Forum for Social Economics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rfse20

Socioeconomic Status & Health Disparities: Utilizing a Composite Index across Health Datasets

Iris Buder, Jacob Jennings, Dae Hyun Kim & Norman Waitzman

To cite this article: Iris Buder, Jacob Jennings, Dae Hyun Kim & Norman Waitzman (2023): Socioeconomic Status & Health Disparities: Utilizing a Composite Index across Health Datasets, Forum for Social Economics, DOI: 10.1080/07360932.2023.2241650

To link to this article: https://doi.org/10.1080/07360932.2023.2241650

Published online: 04 Aug 2023.

https://doi.org/10.1080/07360932.2023.2241650

Routledge
Taylor & Francis Group

Check for updates

Socioeconomic Status & Health Disparities: Utilizing a Composite Index across Health Datasets

Iris Buder[a], Jacob Jennings[b], Dae Hyun Kim[c] and Norman Waitzman[d]

[a]Department of Economics & Finance, Idaho State University, Pocatello, ID, USA; [b]Department of Economics, California State University, Chico, CA, USA; [c]Department of Health Management and Policy, Georgetown University, Washington, DC, USA; [d]Department of Economics, University of Utah, Salt Lake City, UT, USA

ABSTRACT
Research indicates that the relationship between socioeconomic status (SES) and health outcomes are robust, though no 'gold standard' as to what best captures SES exists. Many studies use individual proxies to control for SES, but there are a number of limitations to doing so. Additionally, little research has been conducted as to how to develop a US-based composite SES index that can be utilized across various datasets. The aim of the study is to generate a weighted SES index, following the National Assessment of Educational Progress (NAEP)'s call for an understanding of SES groupings and the guidance of the Bureau of Justice's SES index. After generating a composite SES index, we evaluate whether or not each unique index delivers a similar classification of SES across four national datasets and empirically compare the resulting grouping across datasets utilizing multiple regression analyses. Using weighted descriptive statistics and weighted logistic regression analyses, we reveal the key variations in the distribution of determinants of the socioeconomic index. While previous studies conflated race and ethnicity with SES to resulting health indicators, this study captures both, highlighting the importance of each covariate and a composite index. Our findings suggest similar outcomes with regards to the assigned SES classifications (i.e. low, middle, and high SES) across the four national datasets.

KEYWORDS : Socioeconomic status, health disparities, composite index, comparative study

1. Introduction

The concept of socioeconomic status (SES) serves as a proxy for the ability to create (or consume) goods that are of value to society by a person's, families', or households' place within society and, thus, has become a means of identifying social and economic inequalities (Hauser & Warren, 1997; Oakes & Rossi, 2003). Research consistently indicates the importance of SES when analyzing a range of economic and health outcomes and disparities. Differences in SES (e.g. differences in education or income) are shown to have significant impacts on a multitude of health outcomes and health disparities, both directly (e.g. through differing work conditions, access to nutrition, ability to exercise) and indirectly *via* mental health outcomes, stress levels, and immune responses (Webb et al., 2017; Frank & Mustard, 1994).

While the concept of socioeconomic status was coined in 1883 by the American sociologist Lester Ward, little attention is given to the literature regarding the measurement and collective properties of SES groupings in the United States (Oakes & Rossi, 2003). SES is typically measured by income (or wealth), educational attainment, social class/standing, occupations or occupational prestige, or even through measures of social participation (Hauser & Warren, 1997). Nonetheless, lacking in the US is an SES index allowing for the conceptualization of low, middle, and high SES groupings based on a composite collective, though such composite measures of SES are readily found in Europe and Canada (Vincent & Sutherland, 2013). Thus, while a significant amount of research has focused on the impact of socioeconomic status (as measured by a single factor such as education or income), gender, race/ethnicity on health disparities as independent cofactors and through the intersectionality, gaps in the literature remain, given the lack of utilization and consensus on a US-based composite measure for SES (Adler & Rehkopf, 2008).

While SES provides an indicator of an individual's social standing and is typically measured as a combination of education, income, and occupation, to date, there is no 'gold standard' regarding the 'composite' definition, or index, of SES (Oakes & Rossi, 2003; Berzofsky et al., 2015). Rather, the measurement of SES tends to encompass various individual or loosely combined indicators (Oakes & Rossi, 2003; Braveman et al., 2005; Freeman et al., 2016; Berzofsky et al., 2015). In addition, the terms utilized to describe SES (i.e. education, occupation, wealth, income) are often used interchangeably, creating more confusion in the interpretation of findings (Freeman et al., 2016). As such, in the United States, SES is often, implicitly or explicitly, equated with income or education rather than the interaction of each of the component parts (i.e. education, income and occupation). Another inconsistency in SES measurement is the choice of components to include in a measurement of socioeconomic position, as no consensus regarding either a 'nominal definition of SES' or 'widely accepted SES measurement tool exists' (Oakes & Rossi, 2003, pg. 770). Inconsistent and loose categorization compound the issues with regards to comparison of various health, economics, or social disparities across studies or countries.

To our knowledge, no comparable composite indicator for SES groups exist in the United States and minimal progress has been made to create such an index (Oakes & Rossi, 2003; Berzofsky et al., 2015), even though the development of such a socioeconomic status index could be utilized for a multitude of research and policy questions with accessible national and private data (Krieger et al., 1997). The use, and adoption, of a composite SES index, rather than utilizing the disaggregated single components, can help capture the intricacies of SES and its multifactorial nature in future analyses of health outcomes (Oakes & Rossi, 2003). Building on the National Assessment of Educational Progress (NAEP) and Cowan et al.'s call for a better understanding of grouped SES (2012), the aim of this paper is to compare and contrast an adopted SES index from the Bureau of Justice Statistics (BJS), from which classifications of low, middle, and high SES categorizations can be defined, across four national survey datasets utilizing weighted multivariate logistic regression analyses.

2. Background

Researchers across the social sciences have shown that the geographically and socioeconomically disadvantaged have greater probabilities of reduced wealth, health, and social opportunity, among other adverse outcomes (Adler & Rehkopf, 2008; Deaton, 2003; Marmot & Allen, 2014). Lower socioeconomic status (as assessed by either income, education, or education rather than a composite measure) has consistently been linked to a number of adverse health outcomes (e.g. low birthweight, cardiovascular disease, diabetes, certain cancers), poorer health access, and higher mortality rates (Adler & Newman, 2002; Marmot et al., 2008; Fiscella & Williams, 2004). Concurrently, many social sciences studies, including studies in sociology, economics, and public health, recognize the important role that institutional arrangements have on individuals and that institutional arrangements created within society directly influences an individual's health (Olafsdottir, 2007). As such, when analyzing the social determinants of health, the role of socioeconomic status plays a central role (Adler & Rehkopf, 2008).

However, while there has been a growing literature on how SES is associated with health outcomes, how to define, or measure, SES continues as a complex issue in both theory and practical use (Oakes & Rossi, 2003). Hence, how socioeconomic status is measured remains not well-defined and how individuals are classified (i.e. low, middle, and high SES) remains unclear. One aspect of the issue is in the intricacy and multi-layered nature of one's SES, particularly when accounting for societal origins or pre-determined stratification, and there is no prevailing or agreed upon way of weighting the three primary components of SES (i.e. education, income, and occupation) to form SES groupings (i.e. low, middle, and high SES) (Williams et al., 2016; Braveman et al., 2005; Berzofsky et al., 2015).[1] Additionally, while the stratification literature recognizes and takes into the account differentials by social class, studies regarding health disparities in the Unites States typically do not focus on class-based differences, partly due to the complexities and limitations of utilizing SES, but rather focus on disparities by race/ethnicity, gender, income, education, or occupation (Kawachi, 2013). Indices of SES are prominently used in connection to health in Europe and nations such as Canada (Vincent & Sutherland, 2013). However, less has been done in connecting a combined variable to health in the US. Well-known social scales (i.e. occupational prestige scores) such as the Duncan Index, Nam Powers Boyd, and Nakao & Treas' Occupational Score, amongst others, are clearly useful, but have not been translated into SES classifications such as low, middle and high SES from the scores. Additionally, these (occupational prestige) scores are heavily weighted towards occupation, giving less credence to the intergenerational persistence of other components of SES (Hout, 2018).

Another compounding issue is that studies regarding SES and the associated health gaps are frequently obfuscated by studies emphasizing differences in culture, race/ethnicity, and health outcomes (Williams et al., 2016). Stark disparities in mortality, morbidity, opportunity, and other health indicators exist when separated by race/ethnicity. In the United States, in particular, there is a 'conceptual assumption'

[1]For instance, using census data, many geographic or area-based measures have begun to emerge adding to our understanding of regional disparities in access and mobility.

that when analyzing health disparities, race/ethnicity is the 'most meaningful category' whereby race/ethnicity may become a proxy for socioeconomic position partly due to the lack of nationally-available data on socioeconomic position (Nuru-Jeter et al., 2018, p. 171). However, such analyses in racial/ethnic differences in health outcomes should be analyzed in conjunction with structural determinants and historically persistent inequities (Bell et al., 2020; Chetty et al., 2020; Daw, 2017). These contextual factors, prevalent in stratification economic theory, which determine outcomes and gaps are, however, often omitted (Darity et al., 2014). It should also be noted that ordinal categorizations of race/ethnicity (i.e. White non-Hispanic, Black non-Hispanic, etc.) are imperfect in capturing the multi-dimensional nature of race/ethnicity and its' separation into social categories for statistical analysis (Williams et al., 2016; Nuru-Jeter et al., 2018).

In particular, using race/ethnicity as a 'category' (e.g. White non-Hispanic and Black non-Hispanic) in studying health disparities is further complicated by its attempts to capture biological, genetic, and long-term environmental and ancestral exposures (Williams et al., 2010). When accounting for race/ethnicity, challenges remain in the statistical significance a researcher assigns to any of these factors and their influence upon health and related socioeconomic outcomes and gaps. While many health studies emphasize 'biological' and 'genetic' explanations of disparities by race/ethnicity (Williams et al., 2016), we argue that that omits important economic and social factors. Furthermore, a number of issues arise when solely focusing on race and ethnicity. First, research shows that variations in genetics do not match, or divide into, traditional racial categorization and individual genetic variation is greater than that of population subgroups (Williams et al., 2016). Second, racial and ethnic classifications are not only subject to change, but also reflect the systemic exploitation, oppression, and social inequality (Williams et al., 2016). Third, studies have shown that racial categories, and the resulting classifications, are from socially and historically driven sources. As such, researchers should be cognizant that the classification of racial/ethnic categories have 'historically captured not cultural practices and beliefs but societally imposed stigmatization and marginalization that have been consequential for all aspects of life' (Williams et al., 2016, p. 2).

Though data for the 'big three' indicators of socioeconomic status (i.e. education, income, and occupation) are readily available in most national survey data, there are challenges to their inclusion in the health services research either as an individual or grouped variables. There are several reasons as to why this is the case. First, educational attainment is more easily obtained and thus is often used as a proxy for income, and thus SES. Nonetheless, individuals with similar education may have widely varying occupations or wealth. Second, high correlation between income and education, resulting in concerns about multicollinearity, lead many studies to include only one. However, using only one variable may be insufficient as, for instance, income or wealth varies across social groups (i.e. racial/ethnic, gender, and age groups), even when individuals have similar education levels (Williams et al., 2016; Oakes & Rossi, 2003; Bell et al., 2020). Lastly, there is a lack of standardization of occupational categories within the United States. This lack of consensus leads many researchers not to control for occupational prestige when analyzing health disparities (Braveman et al., 2005). Essentially, a one variable SES proxy does not fully account for differing health disparity gradients or what researchers have termed a significant 'non-equivalence' by race and ethnicity (Williams et al., 2016; Nuru-Jeter et al., 2014).

As such, a composite measure is advantageous for it provides 'a single summary useful for reporting, greater reliability, and representing the full range of SES factors' (Cowan et al., 2012, p. 5). The lack of an agreed upon 'grouped' index leads to the inability to create groupings based on socioeconomic status (i.e. low, middle, or high SES) and thus research remains limited with regards to inequalities and inequities as it continues to rely on individual SES components (i.e. education, income, or occupation), race/ethnicity, and gender. While studies regarding inequality and stratification appropriately control for SES through the use of income or education as a covariate, Adler and Rehkopf (2008) note that such adjustments are insufficient given the evidence of the independent effects of the various components of SES and single cofactors of education, income/wealth, or occupation will not, generally speaking, capture what is truly mean by social class. In addition to understanding the importance of racial and ethnic equity and access gaps, researchers can utilize a composite SES index, which would include social and economic variables, as a proxy for social disadvantage, deprivation, or position. Nuru-Jeter et al. (2018), suggest that due to the lack of consensus and wide variability regarding the 'conceptualization' of socioeconomic position, this has led to the lack of 'comparability' across studies and also 'misclassification' of important risk factors, stemming from the erratic manner in which both individual variables and combined measures are theorized and empirically analyzed regarding socioeconomic position.

The benefits of a composite SES indicator, in addition to the findings of individual proxies, are akin to the benefits of neighborhood-level, geographic, census-based, and area deprivation indexes and studies, but are also subject to their criticisms. Area-based measures allow an additional study of geographic communities and give policy makers additional tools in prediction and prescription. Higher regional poverty indirectly corresponds to both lower individual and community SES (Demissie et al., 2000). Still, regional or geographic indicators are also limited in that individual SES varies by region or census tract. Often, difficulties arise in gathering SES information due to privacy concerns based upon census reporting. The information attained is further limited to averaged SES values. Finally, the relationships between area-based indicators and individual SES proxies deviate between urban and non-urban regions (Xie et al., 2020).

The use of a standardized measurement and agreed upon individual SES proxy has many advantages, and certain disadvantages. One advantage is that a standard measurement/definition permits inter- and intra-country comparison where the same variables are accounted for (Freeman et al., 2016). On the other hand, a single indicator may omit or dampen the interaction between its factors. Proponents of single indicators argue that high correlation exists between income, education, wealth, occupation, housing, poverty, and community features. In many cases, the dataset and availability of the various social variables may also dictate the use of a single variable. For example, income has been used in both crime and victimization literature and health outcomes as a single variable proxy of SES (Rennison & Planty, 2003). Reducing potential measurement error and increasing regression precision is another reason for use of a single variable rather than a composite variable. Another benefit of an individual proxy of SES, is that examining outcomes based on a single variable is often clearer and conveys a greater degree of transparency (Cowan et al., 2012). Additionally, a composite SES index, according to Oakes and Rossi (2003), is best to utilize when tackling ‘stratified analyses, graphical presentations, and explanations to lay audiences’ which are the often the norm in health research.

In contrast, there are multiple drawbacks to using one variable such as income or occupation represent SES. One issue is that income is often dictated by one’s age and is likely to be non-linear in nature, which may miss the presence of pre-existent wealth or other buffering factors. For instance, a retiree with low income may be relatively well-off due to their asset holdings and housing. Income is also ‘less stable’ than variables like occupation or education and may raise the possibility of a misleading statistical relationship to the dependent variable (Berzofsky et al., 2015; Shavers, 2007). Education, when isolated, as with other individual SES proxies, also often fails to explain persistent race and ethnic health disparities (Bell et al., 2020). We argue with Cowen et al, that the use of a composite indicator does not prevent reporting on individual associations among SES variables (2012) while adding important supplemental information. Omitting SES variables such as income, occupation, education, and wealth entails missing long-term proxies for access, social mobility, resources, knowledge, and other societal safeguards. As such, the components contributing to SES are both individually and conjunctively important. Therefore, analyses conducted utilizing the individual components of SES (i.e. income, education, and occupation separately) and their relationship to health gaps are, of course, important but are likely insufficient when it comes to adjusting for ‘socioeconomic status’ (Adler & Rehkopf, 2008). A composite index has the advantage of combining several of the component variables encompassing SES into one variable, thus allowing clear interpretation and social grouping. Additionally, a single measure of SES is at odds with that conventional definition of a grouped variable (Cowan et al., 2012).

3. Data & Methodology

3.1. Data

We compare and contrast the adopted SES indexes across four major national surveys, which provide a cross-comparison of socioeconomic and health data: the 2019 Behavioral Risk Factor Surveillance Survey (BRFSS), 2019 IPUMS USA, the 2019 National Health Interview Survey (NHIS), and the 2019 Medical Expenditure Panel Survey (MEPS).

BRFSS data is an annual cross-sectional state-based telephone interview survey. Approximately 400,000 individuals are interviewed each year in every US state, including the District of Columbia and 3 US territories (CDC, 2019). BRFSS data account for the noninstitutionalized adult population and gathers information on health-related risk behaviors (e.g. smoking, alcohol intake, and physical activity), chronic health conditions and injuries (e.g. diabetes, heart disease, asthma), preventable infectious diseases and the use of preventive services among the adult population (CDC, 2019).

IPUMS USA data collects annual census-level and survey data, establishing the largest database of census microdata that is accessible to the public. Data is available for individuals and households and is drawn from the American population yearly from sixteen federal censuses (Ruggles et al., 2022). IPUMS USA contains data on a number of population characteristics and demographics (e.g. immigration, occupational structure, education). For individual-level analyses, person weights were applied to the estimation in order to account for an accurate representation of the population.

Additional survey interview data is collected from the National Health Interview Survey (NHIS). NHIS gathers, supervises, and publishes annual US health data since 1957. This cross-sectional survey collects and utilizes a sampling design encapsulating the noninstitutionalized and nonmilitary population in the U.S., with roughly a 70% response rate for nearly 87,500 individuals and 35,000 households annually interviewed (NCHS, 2019).

Lastly, we utilize the Agency for Healthcare Research's Medical Expenditure Panel Survey (MEPS). MEPS data are gathered annually and are a representative survey of US households. While smaller than BRFSS and NHIS, MEPS roughly covers 12,000-14,000 nonmilitary and noninstitutionalized civilian households (i.e. 31,000-35,000 individuals). Further, MEPS has a 53.5-66.3% response rate for the five survey times within a two-year period (MEPS, 2018).

3.2. Indexing Socioeconomic Status Classifications

To measure socioeconomic standing (i.e. low, middle, and high SES), the BJS created three potential SES index options: Index 1 and 2 include aspects of education, income, employment status, and housing status, while Index 3 excludes housing status (Berzofsky et al., 2015). Table 1 provides a replication of the scaling and composition of the SES index. For all three indices, education and income had a possible range of 0-3 while employment and housing ranged between 0 and 1 (with housing being omitted from Index 3). Each of these individual components (education, income, employment status, and housing status) were then scaled to establish an index value between 0 and 8 for Index 1 and 2. For Index 3, the values were measured between 0 and 7 (as housing was excluded from the SES composite). Given data limitations (i.e. no information on housing status across the four national surveys), this analysis most closely followed Index 3 of the BJS SES indexing.

Table 2 shows how each category follows the BJS categories and indexing; with a few exceptions (i.e. the BRFSS data, most notably) the categories utilized match the BJS report. Regarding education, those who had less than a high school degree were indexed as 0, those with high school, some college, or associate's degree were indexed at 1, those with a Bachelor's degree were indexed at 2, and those with a Master's, professional, or doctorate degree were indexed as 3. Income was measured by poverty status, relative to the Federal poverty line (FDL) and indexed as follows: 0 for those 100% or less than FDL, 1 for those between 101%–200% of FDL, 2 for those between 201%–400% of FDL, and 3 for those 401% or greater than FDL. One of the limitations of the BRFSS data is that income, as a percentage of the FDL, needed to be manually calculated. Following Hest (2019), we computed the FDL using the midpoint formula. Lastly, occupation was measured *via* employment status and indexed as follows: those who were unemployed in the past 6 months received a 0 and those employed for the past 6 months received a 1.

After categorizing and indexing each of the answers according to the BJS method, the sum of the composite scores were computed (ranging from 0-7). Consequently, those with a composite score between 0 and 2 were classified as low SES, those

Table 1. Replication of BJS composition of SES classification—Index 1, Index 2, & Index 3.

	Index 1	Index 2	Index 3
Education	Possible range (0–3) where: ▪ 0: Less than high school ▪ 1: High school, some college, associate's degree ▪ 2: Bachelor's degree ▪ 3: Master's, professional, doctorate degree		
Income (percentage of Federal poverty level)	Possible range (0–3) where: ▪ 0: 100% or less ▪ 1: 101%–200% ▪ 2: 201%–400% ▪ 3: 401% or greater		
Employment	Possible range (0–1) where: ▪ 0: Unemployed past 6 months ▪ 1: Employed past 6 months		
	Index 1	Index 2	Index 3
Housing	Possible range: 0–1 ▪ 0: Rent or no cash rent ▪ 1: Own	Possible range: 0–1 ▪ 0: Public housing ▪ 1: Non-public housing	N/A
Possible range	**0–8**	**0–8**	**0–7**

Table 2. SES indexing across the datasets.

	Education	Income (Federal Poverty line)	Employment
BJS	Categories & Indexing: 0: less than high school 1: high school, some college, associate's degree 2: Bachelor's degree 3: Master's, professional, doctorate degree	Categories & Indexing: 0: 100% or less 1: 101%–200% 2: 201%–400% 3: 401% or greater	Categories & Indexing: 0: unemployed past 6 months 1: employed past 6 months
NHIS	Categories & Indexing 0: less than high school 1: high school, some college 2: Bachelor's degree 3: Master's, professional, doctorate degree	Categories & Indexing: 0: 100% or less 1: 101%–200% 2: 201%–400% 3: 401% or greater	Categories & Indexing - employment status of last week: 0: not working at a job and not looking 1: working for pay; with a job but not at work; looking for work; working but not for pay
MEPS	Categories & Indexing 0: less than high school 1: high school, other degree 2: Bachelor's degree 3: Master's, professional, doctorate degree	Categories & Indexing: 0: 100% or less 1: 101%–200% 2: 201%–400% 3: 401% or greater	Categories & Indexing: 0: Not employed 1: Employed at round 5/3; job to return to
BRFSS	Categories & Indexing* 0: less than high school 1: high school 2: other degree; attended college or technical school 3: graduated from college or technical school	Categories & Indexing**: 0: 100% or less 1: 101%–200% 2: 201%–400% 3: 401% or greater	Categories & Indexing: 0: out of work for 1 year or more; out of work for less than 1 year; a homemaker; a student; retired; unable to work 1: employed for wages; self-employed
IPUMS	Categories & Indexing 0: less than high school 1: high school, other degree 2: Bachelor's degree 3: Master's, professional, doctorate degree	Categories & Indexing: 0: 100% or less 1: 101%–200% 2: 201%–400% 3: 401% or greater	Categories & Indexing: 0: Unemployed; not in labor force. 1: Employed

*Questions regarding education in BRFSS data do not break it down to distinguishing between master's and bachelor's degree. As such, the categories do not match up well to NHIS and MEPS data with regard to education.

**Federal poverty line statistics needed to be manually computed.

with a composite score between 3 and 5 were classified as middle SES, and those with a composite score of 6 or 7 were classified as high SES.

3.3. Statistical Analyses

To evaluate how well the adopted BJS SES classifications held among four national surveys multiple empirical strategies were examined. First, the weighted descriptive statistics of the three datasets are presented, paying particular attention to the breakdown of socioeconomic classifications within each dataset. We further break the socioeconomic classifications down by race/ethnicity and gender to investigate the comparability between the four datasets utilized.

Next, we run logistic regression analyses to further assess the comparability of the BJS SES classifications and how well it does, or does not, hold across the four datasets. To determine whether or not the SES classifications show similar results, an identical regression model is run over the four datasets. As such, we sought a dependent variable that could be found in all four datasets. The probability of having health insurance was utilized for the regression analysis, with SES being one of the many covariates utilized in the regression analysis. This model was chosen out of necessity to run identical regression analyses over the four datasets. However, use of insurance status captures important aspects of health research (i.e. access to healthcare measured by the probability of having health insurance). Additionally, we ran logistic regression models on the probability of having been diagnosed with diabetes for MEPS, NHIS and BRFSS data to verify that the impact of SES remains consistent, which they did.[2]

Each logistic regression controlled for the same covariates in order to ascertain the comparisons directly from each dataset. Covariates used for the regression analysis include marital status, age, race/ethnicity, gender, region, and socioeconomic status (measured through the composite score). Each sample was restricted to the working age population (i.e. between the ages of 18-64), with age categorized as individuals aged 18—34 (reference), aged 35—54, and those aged 55—64. The analyses controlled for gender (with individuals responding as female (reference) or male), race/ethnicity (classified as White non-Hispanic (reference), Black non-Hispanic, Asian non-Hispanic, or Hispanic), marital status (classified as single (reference), widowed/divorced/separated, or married), and region (Northeast, Midwest, West, and South (reference)). As a sensitivity analysis, provided in Appendix A, we included income and education as covariates (i.e. single-variable factors) rather than the SES groupings where income was classified as either between $0 - $34,999, between $35,000 - $74,999, and over $75,000 (reference), which we defined as either low, middle, or high income. Covariates chosen for the statistical analysis follow primary SES, health, and equity gaps in the literature previously discussed (i.e. Braveman et al., 2005, Williams et al., 2016, and Nuru-Jeter et al., 2018).

As each of the datasets used have intricate, and complex, sampling methods, to correct for non-responses and various selection probabilities, survey weights are

[2]Results of the weighted logistic regression analyses of the probability of being diagnosed with diabetes for MEPS, BRFSS, and NHIS data are available upon request.

used throughout all analyses to verify representative sampling of the population. Collecting national survey data incorporates aspects of stratification, clustering, and oversampling of certain subpopulation (e.g. minority groups). Given these complexities, sample weights need to be utilized in order to generate representative estimates (Blewett et al., 2022). After imposing the restrictions on each sample, the resultant final sample size for the 2019 NHIS data was 21,309, for the 2019 BRFSS data was 144,302, for the IPUMS data was 1,919,589, and for the MEPS data was 15,058. All analyses were conducted utilizing SAS V9.

4. Results

Weighted descriptive statistics for the various components of obtaining SES categories, the SES categories based on the BJS-SES definition, and the demographic characteristics of each dataset are provided in Table 3. Across the four datasets, we observed similar distributions of demographic characteristics, though slight variations are present. For example, the BRFSS and IPUMS sample have slightly lower distributions for White non-Hispanic individuals (59%) compared to the MEPS and NHIS sample (62%), while the BRFSS has slightly higher distributions of Hispanic individuals (22%) compared to the MEPS, NHIS, and IPUMS datasets (19%). Regarding health insurance, 91% of the MEPS individuals had health insurance, while it was slightly lower for the IPUMS dataset (87%), NHIS dataset (86%), and BRFSS dataset (84%). Similar age distributions were observed across the MEPS, NHIS, and IPUMS datasets, while a slightly higher distribution of those aged 18-34 (42%) and those aged 55—64 (17%) was observed in the BRFSS data. Lastly, a higher distribution of men was observed in the BRFSS data (53%), and a lower distribution among those who are widowed, divorced, or separated in the NHIS data (11%). Though it is beyond the scope of this paper to analyze the reasons why the differences in demographic distribution vary across the datasets, it should be noted.

As can be seen from Table 3, the components for SES and the distribution of SES classification (i.e. low SES, middle SES, and high SES) reveal moderate variation between the four datasets. The distribution of the components of SES (i.e. education, employment status, and poverty status) is most similar for the MEPS and NHIS data. Comparability of components can also be seen for education and poverty as MEPS and NHIS show similar distributions for each component, while BRFSS and IPUMS data are modestly different. One of the reasons why BRFSS education and poverty status distributions differ is due to the definitions utilized to measure the scoring being different from MEPS, NHIS and IPUMS data. While the distribution of occupation (measured by employment status) and its components are very close for the NHIS and BRFSS data (77% and 78%, respectively), it deviates slightly in the IPUMS dataset (74%) and the MEPS dataset (80%).

Looking at the resulting composite score for SES classifications (i.e. low SES, middle SES, and high SES), we note the following based on the datasets utilized. NHIS and BRFSS results show slightly higher number of individuals in low SES groups (19% and 18%, respectively) in comparison to MEPS and IPUMS data (16% and 17%, respectively). The percent in middle SES classifications is lower in BRFSS and IPUMS data (50% and 54%, respectively) in comparison to MEPS and NHIS data (60% and 61%, respectively). Lastly, the percentage of individuals in high SES classification is

Table 3. Weighted descriptive statistics (MEPS, NHIS, BRFSS, IPUMS).

	MEPS ($N = 15{,}058$)	NHIS ($N = 21{,}309$)	BRFSS ($N = 144{,}302$)	IPUMS ($N = 1{,}919{,}589$)
Education	0: 10.94%	0: 11.03%	0: 10.72%	0: 8.83%
	1: 54.91%	1: 59.01%	1: 25.71%	1: 35.18%
	2: 21.88%	2: 19.23%	2: 31.46%	2: 45.02%
	3: 12.27%	3: 10.72%	3: 32.10%	3: 10.97%
Poverty	0: 10.21%	0: 11.35%	0: 16.52%	0: 13.97%
	1: 14.78%	1: 17.87%	1: 21%	1: 14.34%
	2: 28.90%	2: 30.46%	2: 36.32%	2: 28.34%
	3: 46.12%	3: 40.33%	3: 26.18%	3: 43.34%
Employment	0: 20.28%	0: 23.32%	0: 22.06%	0: 26.20%
	1: 79.72%	1: 76.68%	1: 77.94%	1: 73.80%
SES	Low SES: 16.12%	Low SES: 19.22%	Low SES: 18.20%	Low SES: 16.94%
Classification	Middle SES: 59.61%	Middle SES: 60.54%	Middle SES: 49.92%	Middle SES: 53.65%High SES:
	High SES: 24.27%	High SES: 20.24%	High SES: 31.89%	29.41%
Race/ethnicity	White NH*: 61.37%	White NH: 62%	White NH: 59.48%	White NH: 59.20%
	Black NH: 12.85%	Black NH: 12.61%	Black NH: 12%	Black NH: 12.88%
	Asian NH: 6.86%	Asian NH: 6.40%	Asian NH: 6.60%	Asian NH: 6.36%
	Hispanic: 18.93%	Hispanic: 19%	Hispanic: 21.93%	Hispanic: 18.51%
Gender	Female: 50.97%	Female: 49.34%	Female: 47.43%	Female: 50.20%
	Male: 49.03%	Male: 50.66%	Male: 52.57%	Male: 49.80%
Marital Status:	Married: 51.27%	Married: 51.66%	Married: 49.96%	Married: 48.38%
	W/D/S**: 14.22%	W/D/S: 10.88%	W/D/S: 13.61%	W/D/S: 14.27%
	Single: 34.51%	Single: 37.46%	Single: 29.43%	Single: 37.35%
Region	Northeast: 17.47%	Northeast: 17.47%	Northeast: 13.10%	Northeast: 17.24%
	Midwest: 20.85%	Midwest: 21.23%	Midwest: 22.25%	Midwest: 20.60%
	South: 37.91%	South: 37.83%	South: 23.25%	South: 38.07%
	West: 23.77%	West: 23.47%	West: 24.76%	West: 24.10%
Age	Aged 18–34: 36.55%	Aged 18–34: 37.44%	Aged 18–34: 41.77%	Aged 18–34: 37.74%
	Aged 35–54: 41.83%	Aged 35–54: 41.07%	Aged 35–54: 41.77%	Aged 35–54: 41.17%
	Aged 55–64: 21.62%	Aged 55–64: 21.49%	Aged 55–64: 16.46%	Aged 55–64: 21.09%
Income	Low: 22.03%	Low: 22.40%	Low: 30.53%	Low: 54.74%
	Middle: 27.37%	Middle 30.56%	Middle: 27.82%	Middle: 27.70%
	High: 50.54%	High: 47.14%	High: 41.65%	High: 17.57%
Health Insurance	Yes: 90.78%	Yes: 85.70%	Yes: 84.01%	Yes: 86.79%
	No: 9.22%	No: 14.30%	No: 15.99%	No: 13.21%

*NH: abbreviation for non-Hispanic.
**W/D/S: abbreviation for widowed, divorced, or separated.

Table 4. Weighted regression results for probability of having health insurance (odds ratios).

	MEPS	NHIS	BRFSS	IPUMS
Married (1 = yes, 0 = no)	1.41**	1.47**	1.38**	1.33**
Widowed, Divorced, or Separated (1 = yes, 0 = no)	0.97**	1.01**	0.85**	0.91**
Aged 35–54 (1 = yes, 0 = no)	0.86**	1.03**	0.93**	0.94**
Aged 55–64 (1 = yes, 0 = no)	1.28**	1.50**	1.42**	1.54**
Black non-Hispanic (1 = yes, 0 = no)	0.78**	1.07**	0.85**	0.88**
Asian non-Hispanic (1 = yes, 0 = no)	0.80**	1.17**	0.80**	0.92**
Hispanic (1 = yes, 0 = no)	0.25**	0.36**	0.41**	0.39**
Female (1 = yes, 0 = no)	1.82**	1.35**	1.41**	1.44**
Northeast (1 = yes, 0 = no)	2.74**	2.43**	1.61**	2.32**
Midwest (1 = yes, 0 = no)	1.81**	1.70**	1.21**	1.70**
West (1 = yes, 0 = no)	2.42**	2.19**	1.60**	1.89**
LOW SES (1 = yes, 0 = no)	**0.14****	**0.12****	**0.11****	**0.12****
MIDDLE SES (1 = yes, 0 = no)	**0.27****	**0.25****	**0.28****	**0.23****

**Denotes statistical significance at the 1% level.

highest in the BRFSS and IPUMS datasets (32% and 29%, respectively), compared to those in the MEPS and NHIS dataset (24% and 20%, respectively).

Odds-ratios (OR) results from the weighted logistic regression analyses can be found in Table 4. While the magnitudes for the covariates vary slightly by dataset utilized, signs are consistent for all covariates across each dataset with the exception of Black non-Hispanic, Asian non-Hispanic individuals, those who were classified as widowed, divorced, or separated, and those aged 35—54 in the NHIS dataset. All covariates are statistically significant at the 1% level. Specifically looking at the magnitude of the probability of having health insurance by socioeconomic status classification, we see that across each dataset, the probability of having health insurance is lower for those of low and middle SES, in comparison to high SES. The magnitude of the ORs across each dataset are starkly similar: for low SES the ORs are 0.14, 0.12, 0.11, and 0.12 for MEPS, NHIS, BRFSS and IPUMS results, respectively. For middle SES, the ORs are 0.27, 0.25, 0.28 and 0.23 for MEPS, NHIS, BRFSS and IPUMS results, respectively. As such, the results of the estimates confirm the notion that those of low and middle SES, in comparison to those of high SES, have a lower probability of having health insurance coverage. Additionally, the magnitude of the impact is similar across each dataset.

The odds-ratios in Table 4 also provide an analysis of important covariates in marital status, age, and race/ethnicity. Similar patterns in regard to race emerge from a comparison of the data sources, with the exception of the NHIS dataset. In comparison to White Non-Hispanic (NH), for Black NH the ORs are 0.78, 1.07, 0.85, and 0.88 for MEPS, NHIS, BRFSS, and IPUMS results, respectively. For Asian NH, the ORs are

0.80, 1.17, 0.80, and 0.92 for MEPS, NHIS, BRFSS, and IPUMS results, respectively. Finally, for Hispanics, the ORs are 0.25, 0.36, 0.41, and 0.39 for MEPS, NHIS, BRFSS, and IPUMS results, respectively. Again, each covariate is statistically significant at the 1% level and the magnitudes are similar across the data, with the exception of the NHIS dataset.

To enhance our study, in addition to examining composite SES scores, we utilize single-dimension indicators of education and income in separate analyses for each dataset. Results are provided in Appendix A. Similar outcomes were found when a single composite SES index was utilized, though, compared to the single composite index results, a few inconsistencies were observed across the datasets. Specifically, inconsistent results were found for the widowed, divorced, or separated coefficients in the MEPS and NHIS datasets, aged 35—54 for the NHIS data, and Black non-Hispanic and Asian non-Hispanic for the NHIS data. With regards to the single indicators (i.e. education and income), consistent signs were observed across all four datasets. Additionally, the magnitude of the impact of income (measured as low, middle, or high (reference)) was starkly similar across the four datasets. While consistent signs were observed on the education coefficients, unsurprisingly, the magnitudes differed across the datasets, stemming in part of the different definitions utilized for education classifications.

4.1. Limitations

While we attempt to minimize the limitations in the study, there are, naturally, a number of challenges faced. First, self-reported survey data is utilized for the analyses. Though we utilize the appropriate statistical weights for the analyses, non-response bias remains a potential issue with survey data. Second, we modeled the SES classifications after the BJS definitions regarding education, income and employment status. While the categories and indexing for education match between BJS, NHIS, and IPUMS data, they are not a 100% match for MEPS and BRFSS data. For example, for the education scores, the BJS categories are 'less than high school', 'high school, some college, or associate's degree', 'Bachelor's degree', and 'Master's, professional, or doctorate degree'. However, in BRFSS data, education could not be tiered out beyond college graduate (i.e. cannot distinguish between Bachelor's degree and Master's or professional degree). Furthermore, among these national datasets, how each component of SES is defined (i.e. education, income, and employment status) varies slightly. As such, adopting the BJS classifications to all four datasets means that there are slight deviations on how each component is weighted by the definition used. However, despite the differences in dataset design, the authors conclude that given the impact of SES classifications on the probability of having insurance being similar, or even identical in some cases, among the datasets utilized, the SES composite measure adopted by the BJS terminology can translate over to any of these four datasets.

Third, while a number of national datasets are available for public use and research (e.g. the Health and Retirement Survey (HRS) or the American Time Use Survey (ATUS)), one is not able to use these datasets because they do not allow for the classification of SES due to the lack of data garnered on the components of SES (i.e. education, occupation, and income). For example, we were not able to utilize either the HRS or ATUS data as there was insufficient information on 'poverty status' in both of the datasets. Additionally, given the BJS definition and how to integrate informal work such as stay at home spouses, we essentially are labeling those individuals are 'not in the labor force' and they are assigned an occupation score of 0.

Fourth, given the lack of health-related variables in the IPUMS data that matched health-related questions asked in MEPS, NHIS, or BRFSS data, we were limited to running the logistic regression analyses on the probability of having health insurance. However, running other analyses on the same covariates but with the probability of being diagnosed with diabetes for the MEPS, NHIS, and BRFSS data showed similar outcomes, thus solidifying the notion that the SES composite index created holds across the datasets. Furthermore, studies show a significant relationship between health insurance status and health disparities, with those that are uninsured (or underinsured) often experiencing worse health outcomes, lower quality of care, and/or less access to care (McWilliams, 2009).

Fifth, while SES is classically defined as a composite measure of income, occupation and education, wealth is often included in the composite score as well. However, wealth is often omitted in more basic measures of SES (Berzofsky et al., 2015; Hout, 2018). When data is available, wealth in conjunction with housing better describes social factors and buffers a household may have. With the even steeper rise in wealth inequality in contrast to income inequality, wealth is an important indicator of resources available to a smaller subset of the population. Wealth as a safety net for health and income shocks that is rarely present for lower SES groups. Given the BJS definition and the limitations of datasets in not asking direct questions regarding wealth, we are not able to include wealth in the composite score of SES.

Finally, while we believe this is a step towards a socially-grouped or multifactorial proxy for the intersection of socioeconomic variables, it remains a starting point. Our SES-composite index, like single variables, are limited in their ability to fully encompass all institutional arrangements, such as the endogenous relationship that exists between policy-setting and how that, in turn, impacts one's SES. Furthermore, the decisions on variable inclusion are subject to availability and variable weighting based upon BJS and our own subjective decisions. While clearly, not the 'gold standard', we believe this is a firm move toward a single summary indicator capturing many of the combined aspects of class and SES.

5. Discussion & Conclusions

In this study, we created, and evaluated, the SES index from the Bureau of Justice Statistics (BJS) into low, middle, and high SES classifications and applied this definition to four national datasets. Next, we investigated whether or not the definitions 'held' across the datasets (i.e. had similar SES distributions) and whether or not the impact of SES classifications across the datasets were similar. Through the use of weighted descriptive statistics and weighted logistic regression analyses, we reveal the key variations in the distribution of determinants of the socioeconomic index. We find remarkably similar outcomes regarding the assigned SES classifications in the logistic regression analyses performed. Strong similarities across datasets were present when comparing race and ethnicity and other covariates. We also find significant disparities in both the SES composite indexes and race/ethnicity, as can be seen when analyzing the descriptive statistics or the likelihood of having health insurance. While previous studies conflated race and ethnicity with SES to resulting health and health access, this study captures both highlighting the importance of each covariate and a composite index. As with neighborhood-level or area deprivation-based indexes, a composite index gives the researcher additional information on collective traits of class. Furthermore, it provides a 'single summary' indicator called for by the NAEP (Cowan et al., 2012), while overcoming privacy and reporting issues with Census reporting and the discrepancies between urban and non-urban regions (Zie et al., 2020). To date, no study to our knowledge has compared one SES composite index across four national survey datasets and found comparable results.

We see that the primary advantage of an SES index lies in capturing the cumulative structural determinants of SES (Braveman et al., 2005). The cumulative structural determinants that present themselves with this composite measure cannot be captured through the use of a single-variable indicator, such as education and income as shown in Appendix A, unless a number of interaction terms are utilized in analyses (e.g. income level*education level*employment status). Such interaction terms are, of course, not unheard of, however, they present challenges of overfitting and in the complexity of interpreting all the various interactions that may be present. As such, they are, typically, underutilized in regression analyses. However, through the use of the cumulative SES index, one is able to capture the interaction of education, income, and employment status. Additionally, a composite index allows comparison across datasets removing the variability in the conception of SES allowing a uniform study of risk factors contributing to inequities (Nuru-Jeter et al., 2018). An SES index also controls for various factors of SES without the need to qualify which variables were used and reducing the discussion of unobserved SES differences in the outcomes (Braveman et al., 2005). Though, the need for a composite indicator by both researchers and healthcare providers is established, an agreed upon indicator, comparable across data sources, is still missing.

Our composite SES index highlights the social and economic factors in determining, specifically in this study, the probability of having health insurance through the interaction of education, income (as measured by poverty status), and employment status. Typically, given the complications of deriving a measure of SES, these interactions are often neglected in studies. Our findings, unsurprisingly, align with the findings utilizing individual proxies, as shown in Appendix A, where lower income groups are less likely to have insurance and access to preventative and life-saving care (Lasser et al., 2006). Furthermore, given the known intersections of social class, race/ethnicity, and gender that are present when determining health and other social locational disparities, researchers will be able to build off of this foundation to analyses the intersections between the SES groupings, as described, and race/ethnicity and gender.

We show that the composite indices created and adopted add greater information as to what constitutes SES for a group, given the complexities of measuring SES and its multifactorial nature in health outcomes, or health access, analyses (Braveman et al., 2005). Additionally, while, generally speaking, the health literature focuses on health disparities based on race/ethnicity, this may contribute to the systemic oppression as these racial/ethnic classifications reflect exploitation, oppression, and social inequality (Williams et al., 2016). As such, studies regarding health disparities should control for race/ethnicity in addition to controlling for the composite measures of SES, and not individual components (i.e. income, education, and occupation). From a policy stance, the use of composite index can work in conjunction with individual SES proxies. Not only do comparisons of health indicators by composite SES grouping allow for better targeting and future reduction of disparities, but providing a methodology for how to create an SES index will be fruitful for researchers seeking to broaden our perception of health inequities.

References

Adler, N., & Newman, K. (2002). Socioeconomic disparities in health: Pathways and policies. *Health Affairs (Project Hope)*, *21*(2), 60–76. https://doi.org/10.1377/hlthaff.21.2.60

Adler, N., & Rehkopf, D. (2008). U.S. disparities in health: Descriptions, causes, and mechanisms. *Annual Review of Public Health*, *29*, 235–252. https://doi.org/10.1146/annurev.publhealth.29.020907.090852

Bell, C. N., Sacks, T. K., Tobin, C. S. T., & Thorpe, R. JJr, (2020). Racial non-equivalence of socioeconomic status and self-rated health among African Americans and Whites. *SSM-population health.*, *10*.

Berzofsky, M., Creel, D., Moore, A., Smiley-McDonald, H., & Krebs, C. (2015). Measuring socioeconomic status (SES) in the NCVS: Background, options, and recommendations. *Bureau of Justice Statistics U.S. Department of Justice*.

Blewett, L., Drew, J., King, M., Williams, K., Del Ponte, N., & Convey, P. (2022). *IPUMS health surveys: National Health Interview Survey*, Version 7.2 [dataset]. IPUMS., 2022. https://doi.org/10.18128/D070.V7.2

Braveman, P., Cubbin, C., Egeter, S., Chideya, S., Marchi, K., Metzler, M., & Posner, S. (2005). Socioeconomic status in health research: one size does not fit all. *The Journal of the American Medical Association*, *294*(22), 2879–2888.

Centers for Disease Control and Prevention (CDC). (2019). *Behavioral risk factor surveillance system survey data*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.,

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and economic opportunity in the United States: An intergenerational perspective. *The Quarterly Journal of Economics*, *135*(2), 711–783. https://doi.org/10.1093/qje/qjz042

Cowan, C., Hauser, R., Kominski, R., Levin, H., Lucas, S., Morgan, S., Spencer, M., & Chapman, C. (2012). *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation*. Recommendations to the National Center for Education Statistics.

Darity, W., Hamilton, D., & Stewart, J. (2014). A Tour de Force in Understanding Intergroup Inequality: An Introduction to Stratification Economics. *Review of Black Political Economy*, *42*(1-2). https://doi.org/10.1007/s12114-014-9201-2

Daw, J. (2017). Contribution of four comorbid conditions to racial/ethnic disparities in mortality risk. *American Journal of Preventive Medicine*, *52*(1S1), S95–S102. https://doi.org/10.1016/j.amepre.2016.07.036

Deaton, A. (2003). Health, inequality, and economic development. *Journal of Economic Literature*, *41*(1), 113–158. https://doi.org/10.1257/jel.41.1.113

Demissie, K., Hanley, J. A., Menzies, D., Joseph, L., & Ernst, P. (2000). Agreement in measuring socio-economic status: Area-based versus individual measures. *Chronic Dis Can*, *21*(1), 1–7.

Fiscella, K., & Williams, D. R. (2004). Health disparities based on socioeconomic inequities: Implications for urban healthcare. *Academic Medicine : journal of the Association of American Medical Colleges*, *79*(12), 1139–1147. https://doi.org/10.1097/00001888-200412000-00004

Frank, J. W., & Mustard, J. F. (1994). The determinants of health from a historical perspective. *Daedalus*, *123*(4), 1–19.

Freeman, A., Tyrovolas, S., Koyanagi, A., Chatterji, S., Leonardi, M., Ayuso-Mateos, J., Tobiasz-Adamczyk, B., Koskinen, S., Rummel-Kluge, C., & Haro, J. (2016). The role of socio-economic status in depression: Results from the COURAGE (aging survey in Europe). *BMC Public Health*, *16*(1), 1098. https://doi.org/10.1186/s12889-016-3638-0

Hest, R. (2019). *Four Methods for Calculating Income as a Percent of the Federal Poverty Guideline (FPG) in the Behavioral Risk Factor Surveillance System (BRFSS). State Health Access Data Assistance Center*.
Hauser, R., & Warren, J. (1997). Socioeconomic indexes for occupations: a review, update, and critique. *Sociological Methdology*, *27*, 177–298.
Hout, M. (2018). Americans' occupational status reflects the status of both of their parents. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(38), 9527–9532. https://doi.org/10.1073/pnas.1802508115
Kawachi, I. (2013). Editorial: Isn't all epidemiology social? *American Journal of Epidemiology*, *178*(6), 841–842. https://doi.org/10.1093/aje/kwt146
Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in US public health research. *Annual Review of Public Health*, *18*, 341–378. https://doi.org/10.1146/annurev.publhealth.18.1.341
Lasser, K. E., Himmelstein, D. U., & Woolhandler, S. (2006). Access to Care, Health Status, and Health Disparities in the United States and Canada: Results of a Cross-National Population-Based Survey. *Journal of Public Health*, *96*(7), 1300–1307. https://doi.org/10.2105/AJPH.2004.059402
Marmot, M., & Allen, J. J. (2014). Social determinants of health equity. *American Journal of Public Health*, *104 Suppl 4*(Suppl 4), S517–S9. https://doi.org/10.2105/AJPH.2014.302200
Marmot, M., Friel, S., Bell, R., Houweling, T. A., & Taylor, S. (2008). Closing the gap in a generation: Health equity through action on the social determinants of health. *Lancet (London, England)*, *372*(9650), 1661–1669. https://doi.org/10.1016/S0140-6736(08)61690-6
McWilliams, J. M. (2009). Health consequences of uninsurance among adults in the United States: Recent evidence and implications. *The Milbank Quarterly*, *87*(2), 443–494. https://doi.org/10.1111/j.1468-0009.2009.00564.x
Medical Expenditure Panel Survey (MEPS). (2018). *Agency for healthcare research and quality, Rockville, MD*. Retrieved August, from https://www.ahrq.gov/data/meps.html
National Center for Health Statistics (NCHS). (2019). *National Health Interview Survey*. https://www.cdc.gov/nchs/nhis/index.htm
Nuru-Jeter, A. M., Michaels, E. K., Thomas, M. D., Reeves, A. N., Thorpe, R. J., Jr,., & LaVeist, T. A. (2018). Relative roles of race versus socioeconomic position in studies of health inequalities: A matter of interpretation. *Annual Review of Public Health*, *39*, 169–188. https://doi.org/10.1146/annurev-publhealth-040617-014230
Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: Current practice and steps toward a new approach. *Social Science & Medicine (1982)*, *56*(4), 769–784. https://doi.org/10.1016/s0277-9536(02)00073-4
Olafsdottir, S. (2007). Fundamental Causes of Health Disparities: Stratification, the Welfare State, and Health in the United States and Iceland. *Journal of Health and Social Behavior*, *48*, 239–253.
Rennison, C., & Planty, M. (2003). Nonlethal intimate partner violence: Examining race, gender, and income patterns. *Violence and Victims*, *18*(4), 433–443. https://doi.org/10.1891/vivi.2003.18.4.433
Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., & Sobek, M. (2022). *IPUMS USA: Version 11.0*. Minneapolis. MN. https://doi.org/10.18128/D010.V11.0
Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, *99*, 1013–1023.
Vincent, K., & Sutherland, J. M. (2013). *A review of methods for deriving an index for socioeconomic status in British Columbia*. Vancouver, BC: UBC Centre for Health Services and Policy Research. Retrieved from: http://healthcarefunding.ca/files/2013/04/Review-of-Methods-for-SES-Index-for-BC.pdf
Webb, S., Janus, M., Duku, E., Raos, R., Brownell, M., Forer, B., Guhn, M., & Muhajarine, N. (2017). Neighbourhood socioeconomic status indices and early childhood development. *SSM - Population Health*, *3*, 48–56. https://doi.org/10.1016/j.ssmph.2016.11.006
Williams, D. R., Priest, N., & Anderson, N. B. (2016). Understanding associations among race, socioeconomic status, and health: Patterns and prospects. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *35*(4), 407–411. https://doi.org/10.1037/hea0000242

Williams, D. R., Mohammed, S. A., Leavell, J., & Collins, C. (2010). Race, socioeconomic status and health: Complexities, ongoing challenges and research opportunities. *Annals of the New York Academy of Sciences*, *1186*, 69–101. https://doi.org/10.1111/j.1749-6632.2009.05339.x

Xie, S., Hubbard, R. A., & Himes, B. E. (2020). Neighborhood-level measures of socioeconomic status are more correlated with individual-level measures in urban areas compared with less urban areas. *Annals of Epidemiology*, *43*, 37–43.e4. https://doi.org/10.1016/j.annepidem.2020.01.012

Appendix A: Weighted Regression Results for Probability of Having Health Insurance (Odds Ratios) with Income & Education as Independent Variables

	MEPS	NHIS	BRFSS	IPUMS
Married (1 = yes, 0 = no)	1.13**	1.31**	1.20**	1.32**
Widowed, Divorced, or Separated (1 = yes, 0 = no)	1.03**	1.08**	0.91**	0.82**
Aged 35–54 (1 = yes, 0 = no)	0.90**	1.07**	0.97**	0.89**
Aged 55–64 (1 = yes, 0 = no)	1.36**	1.62**	1.56**	1.50**
Black non-Hispanic (1 = yes, 0 = no)	0.86**	1.17**	0.94**	0.87**
Asian non-Hispanic (1 = yes, 0 = no)	0.79**	1.13**	0.76**	0.92**
Hispanic (1 = yes, 0 = no)	0.28**	0.41**	0.50**	0.42**
Female (1 = yes, 0 = no)	1.77**	1.32**	1.39**	1.55**
Northeast (1 = yes, 0 = no)	2.68**	2.37**	1.61**	2.29**
Midwest (1 = yes, 0 = no)	1.85**	1.68**	1.23**	1.70**
West (1 = yes, 0 = no)	2.40**	2.08**	1.55**	1.87**
Income (low) (1 = yes, 0 = no)	**0.29****	**0.30****	**0.24****	**0.24****
Income (middle) (1 = yes, 0 = no)	**0.49****	**0.39****	**0.44****	**0.50****
High School Degree (1 = yes, 0 = no)	**1.80****	**1.80****	**1.99****	**1.69****
Some college/Bachelor's Degree (1 = yes, 0 = no)	**2.67****	**3.13****	**2.71****	**2.76****
Master's or Doctoral Degree (1 = yes, 0 = no)	**5.14****	**4.83****	**1.62****	**4.68****