Impact of Individual versus Geographic-Area Measures of Socioeconomic Status on Health Associations Observed in the Behavioral Risk Factor Surveillance System

Lena Leszinsky, Sherrie Xie, BS, Avantika Diwadkar, MS, Rebecca E. Greenblatt, BS, Rebecca A. Hubbard, PhD, Blanca E. Himes, PhD
Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, United States

Abstract

Efforts to enhance Electronic Health Record (EHR) data for the study of conditions in which social and economic variables play a prominent role include linking clinical data to sources of external information via patient-specific geocodes. This approach is convenient, but whether geographic-area-level information from secondary sources is adequate as a surrogate of individual-level information is not fully understood. We used Behavioral Risk Factor Surveillance System (BRFSS) epidemiologic data to compare associations of individual income, median aggregate income, and Area Deprivation Index (ADI)—a validated score of U.S. socioeconomic deprivation—with various health outcomes. Median income and ADI assigned according to respondent area of residence were significantly associated with various health outcomes, but with substantially lower effect sizes than those of individual income. Our results show the limited ability of median income and ADI at the level of metropolitan/micropolitan statistical areas versus individual income for use as measures of socioeconomic status.

Introduction

Epidemiologic studies include efforts to understand the prevalence of diseases and identify associated risk factors. Via careful sampling designs and analyses, results from epidemiologic studies may appropriately represent distributions and patterns that would be observed had an underlying population been sampled in its entirety. In the United States, the Centers for Disease Control and Prevention (CDC) leads various observational studies to understand trends in health at a national level. One of these is the Behavioral Risk Factor Surveillance System (BRFSS), a yearly cross-sectional telephone survey of adults conducted across the 50 U.S. states, the District of Columbia, and the U.S. territories of Guam, Puerto Rico and the Virgin Islands[1,2]. BRFSS goals include monitoring of health risk behavior and chronic health conditions among adults, and BRFSS data has been used to understand multiple conditions, including obesity,[3] cigarette smoking,[4] influenza vaccination among adults with asthma,[5] sex-specific determinants of asthma[6] and Chronic Obstructive Pulmonary Disease (COPD) emergency department visits/hospitalizations[7]. In recent years, BRFSS provides publicly available Selected Metropolitan/Micropolitan Area Risk Trends (SMART) data, which contains information from respondents in metropolitan/micropolitan statistical areas (MMSAs) with populations of at least 10,000 people, along with their MMSA of residence.

Electronic health records (EHRs) provide a rich source of information that is increasingly used for secondary purposes, including the conduct of research studies that seek to relate various health outcomes to patient characteristics. While EHR data offers convenient and low-cost access to data for a large number of individuals, it suffers from bias and missingness compared to epidemiologic study data given that the primary purpose of the EHR is to record clinical procedures and patient information according to the needs of health providers and administrators, rather than represent characteristics of underlying populations[8]. By taking into account the complex and biased nature of EHRs, they can be successfully used for population health studies[9,10]. We have previously shown that EHR-derived data can be enhanced by using patient residential addresses and other location data to link clinical data to rich and diverse sources of social, economic and environmental variables that capture information that is not contained in the EHR[11,12]. For example, the Area Deprivation Index (ADI), a validated score of socioeconomic deprivation in U.S. geographic areas, can be assigned to individual patients via their residential geocode to obtain an estimate of their socioeconomic status[13]. The ADI, which consists of a linear combination of 17 American Community Survey (ACS) socioeconomic variables, including housing, education, income and unemployment, has been associated with mortality and hospital readmission risk[14–16]. Composite measures of area-level socioeconomic status like ADI are typically preferred to individual area-level variables (e.g., median income) for health studies as they have been shown to have more robust associations with health outcomes[17,18].

While enhancing EHR-derived data with external sources of information—such as ADI to represent socioeconomic status—is convenient, validation of specific approaches remains an outstanding issue. Because epidemiologic studies collect individual-level data using fixed and validated protocols, they can be used to test whether variables assigned at a group-level according to a geographic area are associated similarly as individual-level variables without need for further data collection[19]. Here, we compare associations of various BRFSS outcomes with 1) BRFSS individual income, and MMSA-level 2) ACS median income and 3) ADI assigned to BRFSS respondents, to determine whether individual income, ACS median income and ADI capture similar socioeconomic information at the level of MMSAs. We also provide a web application for the analysis and visualization of relationships among chronic disease, risk factor and socioeconomic variables of BRFSS respondents.

Methods

Study Population. BRFSS SMART data corresponding to 1,816,427 participants aged at least 18 years who were surveyed via landline and mobile telephone surveys from January 2011 to December 2017 was obtained from the BRFSS website (https://www.cdc.gov/brfss/).

Variable Selection. Variables selected for analyses were available in each of the seven years considered. Respondents' smoking status was recorded based on self-report of being a *current smoker*, *former smoker*, or *never smoker*. A dichotomous smoking variable was also created in which respondents were classified as *smokers* if they were *current smokers* and classified as *non-smokers* otherwise. BMI was split into five levels: *not overweight or obese* (<25.0kg/m^2), *overweight* (25.0 to <30.0 kg/m^2), *grade 1 obesity* (30.0 to <35.0 kg/m^2), *grade 2 obesity* (35.0 to <40.0 kg/m^2), and *grade 3 obesity* (>40.0 kg/m^2). A dichotomous obesity variable was also created in which respondents were classified as *obese* if they had BMI >30.0 kg/m^2 and classified as *not obese* otherwise. Education was re-leveled into three groups: *no high school*, *some high school,* and *college/some higher education*. Individual annual income was re-leveled into three groups: *less than $25,000*, *$25,000-$75,000*, and *more than $75,000*. Race/ethnicity was re-leveled into five groups: *White, Asian/Pacific Islander, Black, Hispanic,* and *Native American*. Respondents were considered to *have health insurance* if they reported having either a pre-paid health insurance plan, a traditional health insurance plan, a government plan, or coverage from the Indian Health Service. Respondents who reported being vaccinated for the seasonal flu in the past year were recoded as having *received flu shot.* Health outcomes considered were: 1) *asthma*, based on affirmative responses to the questions "Have you ever been told by a doctor, nurse, or other health professional that you had asthma?" and "Do you still have asthma?"; 2) *Coronary Heart Disease (CHD)*, based on affirmative response to having had CHD or a myocardial infarction; 3) *Chronic Obstructive Pulmonary Disease (COPD)*, based on self-reported doctor's diagnosis of COPD, emphysema, or chronic bronchitis; 4) *depressive disorder*, based on self-reported doctor's diagnosis of depression, major depression, dysthymia or minor depression; 5) *diabetes*, based on self-reported doctor's diagnosis of diabetes, excluding pregnant women who reported being diabetic only during pregnancy; and 6) *self-rated fair or poor health*, based on a dichotomous re-leveled self-rated health question in which respondents rated their health as *good, better, fair, or poor*. Subjects were excluded for missingness in any of these listed variables, resulting in 1,225,946 respondents with complete BRFSS data.

ACS Median Income and Area Deprivation Index Measures. Data from ACS 5-year estimates from 2013-2017 for median income and other variables needed to compute ADI were obtained from the U.S. Census website with the R *tidycensus* package[20,21]. Median income was re-leveled into four categories according to percentile rank. The bottom quartile of median incomes contained incomes less than $53,711, the second had incomes ranging from $53,711 to $59,046, the third from $59,047 to $65,757, and the fourth greater than $65,757. ADI was computed as the linear combination of 17 MMSA-level socioeconomic variables using a well-established formula[22]. ADI was re-leveled into four categories according to percentile rank. The first ADI quartile ranged from 0 to 10, the second from 11 to 22, the third from 23 to 35, and the fourth from 36 to 99; an ADI score of 99 corresponds to the highest level of "disadvantage." ACS median income and ADI categories were assigned to BRFSS participants according to their MMSAs, resulting in 976,665 complete cases.

Statistical Analysis. Statistical analyses were performed in R[23]. Logistic regression models that considered survey design were created with the R *survey* package using weights calculated by the BRFSS SMART survey weighting methodology. Adjusted odds ratios were obtained for each health outcome using multivariable logistic regression models. To determine whether the magnitude of the association between area-level socioeconomic status measures and health outcomes was similar to their association with individual income, we estimated logistic regression models for each outcome with race, sex, age, and either 1) individual income and education, 2) ACS median income or 3) ADI as independent variables. Results were considered statistically significant at alpha level 0.05. Subsequently, we fit logistic regression models for each outcome with race, sex, age, individual income, education, ACS median income, and ADI as independent variables, to compare the independent contributions of individual and area-level socioeconomic status variables. To measure the influence of individual and area-level socioeconomic status variables on the association of obesity and smoking with BRFSS health outcomes, we compared ORs for obesity and smoking obtained from four multivariable models: model 1 included demographic variables only (race/ethnicity, sex, age); model 2 additionally included ADI; model 3 additionally included ACS median income; model 4 additionally included ADI, ACS median income, individual income and education. To further explore variable relationships between one health outcome (COPD) and socioeconomic status variables, we compared results of multivariable models with the following as predictors 1) demographic variables and ACS median income, 2) demographic variables and ADI 3) demographic variables, ACS median income, ADI, individual income and education.

Application Development. We used the R *Shiny* package[24] to create a web application that is available at [http://*prevalencemaps.org*](http://prevalencemaps.org). Data displayed in maps and MMSA-specific graphs were weighted using survey weights from the BRFSS source data and created with the R *rgeos, rgdal*, and *sf* packages. The app code was saved on a DigitalOcean droplet containing an RStudio Connect server that uses various R packages, including *Leaflet* to display an interactive map[24,25]. Full code is available at https://github.com/HimesGroup/prevalencemaps.

Results

Demographic characteristics and prevalence of the outcomes considered for BRFSS respondents are provided in Table 1. Weighted percentages reflect the expected national distribution. Raw counts demonstrate that older adults and women were overrepresented among respondents with complete data. Analysis of variable distributions across the seven BRFSS survey years considered found nearly identical distributions of all, except for the variable *has health insurance*, which steadily rose from 84.27% in 2011 to 89.88% in 2017, consistent with the signing of the Affordable Care Act in 2010 and its subsequent implementation.

Table 1. Overall characteristics of the 976,665 BRFSS respondents from survey years 2011-2017 who were complete cases and included in analyses.

	N (weighted %)
Sex	
Female	544,488 (49.10)
Race/Ethnicity	
White	785,606 (65.10)
Asian/Pacific Islander	23,775 (4.90)
Black	83,372 (12.80)
Hispanic	79,818 (16.90)
Native American	4094 (0.40)
Education	
No High School	18,171 (3.70)
Some High School	288,315 (33.80)
College/Some Higher Education	670,179 (62.50)
BMI	
Normal	336,216 (35.20)
Overweight	358,223 (36.20)
Grade 1 Obesity	175,253 (17.70)
Grade 2 Obesity	65,361 (6.60)
Grade 3 Obesity	41,612 (4.20)
Individual Income	
Less than $25,000	247,650 (27.30)
$25,000 to $75,000	407,148 (39.70)
More than $75,000	321,867 (33.00)
Smoking	
Never Smoked	545,153 (58.10)
Former Smoker	285,243 (24.80)
Current Smoker	146,269 (17.10)
Age	
18-24	44,836 (10.60)
25-34	101,904 (17.70)
35-44	130,051 (18.00)
45-54	176,841 (19.20)
55-64	220,578 (16.90)
65+	302,455 (17.60)
Has Health Insurance	892,912 (86.40)
Self-Reported Good or Better Health	814,814 (83.80)
Received Flushot	455,080 (38.30)
Asthma	90,156 (8.80)
CHD	55,873 (4.10)
COPD	69,585 (5.60)
Depressive Disorder	184,874 (17.10)
Diabetes	122,119 (10.50)

Individual income is more strongly associated with health outcomes than area-level measures of socioeconomic status. Comparison of logistic regression models that included demographic variables and either 1) individual income and education 2) ACS median income, or 3) ADI found that ORs corresponding to individual income *less than $25,000* versus *more than $75,000*, as well as the first versus fourth quartile of ACS median income, were significant for all outcomes (Figure 1). The fourth quartile of ADI (i.e., most deprived areas) differed significantly relative to the first quartile in all models except for asthma (Figure 1). Comparison of fully adjusted models that included individual income, ACS median income, and ADI found that their effects remained significant for all outcomes except for the association of ACS median income with *asthma* (OR: 0.95, 95% CI 0.89-1.01), *diabetes* (OR: 1.03, 95% CI 0.96-

1.10), *smoking* (OR: 1.00, 95% CI 0.95-1.06), and *fair or poor health* (OR: 1.06, 95% CI 1.00-1.12), and the association of ADI with *COPD* (OR: 1.05, 95% CI 0.96-1.13). Significant associations of ACS median income with outcomes that were lost in the fully adjusted model were for *smoking*, *fair or poor health*, *diabetes* and *asthma*. In the case of ADI, its association with *COPD* was lost in the fully adjusted model, while it became associated with *asthma* in the fully adjusted model. ORs of individual income remained similar in magnitude with all outcomes and were largest for self-reported *fair or poor health*, *COPD*, *smoking* and *depressive disorder*. ACS median income had the greatest effect on *no flu shot*. ADI had the greatest effects on *CHD*, *smoking* and *obesity* albeit with much smaller ORs than the individual income effects.

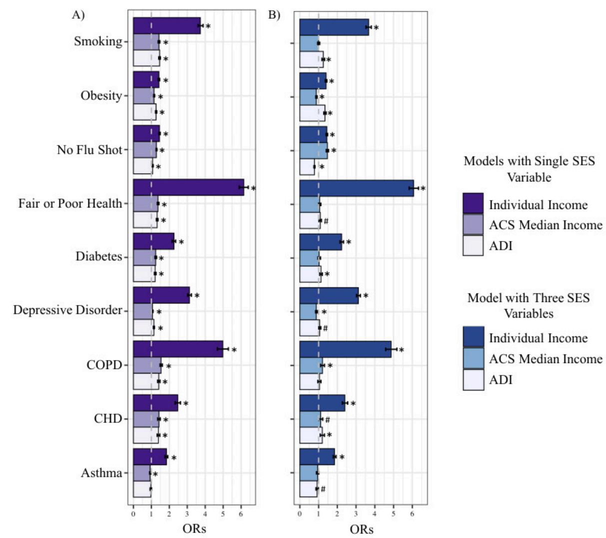

Figure 1. A) Association of individual income, ACS median income *or* ADI with various outcomes shows that individual income has greatest ORs. B) In models that include individual income, ACS median income *and* ADI, individual income maintains greatest ORs, while effects of ACS median income and ADI tend to decrease compared to A). #p<0.05 *p<0.001

Confounding effects of socioeconomic status variables on obesity and smoking. Obesity and smoking are known risk factors for various health outcomes[26] and both have been associated with SES variables as observed in Figure 1. Comparison of ORs for obesity and smoking obtained from four multivariable models (model 1 included demographic variables only; model 2 additionally included ADI; model 3 additionally included ACS median income; model 4 additionally included ADI, ACS median income, individual income and education) found that inclusion of ADI or ACS median income did not significantly change the magnitude of ORs of either smoking or obesity with the health outcomes considered, compared with the baseline model that included demographic variables only (Figure 2). Inclusion of individual income and education (model 4) decreased many of the ORs for obesity and smoking, and although the obesity effect sizes remained similar in magnitude compared to models 1, 2 and 3, several of the smoking ORs decreased substantially (Figure 2). Specifically, the largest shifts in ORs were observed for *COPD*, *fair or poor health* and *depressive disorder*, which is consistent with Figure 1, where income had the greatest effect on these three outcomes, as well as on *smoking*.

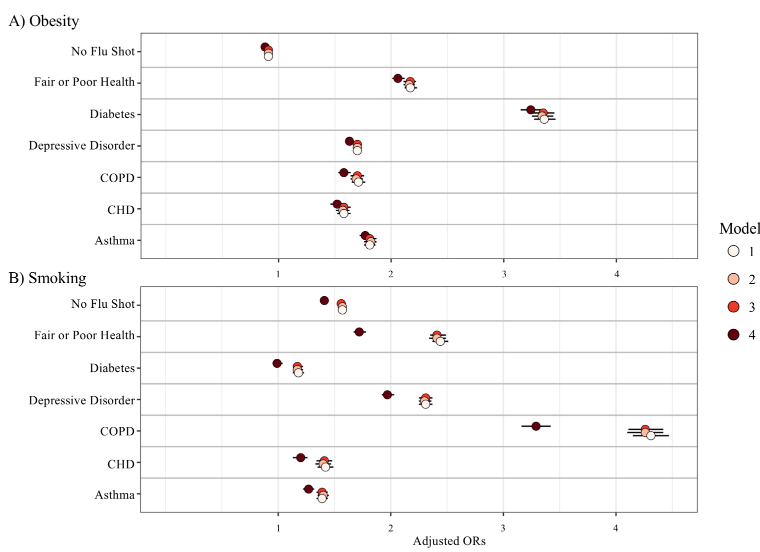

Figure 1. Confounding of SES measures on the relationships between A) obesity and B) smoking and listed health outcomes. Model 1 independent variables are race, sex, age and obesity or smoking; Model 2 also includes ADI; Model 3 also includes ACS median income; Model 4 also includes ADI, ACS median income, individual income, and education. Smoking is a dichotomous variable based on being a *current smoker*. Obesity is a dichotomous variable based on having BMI > 30 kg/m^2.

Influence of socioeconomic status variables on association of COPD with demographic and health risk factors. Among the relationships in the multivariable models considered in Figure 2, the association between smoking and COPD was most affected by inclusion of socioeconomic status variables. Logistic regression models for COPD found that it was more likely to occur in women, people who self-identified as White, current or former smokers, obese individuals and those of older age (Table 2), all of which are consistent with published COPD trends[7,27]. In the model that included ADI as the sole socioeconomic status variable, those living in more disadvantaged MMSA's (quartiles 3 and 4) were more likely to have COPD compared to those living in the least disadvantaged MMSA's (quartile 1). In the model that included ACS median income as the sole socioeconomic status variable, those living in MMSAs with lower median incomes were significantly more likely to have COPD compared to those in MMSAs with higher median incomes. The fully adjusted model that included ADI, ACS median income, individual income and education, found that ADI was no longer significant, ACS median income had reduced but still significant effects, and individual income had a strong effect on COPD. An increased association of COPD with Native American race/ethnicity versus White that was significant in the model with ADI or ACS median income alone was not significant in the model with all socioeconomic status variables, while there was a decreased risk of COPD with Black race/ethnicity in the model with all socioeconomic status variables that was not present when including ADI or ACS median income alone.

Web application. Users can view yearly 2011-2017 BRFSS results of associations among chronic diseases, risk factor, and socioeconomic variables, as well as the overall seven-year span of the BRFSS respondents included in analyses. A map feature enables users to view the geographic distribution of these variables across the U.S. at the resolution of MMSAs (Figure 3) and select an MMSA to view its specific data. Barplots are provided to visualize general trends among user-selected variables.

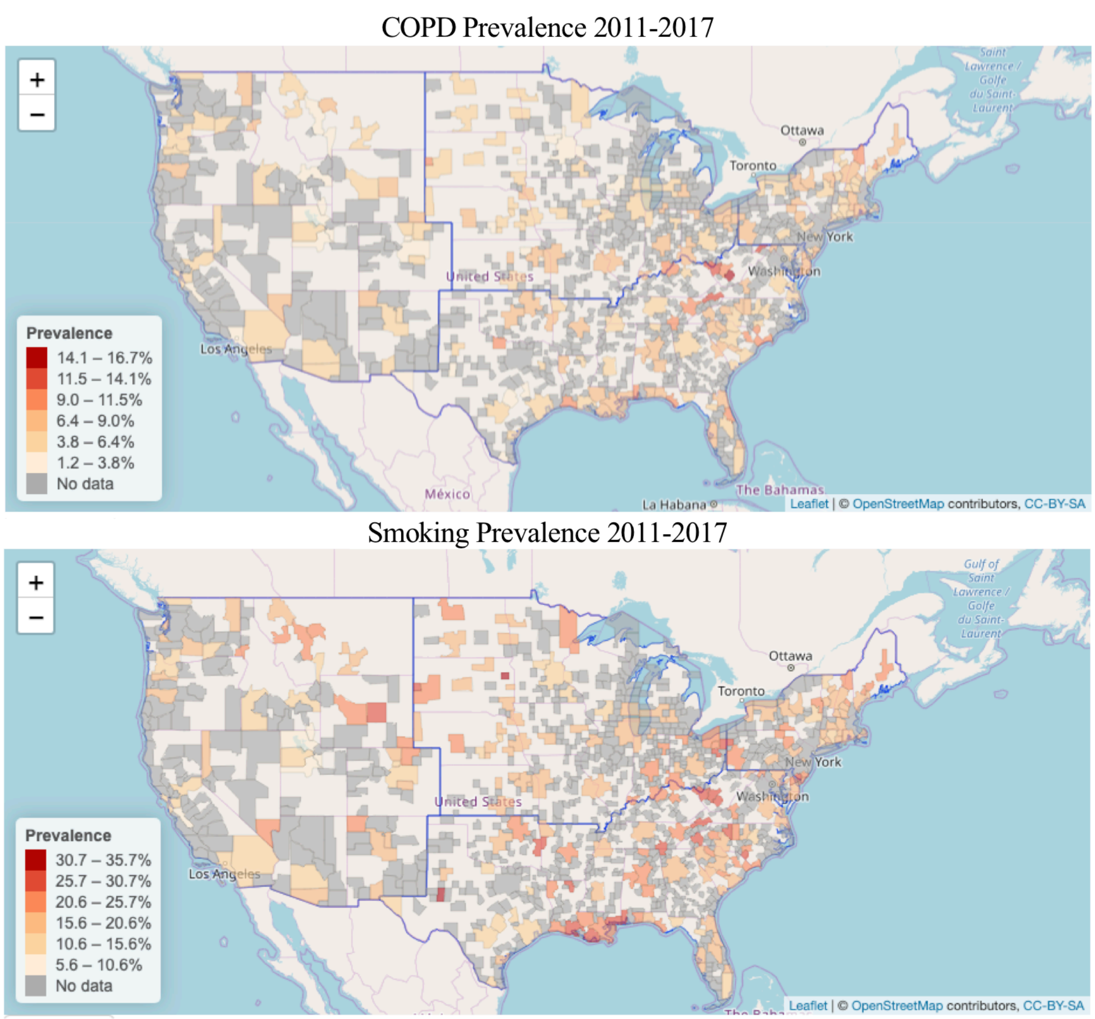

Figure 2. Maps of COPD and smoking prevalence among 2011-2017 BRFSS respondents in MMSAs with available data (*http://prevalencemaps.org*).

Table 2. Factors Associated with COPD in BRFSS Multivariable Analysis. Adjusted odds ratios (ORs) were derived from adjusted survey logistic regression models with COPD as the outcome. *p<0.05; **p<0.001

	ADI Model	ACS Median Income Model	Full SES Model
	Adjusted ORs (95% CI)	Adjusted ORs (95% CI)	Adjusted ORs (95% CI)
Sex			
Male	Reference	Reference	Reference
Female	1.48 (1.43, 1.54)**	1.48 (1.43, 1.54)**	1.37 (1.32, 1.42)**
Race/Ethnicity			
White	Reference	Reference	Reference
Asian/Pacific Islander	0.77 (0.65, 0.91)*	0.76 (0.64, 0.90)*	0.72 (0.61, 0.86)**
Black	1.04 (0.98, 1.10)	1.03 (0.97, 1.09)	0.80 (0.76, 0.86)**
Hispanic	0.71 (0.66, 0.77)**	0.69 (0.64, 0.75)**	0.48 (0.44, 0.52)**
Native American	1.38 (1.13, 1.69)*	1.36 (1.11, 1.67)*	1.08 (0.88, 1.33)
Smoking			
No smoking	Reference	Reference	Reference
Former smoker	2.99 (2.86, 3.12)**	2.98 (2.85, 3.11)**	2.89 (2.76, 3.02)**
Current smoker	7.03 (6.71, 7.36)**	7.03 (6.71, 7.37)**	5.46 (5.20, 5.73)**
BMI			
Not overweight or obese	Reference	Reference	Reference
Overweight	0.90 (0.86, 0.94)**	0.91 (0.87, 0.95)**	0.91 (0.87, 0.96)**
Grade 1 Obesity	1.23 (1.17, 1.29)**	1.24 (1.18, 1.30)**	1.19 (1.14, 1.26)**
Grade 2 Obesity	1.82 (1.71, 1.93)**	1.83 (1.72, 1.94)**	1.68 (1.58, 1.79)**
Grade 3 Obesity	2.85 (2.65, 3.06)**	2.87 (2.67, 3.09)**	2.47 (2.29, 2.66)**
Age			
18-24	Reference	Reference	Reference
25-34	0.97 (0.85, 1.12)	0.98 (0.85, 1.12)	1.10 (0.96, 1.26)
35-44	1.27 (1.12, 1.45)**	1.28 (1.12, 1.45)**	1.58 (1.39, 1.79)**
45-54	2.38 (2.11, 2.68)**	2.38 (2.11, 2.68)**	2.87 (2.55, 3.24)**
55-64	3.91 (3.48, 4.39)**	3.9 (3.47, 4.38)**	4.51 (4.01, 5.07)**
65+	6.31 (5.62, 7.08)**	6.26 (5.58, 7.03)**	6.21 (5.53, 6.97)**
ADI			
Q1 (0-10)	Reference	-	Reference
Q2 (11-22)	1.06 (1.01, 1.12)*	-	0.95 (0.89, 1.01)
Q3 (23-35)	1.24 (1.18, 1.31)**	-	1.04 (0.96, 1.12)
Q4 (36-99)	1.27 (1.21, 1.33)**	-	0.99 (0.91, 1.08)
ACS Median Income			
Q1 (<$53,711)	-	1.41 (1.35, 1.47)**	1.22 (1.13, 1.33)**
Q2 ($53,711 to $59,046)	-	1.34 (1.28, 1.40)**	1.21 (1.13, 1.31)**
Q3 ($59,047 to $65,757)	-	1.21 (1.15, 1.27)**	1.16 (1.09, 1.24)**
Q4 (>$65,757)	-	Reference	Reference
Income			
< $25,000	-	-	3.47 (3.27, 3.70)**
$25,000 to $75,000	-	-	1.81 (1.71, 1.92)**
> $75,000	-	-	Reference
Education			
Less than high school	-	-	Reference
Some High School	-	-	0.86 (0.78, 0.95)*
College or Some Higher Education	-	-	0.73 (0.66, 0.80)**

Discussion

We used publicly available data on 976,665 complete cases from the 2011-2017 BRFSS, along with ACS median income and ADI computed with data from ACS 5-year estimates for 2013-2017, which we assigned to BRFSS participants according to their MMSAs, to compare associations of individual-level versus geographic-area-level SES variables with various health outcomes. The goal of this comparison was to shed light on the appropriateness of using geographic-area-level information from secondary sources as a surrogate of individual-level information, an appealing approach that is being increasingly used to enhance EHR data with social, economic and environmental variables from external data sources via patient-specific geocodes[11,12,28]. We focused on income, an important individual-level indicator of socioeconomic status that is associated with various health outcomes but not typically recorded in the EHR, and two related area-level measures of socioeconomic status, ACS median income and ADI, that can be integrated with EHR data via geographic linkage. Overall, we found that ACS median income and ADI determined at the MMSA level were poor surrogates for individual income and did not sufficiently account for its confounding effects on the association between risk factors (i.e., smoking and obesity) and a number of health outcomes.

While ACS median income and ADI variables were significantly associated with several of the health outcomes considered, the effect sizes were considerably smaller compared to individual income, which was significantly associated with all health outcomes/factors considered (i.e., smoking, obesity, receiving flu shot, self-rated fair or poor health, diabetes, depressive disorder, COPD, CHD, asthma). Our analysis examining the confounding effects of individual- vs. MMSA-level socioeconomic status variables on risk factor-health outcome associations illustrates the potential consequences of using an area-level measure to represent individual-level variables. Individual income was found to be an important confounder of the association between smoking and many health outcomes, most notably with COPD, where adjusting for income strongly attenuated the effect of smoking. However, the inclusion of ACS median income or ADI did little to affect the association of smoking with most outcomes, suggesting that ACS median income and ADI determined at the MMSA level are inadequate measures to account for the cofounding effects of socioeconomic status in EHR or other health studies. Given that previous studies have found that ADI was significantly associated with health outcomes at the level of census blocks, future studies using epidemiologic datasets that contain geographic information for respondents in areas smaller than MMSAs are warranted. For contrast, census blocks are substantially smaller on average than MMSAs and there are 11,078,297 census blocks compared to 942 MMSAs in the United States[29].

We provided further details on the regression results for COPD, given that it had the greatest change in relationships with risk factors when individual income was included in a model. The associations we observed between COPD and smoking, demographic factors and individual income have been observed previously and serve as a check of consistency for our BRFSS-based results[30–35]. Interestingly, when not accounting for individual income, associations with race/ethnicity changed: Native American respondents had greater risk of COPD than White respondents in unadjusted models only, suggesting that individual SES contributes to COPD for those respondents. Conversely, Black respondents had similar risk of COPD as white respondents in unadjusted models, but when individual income was included, Black respondents had decreased risk. Further studies are needed to understand these relationships between income, race/ethnicity and COPD.

To facilitate exploration of various BRFSS health outcomes and demographic factors, we designed a Shiny web application ([http://*prevalencemaps.org*](http://prevalencemaps.org)) that displays their geographic distribution across the United States and shows bivariate and multivariable relationships. This app differs from existing online resources that utilize BRFSS data. For example, Chronic Disease Indicators is a CDC web application that provides an interactive map of multiple diseases and risk factors stratified by specific indicators, years and data types covering the 50 U.S. states, the District of Columbia, and the U.S. territories of Guam, Puerto Rico and the Virgin Islands[36]. Although it includes a large number of variables, prevalence is displayed at a state, rather than MMSA, level. Another interactive map application, the 500 Cities Project, displays BRFSS measures at the census tract level for the 500 largest U.S. cities but does not contain information on other geographic locations[37,38]. Other tools that facilitate analysis of BRFSS data, such as VitalWeb, are not freely available[39]. Thus, beyond serving as a source of data and results to ensure reproducibility of the work presented here, our web application provides a user-friendly resource for the exploration of relationships among BRFSS variables, including geospatial trends at the level of MMSAs.

In addition to being constrained to use ACS median income and ADI at the level of MMSAs given that MMSAs are the smallest geographic location available for BRFSS respondents, limitations of our study include potential error in self-reported measures of BRFSS, such as obesity and income[40]. While these errors cannot be discounted, several relationships we observed are consistent with those in published studies, and thus, do not affect the major question addressed regarding individual income versus geographic-area-level socioeconomic status variables. Future studies using epidemiologic data are needed to explore the utility of neighborhood- versus individual-level measures of income and other variables that are not recorded in the EHR to improve the scope of secondary studies that address individual and population health outcomes.

Conclusion

To better understand whether geographic-area information from secondary sources is helpful as a surrogate of individual-level information, we used BRFSS data to compare associations of individual income and two geographic-area-level socioeconomic status variables with various health outcomes. Our results show that use of ACS median income or ADIs assigned according to MMSA of residence are significantly associated with various BRFSS outcomes, but effect sizes are much smaller than those of individual income. Furthermore, adjusting for individual income substantially decreased known confounding between smoking and health outcomes such as COPD, while adjusting for ACS median income or ADI had little effect, suggesting that these two variables measured at the MMSA level do a poor job accounting for the confounding effects of socioeconomic status. Relationships among BRFSS health outcomes and risk factors can be further explored and visualized using a web application developed and made available at [http://*prevalencemaps.org*](http://prevalencemaps.org).

References

1. CDC; BRFSS https://www.cdc.gov/brfss/index.html [Accessed March 20, 2020].

2. Mokdad AH. The Behavioral Risk Factors Surveillance System: past, present, and future. Annu Rev Public Health. 2009;30:43–54.

3. Ezzati M, Martin H, Skjold S, Vander Hoorn S, Murray CJL. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. J R Soc Med. 2006 May;99(5):250–7.

4. Arday DR, Tomar SL, Nelson DE, Merritt RK, Schooley MW, Mowery P. State smoking prevalence estimates: a comparison of the Behavioral Risk Factor Surveillance System and current population surveys. Am J Public Health. 1997 Oct;87(10):1665–9.

5. Lu P, Euler GL, Callahan DB. Influenza Vaccination Among Adults with Asthma: Findings from the 2007 BRFSS Survey. Am J Prev Med. 2009 Aug 1;37(2):109–15.

6. Greenblatt R, Mansour O, Zhao E, Ross M, Himes BE. Gender-specific determinants of asthma among U.S. adults. Asthma Res Pract. 2017 Jan 24;3(1):2.

7. Kumbhare SD, Beiko T, Wilcox SR, Strange C. Characteristics of COPD Patients Using United States Emergency Care or Hospitalization. Chronic Obstr Pulm Dis Miami Fla. 2016 Mar;3(2):539–48.

8. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc JAMIA. 2013 Jan 1;20(1):117–21.

9. Kruse CS, Stein A, Thomas H, Kaur H. The use of Electronic Health Records to Support Population Health: A Systematic Review of the Literature. J Med Syst. 2018 Sep 29;42(11):214.

10. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ. 2018 Apr 30;361:k1479.

11. Xie S, Himes BE. Approaches to Link Geospatially Varying Social, Economic, and Environmental Factors with Electronic Health Record Data to Better Understand Asthma Exacerbations. AMIA Annu Symp Proc AMIA Symp. 2018;2018:1561–70.

12. Xie S, Greenblatt R, Levy MZ, Himes BE. Enhancing Electronic Health Record Data with Geospatial Information. AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci. 2017;2017:123–32.

13. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. N Engl J Med. 2018 Jun 28;378(26):2456–8.

14. Kind AJ, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, et al. Neighborhood Socioeconomic Disadvantage and 30 Day Rehospitalizations: An Analysis of Medicare Data. Ann Intern Med. 2014 Dec 2;161(11):765–74.

15. Jencks SF, Schuster A, Dougherty GB, Gerovich S, Brock JE, Kind AJH. Safety-Net Hospitals, Neighborhood Disadvantage, and Readmissions Under Maryland's All-Payer Program: An Observational Study. Ann Intern Med. 2019 Jul 16;171(2):91–8.

16. Singh GK, Azuine RE, Siahpush M, Kogan MD. All-cause and cause-specific mortality among US youth: socioeconomic and rural-urban disparities and international patterns. J Urban Health Bull N Y Acad Med. 2013 Jun;90(3):388–405.

17. Lòpez-De Fede A, Stewart JE, Hardin JW, Mayfield-Smith K. Comparison of small-area deprivation measures as predictors of chronic disease burden in a low-income population. Int J Equity Health. 2016 Jun 10;15:89.

18. Krieger N, Chen JT, Waterman PD, Soobader M-J, 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. Am J Epidemiol. 2002 Sep 1;156(5):471–82.

19. Xie S, Hubbard RA, Himes BE. Neighborhood-level measures of socioeconomic status are more correlated with individual-level measures in urban areas compared with less urban areas. Ann Epidemiol. http://www.sciencedirect.com/science/article/pii/S1047279719306088. [Accessed March 20, 2020]

20. United States Census Bureau; American Community Survey (ACS). https://www.census.gov/programs-surveys/acs. [Accessed March 20, 2020].

21. Walker, K. Load US Census Boundary and Attribute Data as tidyverse and sf-Ready Data Frames. https://walkerke.github.io/tidycensus/. [Accessed March 20, 2020].

22. Singh GK. Area Deprivation and Widening Inequalities in US Mortality, 1969–1998. Am J Public Health. 2003 Jul;93(7):1137–43.

23. R: The R Project for Statistical Computing. https://www.r-project.org/. [Accessed March 20, 2020].

24. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J, RStudio, et al. shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny. [Accessed March 20, 2020].

25. Cheng J, Karambelkar B, Xie Y, Wickham H, Russell K, Johnson K, et al. leaflet: Create Interactive Web Maps with the JavaScript "Leaflet" Library. https://CRAN.R-project.org/package=leaflet. [Accessed March 20, 2020].

26. Sturm R. The effects of obesity, smoking, and drinking on medical problems and costs. Health Aff Proj Hope. 2002 Apr;21(2):245–53.

27. Barnes PJ. Sex Differences in Chronic Obstructive Pulmonary Disease Mechanisms. Am J Respir Crit Care Med. 2016 Apr 15;193(8):813–4.

28. Bhavsar NA, Gao A, Phelan M, Pagidipati NJ, Goldstein BA. Value of Neighborhood Socioeconomic Status in Predicting Risk of Outcomes in Studies That Use Electronic Health Record Data. JAMA Netw Open. 2018 07;1(5):e182716.

29. United States Census Bureau. Tallies. https://www.census.gov/geographies/reference-files/time-series/geo/tallies.html. [Accessed March 20, 2020]

30. Prescott E, Lange P, Vestbo J. Socioeconomic status, lung function and admission to hospital for COPD: results from the Copenhagen City Heart Study. Eur Respir J. 1999 May;13(5):1109–14.

31. Eisner MD, Blanc PD, Omachi TA, Yelin EH, Sidney S, Katz PP, et al. Socioeconomic status, race and COPD health outcomes. J Epidemiol Community Health. 2011 Jan;65(1):26–34.

32. Prescott E, Vestbo J. Socioeconomic status and chronic obstructive pulmonary disease. Thorax. 1999 Aug 1;54(8):737.

33. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2016;138(1):16–27.

34. Godtfredsen NS, Vestbo J, Osler M, Prescott E. Risk of hospital admission for COPD following smoking cessation and reduction: a Danish population study. Thorax. 2002 Nov 1;57(11):967.

35. Di Stefano A, Caramori G, Oates T, Capelli A, Lusuardi M, Gnemmi I, et al. Increased expression of nuclear factor-kappaB in bronchial biopsies from smokers and patients with COPD. Eur Respir J. 2002 Sep;20(3):556–63.

36. CDC. Chronic Disease Indicators (CDI) | DPH. https://www.cdc.gov/cdi/index.html. [Accessed March 20, 2020].

37. CDC, Robert Wood Johnson Foundation, CDC Foundation. The 500 Cities project. https://www.cdc.gov/500cities/. [Accessed March 20, 2020]

38. Wang Y, Holt JB, Zhang X, Lu H, Shah SN, Dooley DP, et al. Comparison of Methods for Estimating Prevalence of Chronic Diseases and Health Behaviors for Small Geographic Areas: Boston Validation Study, 2013. Prev Chronic Dis. 2017 Oct 19;14:170281.

39. Expert Health Data Programming, Inc. Vitalnet BRFSS. https://www.ehdp.com/brfss/. [Accessed March 20, 2020]

40. Le A, Judd SE, Allison DB, Oza-Frank R, Affuso O, Safford MM, et al. The Geographic Distribution of Obesity in the US and the Potential Regional Differences in Misreporting of Obesity. Obes Silver Spring Md. 2014 Jan;22(1):300–6.