

Area Characteristics and Individual-Level Socioeconomic Position Indicators in Three Population-Based Epidemiologic Studies

A.V. DIEZ ROUX, MD, PHD, C.I. KIEFE, MD, PHD, D.R. JACOBS, Jr, PHD,
M. HAAN, MPH, DRPH, S. A. JACKSON, PHD, F.J. NIETO, MD, PHD, C.C. PATON, MSPH,
AND R. SCHULZ, PHD

PURPOSE: There is growing interest in incorporating area indicators into epidemiologic analyses. Using data from the 1990 U.S. Census linked to individual-level data from three epidemiologic studies, we investigated how different area indicators are interrelated, how measures for different sized areas compare, and the relation between area and individual-level social position indicators.

METHODS: The interrelations between 13 area indicators of wealth/income, education, occupation, and other socioenvironmental characteristics were investigated using correlation coefficients and factor analyses. The extent to which block-group measures provide information distinct from census tract measures was investigated using intraclass correlation coefficients. Loglinear models were used to investigate associations between area and individual-level indicators.

RESULTS: Correlations between area measures were generally in the 0.5–0.8 range. In factor analyses, six indicators of income/wealth, education, and occupation loaded on one factor in most geographic sites. Correlations between block-group and census tract measures were high (correlation coefficients 0.85–0.96). Most of the variability in block-group indicators was between census tracts (intraclass correlation coefficients 0.72–0.92). Although individual-level and area indicators were associated, there was evidence of important heterogeneity in area of residence within individual-level income or education categories. The strength of the association between individual and area measures was similar in the three studies and in whites and blacks, but blacks were much more likely to live in more disadvantaged areas than whites.

CONCLUSIONS: Area measures of wealth/income, education, and occupation are moderately to highly correlated. Differences between using census tract or block-group measures in contextual investigations are likely to be relatively small. Area and individual-level indicators are far from perfectly correlated and provide complementary information on living circumstances. Differences in the residential environments of blacks and whites may need to be taken into account in interpreting race differences in epidemiologic studies.

Ann Epidemiol 2001;11:395–405. © 2001 Elsevier Science Inc. All rights reserved.

KEY WORDS: Socioeconomic Status, Social Class, Neighborhoods, Race, Ethnicity.

From the Division of General Medicine (AVDR), Columbia College of Physicians and Surgeons, and Division of Epidemiology, Joseph L. Mailman School of Public Health, Columbia University, New York, NY; CARDIA Coordinating Center and Center for Outcomes and Effectiveness Research and Education (CIK), University of Alabama at Birmingham, Birmingham, AL; Division of Epidemiology (DRJ), University of Minnesota School of Public Health, Minneapolis, MN; Center for Aging and Health (MH), Department of Epidemiology & Preventive Medicine, University of California School of Medicine, San Francisco, CA, USA; Department of Public Health Sciences (SAJ), Wake Forest University School of Medicine, Winston-Salem, NC, USA; Department of Epidemiology (FJN), The Johns Hopkins University School of Hygiene and Public Health, Baltimore, MD, USA; Health and Human Services (CCP), Injury and Violence Prevention Unit, Raleigh, NC, USA; University Center for Social and Urban Research (RS), School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA.

Address correspondences to: Ana V. Diez Roux M.D., Ph.D., Division of General Medicine, Columbia Presbyterian Medical Center, 622 West 168th Street, PH9 East Rm 105, New York, NY 10032. Tel.: 212-305-5097; Fax: 212-305-9349. E-mail: ad290@columbia.edu

Received 24 April 2000; revised 23 January 2001; accepted 31 January 2001

There has been growing interest in the use of area-level measures of socioeconomic characteristics in studies of social inequalities and health (1–3). Area-based measures have been used in ecological studies relating area characteristics to morbidity and mortality rates (4–7). Area-based measures have also been proposed as alternate indicators of socioeconomic position in studies of individuals when individual-level measures are unavailable (8,9). In addition, area measures have been increasingly studied as indicators of contextual area characteristics that may be related to health independent of individual level variables (9–23).

In the United States, health researchers using area-based indicators have often relied on existing sources such as the United States Census. Traditionally, the census areas most commonly employed in health research have been census tracts (13–15,18,24–28) or groups of census tracts (6,7,11,19,20,29–31). Block-groups (subdivisions of census tracts) have also been used (8,9,16). A variety of census variables have been used to characterize areas at both the

Selected Abbreviations and Acronyms

CARDIA = Coronary Artery Disease Risk Development in Young Adults Study
 ARIC = Atherosclerosis Risk in Communities Study
 CHS = Cardiovascular Health Study
 ICC BG = Intraclass correlation coefficient for multiple measures on a block-group
 ICC CT = Intraclass correlation coefficient for multiple block-groups within a census tract
 (Because these abbreviations are nonstandard, they are spelled out on their first usage.)

census tract and block-group level, including aggregate measures of education, income, poverty, and occupation.

The interrelations between area indicators has been a topic of much interest to geographers, sociologists, and demographers (32-35). However, despite the growing use of area measures in epidemiologic research, there is little information in the U.S. public health literature on how indicators measured for different sized areas are interrelated and on the relationship between area measures and individual-level social position indicators. We investigated these questions using area data from the 1990 U.S. Census linked to individual-level data from three large epidemiologic studies. Specifically, we describe: i) the interrelations between multiple indicators of area characteristics available in the U.S. census as well as the extent to which block-group and census tract measures provide distinct information; and ii) the associations between traditional individual-level indicators used in epidemiologic studies and area measures. In addition, because the three studies included participants from two race/ethnic groups, we also investigated whether the interrelation between individual and area indicators varied by race/ethnicity.

METHODS
Sources of Data

The data used in these analyses came from the 1990 U.S. Census and from three population-based epidemiologic studies: the Coronary Artery Disease Risk Development in Young Adults (CARDIA) Study, the Atherosclerosis Risk in Communities (ARIC) Study, and the Cardiovascular Health Study (CHS). The CARDIA sample consisted of persons aged 18-30 years at baseline selected by probability sampling in Birmingham, AL; Chicago, IL; Minneapolis MN; and Oakland, CA (36). Black and white participants were recruited at each field center. Because earlier addresses were unavailable, the addresses and individual-level socioeconomic indicators used in these analyses were obtained from the year 10 follow-up (1995-1996). Retention of the cohort at year 10 was 79%.

The ARIC sample consisted of persons aged 45 to 64 years at baseline selected by probability sampling in Forsyth County, NC; Jackson, MS; the northwestern suburbs of Minneapolis, MN; and Washington County, MD (37). Three samples reflect the demographic composition of the communities (virtually all white in Washington County and suburbs of Minneapolis and 85% white in Forsyth County). The Jackson sample was entirely black. The addresses and socioeconomic data used in these analyses were obtained at baseline (1987-1989). The CHS sample consisted of persons aged 65 years or older at baseline selected from Medicare eligibility lists in Forsyth County, NC; Washington County, MD; Sacramento County, CA; and Pittsburgh, PA. (38,39). Three of the four field centers recruited black and white participants. The Washington County sample was virtually all white. Because earlier addresses were not available, addresses and individual-level socioeconomic data used in these analyses correspond to the latest information available on file in early 1998 (4-8 years after baseline).

Study participants were asked to select their total combined family income from a list of categories and to report the highest grade or year of school completed. Income was categorized into six groups (under \$12,000; 12,000-15,999; 16,000-24,999; 25,000-34,999; 35,000-49,999; 50,000 or more). Education was categorized into three groups: incomplete high school, complete high school or GED but no college degree, and complete 4-year college. Occupation was not collected in a comparable fashion across studies.

Two census-defined areas were investigated: census tracts and block groups. Census tracts are subdivisions of a county with an average size of 4000 residents. When first delineated, they were designed to be homogeneous with respect to economic status and living conditions. In non-metropolitan areas, census tracts are replaced by block numbering areas (BNAs). Each census tract(or BNA) is subdivided into block-groups (average size 1000 individuals) (40,41). Study participants were geocoded to their area of residence by a commercial firm.

Information on census tracts and block-groups in the CARDIA, ARIC, and CHS sites was obtained from the 1990 census. Selected variables were chosen a priori from available census measures to reflect the constructs of area income and wealth, education, occupation and employment (42,43), and socioenvironmental characteristics related to area crowding, stability, and housing. The 13 variables investigated (for block-groups and census tracts) are listed in Table 1.

CARDIA, ARIC, and CHS sites included a total of 1910 census tracts and 5964 block-groups. Census tracts and block-groups were excluded if they had populations of less than 100 persons, had less than 30 housing units, or had 33% or more of their population living in group quar-

TABLE 1. Mean block-group characteristics by quintiles of block-group summary score ($n = 5494$)^a

	Lowest quintile (−11.6 to −4.39)	Quintile II (−4.4 to −1.79)	Quintile III (−1.8 to 0.69)	Quintile IV (0.7 to 3.89)	Highest quintile (3.9 to 16.7)
Income/wealth					
Median household income (U.S.\$)	14,600	22,800	28,200	33,000	45,700
Percentage persons below poverty	40	22	13	9	6
Percentage households receiving interest, dividend, or net rental income ^b	12	23	35	45	60
Median value of owner-occupied housing units (U.S.\$)	46,100	61,500	79,400	105,200	193,300
Percentage of housing units that are owner occupied	35	50	59	61	65
Education					
Percentage adults 25 years or older with complete high school	48	62	72	81	92
Percentage adults 25 years or older with complete college	5	9	14	23	49
Occupation/employment					
Percentage employed persons 16 years of age and older in executive, managerial or professional occupations ^c	10	16	21	28	48
Percentage unemployment ^d	20	13	8	6	4
Other socioenvironmental					
Percentage of housing units unoccupied	13	9	6	5	5
Percentage of housing units boarded up ^e	2	1	0	0	0
Percentage occupied housing units with > one person per room	13	10	6	4	2
Percentage persons living in same house for last 5 years	57	57	57	55	53

^aBlock-group score constructed by summing Z-scores for block-group log median household income; percentage households receiving interest, dividend, or net rental income; log median value of housing units; percentage adults with complete high school; percentage adults with complete college; and percentage persons in executive, managerial, or professional specialty occupations. Z-scores and quintiles of score based on distribution observed across all three studies.

^bIncludes interest on savings or bonds, dividends from stockholdings or membership in associations, net income from rental of property to others, and receipts from boarders or lodgers, net royalties, and periodic payments from an estate or trust fund.

^cCorresponds to top category of the six summary occupational categories used in the 1990 Census (42,43).

^dAmong civilians 16 years and older in labor force. Persons in labor force include employed (at work or with a job but temporarily away) and unemployed (no job but looking for work in the last 4 weeks and available to accept a job).

^eBoarded-up status was obtained for all vacant housing units. Boarded-up units have windows and doors covered by wood, metal, or masonry to protect the interior and to prevent entry into the building.

ters (such as prisons, nursing homes, college dorms, etc.). This left 1772 census tracts and 5494 block groups within them for analysis.

Of the 3950 CARDIA participants attending the year 10 follow-up, 3531 matched to block-groups and census tracts that fulfilled the inclusion criteria (because nearly 50% of CARDIA participants no longer lived in the study sites at the year 10 follow-up, persons residing outside study sites were included as long as they matched to areas not meeting the exclusion criteria). Two participants were excluded, because they had no information on individual-level socioeconomic indicators, leaving 1797 whites and 1732 blacks for analysis. Of the 15,792 ARIC participants at baseline, 14,158 matched to nonexcluded census areas. Fifty participants were excluded, because they lacked information on socioeconomic indicators or belonged to race/ethnic groups other than black or white, leaving 10,229 whites and 3879 blacks for analysis. Of the 5888 CHS participants at baseline, 4553 matched to nonexcluded census areas. Twenty-nine participants were excluded, because they lacked information on socioeconomic indicators or be-

longed to race/ethnic groups other than black or white, leaving 3804 whites and 720 blacks for analysis.

Statistical Methods

Interrelations between area indicators and comparison of census-tract and block-group measures. Analyses of area measures were based on all census tracts and block-groups included in study sites, regardless of whether they included a sampled participant. After examination of correlations, factor analysis was used as an exploratory technique to identify subsets of variables that it would be meaningful to combine into a summary score (44). Variables with extremely skewed distributions (skewness > 1.5) (45) (median household income, median house value, unemployment, percentage below poverty, percentage of housing units boarded up, crowding, and percentage housing units occupied) were log transformed. Transformation of skewed variables increases consistency of the index over time or across geographic areas (46). Scree plot inspection of principal components was used to determine the number of fac-

tors to be extracted (47,48). Principal factor analysis with varimax orthogonal rotation was used for factor extraction (47,48). Analyses were performed for block-groups, and census tracts and were stratified by site.

Agreement between indicators for census tracts and block-groups was investigated by comparing measures for block-groups to measures for the census tract to which each block group belonged. Corresponding census tract measures were appended to each block-group. We examined Spearman's correlation coefficients between measures at both levels and intraclass correlation coefficients (49) for multiple measures on a block-group (ICC BG), the multiple measures being the block-group specific measure and the corresponding census tract measure. The Spearman's correlation measures the extent to which the ranking of block-groups based on block-group and census tract measures agree. The intraclass correlation coefficient (where "classes" are individual block groups) measures the extent to which there is absolute agreement between the block-group measure and the corresponding census tract measure. A value of ICC BG close to 1 signifies strong absolute agreement between the block-group measure and the census tract measure. Because block-groups are nested within census tracts, we also estimated the intraclass correlation coefficient for multiple block-groups within a census tract (ICC CT). Here the multiple measures are the measures on different block-groups within a census tract. This intraclass correlation coefficient (where "classes" are census tracts) is an estimate of the proportion of total variability in block-group indicators that is between census tracts (50). ICC CTs close to 1 imply that there is relatively little variability in block-groups measures within census tracts (i.e., strong agreement between measures for multiple block-groups within a census tract).

Relation between individual-level and area indicators.

The relation between individual-level and area indicators was investigated by examining the cross-classification of individual-level indicators and categories of area indicators. Loglinear models were used to obtain a summary measure of the strength of the association between individual-level indicators and area score category (51,52). The logarithm of the count in each cell of the individual social indicator-area quintile cross classification was modeled as a function of the individual-level indicator, area quintile, and their interaction. Individual-level indicators were included in the models as ordinal variables (1-6 for income, and 1-3 for education), and area quintiles were included as categorical variables. This allowed estimation of the odds ratio of being in each of the four lowest area quintiles (compared to the most advantaged quintile) per each unit decrease in individual level income or educational category (e.g., the percentage increase in the odds of living in the lowest area

quintile per one category decrease in individual-level income). The assumption of linearity of the log odds by categories of income or education was examined by fitting models with income or education as categorical variables and inspecting the regression coefficients for each category. Linearity was not found to be seriously violated. There was no evidence of gender differences in the association between individual and area indicators. However, gender, and its interactions with individual indicators and area category, were included to control for gender differences in the distribution of the socioeconomic indicators. Models were run separately by race/ethnicity and study.

To investigate differences in the likelihood of living in different types of areas by race/ethnicity, log linear models were also used to estimate the relative odds of being in each area category (compared to the most advantaged area category) in blacks versus whites within categories of income or education. The log of the count of the race*gender*area category cross classification was modeled as a function of race, gender, area category, and the gender*area and race*area interactions. These models allowed estimation of the relative odds of living in each area category (compared to the best-off category) in blacks versus whites within categories of individual-level income or education. Because there was no evidence of heterogeneity by income or education, associations between race and area category were also estimated for all income or educational categories combined (after adjusting for income or education). Results regarding associations between area and individual-level indicators were similar for census tracts and block groups, so only results for block-groups are reported.

RESULTS

Interrelations Between Census Indicators and Comparison of Census Tract and Block-Group Measures

Means and medians for the census indicators were similar for census tracts and block-groups (not shown), although the range (1st-99th percentiles) was slightly greater for block-groups. At the block-group level, wealth/income, education, and occupation variables tended to be fairly highly correlated (Spearman's correlations (r): 0.5-0.8). However, percentage of housing units that were owner occupied was less strongly associated with education and occupation (most $r < 0.4$) and was weakly correlated with house value ($r = 0.1$). The stability indicator was generally weakly correlated with all variables (all $r < 0.2$) except percentage of owner occupied houses ($r = 0.5$). Correlations between socioenvironmental indicators (housing units occupied, boarded up, and crowding), and all other indicators were in the intermediate range (0.3-0.6). Similar patterns were observed for census tracts; although, as expected, correlations

for census tracts were generally slightly stronger than those observed for block-groups (not shown).

Although results of factor analyses were not identical across sites, common patterns emerged. Scree plot examination revealed that the extraction of two components was appropriate in most cases. These two components accounted for 65–80% of the total variance in the 13 original variables across areas. In factor analyses, six variables related to wealth/income, education and occupation (log median household income, log median value of housing units, percentage of households receiving interest, dividend, or net rental income, percentage of adults with complete high school, percentage of adults with complete college, and percentage of persons in executive, managerial, or professional occupations) loaded strongly (loadings ≥ 0.6) on the first factor in most sites. The stability indicator and percentage owner-occupied houses generally loaded strongly (factor loadings ≥ 0.6) on the second factor. Similar patterns were documented for census tracts, although additional factors, such as percentage below poverty, crowding, percentage houses boarded up, and unemployment, often loaded on the first factor as well.

A summary score of area socioeconomic characteristics (reflecting patterns common across area sizes and sites), was constructed by summing Z-scores for the six variables that usually loaded together on the first or socioeconomic-position related factor, with an increasing index signifying increasing socioeconomic advantage. Z-scores were based on the mean and standard deviation of the variables for all sites combined. The internal consistency (Cronbach's alpha) of the area score for the entire sample was 0.94 for census tracts and 0.92 for block-groups. Internal consistency was high (> 0.90) across all sites. Table 1 summarizes block-group characteristics by quintiles of block-group score. Similar patterns (not shown) were observed for categories based on census tract scores.

Overall, census tract and block-group measures tended to be highly correlated (Spearman's correlation coefficients ≥ 0.85). Absolute agreement between block-group measures and the corresponding census tract measure were also high (all ICC BG ≥ 0.87). Eighty-six percent of the variation in block-group score was between census tracts (with ICC CT ranging from 0.71 to 0.92 for score components) (Table 2).

Relation Between Individual-Level and Area Indicators

Figure 1 shows the distribution of categories of block-group score (based on quintiles) within categories of individual-level income. The percentage of persons living in more disadvantaged block-groups tended to decrease as individual-level income increased. However, there was also evidence of heterogeneity in block-group score within categories of individual-level income. Blacks were much more likely than whites to live in disadvantaged areas, regardless of their in-

TABLE 2. Measures of agreement between census tract and block-group indicators and intraclass correlation coefficients for block groups within census tracts for the area score and its components^a

	Spearman's correlation	ICC BG ^b	ICC CT ^c
Area score	0.94	0.94	0.86
Log median household income	0.88	0.87	0.72
Percentage households w/ interest, dividend, or rental income	0.91	0.91	0.76
Log median value of housing units	0.96	0.96	0.92
Percentage persons 25 or over with complete high school	0.91	0.90	0.76
Percentage persons 25 or over with complete college	0.90	0.93	0.84
Percentage persons in executive, managerial, or professional specialty occupations	0.85	0.88	0.71

^aBased on 5494 block-groups nested within 1,772 census tracts.

^bICC BG: intraclass correlation coefficient for multiple measures on a block-group, the multiple measures being the block-group specific measure and the measure for the census tract to which the block-group belongs. A value of this intraclass correlation coefficient close to one signifies strong absolute agreement between the block-group measure and the corresponding census tract measure for that block-group.

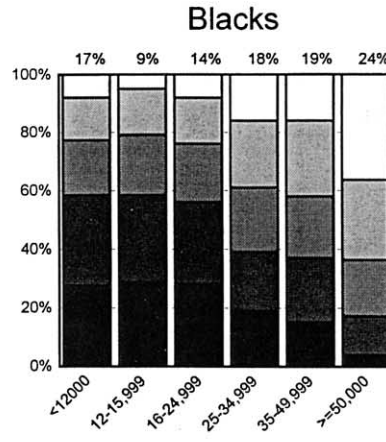
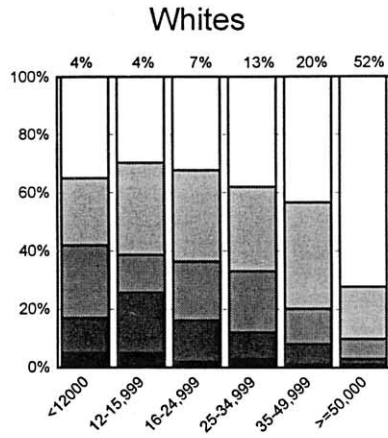
^cICC CT: intraclass correlation coefficient for block-groups nested within census tracts. Here, the multiple measures are the measures on different block-groups within a census tract. This intraclass correlation coefficient is an estimate of the proportion of total variability in block-group indicators that is between census tracts. Intraclass correlation coefficients close to one imply that there is relatively little variability in block-groups measures within census tracts (i.e., strong agreement between measures for multiple block-groups within a census tract).

come. Similar differences by race/ethnicity were present in men and women (data not shown). Similar patterns were observed for individual-level education categories.

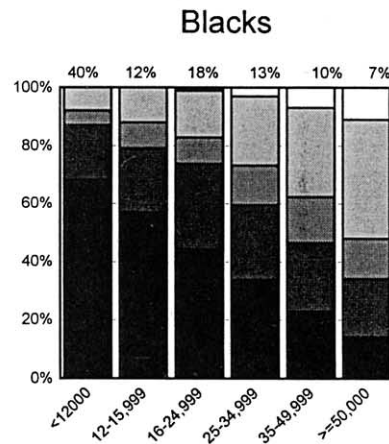
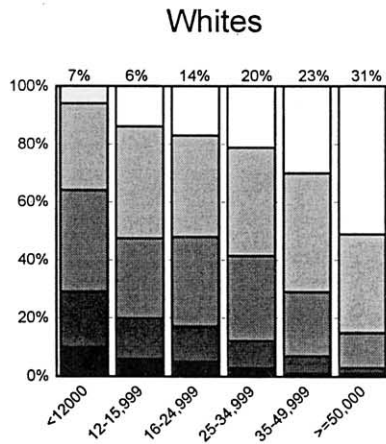
Table 3 shows the relation between individual-level and area-level income measures (income is shown, because it can be similarly categorized at both levels). There is substantial variability in block-group income within categories of individual-level income. Table 3 also illustrates the degree of misclassification that would occur if block-group median income were used as a proxy for individual-level income (or vice versa). For example, over 90% of white ARIC participants with individual-level incomes below \$16,000 lived in block groups where the median income was \$16,000 or more. The degree of misclassification resulting from the use of area measures as proxies for individual-level measures would differ by race and income level.

Table 4 shows the odds ratios of being in each block-group quintile (vs. the top quintile or best-off category) per unit decrease in individual-level income. Very similar results were obtained for individual-level education (not shown). The strength of the association between block-group score and individual income or education was generally similar across studies and in both race ethnic groups.

CARDIA



ARIC



CHS

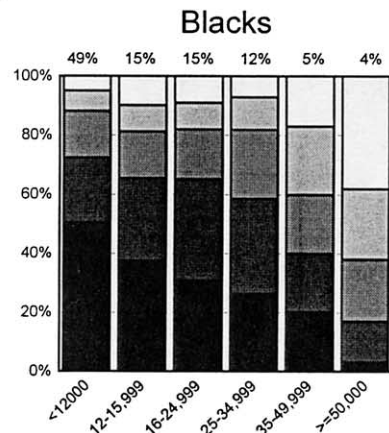
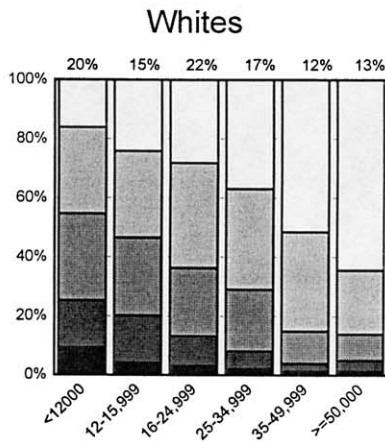


FIGURE 1. Distribution of categories of block group score (based on quintiles) by individual-level income (U.S. \$). Each bar shows how persons of a given individual-level income are distributed across the five quintiles of block-group. Shaded areas within bars correspond to neighborhood quintiles from lowest quintile (darkest) to highest quintile (lightest). For example, as individual-level income increases, the percentage of persons in the best-off neighborhood category (white in the bars) increases. Numbers across the top show the percentage of the total sample in each of the individual-level income categories.

TABLE 3. Percentage distribution of block group income within categories of individual-level income

Individual-level income (U.S.\$)	Study and race/ethnicity	Block group median household income (US\$) (percent distribution)		
		<\$16,000	\$16–34,999	≥\$35,000
<\$16,000	CARDIA			
	Whites (n = 140)	7.9	64.3	27.9
	Blacks (n = 445)	28.1	62.2	9.7
	ARIC			
	Whites (n = 1199)	3.3	65.3	31.4
	Blacks (n = 1824)	55.7	40.7	3.7
CHS	Whites (n = 1,308)	5.1	66.6	28.3
	Blacks (n = 452)	44.3	48.9	6.9
\$16-34,999	CARDIA			
	Whites (n = 363)	3.9	58.7	37.5
	Blacks (n = 538)	19.0	65.4	15.6
	ARIC			
	Whites (n = 3,293)	0.9	54.4	44.7
	Blacks (n = 1,059)	31.4	58.2	10.4
CHS	Whites (n = 1,466)	1.8	55.1	43.2
	Blacks (n = 188)	25.5	61.2	13.3
≥ \$35,000	CARDIA			
	Whites (n = 1,277)	1.2	29.8	69.0
	Blacks (n = 725)	8.0	52.0	40.0
	ARIC			
	Whites (n = 5,289)	0.2	33.6	66.2
	Blacks (n = 610)	17.4	55.2	27.4
CHS	Whites (n = 917)	1.1	41.9	57.0
	Blacks (n = 64)	15.6	53.1	31.3

For example, the odds ratio of living in the lowest block-group quintile (vs. the highest) associated with a unit decrease in income category ranged from 2.1 to 2.4 in whites and from 2.0 to 2.9 in blacks. Table 5 shows the odds ratios of living in each block-group quintile (as compared to the highest quintile) in blacks versus whites. The strength of these associations did not differ significantly across income strata, so only income-adjusted results are shown. Blacks were much more likely than whites to live in disadvantaged block-groups, even after adjustment for individual-level income. Black–white differences were most pronounced in ARIC and least pronounced (but still substantial) in CARDIA. Similar patterns were observed after adjustment for education (not shown).

DISCUSSION

In these samples, correlations between block-group measures of income/wealth, education, and occupation were generally in the 0.5–0.8 range. Slightly higher correlations were observed for census tract measures. Two factors, reflecting a socioeconomic dimension and a residential stability dimension, emerged in factor analyses of the 13 variables. We found generally high agreement between

continuous block-group and census tract measures of the same construct, and most of the variability in block-group measures was between census tracts. Although area-based and individual-level socioeconomic position measures were associated, there was evidence of important heterogeneity in area of residence within individual-level income or education categories. Regardless of their income or education, blacks participating in the studies were much more likely to live in more disadvantaged areas than their white counterparts.

The emergence of a socioeconomic factor and a residential stability factor in factor analyses of area data is consistent with results reported by others in the United States (53–56). We combined the six area variables that loaded strongly on the first factor into a summary score of area socioeconomic environment. The need for (and components of) such a summary score depend on the objective being pursued (57–59). Our objective was to develop a summary of area socioeconomic characteristics that could be used in analyses of these data involving the construct of area socioeconomic environment. This type of summary may be useful when interest lies in the causal investigation of the contextual effect of area socioeconomic environment on health generally, when all potential indicators are correlated, and when there is no theoretical reason to choose

TABLE 4. Odds ratios of being in each block-group quintile (compared to the most advantaged block-group quintile) per unit decrease in individual-level income^a

	Odds ratios (95% CI)	
	Whites	Blacks
CARDIA		
QV	1.0	1.0
QIV	1.5 (1.4-1.7)	1.3 (1.2-1.4)
QIII	1.8 (1.6-2.0)	1.5 (1.3-1.7)
QII	2.1 (1.9-2.4)	1.8 (1.6-2.0)
QI	2.1 (1.7-2.6)	2.0 (1.8-2.2)
ARIC		
QV	1.0	1.0
QIV	1.4 (1.4-1.5)	1.5 (1.3-1.7)
QIII	1.7 (1.7-1.8)	1.6 (1.4-1.9)
QII	2.1 (2.0-2.2)	2.0 (1.7-2.4)
QI	2.4 (2.2-2.6)	2.9 (2.5-3.4)
CHS		
QV	1.0	1.0
QIV	1.3 (1.3-1.4)	1.2 (0.9-1.4)
QIII	1.6 (1.5-1.7)	1.4 (1.2-1.7)
QII	1.8 (1.7-2.0)	1.5 (1.2-1.8)
QI	2.2 (1.9-2.5)	2.1 (1.7-2.5)

^aThe odds ratios shown correspond to the relative change in the odds of being in the corresponding block-group quintile (vs. the most advantaged quintile, quintile V) per unit decrease in individual-level income. For example, the odds ratio of 2.1 for QI in CARDIA whites implies that the odds of living in block-groups in the worst quintile (vs. quintile V) increased 2.1 times per each unit decrease in income category. CI: Confidence interval; QI-QV: first through fifth quintile of block-group score.

one indicator over another. However, the dependency of factor analytic results on the set of variables initially included and on correlations in a particular sample may limit the generalizability of our results. The development of a valid summary score that could be used in the United States as a whole (or in other samples) and with other specific purposes would require different analyses.

Although census tracts were originally delineated to be homogeneous in socioeconomic characteristics (60), it has been suggested that their use may obscure important variations within them (8,9,61,62). We found high agreement between block-group and census tract measures, and the variability in block-group measures within census tracts was small in relation to between tract variability. Although the actual empirical consequences of using census tract or block-group measures in the investigation of contextual effects can only be directly determined in studies relating these measures to health outcomes, our analyses suggests that differences in results using census tracts or block-groups are likely to be relatively small. However, some variability in block-groups within census tracts remains. The extent to which block-groups provide information that is not captured by census tract measures may differ for different area characteristics or across larger contexts (e.g., cities,

TABLE 5. Odds ratios (95% confidence intervals) of being in each block-group quintile (compared to quintile V) in blacks vs. whites by study^a

	Odds ratios blacks vs. whites
CARDIA	
QV	1.0
QIV	2.6 (2.2-3.2)
QIII	4.8 (3.9-6.0)
QII	9.7 (7.5-12.5)
QI	26.5 (17.6-40.1)
ARIC	
QV	1.0
QIV	5.2 (4.1-6.5)
QIII	4.2 (3.2-5.4)
QII	27.9 (21.7-36.0)
QI	129.0 (98.1-169.8)
CHS	
QV	1.0
QIV	1.2 (0.8-1.7)
QIII	3.0 (2.1-4.2)
QII	9.0 (6.5-12.6)
QI	36.2 (25.5-51.6)

^aThe odds ratios correspond to the odds of living in each quintile (vs. the most advantaged quintile) in blacks vs. whites. For example, the odds ratio of 26.5 for QI in CARDIA participants means that the odds of living in QI (vs. V) was 26.5 times greater in blacks than in whites. Odds ratios are adjusted for individual-level income. QI-QV: first through fifth quintile of block-group score.

regions). Therefore, in some situations, and particularly if the size of the area whose characteristics are hypothesized to be related to the outcome is closer to block-groups than census tracts for theoretical reasons, the use of block-group measures still may be appropriate.

Previous methodological studies have focused on the validity of using area-based measures as proxies for individual-level indicators (63-65) and have compared the strengths of the associations of both types of indicators with health outcomes (8,9,63-67). We directly examined the relation between area and individual-level measures as indicators of two interrelated but different constructs. The presence of variability in individual-level indicators within areas is the source of the limitation inherent in using area measures as proxies for their individual-level analogues. On the other hand, the variability in characteristics of area of residence within categories of individual-level socioeconomic indicators suggests that area variables may provide information on living circumstances, which is not captured by individual-level data. Our results highlight the limitations of using measures at one level as proxies for measures at another level in causal epidemiologic investigations and suggest that it may often be analytically possible to separate out the contributions of measures at both levels to outcomes.

The strength of the association between individual and area measures (in relative terms; i.e., as quantified by the rel-

ative odds) was generally similar in all three studies and in both race/ethnic groups. However, there were large differences in area characteristics by race/ethnicity. The sampling schemes employed may have magnified these differences. In ARIC, blacks were drawn predominantly from one field center that generally represented more disadvantaged areas. In CARDIA, the types of subjects recruited and the areas from which they were recruited changed over time in order to comply with study quotas (68). Selection bias may also have resulted from nonparticipation and losses to follow-up. If these factors resulted in selection of black and white participants differentially with regard to their areas of residence, our estimates of black–white differences could be biased. Similarly, if participation and follow-up rates resulted in selection of participants based on *both* individual-level indicators and area characteristics, our estimates of associations between these indicators may be biased. For example, if selection factors resulted in the inclusion of high-income persons from high-income areas and low-income persons from low-income areas, our results may overestimate the strength of the association between area and individual-level indicators in the reference population.

Despite these limitations, our results regarding black–white differences in areas of residence are generally consistent with analyses on other samples (69–71). Although differences in the residential environments of blacks and whites in the United States are well-established in the sociological literature (69–71), they have only recently been highlighted in epidemiology. Even if differences are magnified in these samples by selection factors, our results suggest that these area differences may need to be taken into consideration when comparing blacks and whites in the context of “population-based” epidemiologic studies such as these. They also illustrate the limitations of drawing conclusions regarding biological differences between race/ethnic groups in these studies after adjusting for single indicators of socioeconomic position.

One limitation of these analyses pertains to the timing of available addresses. For ARIC, addresses, census variables, and individual-level indicators were closely matched in time (1987–1990). For CARDIA and CHS, addresses and individual-level socioeconomic indicators pertained to the mid- 1990s, but area measures were obtained from the 1990 census. Area characteristics may have changed over the 1990s. However, unless this change was substantial and differential by individual-level characteristics, it is unlikely to have affected the associations between area and individual-level indicators that we observed. Unfortunately it was not possible to validly estimate family size-adjusted income in these studies. However, it is unlikely that failure to adjust for family size would have substantially affected the associations between area and individual-level measures or the race differences that we observed.

The increasing interest in examining contextual area effects on health and the reliance on standard datasets such as the U.S Census to characterize areas highlights the need to examine empirically the relations between indicators available at different levels. Together with the theoretical model being tested and the specific question addressed, these descriptive analyses may inform the selection of variables and the interpretation of results obtained in contextual investigations. A distinct question, however, is the use of area indicators to monitor social inequalities in health generally (57–59,72,73). The implications for monitoring of using measures at different levels, different area-based indicators, and different sized-areas can only be addressed by directly examining the empirical associations of the different indicators with the outcomes of interest.

This work was supported by R29 HL59386 (Dr. Diez-Roux) from the National Heart, Lung, and Blood Institute. CHS was supported by contracts N01-HC-85079—N01-HC-85086 from the National Heart, Lung, and Blood Institute, and Georgetown Echo RC-HL 35129 JHU MRI RC-HL 15103. The ARIC Study was supported by Contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022 from the National Heart, Lung, and Blood Institute. The CARDIA Study was supported by Contracts N01-HC-48047, N01-HC-48048, N01-HC-48049, N01-HC-48050, and N01-HC-95095 from the National Heart, Lung, and Blood Institute. The authors thank the following: Sharon Merkin for assistance with the analyses; the CARDIA investigators and staff, and especially Dan Garside, Karen Virmig, Jerry Hamilton, Debbie Parker, Heather McGrath, and Rex Bomhold for their assistance in geocoding the CARDIA data; the ARIC investigators and staff for their important contributions; and the following CHS investigators and staff: Forsyth County, NC: Bowman Gray School of Medicine of Wake Forest University: Gregory L. Burke, Sharon Jackson, Alan Elster, Curt D. Furberg, Gerardo Heiss, Dalane Kitzman, Margie Lamb, David S. Lefkowitz, Mary F. Lyles, Cathy Nunn, Ward Riley, John Chen, Beverly Tucker; Forsyth County, NC-Wake Forest University-ECG Reading Center: Farida Rautaharju, Pentti Rautaharju; Sacramento County, CA; University of California, Davis: William Bonekat, Charles Bernick, Michael Buonocore, Mary Haan, Calvin Hirsch, Lawrence Laslett, Marshall Lee, John Robbins, William Seavey, Richard White, Washington County, MD, The Johns Hopkins University: M. Jan Busby-Whitehead, Joyce Chabot, George W. Comstock, Adrian Dobs, Linda P. Fried, Joel G. Hill, Steven J. Kittner, Shiriki Kumanyika, David Levine, Joao A. Lima, Neil R. Powe, Thomas R. Price, Jeff Williamson, Moyses Szklo, Melvyn Tockman; MRI Reading Center, Washington County, MD, The Johns Hopkins University: Norman Beauchamp, R. Nick Bryan, Douglas Fellows, Melanie Hawkins, Patrice Holtz, Naiyer Iman, Michael Kraut, Cynthia Quinn, Grace Lee, Carolyn C. Meltzer, Larry Schertz, Earl P. Steinberg, Scott Wells, Linda Wilkins, Nancy C. Yue; Allegheny County, PA, University of Pittsburgh: Diane G. Ives, Charles A. Jungreis, Laurie Knepper, Lewis H. Kuller, Elaine Meilahn, Peg Meyer, Roberta Moyer, Anne Newman, Richard Schulz, Vivienne E. Smith, Sidney K. Wolfson; Echocardiography Reading Center (Baseline), University of California, Irvine: Hoda Anton-Culver, Julius M. Gardin, Margaret Knoll, Tom Kurosaki, Nathan Wong; Echocardiography Reading Center (Follow-Up), Georgetown Medical Center: John Gottdiener, Eva Hausner, Stephen Kraus, Judy Gay, Sue Livengood, Mary Ann Yohe, Retha Webb; Ultrasound Reading Center, New England Medical Center, Boston: Daniel H. O’Leary, Joseph F. Polak, Laurie Funk; Central Blood Analysis Laboratory, University of Vermont: Elaine Cornell, Mary Cushman, Russell P. Tracy; Pulmonary Reading Center, University of Arizona-Tucson: Paul Enright; Coordinating Center, University of Washington, Seattle:

Alice Arnold, Annette L. Fitzpatrick, Richard A. Kronmal, Bruce M. Psaty, David S. Siscovick, Will Longstreth, Patricia W. Wahl, David Yanez, Paula Diehr, Corrine Dulberg, Bonnie Lind, Thomas Lumley, Ellen O'Meara, Jennifer Nelson, Chuck Spiekerman; *NHLBI Project Office*: Robin Boineau, Teri A. Manolio, Peter J. Savage, Patricia Smith.

REFERENCES

- Krieger N, Williams DR, Moss NE. Measuring social class in U.S. public health research: Concepts, methodologies, and guidelines. *Ann Rev Public Health* 1997;18:341-378.
- Moss N, Krieger N. Measuring social inequalities in health. *Pub Health Rept* 1995;110:302-305.
- Syme SL, Moss N, Krieger J. Recommendations of the conference "measuring social inequalities in health." *Int J Health Serv.* 1996; 26:521-527.
- Townsend P, Phillimore P, Beattie A. Health and deprivation. Inequality and the north. London: Routledge; 1988.
- Carstairs V, Morris R. Deprivation and health in Scotland. Aberdeen: Aberdeen University; 1991.
- Wing S, Casper M, Riggan W, Hayes C, Tyroler HA. Socioenvironmental characteristics associated with the onset of decline of ischemic heart disease mortality in the United States. *Am J Publ Health.* 1988;78:923-926.
- Wing S, Barnett E, Casper M, Tyroler HA. Geographic and socioeconomic variation in the onset of decline of coronary heart disease mortality in white women. *Am J Publ Health.* 1992;82:204-209.
- Krieger N. Women and social class: A methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. *J Epidemiol Comm Health.* 1991;45:35-42.
- Krieger N. Overcoming the absence of socioeconomic data in medical records: Validation and application of a census-based methodology. *Am J Publ Health.* 1992;92:703-710.
- MacIntyre S, MacIver S, Sooman A. Area, class, and health: Should we be focusing on places or people? *J Soc Pol.* 1993;22:213-234.
- Haan M, Kaplan G, Camacho T. Poverty and health: Prospective evidence from the Alameda County Study. *Am J Epidemiol.* 1987;125: 989-998.
- Carstairs V, Morris R. Deprivation and mortality. An alternative to social class? *Comm Med.* 1989;11:210-219.
- O'Campo P, Gielen A, Faden R, Xue X, Kass N, Wang M. Violence by male partners against women during the child-bearing year: A contextual analysis. *Am J Publ Health.* 1995;1092-1097.
- O'Campo P, Xue X, Wang M, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: A multilevel analysis. *Am J Publ Health.* 1997;87:1113-1118.
- Anderson RT, Sorlie P, Backlund E, Johnson N, Kaplan GA. Mortality effects of community socioeconomic status. *Epidemiology.* 1997;8: 42-47.
- Diez Roux AV, Nieto FJ, Muntaner C, Tyroler HA, Comstock GW, Shahar E, Cooper LS, Watson RL, Szklo M. Neighborhood environments and coronary heart disease. *Am J Epidemiol.* 1997;146:48-63.
- Hart C, Ecob R, Davey Smith G. People, places, and coronary heart disease risk factors: A multilevel analysis of the Scottish Heart Health Study archive. *Soc Sci Med.* 1997;45:893-902.
- Robert S. Community-level socioeconomic status effects on adult health. *J Health Social Behav.* 1998;39:18-37.
- Yen I, Kaplan G. Poverty area residence and changes in physical activity level. *Am J Publ Health.* 1998;88:1709-1712.
- Yen I, Kaplan G. Neighborhood social environment and risk of death: Multilevel evidence from the Alameda County study. *Am J Epidemiol.* 1999;149:898-907.
- Davey Smith G, Hart C, Watt G, Hole D, Hawthorne V. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: The Renfrew and Paisley study. *J Epidemiol Comm Health.* 1998;52:399-405.
- Yen IH, Syme SL. The social environment and health: A discussion of the epidemiologic literature. *Ann Rev Publ Health.* 1999;20:287-308.
- Robert S. Socioeconomic position and health: The independent contribution of community socioeconomic context. *Ann Rev Sociol.* 1999;25:489-516.
- Dayal HH, Power RN, Chiu C. Race and socioeconomic status in survival for breast cancer. *J Chron Dis.* 1982;35:675-683.
- Centerwall B. Race, socioeconomic status, and domestic homicide, Atlanta, 1971-72. *Am J Publ Health.* 1984;74:813-815.
- Kitagawa EM, Hauser PM. Differential mortality in the United States: A study in socioeconomic epidemiology. Cambridge, MA: Harvard University Press; 1973.
- Yeracaris CA, Jay HK. Socioeconomic differentials in selected causes of death. *Am J Publ Health.* 1978;68:342-351.
- Leclere F, Rogers R, Peters K. Neighborhood social context and racial differences in women's heart disease mortality. *J Health Soc Behav.* 1998;39:91-107.
- Jenkins CD, Tuthill RW, Tannenbaum SI, Kirby C. Social stressors and excess mortality from hypertensive diseases. *J Hum Stress.* 1979; September:29-40.
- Struening EL, Rabkin JG, Cohen P, Raabe G, Muhlin GL, Cohen J. Family, ethnic, and economic indicators of low birth weight and infant mortality: A social area analysis. *Ann NY Acad Sci.* 1973;218: 87-102.
- Roberts EM. Neighborhood social environments and the distribution of low birthweight in Chicago. *Am J Publ Health.* 1997;87:597-603.
- Sampson RJ, Morenoff JD. Ecological perspectives on the neighborhood context of urban poverty. In: Brooks-Gunn J, Duncan GJ, Aber JL, eds. *Neighborhood Poverty: Volume II. Policy Implications in Studying Neighborhoods.* New York: Russell Sage; 1997.
- White MJ. *American Neighborhoods and Residential Differentiation.* New York: Russell Sage; 1987.
- Conference on Small Area Statistics, Montreal, 1972: Social indicators for small areas; papers. [Washington] U.S. Bureau of the Census, 1973.
- Bixhorn H, Midlin A. Composite social indicators: Methodology and results in Washington, DC. In: U.S. Bureau of the Census. *Social Indicators.* Census Tract Papers GE46. Washington DC.: U.S. Government Printing Office; 1973: 3-17.
- Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley S, Jacobs DR, Liu K, Savage PJ. CARDIA: Study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol.* 1988; 41:1105-1116.
- ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol.* 1989;129: 687-702.
- Fried LP, Borhani NO, Enright P, Furberg CD, Julius JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, O'Leary DH, Psaty B, Rautaharju P, Tracy RP, Weiler PG. for the CHS Research Group. The Cardiovascular Health Study: Design and rationale. *Ann Epidemiol.* 1991;1:263-276.
- Tell GT, Fried LP, Hermanson B, Manolio T, Newman AB, Borhani NO. Recruitment of adults 65 years and older as participants in the Cardiovascular Health Study. *Ann Epidemiol.* 1993;3:358-366.
- U.S. Bureau of the Census. Summary Tape File 1 on CD-Rom Technical Documentation. Washington, DC: U.S. Department of Commerce, Economics, and Statistics; 1992.
- U.S. Bureau of the Census. Summary Tape File 3 on CD-Rom Technical Documentation. Washington, DC: U.S. Department of Commerce, Economics, and Statistics; 1992.
- U.S. Bureau of the Census. Classified Index of Industries and Occupations. C3.223/22:990-CPHR-4; 1990.

43. U.S. Bureau of the Census. Alphabetical Index of Industries and Occupations. C3.223/22:990-CPHR-3; 1990.
44. Folwell K. Single measures of deprivation. *J Epidemiol Comm Health.* 1995;49 (Suppl 2):S51–S56.
45. SAS Institute Inc. SAS Procedures Guide, Version 6, Third Ed. Cary, NC: SAS Institute Inc.; 1990; 705 pp.
46. Gilthorpe MS. The importance of normalization in the construction of deprivation indices. *J Epidemiol Comm Health.* 1995;49 (Suppl 2): S45–S50.
47. Kline P. *An Easy Guide to Factor Analysis.* New York: Routledge; 1994.
48. DeVellis RF. *Scale development: Theory and applications.* Newbury Park: Sage; 1991.
49. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health statistics measures: Statistics and strategies for evaluation. *Control Clin Trials.* 1991;12:142S–158S.
50. Snijders T, Bosker R. *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage; 1999.
51. Stokes ME, Davis CS, Koch GC. *Categorical data analysis using the SAS system.* Cary, NC: SAS Institute Inc.; 1995: 499 pp.
52. SAS Institute Inc. *SAS/STAT Software: Changes and enhancements through Release 6.12.* Cary, NC: SAS Institute Inc., 1997; 705 pp.
53. Kasl SV, Harburg E. Mental health and the urban environment: Some doubts and second thoughts. *J Health Soc Behav.* 1975;16:268–282.
54. Johnston RJ. Residential area characteristics: Research methods for identifying urban subareas—Social area analysis and factorial ecology. In: Herbert DT, Johnston RJ, eds. *Spatial Processes and Form, Vol. I.* New York: Wiley; 1976:193–236.
55. Sampson R, Raudenbush S, Earls F. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science.* 1997;277: 918–924.
56. Harburg E, Erfurt JC, Chape C, Hauenstein LS, Schull WJ, Schork MA. Socioecological stressor areas and black–white blood pressure: Detroit. *J Chron Dis.* 1973;26:595–611.
57. Carstairs V. Deprivation indices: Their interpretation and use in relation to health. *J Epidemiol Comm Health.* 1995;49 (Suppl 2):S3–S8.
58. Morris R, Carstairs V. Which deprivation? A comparison of selected deprivation indices. *J Publ Health Med* 1991;13:318–326.
59. Lee P, Murie A, Gordon D. *Area Measures of Deprivation: A Study of Current Methods and Best Practices in the Identification of Poor Areas in Great Britain.* Birmingham, UK: The University of Birmingham; 1995.
60. Kaplan CP, Van Valey TL. *Census 1980: Continuing the Factfinder Tradition.* Washington, DC: U.S. Bureau of the Census; 1980.
61. U.S. Bureau of the Census. *Census Use Study: Health Information System-II. Report 12.* Washington, DC: U.S. Government Printing Office; 1971.
62. Fernandez RM, Kulik JC. A multilevel model of life satisfaction: Effects of individual characteristics and neighborhood composition. *Am Sociol Rev.* 1981;46:840–850.
63. Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: Evidence from national samples. *Am J Epidemiol.* 1998 Sep 1;148(5):475–486.
64. Greenwald HP, Polissar NL, Borgatta EF, McCorke R. Detecting survival effects of socioeconomic status: Problems in the use of aggregate data. *J Clin Epidemiol.* 1994;47:903–909.
65. Geronimus AT, Bound J, Neidert LJ. On the validity of using Census Geocode characteristics to proxy individual socioeconomic characteristics. *JAMA.* 1996;91:529–537.
66. Krieger N, Gordon D. Re: Use of census-based aggregate variables to proxy for socioeconomic group: Evidence from national samples [Letter]. *Am J Epidemiol.* 1999;150:892–894.
67. Davey-Smith G, Ben-Shlomo Y, Hart C. Re: Use of census-based aggregate variables to proxy for socioeconomic group: Evidence from national samples [Letter]. *Am J Epidemiol.* 1999;150:996–997.
68. Hughes GH, Cutter G, Donahue R, Friedman GD, Hulley S, Hunkeler E, Jacobs DR, Jr, Liu K, Orden S, Pirie P, Tucker B, Wagenknecht L. Recruitment in the Coronary Artery Disease Risk Development in Young Adults (Cardia) Study. *Control Clin Trials.* 1987;8: 68S–73S.
69. Massey DS, Fong E. Segregation and neighborhood quality: Blacks, Hispanics, and Asians in the San Francisco metropolitan area. *Social Forces.* 1990;69:15–32.
70. Wilson WJ. *The Truly Disadvantaged: The inner city, the underclass, and public policy.* Chicago: University of Chicago Press; 1987.
71. Williams DR, Collins C. U.S. socioeconomic and racial differences in health: Patterns and explanations. *Ann Rev Sociol.* 1995;21:349–386.
72. Hofer TP, Wolfe RA, Tedeschi PJ, McMahon LF, Griffith JR. Use of community versus individual socioeconomic data in predicting variation in hospital use. *Health Serv Res.* 1998;33:243–259.
73. Carr-Hill R, Rice N. Is enumeration district level an improvement on ward level analysis in studies of deprivation and health? *J Epidemiol Comm Health.* 1995;49:S28–S29.