

Statistical Assessment of Neighborhood Socioeconomic Deprivation Environment in Spatial Epidemiologic Studies

Min Lian^{1,2*}, James Struthers¹, Ying Liu³

¹Division of General Medical Sciences, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

²Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine, St. Louis, MO, USA

³Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

Email: ^{*}mlian@wustl.edu

Received 17 December 2015; accepted 11 June 2016; published 14 June 2016;

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Neighborhood socioeconomic deprivation has been associated with health behaviors and outcomes. However, neighborhood socioeconomic status has been measured inconsistently across studies. It remains unclear whether appropriate socioeconomic indicators vary over geographic areas and geographic levels. The aim of this study is to compare the composite socioeconomic index to six socioeconomic indicators reflecting different aspects of socioeconomic environment by both geographic areas and levels. Using 2000 U.S. Census data, we performed a multivariate common factor analysis to identify significant socioeconomic resources and constructed 12 composite indexes at the county, the census tract, and the block group levels across the nation and for three states, respectively. We assessed the agreement between composite indexes and single socioeconomic variables. The component of the composite index varied across geographic areas. At a specific geographic region, the component of the composite index was similar at the levels of census tracts and block groups but different from that at the county level. The percentage of population below federal poverty line was a significant contributor to the composite index, regardless of geographic areas and levels. Compared with non-component socioeconomic indicators, component variables were more agreeable to the composite index. Based on these findings, we conclude that a composite index is better as a measure of neighborhood socioeconomic deprivation than a single indicator, and it should be constructed on an area- and unit-specific basis to accurately identify and quantify small-area socioeconomic inequalities over a specific study region.

*Corresponding author.

Keywords

Assessment, Neighborhood, Socioeconomic, Deprivation, Spatial Epidemiology

1. Introduction

Health-related behaviors and outcomes display significant geographic variations. Neighborhood socioeconomic environment (SES) has been associated with health-related behaviors [1]-[4], incidence [5]-[7] and poor prognosis [8] of diseases, and premature mortality [5] [9]-[12]. Population-based data sources from local and federal governments (e.g. U.S. Census) provide a number of SES-related data elements and are commonly used to assess the role of neighborhood SES in health behaviors and outcomes. However, there is no consensus on which neighborhood measures, at which geographic level should be used to examine socioeconomic disparities in health behaviors and outcomes. Neighborhood SES has been defined inconsistently across studies, which may contribute to inconsistent findings regarding the relationships between neighborhood SES and health behaviors and outcomes [13]. Various single SES indicators at different geographic levels (e.g. county, census tract, block group) have been used as neighborhood SES measures. It remains unclear regarding appropriate SES indicators for a specific geographic region at a specific geographic level.

Neighborhood SES is a complex concept consisting of multiple aspects of socioeconomic resources. A variety of single-variable measures makes it possible to develop a composite index to comprehensively assess neighborhood SES environment. We propose that, compared with single-variable measures, a composite index can more accurately reflect neighborhood deprivation by capturing more dimensions of socioeconomic resources.

In this study, we apply 2000 U.S. Census data to identify individual socioeconomic variables that significantly reflect socioeconomic deprivation across four geographic areas at three geographic levels. We compare composite indexes with six socioeconomic indicators reflecting different aspects of socioeconomic deprivation environment.

2. Methods

2.1. Data Source

U.S. Census data have been widely applied to assess neighborhood socioeconomic context. For the 2000 census and before, the Census Bureau collected population and housing data from all households and socioeconomic data from about one in six households every ten years at a single point in time. From 2006, these information has been collected over time with households sampled per year by the American Community Survey (ACS) and only the cumulative five-year ACS approximating the sample proportion achieved by the decennial census. Considering ACS margins of error for small areas, we applied 2000 U.S. data for the socioeconomic information of geographic areas. In this study, ethical review was not needed because only public-use area-level Census data were applied.

2.2. Single SES Variables

To capture broad aspects of socioeconomic deprivation context, based on the literature [5] [10] [14]-[16], we selected 21 Census variables at three geographic levels (county, census tract, and block group) (**Table 1**). These variables, which reflect neighborhood socioeconomically deprived resources from six different domains, include 1) education (the percentage of population without high school education); 2) occupation (the percentage of population in working class, the percentage of civilian labor force unemployed); 3) housing conditions (the percentage of household rent, the percentage of vacant household, the percentage of household with at least one person per room, the percentage of female headed households with dependent children, the percentage of household with public assistance, the percentage of household with no car, the percentage of household with no phone, the percentage of occupied household with incomplete plumbing, the percentage of household with no kitchen); 4) income and poverty (income disparity, the percentage of household with low income, the percentage of households below federal poverty line, the percentage of population below federal poverty line); 5) racial composition (the percentage of non-Hispanic African Americans, the percentage of Hispanic population, the

Table 1. Variables selected to comprise deprivation index at three levels in four areas.

	County				Tract				Block Group			
	I ^a	II ^b	III ^c	IV ^d	I	II	III	IV	I	II	III	IV
Education												
% Population with less than high school	*					*				*		
Occupation												
% Population with working class		*			*	*	*	*	*	*	*	*
% Civilian labor force unemployed		*	*	*		*	*	*		*	*	*
Housing Conditions												
% HH ^e rent			*	*				*				
% Vacant HH					*				*			
% HH with ≥1 person/room						*		*		*		*
% Female headed HH with dependent children			*	*			*	*			*	*
% HH on public assistance		*	*	*		*	*	*		*	*	*
% HH with no car			*	*			*	*			*	*
% HH with no phone	*				*			*	*			
% Occupied HH with incomplete plumbing												
% HH with no kitchen												
Income & Poverty												
Income disparity	*	*		*	*	*	*	*	*	*	*	*
% HH with low income	*	*		*	*	*	*	*	*	*	*	*
% HH below federal poverty line	*	*		*	*	*	*	*	*	*	*	*
% Population below federal poverty line	*	*	*	*	*	*	*	*	*	*	*	*
Racial Composition												
% Non-Hispanic African Americans			*	*			*	*			*	*
% Hispanic population						*				*		
% Foreign born						*				*		
Residential Stability												
% Residents aged 65 or older												
% Persons with the same house ≥5 years												
% Total variance explained	33.0	31.6	47.8	43.6	40.7	47.2	44.6	47.1	36.5	41.4	39.8	40.4
Cronbach's Alpha	0.94	0.95	0.92	0.95	0.91	0.96	0.96	0.96	0.88	0.94	0.94	0.94

^aI: the nation; ^bII: California; ^cIII: Georgia; ^dIV: Louisiana; ^eHH: household; * variables selected for constructing the composite index.

percentage of population foreign-born); and 6) residential stability (the percentage of residents aged 65 or older, the percentage of persons with the same house at least five years). To examine the influence of geographic size, we performed the analysis across the nation and three states that have different socioeconomic characteristics and are involved in the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute.

2.3. Statistical Analysis

2.3.1. Development of Neighborhood Socioeconomic Deprivation Index

Using a multivariate common factor analysis with the “varimax” rotation, we examined the internal structure of Census variables and identified their importance. We selected the common factor which predominantly accounted for total variance of all variables. A variable was selected to construct a composite index if its factor

loading on the selected common factor was: 1) no less than 0.5; 2) the largest among its factor loadings across all common factors; and 3) at least 0.1 larger than the second largest factor loading across all common factors. A composite index was constructed by summing all selected variables that were standardized and weighted by their factor scoring coefficients. Cronbach alpha was applied to evaluate the internal consistency of selected variables with bigger value indicating greater internal consistency. A total of 12 composite index scores were independently developed for four geographic areas at three geographic levels, respectively.

2.3.2. Examination of the Agreements

To compare a composite index to single socioeconomic indicators, we selected six commonly-used variables from the aforementioned six domains (one per domain). They included the percentage of population without high school education, the percentage of civilian labor force unemployed, the percentage of households with public assistance, the percentage of population below federal poverty line, the percentage of non-Hispanic African Americans, and the percentage of residents age 65 or older. Regarding potential skewed distributions of Census variables, we categorized the composite index and six single indicators into quintiles (five categories) according to their distributions. The categorization is commonly and broadly applied to assess the effects of environmental exposures on health behaviors and outcomes in epidemiological studies. We examined the agreements between seven variables through computing weighted Kappa coefficients for each pair of these variables [17]. Based on previous literature [18], the degree of agreement was defined as six categories, including 0 (no agreement, $\kappa < 0$), 1 (slight agreement, $\kappa = 0.01 - 0.20$), 2 (fair agreement, $\kappa = 0.21 - 0.40$), 3 (moderate agreement, $\kappa = 0.41 - 0.60$), 4 (substantial agreement, $\kappa = 0.61 - 0.80$), and 5 (perfect agreement, $\kappa > 0.80$). The data management and analysis were performed in SAS System (version 9.3, SAS Institute Inc., Cary, North Carolina).

3. Results

Table 1 shows the component structure of 12 geographic area- and level-specific composite SES indexes. The component of the composite index varied across examined geographic areas. These component variables selected for each of 12 composite indexes account for a large proportion of overall variance of all Census variables (ranged from 31.6% to 47.8%), and have high internal consistencies (Cronbach's alpha ranged from 0.88 to 0.96). At a specific geographic region, the component of the composite index was similar at the census tract- and block group-level but different from that at the county level. The percentage of population below federal poverty line was consistently selected for the composite index, regardless of geographic areas and levels. In contrast, the residential stability domain did not significantly contribute to the composite index at any of geographic areas or levels.

The percentage of population without high school education and the percentage of households with public assistance were the component of the composite index for each of three states, regardless of geographic levels, but not for the nation. The percentage of non-Hispanic African Americans is one of significant contributors to the composite index in Georgia and Louisiana, the states with a relatively high proportion of African American residents.

At the census tract level, the composite indexes had moderate-to-substantial agreements with their components and no-to-moderate agreements with non-component variables (**Table 2**). Across the nation, the composite index showed a substantial similarity (κ category is 4) to its component variable (the percentage of population below federal poverty line), and slight-to-moderate similarities (κ categories range from 0 to 3) to non-component variables. This agreement difference between the composite index and component and non-component variables was also observed in three states. The percentage of population below federal poverty line had no-to-substantial agreements with other socioeconomic indicators (κ categories range from 0 to 4).

4. Discussion

Neighborhood SES has been widely used to assess socioeconomic gradients and inequalities in a variety of health behaviors and outcomes [1]-[12]. However, there is no consensus on the definition of neighborhood SES, and thus various socioeconomic variables have been used across studies. This may explain, at least in part, the inconsistent results of the role of neighborhood SES in health behaviors and outcomes [13].

Using a uniform set of U.S. Census variables, we compared a composite index to six commonly-used socioeconomic indicators from different socioeconomic deprivation domains. The result showed that substantial

Table 2. Weighted Kappa agreement between seven socioeconomic variables at census tract level.

	PNH ^a	PNE ^b	PPA ^c	PPV ^d	PAA ^e	POD ^f
Socioeconomic deprivation index	3 ^g	3	3	4	1	1
	4 ^h	3	4	4	1	0
	3 ⁱ	4	4	4	4	1
	2 ^j	4	4	4	4	1
% Population with less than high school		2	3	3	1	1
		3	3	3	1	0
		2	3	3	1	2
		2	2	3	2	1
% Civilian labor force unemployed			3	3	2	0
			3	3	1	0
			3	3	3	1
			3	4	3	1
% Household on public assistance				3	1	0
				3	2	0
				3	3	1
				3	3	1
% Population below federal poverty line					2	0
					1	0
					3	2
					3	1
% non-Hispanic African Americans						0
						0
						1
						0

^aPNH: % Population with less than high school; ^bPNE: % Civilian labor force unemployed; ^cPPA: % Household on public assistance; ^dPPV: % Population below federal poverty line; ^ePAA: % Non-Hispanic African Americans; ^fPOD: % Residents aged 65 or older; ^gthe nation (1st row); ^hCalifornia (2nd row); ⁱGeorgia (3rd row); ^jLouisiana (4th row). 0: No agreement; 1: Slight agreement; 2: Fair agreement; 3: Moderate agreement; 4: Substantial agreement; and 5: Perfect agreement.

resources of neighborhood SES varied over target regions and geographic units. A composite index was not identical to single SES indicators and more representative of neighborhood SES by capturing broad dimensions of SES resources.

Therefore, geographic area- and level-specific SES indicators should be used to define SES for the study area. In studies examining the role of general neighborhood SES in health behaviors and outcomes, a composite index is a measure of neighborhood SES better than single SES indicators. If we assess the role of a specific SES indicator, such as poverty, it is necessary to examine if that indicator substantially reflects overall SES environment of the studied geographic region at a certain geographic level. Otherwise, the SES indicator selected may not be generalizable to overall neighborhood SES environment. In this study, we only compare the composite SES index to six commonly-used Census variables from different socioeconomic domains. Further research may be necessary to compare neighborhood SES deprivation index to other variables or indexes of interest. However, our findings suggest that the assessment method of neighborhood SES environment should be paid more attention. Researchers should examine specific characteristics of SES environment in their own study regions to design an appropriate strategy in assessing neighborhood SES, instead of simply selecting SES variables applied in previous literature.

Regarding the margins of error of the ACS data, we apply the 2000 Census data which may not benefit recently-initiated studies. However, historic data source sometimes can be useful for prospective studies initiated in an earlier time-point. History of neighborhood exposures and their changes over time should be integrated into advanced statistical modeling to control for spatial uncertainty due to time-varying exposures and confounders for unbiased estimations of neighborhood effects on health behaviors and outcomes. In addition, the main purpose of this study is to address the strategy in assessing small-area neighborhood socioeconomic environment by comparing different socioeconomic variables to a composite index and examining the degree of their agreements using a uniform and reliable data source. Previous study has indicated that selecting different socioeconomic indicators can lead to inconsistent findings [13]; therefore, it is necessary for researchers to select an appropriate approach in accurately assessing neighborhood SES environment.

In conclusion, geographic area- and unit-specific SES measures should be applied to identify and quantify socioeconomic inequalities in health behaviors and outcomes. A multivariate factor analysis with an appropriate rotation method is a useful approach to identify region- and geographic unit-specific SES indicators and construct a composite index. SES resources of the specific geographic area, along with the research question, should be taken into account in selecting a composite index or single indicators as a SES measure.

Acknowledgements

This work was supported in part by a career development award (K07 CA178331) and a research award (R21 CA169807) from the National Cancer Institute at the National Institutes of Health, and a research award (R01 AA021492) from the National Institute on Alcohol Abuse and Alcoholism at the National Institutes of Health. In addition, Y. L. is supported by the Barnes-Jewish Hospital Foundation, St. Louis, Missouri and the Breast Cancer Research Foundation. We also thank for the use of the Health Behavior, Communication and Outreach Core, part of a cancer center grant (P30 CA091842) funded by the National Cancer Institute at the National Institutes of Health. No conflicts of interest were declared.

References

- [1] Dailey, A.B., Kasl, S.V., Holford, T.R., Calvocoressi, L. and Jones, B.A. (2007) Neighborhood-Level Socioeconomic Predictors of Nonadherence to Mammography Screening Guidelines. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, **16**, 2293-2303. <http://dx.doi.org/10.1158/1055-9965.EPI-06-1076>
- [2] Shishehbor, M.H., Gordon-Larsen, P., Kiefe, C.I. and Litaker, D. (2008) Association of Neighborhood Socioeconomic Status with Physical Fitness in Healthy Young Adults: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. *American Heart Journal*, **55**, 699-705. <http://dx.doi.org/10.1016/j.ahj.2007.07.055>
- [3] Mathur, C., Erickson, D.J., Stigler, M.H., Forster, J.L. and Finnegan Jr., J.R. (2013) Individual and Neighborhood Socioeconomic Status Effects on Adolescent Smoking: A Multilevel Cohort-Sequential Latent Growth Analysis. *American Journal of Public Health*, **103**, 543-548. <http://dx.doi.org/10.2105/AJPH.2012.300830>
- [4] Cohen, S.S., Sonderman, J.S., Mumma, M.T., Signorello, L.B. and Blot, W.J. (2011) Individual and Neighborhood-Level Socioeconomic Characteristics in Relation to Smoking Prevalence among Black and White Adults in the Southeastern United States: A Cross-Sectional Study. *BMC Public Health*, **11**, 877. <http://dx.doi.org/10.1186/1471-2458-11-877>
- [5] Krieger, N., Chen, J.T., Waterman, P.D., Soobader, M.J., Subramanian, S.V. and Carson, R. (2002) Geocoding and Monitoring of US Socioeconomic Inequalities in Mortality and Cancer Incidence: Does the Choice of Area-Based Measure and Geographic Level Matter?: The Public Health Disparities Geocoding Project. *American Journal of Epidemiology*, **156**, 471-482. <http://dx.doi.org/10.1093/aje/kwf068>
- [6] Kim, D., Masyn, K.E., Kawachi, I., Laden, F. and Colditz, G.A. (2010) Neighborhood Socioeconomic Status and Behavioral Pathways to Risks of Colon and Rectal Cancer in Women. *Cancer*, **116**, 4187-4196. <http://dx.doi.org/10.1002/cncr.25195>
- [7] Palmer, J.R., Boggs, D.A., Wise, L.A., Adams-Campbell, L.L. and Rosenberg, L. (2012) Individual and Neighborhood Socioeconomic Status in Relation to Breast Cancer Incidence in African-American Women. *American Journal of Epidemiology*, **176**, 1141-1146. <http://dx.doi.org/10.1093/aje/kws211>
- [8] Gerber, Y., Koton, S., Goldbourt, U., Myers, V., Benyamini, Y., Tanne, D., et al. (2011) Poor Neighborhood Socioeconomic Status and Risk of Ischemic Stroke after Myocardial Infarction. *Epidemiology*, **22**, 162-169. <http://dx.doi.org/10.1097/EDE.0b013e31820463a3>

- [9] Bosma, H., van de Mheen, H.D., Borsboom, G.J. and Mackenbach, J.P. (2001) Neighborhood Socioeconomic Status and All-Cause Mortality. *American Journal of Epidemiology*, **153**, 363-371. <http://dx.doi.org/10.1093/aje/153.4.363>
- [10] Singh, G.K. (2003) Area Deprivation and Widening Inequalities in US Mortality, 1969-1998. *American Journal of Public Health*, **93**, 1137-1143. <http://dx.doi.org/10.2105/AJPH.93.7.1137>
- [11] Doubeni, C.A., Schootman, M., Major, J.M., Stone, R.A., Laiyemo, A.O., Park, Y., *et al.* (2012) Health Status, Neighborhood Socioeconomic Context, and Premature Mortality in the United States: The National Institutes of Health-AARP Diet and Health Study. *American Journal of Public Health*, **102**, 680-688. <http://dx.doi.org/10.2105/AJPH.2011.300158>
- [12] Foraker, R.E., Patel, M.D., Whitsel, E.A., Suchindran, C.M., Heiss, G. and Rose, KM. (2013) Neighborhood Socioeconomic Disparities and 1-Year Case Fatality after Incident Myocardial Infarction: The Atherosclerosis Risk in Communities (ARIC) Community Surveillance (1992-2002). *American Heart Journal*, **165**, 102-107. <http://dx.doi.org/10.1016/j.ahj.2012.10.022>
- [13] Zhang-Salomons, J., Qian, H., Holowaty, E. and Mackillop, W.J. (2006) Associations between Socioeconomic Status and Cancer Survival: Choice of SES Indicator May Affect Results. *Annals of Epidemiology*, **16**, 521-528. <http://dx.doi.org/10.1016/j.annepidem.2005.10.002>
- [14] Diez-Roux, A.V., Kiefe, C.I., Jacobs Jr., D.R., Haan, M., Jackson, S.A., Nieto, F.J., *et al.* (2001) Area Characteristics and Individual-Level Socioeconomic Position Indicators in Three Population-Based Epidemiologic Studies. *Annals of Epidemiology*, **11**, 395-405. [http://dx.doi.org/10.1016/S1047-2797\(01\)00221-6](http://dx.doi.org/10.1016/S1047-2797(01)00221-6)
- [15] Messer, L.C., Laraia, B.A., Kaufman, J.S., Eyster, J., Holzman, C., Culhane, J., *et al.* (2006) The Development of a Standardized Neighborhood Deprivation Index. *Journal of Urban Health*, **83**, 1041-1062. <http://dx.doi.org/10.1007/s11524-006-9094-x>
- [16] Lian, M., Schootman, M., Doubeni, C.A., Park, Y., Major, J.M., Torres Stone, R.A., *et al.* (2011) Geographic Variation in Colorectal Cancer Survival and the Role of Small-Area Socioeconomic Deprivation: A Multilevel Survival Analysis of the NIH-AARP Diet and Health Study Cohort. *American Journal of Epidemiology*, **174**, 828-838. <http://dx.doi.org/10.1093/aje/kwr162>
- [17] Feinstein, A.R. and Cicchetti, D.V. (1990) High Agreement But Low Kappa: I. The Problems of Two Paradoxes. *Journal of Clinical Epidemiology*, **43**, 543-549. [http://dx.doi.org/10.1016/0895-4356\(90\)90158-L](http://dx.doi.org/10.1016/0895-4356(90)90158-L)
- [18] Landis, J.R. and Koch, G.G. (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics*, **33**, 159-174. <http://dx.doi.org/10.2307/2529310>