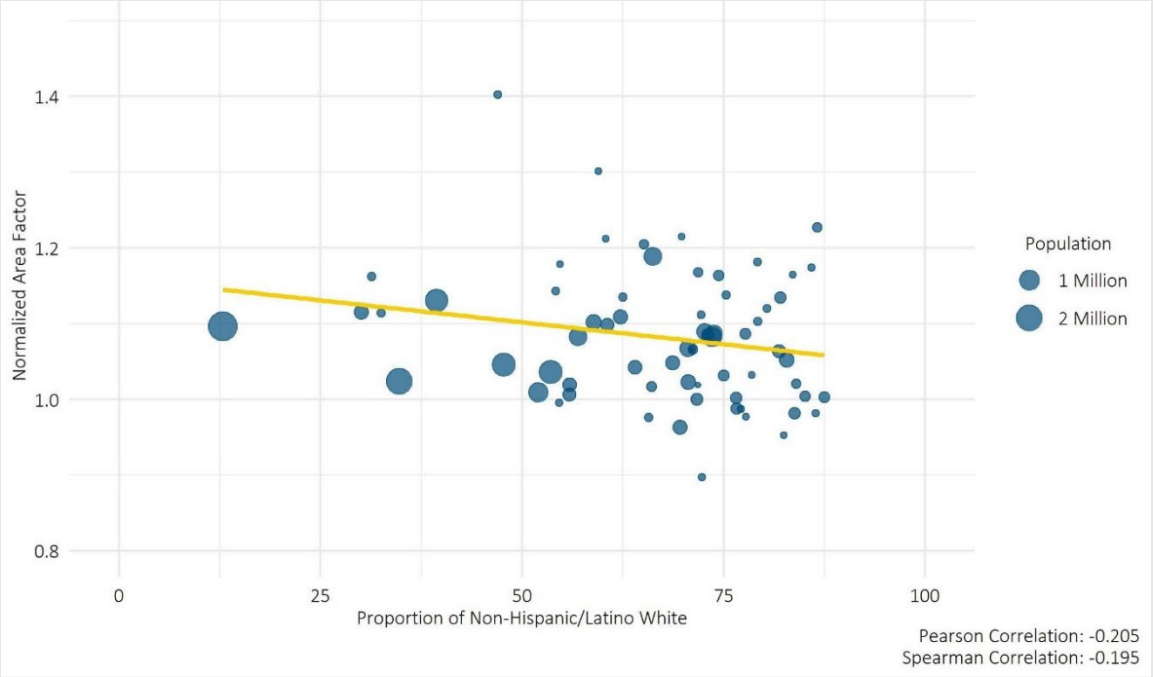
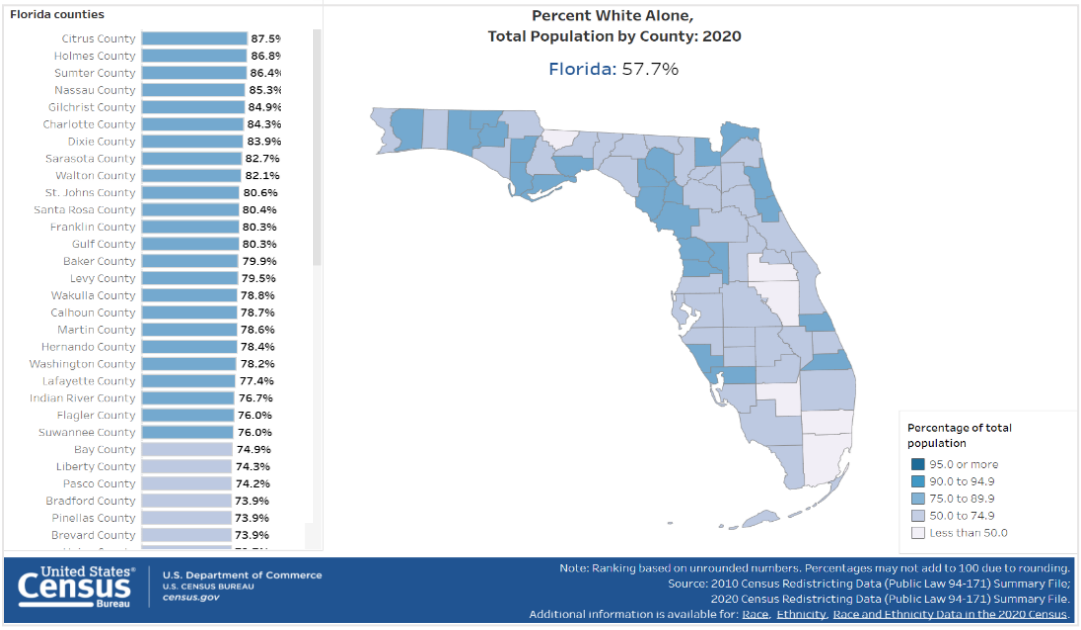


Figure 13

PERCENT WHITE ALONE BY COUNTY IN FLORIDA



Source: U.S. Census Bureau, 2010 Census Redistricting Data (Public Law 94–171) Summary File; 2020 Census Redistricting Data (Public Law 94–171) Summary File.

The following set of bubble plots represent similar relationships but for the small group market. The Hispanic/Latino population is positively correlated with area factor and the Non-Hispanic white population is negatively correlated. Therefore, in both the individual and small group ACA markets, higher Hispanic/Latino population percentages correlate with higher area rating factors and likely greater health premiums. Inversely, a higher Non-Hispanic/Latino white population correlates with lower rating area factors and likely lower health premiums.

Figure 14

SMALL GROUP ACA MARKET AREA FACTORS AND PROPORTION OF HISPANIC/LATINOS BY FLORIDA COUNTY

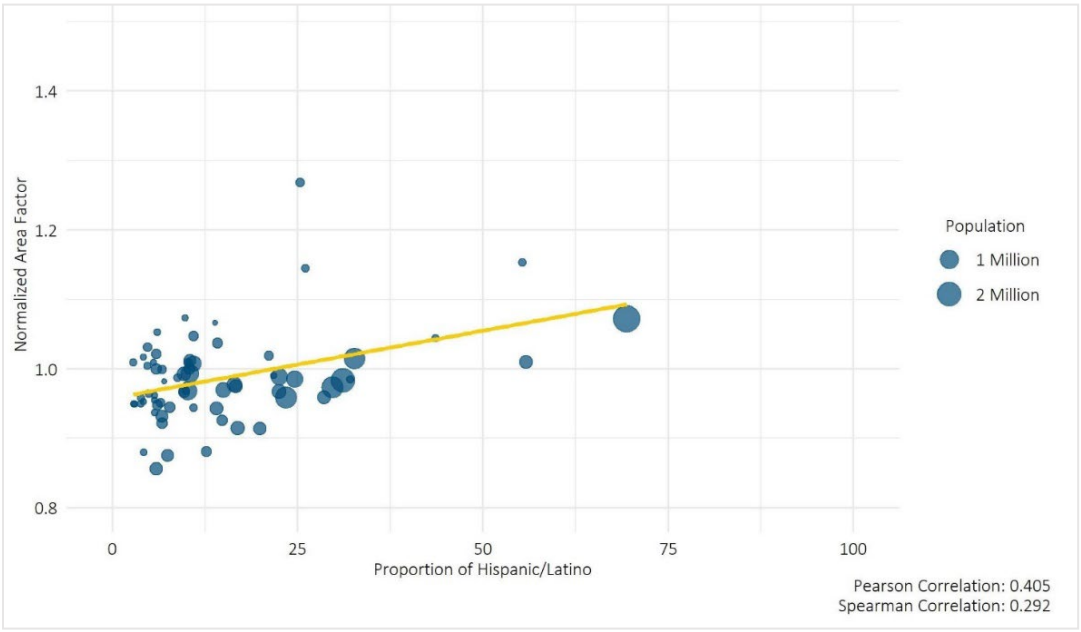
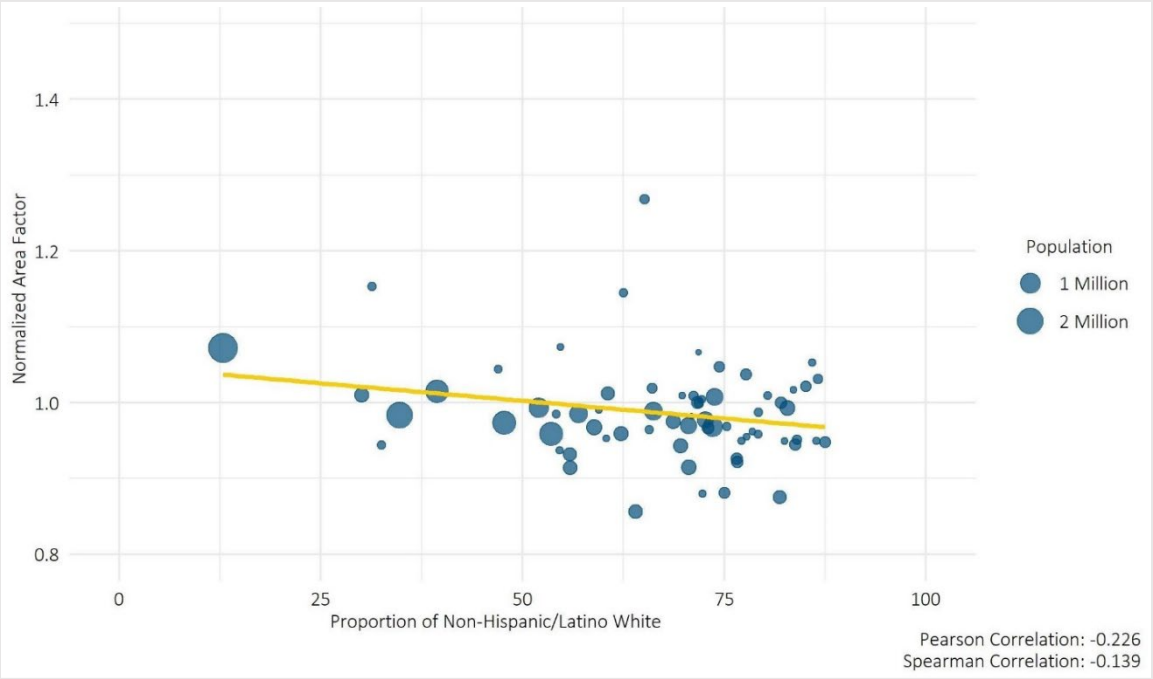


Figure 15

SMALL GROUP ACA MARKET AREA FACTORS AND PROPORTION OF NON-HISPANIC/LATINO WHITES BY FLORIDA COUNTY



Detailed linear regression statistics for bubble plot Figures 10, 12, 14 and 15 are provided in Appendix B of this report.

AREA FACTOR CORRELATION WITH OTHER RACES/ETHNICITIES

Data were also available for Non-Hispanic/Latino Black/African Americans and Asian Americans in Florida. When conducting linear regression analyses, greater proportions of a Non-Hispanic/Latino Black/African American population was slightly negatively correlated with area factors. Such was the case in both the individual and small group ACA markets. Counties in Florida with greater proportions of Non-Hispanic/Latino Black/African Americans were likely to have slightly lower area rating factors.

Results of the linear regression for the Asian American population were similarly slightly negatively correlated. However, Asian Americans makes up less than ten percent of the population in each of Florida's counties, making it difficult to determine a statistically credible observation. Bubble plots for these additional populations are shown in Figures 16 through 19.

Figure 16

INDIVIDUAL ACA MARKET AREA FACTORS AND PROPORTION OF NON-HISPANIC/LATINO BLACK/AFRICAN AMERICANS BY FLORIDA COUNTY

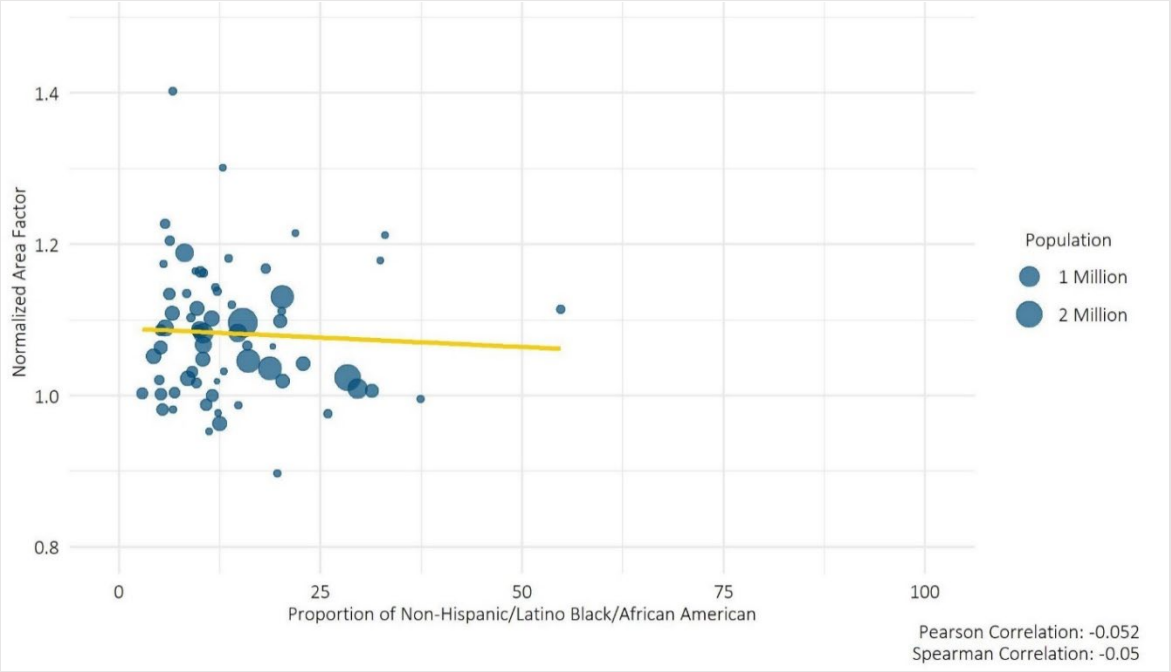


Figure 17

INDIVIDUAL ACA MARKET AREA FACTORS AND PROPORTION OF ASIAN AMERICANS BY FLORIDA COUNTY

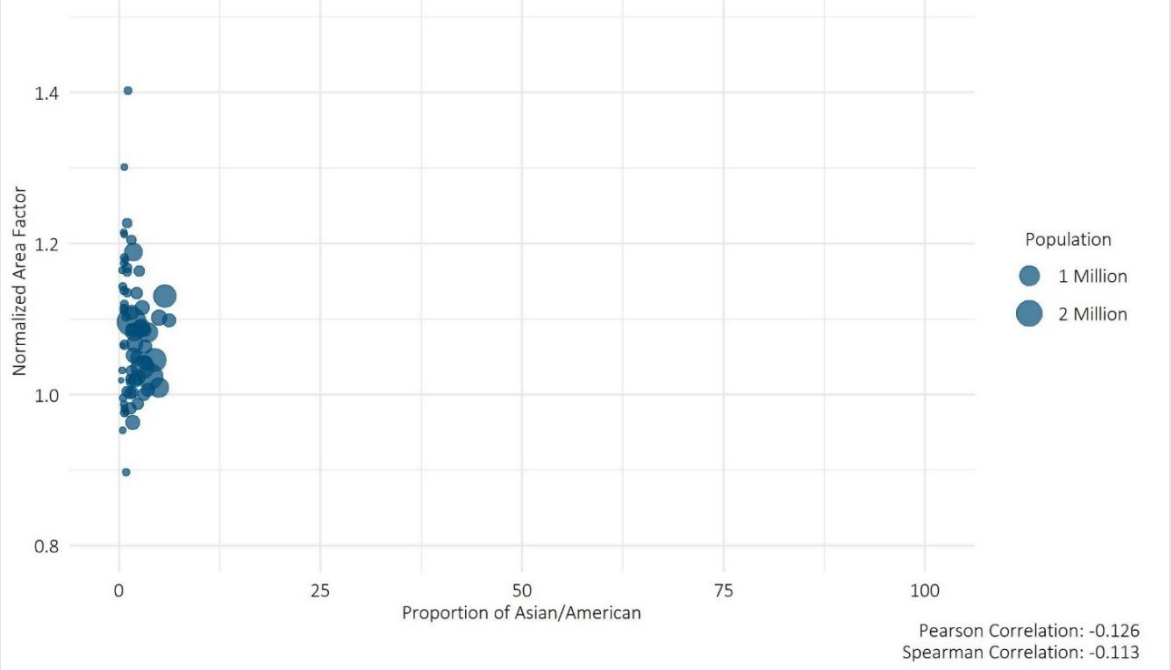


Figure 18

SMALL GROUP ACA MARKET AREA FACTORS AND PROPORTION OF NON-HISPANIC/LATINO BLACK/AFRICAN AMERICANS BY FLORIDA COUNTY

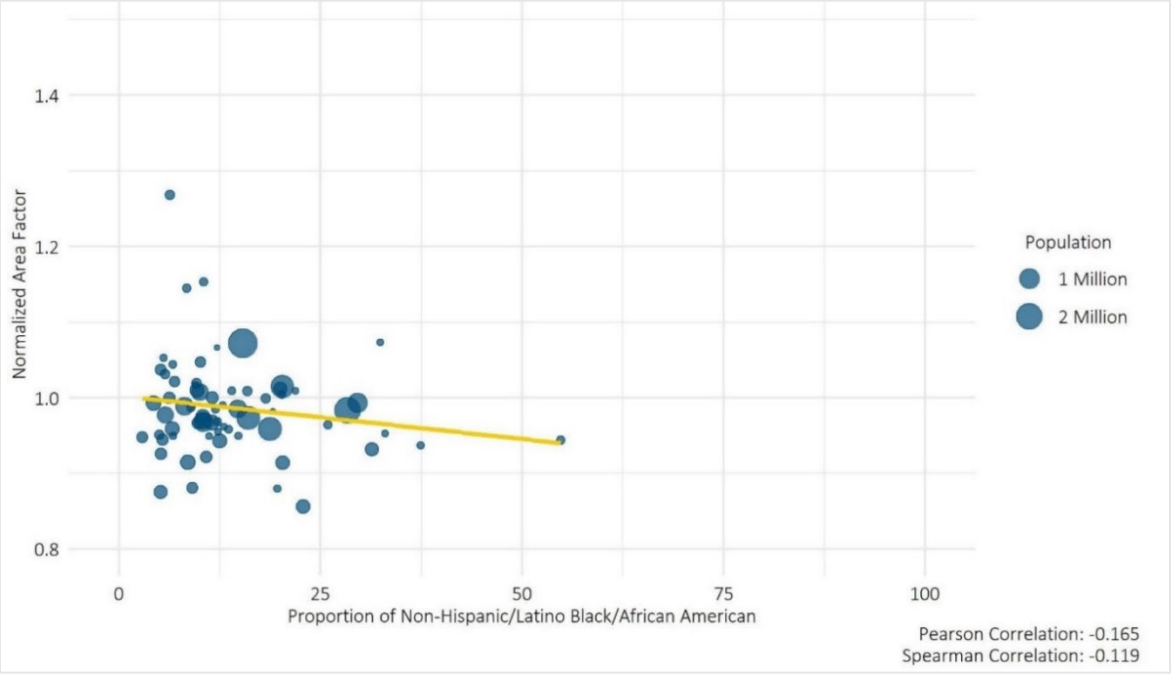
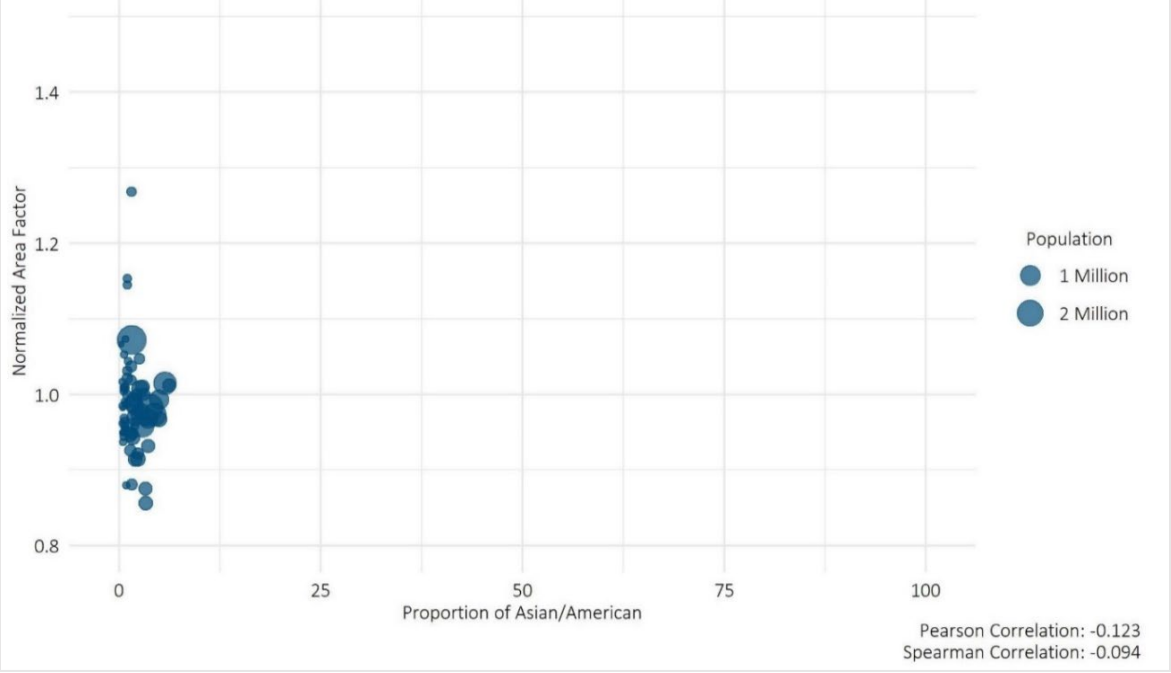


Figure 19

SMALL GROUP ACA MARKET AREA FACTORS AND PROPORTION OF ASIAN AMERICANS BY FLORIDA COUNTY



COMMENTARY AND CONCLUSIONS REGARDING GEOGRAPHIC FACTORS

Different locations' costs vary based provider contracting, cost of living, utilization differences, etc. This analysis studies the correlations of rating factors with race/ethnicity variables but does not determine if those rating factors are associated with higher claim costs. The study of claim costs was outside the scope of this research and is a topic for further research. However, if area factors are indicative of cost, our results indicate the health care delivery system studied could be impacting some demographic groups disparately. Specifically, our analysis concludes that area factors used to adjust ACA premiums are correlated with the Hispanic/Latino and Non-Hispanic/Latino white populations in Florida. Based on the distribution of the current Florida residents, the Hispanic/Latino population is more likely to see a higher area factor, where the Non-Hispanic/Latino white population is more likely to see a lower area factor.

After studying the same relationships for Non-Hispanic/Latino Black/African American and Asian/American populations in Florida, we did not observe the same correlation. This study did not consider federal reimbursement, uninsured rates, practice patterns or other external variables that could drive these results. As more data is made available with the inclusion of demographic variables, this dispersion of premium costs is worth further consideration. In a recent effort, the Centers for Medicare & Medicaid Services (CMS) proposed in the HHS Notice of Benefit and Payment Parameters for 2023 Proposed Rule that EDGE data to be used for risk adjustment include race and ethnicity fields. If pursued, this dataset could be a potential venue for future research.

INDUSTRY FACTORS

In the large group commercial market, health insurers have used Standard Industrial Classification (SIC) codes and SIC code ranges for their industry factors for several years. SIC codes were first introduced in the late 1930s with the last revision by the United States to the codes occurring 35 years ago in 1987. These codes have been further segmented into 6-, 7- and 8-digit codes from the standard 4-digit code by private companies. However, most health insurers still use the 4-digit standard SIC code when rating large groups, which are defined as employers with more than 50 full time employees in most states.[10]

SIC codes are becoming less and less widely used by governmental agencies, like the U.S. Census Bureau. Thus, unfortunately, the publicly available data (the American Community Survey) did not include SIC code and rather used the North American Industry Classification System (NAICS) codes. Because of the level of data provided by the ACS and the availability of crosswalks between the two code systems, the results available to the researchers were rolled up at a fairly high level as shown in Appendix A.

Additionally, large group health insurers do not have to publicly disclose the exact industry factor used in their rates for most states. Colorado appears to be an exception to this rule and rating manuals for the major large group health insurers were found, which is why our analyses relied on these rating manuals rather than an equivalent manual in Florida. Typically, industry factors are not varied by state, which is why the researchers felt this was an appropriate simplification.

This loss of granularity in the industry code classification in the data when matched up with a rating factor makes it difficult to draw any conclusions as to the interactions that may exist between industry factors and rating factors. As can be seen in Figures 18 and 19, incomes and race are relatively evenly distributed across industries.

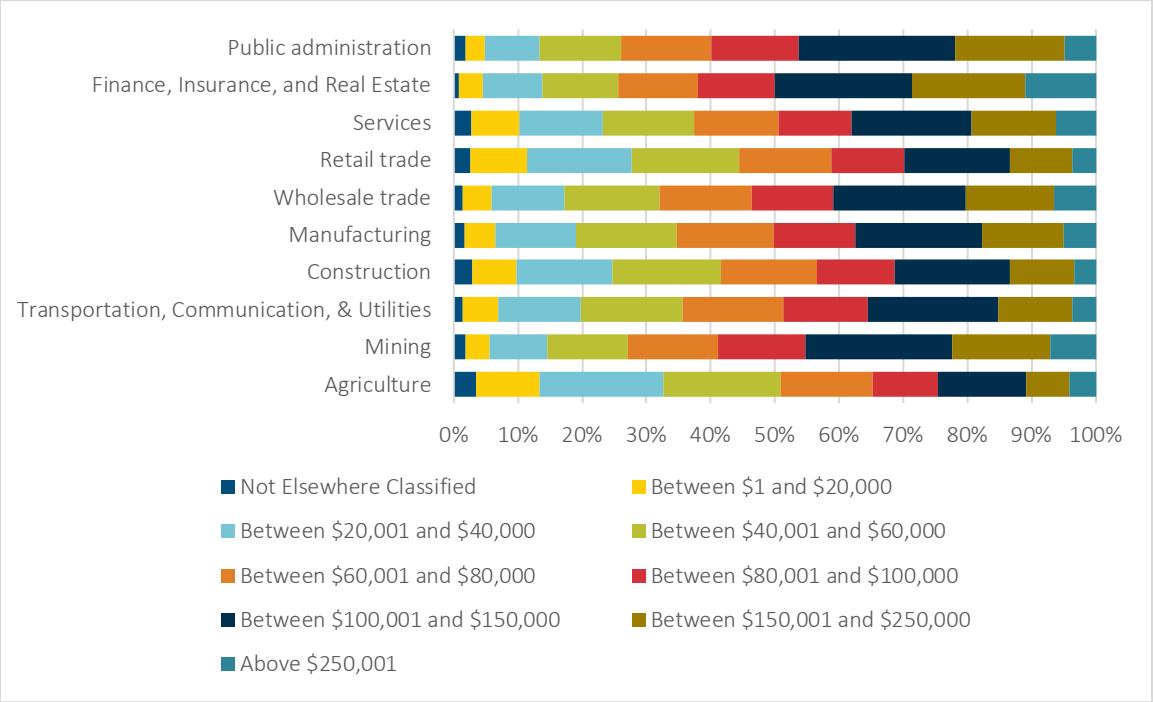
[10] SICCODE.com. (2020, January 21). History of SIC codes. SIC & NAICS Codes, Company Search, Business Lists - SICCODE.com. Retrieved January 7, 2022, from https://siccode.com/page/history-of-sic-codes

Along with the proportion of people across the U.S. within the industry making incomes within certain ranges. All the categories shown have a broad mix of income ranges. This makes it difficult to draw any conclusions about industry factor and socioeconomic status. Table 8 further demonstrates this. Table 8 shows the correlations between income range and industry factor; none are statistically significant.

Thus, we cannot say that there are disparate impacts in industry factor load being charged when examining socioeconomic status. This does not mean that disparate impacts are not occurring. Rather, because industry factor loads are not standardized and are not transparent, the factors studied were aggregated to such a level that detailed statistical analyses could not be performed.

Figure 18

INCOME DISTRIBUTION BY INDUSTRY ACROSS THE U.S.



Source: ASC 5-Year Estimates - Public Use Microdata Sample 2019. Note: The terms for race and ethnicity are from the source and may not reflect the SOA Research Institute's preferred terms for inclusivity.

Table 8

CORRELATION OF INDUSTRY FACTOR AND INCOME RANGE

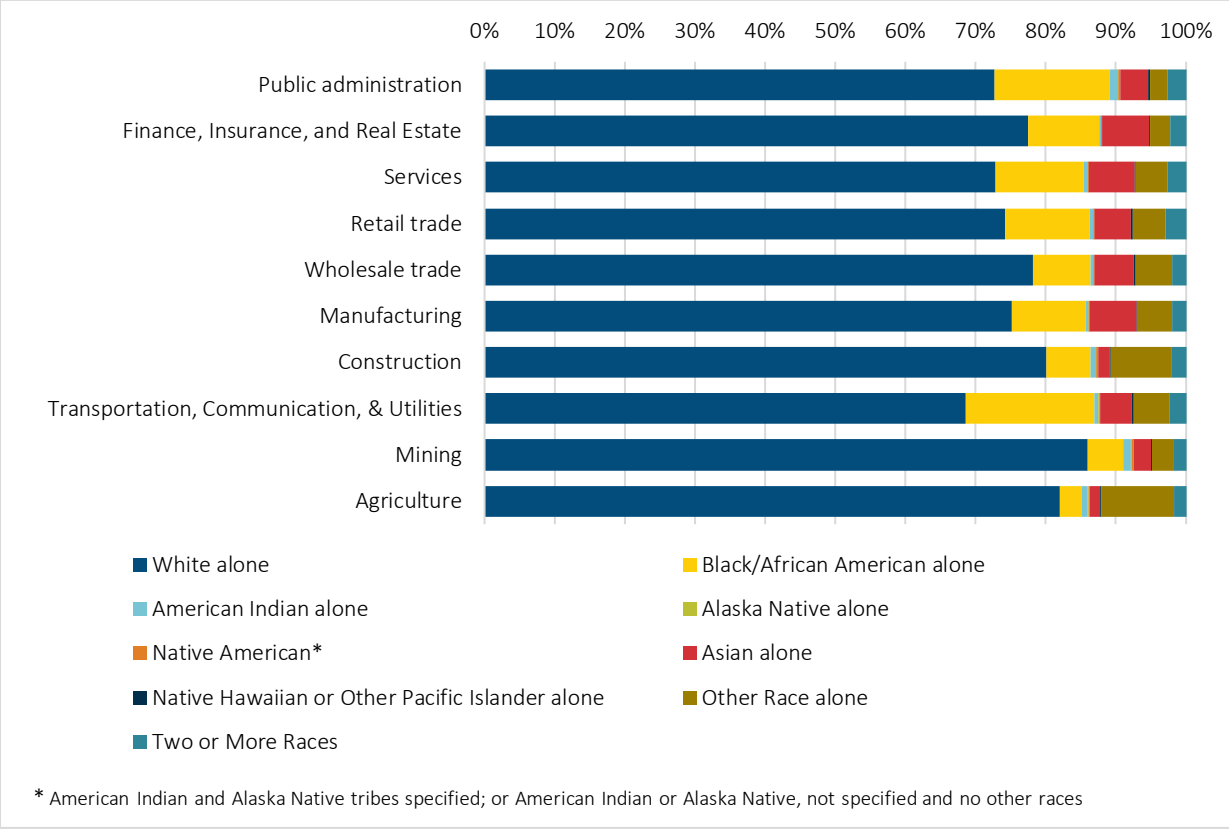
Income Range	Pearson Correlation, r	Spearman Correlation, rho
Between $1 and $20,000	–0.188	–0.265
Between $20,001 and $40,000	–0.301	–0.359
Between $40,001 and $60,000	–0.359	–0.359
Between $60,001 and $80,000	–0.382	–0.359
Between $80,001 and $100,000	–0.380	–0.382
Between $100,001 and $150,000	–0.343	–0.447
Between $150,001 and $250,000	–0.201	–0.191
Above $250,001	–0.046	–0.212

Figure 19 shows broad industry categories along with the proportion people across the U.S. within the industry identifying with a certain racial group as defined by the data source. All the categories shown have a broad mix of racial/ethnic groups represented. This makes it difficult to draw any conclusions about industry factor and race/ethnicity. Table 9 further demonstrates this. Table 9 shows the correlations between race/ethnicity and industry factor; none are statistically significant.

Thus, we cannot say that there are disparate impacts in industry factor load being charged when examining race/ethnicity. This does not mean that disparate impacts are not occurring. Rather, because industry factor loads are not standardized and are not transparent, the factors studied were aggregated to such a level that detailed statistical analyses could not be performed.

Figure 19

RACIAL DISTRIBUTION BY INDUSTRY ACROSS THE U.S.



Source: ASC 5-Year Estimates - Public Use Microdata Sample 2019. Note: The terms for race and ethnicity are from the source and may not reflect the SOA Research Institute's preferred terms for inclusivity.

Table 9

CORRELATION OF INDUSTRY FACTOR AND RACE/ETHINICITY

Race/Ethnicity	Pearson Correlation, r	Spearman Correlation, rho
White alone	0.073	–0.139
Black or African American alone	0.175	0.236
American Indian alone	0.508	0.497
Alaska Native alone	0.476	0.342
American Indian and Alaska Native tribes specified; or American Indian or Alaska Native, not specified and no other races	0.279	0.284
Asian alone	–0.108	–0.067
Native Hawaiian and Other Pacific Islander alone	–0.282	–0.166
Some Other Race alone	–0.532	–0.782
Two or More Races	0.377	0.372

The terms for race and ethnicity are from the source and may not reflect the SOA Research Institute's preferred terms for inclusivity.

Although we were unable to draw any substantive conclusions about the interaction between industry factors and demographic or socioeconomically disadvantaged groups, we did want to note the large variation in industry factors across the rating manuals studied.

Specifically, some rating manuals listed the precise factor used for each 4-digit SIC code while others provided a range of SIC codes, e.g., from 4000 to 4900, with a range of possible values and a median factor, e.g., 0.95 to 1.20 with a median of 0.95, provided. This lack of transparency in some cases compounded with the necessity to roll up industries into broader categories pushed the median used in the analysis toward an average value. This may have masked any possible interactions on a systemic basis.

We also noted that several companies had large swings in industry factor surcharges across related SIC codes while others did not have that same variation. There are several reasons this could be occurring; examining these reasons are outside the scope of this paper. However, if industry factor surcharges were more transparent, future research may be able to determine whether these differences affect certain groups differently.

Limitations

This study used county-level data within the state of Florida in the analysis of tobacco and area rating factors. The statistical analysis was based on aggregate population statistics. Results were determined using race/ethnicity splits mentioned in the observations above. If further combinations of race/ethnicity splits had been studied, results may vary. The same is applicable to the size of the geographic splits.

We do acknowledge that the rate buildup methodology under the Affordable Care Act varies from carrier to carrier. Impacts from other rating factors may be captured, such as differences in networks and general variances from cost of care. There were no adjustments made to the datasets for this impact since it cannot be easily quantified and would vary among carriers. We also did not make any adjustments for the impact of level funding on the small group market.

The impact of external events that could have impacted this study are beyond the scope of this research. This is especially important given the volatility of health care related to the recent impacts of COVID-19. This research has been conducted in accordance with accepted actuarial standards and are fairly stated in accordance with sound actuarial principles. This study is based on actuarial assumptions that we deem to be reasonable and appropriate under the circumstances. Actuarial methods, considerations and analyses used in forming our opinion conform to the appropriate Standards of Practice as promulgated from time to time by the Actuarial Standards Board.

Lastly, we would like to emphasize that the goal of this research is to determine if “ghost factors,” like race, ethnicity or socioeconomic status, are inherently included in allowable rating factors and encourage more nuanced thinking among health actuaries and health insurers when developing their rating factors. With this paper, we advocate neither for nor against the inclusion of tobacco, area or industry factors. Nor do we advocate for or against proposing the inclusion of alternative rating factors.

Potential Future Research

The scope of this project was focused on the study of interactions of tobacco, geographic and industry factors with race/ethnicity and socioeconomic factors such as income. This study was only focused on the states of Florida and Colorado and used only publicly available data. There are many other areas in which this kind of research can be further developed, with a few examples listed below.

EXPANSION OF GEOGRAPHIC AREAS

Geographic and tobacco rating factors were studied from the 2021 ACA Individual and Small Group filings in the state of Florida for this analysis. Florida was chosen as the target for this research because its rating areas correspond to counties, and it is a diverse and heavily populated state. However, this study could be expanded to cover other states or regions, to determine if the relationships observed in Florida are present in other geographic regions. If rating areas are mapped to the counties for which racial/ethnic and socioeconomic data are available, this study could potentially be expanded nationwide.

DETAILED CARRIER-SPECIFIC STUDY

A portion of the data limitations come from combining information across multiple carriers. We had to consider normalization and the different rating methodologies that different carriers use. If we were to instead focus on one carrier that offers across large geographic regions, there would not be as many potential data points. However, the adverse impacts of normalizing across carriers would be removed.

STUDY OF CLAIMS VERSUS RATING FACTORS

This study solely focused on tobacco, geographic and industry rating factors. We did not consider the actual or projected claim costs incurred by enrollees, for which the rating factors are applied. This was outside the scope of our research, but further studies could determine whether the rating factors are accurately capturing variances in medical costs.

These proposed analyses only scratch the surface of other potential research opportunities associated with Diversity, Equity and Inclusion. With further efforts and the increased availability of demographic data, there will be many other potential facets for exploration.

Acknowledgments

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Project Oversight Group and others for their diligent work overseeing questionnaire development, analyzing and discussing respondent answers, and reviewing and editing this report for accuracy and relevance.

Project Oversight Group members:

Kelly Edmiston, PhD

Gabrielle Guzman, FSA, MAAA

Mary Hegemann, FSA, MAAA

Sarah Maune, FSA, EA

Norman Niami, FCAS, MAAA, Affiliate IFoA

Ann Weber, JD

At the Society of Actuaries Research Institute:

Lisa Schilling, FSA, EA, MAAA, FCA, Senior Research Actuary

Appendix A: Industry Code Mapping

Table A.1 provides the mapping of filed industry codes (SIC code) to NAICS 2-Dight Code/NAICS Prefix.

Table A.1

MAP OF SIC CODES TO NAICS 2-DIGHT CODE AND NAICS PREFIX

SIC Codes	NAICS Codes	NAICS Prefixes	Rolled Up Industry	Median Industry Factor
100–999	11	AGR	Agriculture	0.9938
1000–1499	21	EXT	Mining	1.0550
1500–1799	23	CON	Construction	1.0112
2000–3999	31–33	MFG	Manufacturing	0.9977
4000–4999	22, 48–49	UTL, TRN	Transportation, Communication, & Utilities	1.0101
5000–5199	42	WHL	Wholesale Trade	0.9700
5200–5999	44–45	RET	Retail Trade	1.0310
6000–6799	52	FIN	Finance, Insurance and Real Estate	1.0215
7000–8999	51, 54, 61, 62, 71, 81	INF, PRF, EDU, MED, SCA, ENT, SRV	Services	1.0290
9000–9729	92	ADM	Public Administration	1.0346

Appendix B: Statistically Significant Regression Results

B.1 ADULT SMOKING

Table B.1 shows statistically significant (probability less that chosen p-value, 0.05) regression results from the F-test for adult smoking.

Table B.1

SIMPLE LINEAR REGRESSION MODELS OF ADULT SMOKING BY VARIOUS VARIABLES

Variable	Estimate	Std. Error	DF	F-Statistic	Prob.
Median Household Income					
(Intercept)	1.582	0.354	-	-	-
Median Household Income	−0.0003	0.00003	1	61.43	< 0.001
High School Completion					
(Intercept)	1.776	0.385	-	-	-
High School Completion	−0.341	0.053	1	42.022	< 0.001
Some College					
(Intercept)	1.016	0.418	-	-	-
Some College	−0.204	0.027	1	57.791	< 0.001
Food Environment Index					
(Intercept)	3.112	0.348	-	-	-
Food Environment Index	−1.446	0.432	1	11.208	0.001
Asian % of Population					
(Intercept)	0.498	0.540	-	-	-
Asian % of Population	−2.967	0.456	1	42.397	< 0.001

B.2 AREA FACTORS

Tables B.2.1 and B.2.2 show statistically significant (probability less that chosen p-value, 0.05) regression results from the F-test for area factors. Table B.2.1 shows results for the individual market, and Table B.2.2 shows results for the small group market.

Table B.2.1
SIMPLE LINEAR REGRESSION MODELS OF INDIVIDUAL AREA FACTOR BY VARIOUS VARIABLES

Variable	Estimate	Std. Error	DF	F-Statistic	Prob.
Hispanic/Latino					
(Intercept)	0.035	0.013	-	-	-
Hispanic	0.005	0.001	1	17.497	< 0.001
Non-Hispanic/Latino White					
(Intercept)	0.024	0.013	-	-	-
Non-Hispanic/Latino White	−0.004	0.001	1	13.363	< 0.001
Home Ownership Status					
(Intercept)	0.404	0.169	-	-	-
Home Ownership Status	−0.005	0.002	1	5.388	0.023
Food Environment Index					
(Intercept)	0.001	0.014	-	-	-
Food Environment Index	−0.039	0.017	1	5.078	0.028
High School Graduates					
(Intercept)	−0.006	0.014	-	-	-
High School Graduates	−0.004	0.002	1	4.874	0.031
Some College					
(Intercept)	−0.028	0.021	-	-	-
Some College	−0.003	0.001	1	5.707	0.020

Table B.2.2
SIMPLE LINEAR REGRESSION MODELS OF SMALL GROUP AREA FACTOR BY VARIOUS VARIABLES

Variable	Estimate	Std. Error	DF	F-Statistic	Prob.
Hispanic/Latino Proportion					
(Intercept)	0.022	0.007	-	-	-
Hispanic/Latino	0.003	0.001	1	23.553	< 0.001
Non-Hispanic White/Latino Proportion					
(Intercept)	0.012	0.007	-	-	-
Non-Hispanic/Latino White	−0.001	0.001	1	4.082	0.047
Other Proportion					
(Intercept)	0.012	0.007	-	-	-
Other	−0.021	0.009	1	5.951	0.017
High School Completion					
(Intercept)	−0.009	0.010	-	-	-
High School Completion	−0.003	0.001	1	5.983	0.017

References

Hiscock, R., Bauld, L., Amos, A., Fidler, J. A., & Munafo, M. (2012). Socioeconomic status and smoking: a review. ANNALS OF THE NEW YORK ACADEMY OF SCIENCES, 1248(Addiction Reviews), 107–123. https://doi.org/10.1111/j.1749-6632.2011.06202.x

Centers for Disease Control and Prevention. (2021, April 7). Burden of Cigarette Use in the U.S. Centers for Disease Control and Prevention. https://www.cdc.gov/tobacco/campaign/tips/resources/data/cigarette-smoking-in-united-states.html

Jones, J. M. (2021, March 23). Drinking Highest Among Educated, Upper-Income Americans. Gallup.com. https://news.gallup.com/poll/184358/drinking-highest-among-educated-upper-income-americans.aspx

Explore health rankings: Rankings Data & Documentation. County Health Rankings & Roadmaps. (n.d.). Retrieved January 7, 2022, from https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation

Food environment index. County Health Rankings & Roadmaps. (n.d.). Retrieved January 7, 2022, from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/health-behaviors/diet-exercise/food-environment-index

United States Census Bureau. (n.d.). Data Definitions for American Community Survey (ACS) 5-year estimates—Public Use Microdata Sample 2019. MDAT Table Creator. Retrieved January 7, 2022, from https://data.census.gov/mdat/#/search?ds=ACSPUMS5Y2019

Food environment index. County Health Rankings & Roadmaps. (n.d.). Retrieved January 7, 2022, from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/health-behaviors/diet-exercise/food-environment-index

Center for Medicare and Medicaid Services. (n.d.). 2020 marketplace open enrollment period public use files. CMS. Retrieved January 7, 2022, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Marketplace-Products/2020-Marketplace-Open-Enrollment-Period-Public-Use-Files

Friedman, A. S., AB, K., CM, K., T, L., Al., E., K, C., SC, S., KG, V., SD, H., JR, H., BD, S., J, G., AC, L., U, G., BL, C., TJ, L., Manz, K. C., & Pesko, M. F. (2016, July 1). Evidence suggests that the ACA's tobacco surcharges reduced insurance take-up and did not increase smoking cessation: Health Affairs Journal. Health Affairs. Retrieved January 5, 2022, from https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2015.1540

SICCODE.com. (2020, January 21). History of SIC codes. SIC & NAICS Codes, Company Search, Business Lists - SICCODE.com. Retrieved January 7, 2022, from https://siccode.com/page/history-of-sic-codes

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org