American Journal of Epidemiology
Copyright © 2002 by the Johns Hopkins Bloomberg School of Public Health

Vol. 156, No. 5
Printed in U.S.A.
DOI: 10.1093/aje/kwf068

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

Nancy Krieger, Jarvis T. Chen, Pamela D. Waterman, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson

From the Department of Health and Social Behavior, Harvard School of Public Health, Boston, MA.

Received for publication October 18, 2001; accepted for publication May 7, 2002.

Despite the promise of geocoding and use of area-based socioeconomic measures to overcome the paucity of socioeconomic data in US public health surveillance systems, no consensus exists as to which measures should be used or at which level of geography. The authors generated diverse single-variable and composite area-based socioeconomic measures at the census tract, block group, and zip code level for Massachusetts (1990 population: 6,016,425) and Rhode Island (1990 population: 1,003,464) to investigate their associations with mortality rates (1989–1991: 156,366 resident deaths in Massachusetts and 27,291 in Rhode Island) and incidence of primary invasive cancer (1988–1992: 140,610 resident cases in Massachusetts; 1989–1992: 19,808 resident cases in Rhode Island). Analyses of all-cause and cause-specific mortality rates and all-cause and site-specific cancer incidence rates indicated that: 1) block group and tract socioeconomic measures performed comparably within and across both states, but zip code measures for several outcomes detected no gradients or gradients contrary to those observed with tract and block group measures; 2) similar gradients were detected with categories generated by quintiles and by a priori categorical cutpoints; and 3) measures including data on economic poverty were most robust and detected gradients that were unobserved using measures of only education and wealth. *Am J Epidemiol* 2002;156:471–82.

censuses; geographic information system; geostatistics; mortality; neoplasms; population surveillance; poverty; socioeconomic factors

Abbreviations: IRR, incidence rate ratio; RII, relative index of inequality; SEP, socioeconomic position.

Despite growing recognition of the magnitude and persistence of socioeconomic inequalities in health and the need to address them (1–4), few or no socioeconomic data exist in most US public health surveillance databases (5, 6). Only in 1989 did collection of educational data on birth and death certificates become routine (7)—60 years after the last attempt, in 1930, to generate vital statistics stratified by occupational class (8, 9). Even so, in 1997, only 7 percent, 4 percent, and 0 percent of US state registries for cancer, tuberculosis, and acquired immunodeficiency syndrome included data on education—and neither they nor birth and death databases included data on poverty, income, or other aspects of socioeconomic position (SEP) (5). This lack of data hampers meaningful monitoring of socioeconomic inequalities in public health databases.

Reflecting the limitations of available data, the US National Center for Health Statistics' first-ever national chartbook on "Socioeconomic Status and Health," issued in 1998 (1), presented data based solely on birth and death records plus data from the National Health Interview Survey, but it could not include data on cancer incidence or survival, tuberculosis, human immunodeficiency virus/acquired immunodeficiency syndrome, and other health outcomes not assessed in the National Health Interview Survey. Relatedly,

Correspondence to Dr. Nancy Krieger, Department of Health and Social Behavior, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115 (e-mail: nkrieger@hsph.harvard.edu).

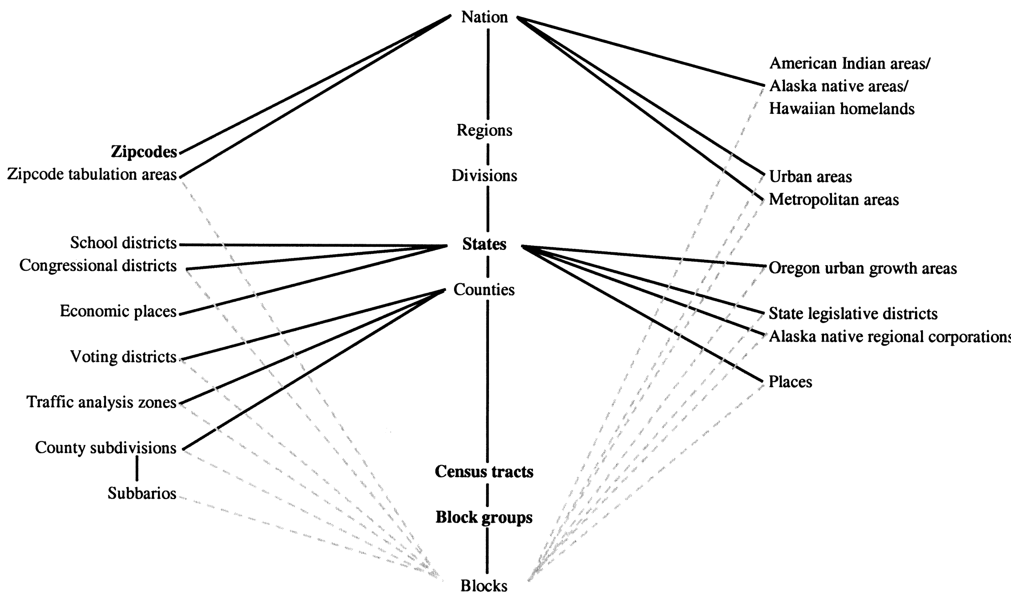

FIGURE 1. Geographic relations in the US Census. The solid lines (—) indicate connections between entities in the basic census hierarchy (from the nation to blocks) and other geographic areas; the dotted lines (– – –) indicate geographic areas that have boundaries coterminous with census blocks (85).

70 percent of the 467 US public health objectives for the year 2010 lack quantitative targets for reducing socioeconomic disparities in health, given a lack of baseline data (10, 11).

One possible solution to these gaps is to combine data from public health surveillance systems with socioeconomic data derived from the US Census. The basic approach is to classify people in public health databases and in the total population by the socioeconomic characteristics of their residential neighborhood, thereby permitting calculation of population-based rates stratified by area-based SEP (6, 12). These area-based geosocial measures—conceptualized as meaningful indicators of socioeconomic context in their own right and not merely "proxies" for individual-level data—can be validly applied to all persons, regardless of age, gender, and employment status (6, 12–15). First employed in US health studies in the 1930s (16–22), the use of such geosocial measures—by which we mean empirically observable social and physical characteristics of areas whose spatial distribution is patterned by human activity—has been greatly facilitated by the past decade's rapid development of geographic information systems technology (23–25). Indeed, National Objective 23-3 of Healthy People 2010 sets the goal of geocoding, by the year 2010, 90 percent of "all major national, state, and local health data systems… to promote nationwide use of geographic information systems (GIS) at all levels" (10).

Obstacles to the use of area-based socioeconomic measures are not only technical, however. They are also conceptual. To date, there exists no consensus in the United States regarding which area-based measures should be used, at which level of geography, to measure or monitor socioeconomic inequalities in health (6, 26). Instead, studies on a variety of outcomes, spanning from birth to death (27–38), have employed markedly different single-variable and composite area-based measures, variously derived from three different geographic levels (figure 1): the census tract (average population = 4,000); the census block group, a subdivision of the census tract (average population = 1,000); and the US Postal Service zip code (average population = 30,000) (25). By contrast, in the United Kingdom, several well-established, theoretically conceived, and validated area-based deprivation measures, such as the Townsend index, permit meaningful comparisons and monitoring of national, regional, and local socioeconomic gradients in health over time (14, 39–42).

Accordingly, we designed the Public Health Disparities Geocoding Project to determine which area-based socioeconomic measures, at which level of geography, would be most appropriate for US public health surveillance systems and research. Considerations pertained to 1) external validity (do the measures find gradients in the direction reported in the literature, i.e., positive, negative, or none, and across the full range of the distribution?); 2) robustness (do the measures detect expected gradients across a wide range of outcomes?); 3) completeness (is the measure relatively unaffected by missing data?); and 4) user-friendliness (how easy is the measure to understand and explain?). Guided by an ecosocial framework (43), we deliberately included data from multiple public health surveillance systems to maximize our ability to assess associations and geosocial health disparities observed for diverse health outcomes manifested at different ages. In this paper, we report our results for mortality rates and cancer incidence.

TABLE 1. People and areas included in a study of geocoding and health disparities, Massachusetts and Rhode Island, 1988–1992

Study base	Massachusetts			Rhode Island		
	No.	Population size		No.	Population size	
		Mean	Range		Mean	Range
Population						
1990 population	6,016,425			1,003,464		
Mortality data* (1989–1991)	156,366			27,291		
Cancer data* (primary invasive neoplasm) (1988–1992)†	140,610			19,808		
Areas						
Block groups	5,603	1,085.4 (665.2)‡	5–10,096	897	1,137.7 (670.8)	7–5,652
Census tracts	1,331	4,571.8 (2,080.0)	18–15,411	235	4,325.3 (1,810.9)	26–9,822
Zip codes	474	12,719.7 (12,244.1)	14–65,001	70	14,335.2 (13,234.8)	63–53,763

* In-state residents only.

† Massachusetts data were from 1988–1992; Rhode Island data were from 1989–1992 (data from 1988 were not available for Rhode Island because of the recency of the registry).

‡ Numbers in parentheses, standard deviation.

MATERIALS AND METHODS

Data sources

The study base comprised populations and areas in Massachusetts and Rhode Island enumerated at or within 2 years of the 1990 US Census (44, 45). Mortality data and cancer incidence data (table 1) were provided by the Massachusetts Department of Public Health and the Rhode Island Department of Health. Use of these data was approved by all relevant institutional review boards/human subjects committees at the Harvard School of Public Health, the Massachusetts Department of Public Health, and the Rhode Island Department of Health. Cause of death was categorized according to the *International Classification of Diseases, Ninth Revision, Clinical Modification* (46), cancer type according to the standard site/histology definitions of the Surveillance, Epidemiology, and End Results Program (47), and gender and race/ethnicity (plus educational level, death only) as reported by next of kin, and/or as recorded by the funeral director (for death data) or abstracted by registry staff from medical records (for cancer data). Mortality outcomes analyzed included all-cause mortality and the top five causes of death in each state by race/ethnicity, yielding nine specific causes of death: heart disease, malignant neoplasm, cerebrovascular disease, pneumonia and influenza, chronic obstructive pulmonary disease, unintentional injury, diabetes mellitus, human immunodeficiency virus, and homicide and legal intervention. Incidence of cancer was analyzed for all cancers combined and for five leading sites: the breast, cervix, colon, lung, and prostate (27, 47).

We obtained 1990 Census data for census tracts and block groups from US Bureau of the Census Summary Tape File 3A and zip code data from Summary Tape File 3B (48). The Census Bureau defines a census tract as a "small, relatively permanent statistical subdivision of a county… designed to be relatively homogeneous with respect to population characteristics, economic status, and living conditions" (25, pp. G-10, G-11); its subdivision, the block group, is the smallest geographic census unit for which census socioeconomic data are tabulated (25, p. G-6). By contrast, zip codes are "administrative units established by the United States Postal Service… for the most efficient delivery of mail, and therefore generally do not respect political or census statistical area boundaries" (48, p. A-13). Spanning from large areas cutting across states to a single building or company with a high volume of mail, "carrier routes for one zip code may intertwine with those of one or more zip codes" such that "this area is more conceptual than geographic" (49, p. 22). To geocode data to the census tract, block group, and zip code levels, we submitted residential addresses from the mortality and cancer databases to a commercial geocoding firm selected for its accuracy (50).

Two criteria central to formulating apt area-based measures of SEP are that they 1) meaningfully summarize important aspects of the specified area's socioeconomic conditions and 2) employ socioeconomic data that can legitimately be compared over time and across regions (6, 11, 14, 26, 39–42). On the basis of a priori conceptual definitions of SEP and social class (6) and evidence from both the United States and the United Kingdom emphasizing the detrimental effects of material deprivation on health (1–4, 51), we developed area-based socioeconomic measures for six domains of SEP—occupational class, income, poverty, wealth, education, and crowding—premised on the understanding that social class, as a social relationship, fundamentally drives the distribution of these manifest aspects of SEP (6).

Table 2 provides information on the 11 single-variable measures and eight composite measures we generated for each state at each level of geography. Among the composite variables, two were US analogs of the United Kingdom

TABLE 2. Constructs and operational definitions for area-based socioeconomic measures,* Massachusetts and Rhode Island, 1988–1992

Construct	Operational definition	Census variable
Occupational class		
Working class (6)	Percentage of persons employed in predominantly working-class occupations, i.e., as nonsupervisory employees. Operationalized as percentage of persons employed in the following eight of 13 census-based occupational groups: administrative support; sales; private household service; other service (except protective); precision production, craft, and repair; machine operators, assemblers, and inspectors; transportation and material moving; handlers, equipment cleaners, and laborers.	P78
Unemployment	Percentage of persons aged 16 years or older in the labor force who are unemployed (and actively seeking work).	P71
Income		
Median household income	Median household income in the year prior to the decennial census ($30,056 for the United States in 1989).	P80A
Low income (67)	Percentage of households with an income <50% of the US median household income (i.e., <$15,000 in 1989).	P80
High income	Percentage of households with an income ≥400% of the US median household income (i.e., ≥$150,000 in 1989).	P80
Gini coefficient	A measure of income inequality regarding the share of income distribution across the population. Calculated using the standard algorithm employed by the US Bureau of the Census to extrapolate the lower and upper ends of the income distribution (86, 87).	P80, P80A, P81
Poverty		
Below US poverty line	Percentage of persons below the federally defined poverty line, a threshold that varies by the size and age composition of the household; on average, it equaled $12,647 for a family of four in 1989 (48).	P117
Wealth		
Expensive homes	Percentage of owner-occupied homes worth ≥$300,000 (400% of the median value of owned homes in 1989).	H61
Educational level		
Low: less than high school	Percentage of persons aged ≥25 years with less than a 12th-grade education.	P57
High: ≥4 years of college	Percentage of persons aged ≥25 years with at least 4 years of college.	P57
Crowding		
Crowded households	Percentage of households containing more than one person per room.	H69, H49
Composite measures		
Townsend index (39–41)	A United Kingdom deprivation measure consisting of a standardized *z* score combining data on percentage of crowding, percentage of unemployment, percentage of no car ownership, and percentage of renters.	H69, H49, H40, H8
Carstairs index (14, 40–42)	A United Kingdom deprivation measure consisting of a standardized *z* score combining data on percentage of crowding, percentage of male unemployment, percentage of no car ownership, and percentage of low social class (equivalent to the following US census categories: transportation and material moving; handlers, equipment cleaners, and laborers; and household service).	H69, H49, H40, P78
Index of Local Economic Resources (52)	A "summary index" based on "white-collar employment, unemployment, and family income" (52).	P78, P71, P107A
SEP1	A composite categorical variable based on percentage below the US poverty line, working class, and expensive homes.	P117, P78, H61
SEP2	A composite categorical variable based on percentage below the US poverty line, working class, and high income.	P117, P78, P80
Factor 1†	A factor pertaining to economic resources. Highly correlated with poverty, median household income, home ownership, and car ownership.	—†
Factor 2†	A factor pertaining to occupation and education. Highly correlated with percentage working class, low education (less than high school), and high education (≥4 years of college).	—†
SEP index	A summary deprivation measure consisting of a standardized *z* score combining data on percentage working class, unemployment, percentage below the US poverty line, low education (less than high school), expensive homes, and median household income‡.	P78, P71, P117, P57, P80

* Created using data from the 1990 US Census (40).

† Variables employed in the factor analysis: percentage working class, unemployment, percentage below poverty line, home ownership, car ownership, no telephone, expensive homes, low education (less than high school), high education (≥4 years of college), household crowding, households with only one room, no kitchen, no private plumbing, median household income, and proportion of total income in the area derived from interest, dividends, and net rent.

‡ Values for "expensive homes" and "median household income" were reversed before the *z* score was computed so that a higher score on the SEP index would correspond to a higher degree of deprivation.

Townsend (39–41) and Carstairs (14, 42) deprivation indices, one used the algorithm for the US Centers for Disease Control and Prevention's Index of Local Economic Resources (52), and five were created exclusively for our study. To mirror the skewed population distribution of socioeconomic resources, we created the variables "SEP1" and "SEP2" to combine simultaneously categorical data on poverty, working class, and either wealth or high income. We generated "factor 1" and "factor 2" by factor analysis with a maximum likelihood approach (53, 54) applied to inputs listed in table 2 (see second footnote in table), using rank values of the census data, rather than impose arbitrary transformations to normalize their often considerably skewed distributions; tied values were assigned an average rank. We selected the two-factor model as the most appropriate description of the underlying factor structure. Correlations between the factors ranged from 0.420 to 0.564 after oblique rotation. Finally, we generated the "SEP index," a standardized z score akin to the Townsend index, using inputs identified by the factor analysis.

Data analysis

Our analytical plan involved five steps. Step 1 was to assess the distribution and missingness of data. Step 2 was to calculate age-standardized average annual mortality rates and cancer incidence rates stratified by the area-based socioeconomic measures at each level of geography for each state (55, p. 54; 56, p. 263). We standardized for age using the year 2000 standard million (57) and age-specific rates generated for 11 age groups (<1, 1–4, 5–14, 15–24, …, 75–84, and ≥85 years). The numerators and denominators of these rates consisted of persons residing in areas identified at the specified geographic level for which data on the specified area-based socioeconomic measure were available. Following standard practice for rates centered around a census (58, 59), we set the total number of person-years in the denominator equal to the population in that socioeconomic stratum enumerated in the 1990 Census multiplied by the relevant number of years of observation. Cutpoints for categorical area-based socioeconomic measures (see Appendix at http://www.aje.oupjournals.org) were based on both their percentile distributions (e.g., quintiles) and a priori considerations (e.g., the federal definition of "poverty areas" as regions where ≥20 percent of the population lives below the US poverty line (60, 61)).

In step 3, we visually inspected and quantified socioeconomic gradients for each outcome using each area-based socioeconomic measure at each geographic level, excluding persons who were geocoded to areas with no population (e.g., geocoded to a zip code not included in the 1990 Census). Based on clear evidence of linear trends (data not shown; available upon request), we followed standard US reporting practices (1) and computed the mortality rate ratio, incidence rate ratio (IRR), and incidence rate difference, comparing rates for people living in areas with the least resources with rates for people living in areas with the most resources; given similar patterns, we report only the IRR. To take into account both the population distribution of the exposure and the magnitude of the rate ratio detected in each socioeconomic stratum, we also calculated the relative index of inequality (RII), a measure of effect that consequently permits meaningful comparison of gradients across different socioeconomic measures (62–64). In step 4, we further restricted analyses to persons geocoded to all three levels of geography. In step 5, we summarized findings across socioeconomic measures and geographic levels, in relation to our a priori considerations regarding external validity, robustness, and completeness of each measure. As a further check on internal validity, we also analyzed mortality using individual-level educational data. All analyses were conducted in SAS (65).

RESULTS

Fully 92.8 percent of the 370,196 mortality and cancer records for Massachusetts and Rhode Island were successfully geocoded to the census block group level, and 99.6 percent were geocoded to both the census tract level and the zip code level. These results were independent of gender, age, race/ethnicity, and, for the mortality data, educational level (table 3). The proportion of areas without the specified socioeconomic measures was also low (typically <1 percent), considering all measures across all levels of geography in both states (data not shown; available upon request). Among the total 368,530 records geocoded to the zip code level, 23,350 (6.3 percent) could not be linked to 1990 Census data because their zip codes either were for nonresidential areas (e.g., government agencies, businesses with a high mail volume, or post offices and post office boxes) or were created or changed after the 1990 Census.

Table 4 (the full version is available on the World Wide Web at http://www.aje.oupjournals.org) presents results of selected analyses generating and comparing all-cause and cause-specific mortality rates and cancer incidence rates, stratified by each area-based socioeconomic measure at each level of geography, for each state. Given the similar findings, we present data for the categorical version of the poverty variable but not the quintile version, for SEP1 but not SEP2, for the SEP index but not factor 1 or factor 2, and for death due to diabetes but not death due to unintentional injury (data not shown; available upon request). Patterns of association were equivalent for analyses restricted to persons geocoded to all three levels of geography (data not shown; available upon request).

As table 4, section a, illustrates, depending on the type of mortality and the area-based socioeconomic measure chosen, estimates of effect comparing Massachusetts mortality rates for persons living in areas with the least resources versus persons living in areas with the most resources ranged from no effect to a substantial effect; similar patterns were observed at each level of geography. For example, across levels of geography, the median value of both the IRR and the RII for all-cause mortality was 1.3–1.4, with most measures performing similarly in detecting (as expected) associations between higher mortality and fewer economic resources (1); the exception was the Gini coefficient, a measure of income inequality (no gradient detected). Similar patterns were evident for Massachusetts mortality due to heart disease, malignant neoplasm (albeit

TABLE 3. Percentages of deaths and cancer cases geocoded to the census block group, census tract, and zip code levels, Massachusetts and Rhode Island, 1988–1992*

	No.		Percent geocoded							
			Block group		Census tract		Zip code		Not geocoded	
	MA†	RI†	MA	RI	MA	RI	MA	RI	MA	RI
			Mortality							
Total	156,366	27,291	93.8	91.1	99.8	95.3	99.9	94.7	0.1	4.7
Gender										
Men	75,051	13,279	93.9	92.3	99.8	95.9	99.9	95.4	0.1	4.1
Women	81,315	14,012	93.7	90.0	99.8	94.6	99.9	94.1	0.1	5.3
Age (years)										
<15	1,904	466	92.9	93.6	98.9	96.1	99.3	95.9	0.7	3.9
15–44	9,702	1,490	94.7	93.0	99.5	96.2	99.6	95.9	0.4	3.8
45–64	23,949	4,032	94.4	93.5	99.7	96.5	99.8	96.3	0.2	3.5
≥65	120,209	21,299	93.6	90.5	99.9	94.9	99.9	94.3	0.1	5.0
Race/ethnicity										
White, non-Hispanic	147,946	25,883	93.7	91.0	99.8	95.2	99.9	94.6	0.1	4.8
Black, non-Hispanic	5,572	853	96.1	94.1	99.7	97.1	99.7	96.6	0.3	2.9
Other, non-Hispanic‡	939	152	93.6	93.4	97.8	96.1	97.9	95.4	2.1	3.9
Hispanic	1,908	246	95.3	96.3	98.8	97.2	99.1	97.2	0.9	2.8
Education (among persons aged ≥25 years)										
0–11 years	36,285	—§	93.2	—	99.9	—	99.9	—	0.1	—
12 years	77,454	—	94.3	—	99.9	—	99.9	—	0.1	—
≥13 years	29,935	—	93.1	—	99.7	—	99.8	—	0.2	—
			Cancer incidence							
Total	140,610	19,809	92.4	91.5	100.0	99.8	100.0	99.8	0	0.2
Gender										
Men	69,334	9,725	92.2	91.6	100.0	99.8	100.0	99.8	0	0.2
Women	71,276	10,084	92.5	91.4	100.0	99.8	100.0	99.8	0	0.2
Age (years)										
<15	904	90	94.9	90.0	100.0	100.0	100.0	100.0	0	0.0
15–44	12,687	1,599	93.1	92.1	100.0	99.7	100.0	99.7	0	0.3
45–64	41,260	5,227	93.5	92.4	100.0	99.7	100.0	99.7	0	0.3
≥65	85,759	12,882	91.7	91.1	100.0	99.8	100.0	99.8	0	0.2
Race/ethnicity										
White, non-Hispanic	131,176	18,789	92.3	91.6	100.0	99.8	100.0	99.8	0	0.2
Black, non-Hispanic	3,716	392	94.0	94.6	100.0	100.0	100.0	100.0	0	0.0
Other, non-Hispanic‡	1,040	88	95.1	95.5	100.0	100.0	100.0	100.0	0	0.0
Hispanic	842	129	96.3	93.0	100.0	99.2	100.0	99.2	0	0.8

* For both Massachusetts and Rhode Island, mortality data were from 1989–1991; cancer data for Massachusetts were from 1988–1992 and data for Rhode Island were from 1989–1992.

† MA, Massachusetts; RI, Rhode Island.

‡ Includes "Asian and Pacific Islander," "American Indian and Alaska Native," and groups classified in the US Census as "other." These ethnic groups together constituted less than 3 percent of the Massachusetts and Rhode Island populations in 1990.

§ Data not available.

TABLE 4. Rates of all-cause mortality according to area-based socioeconomic measures* (census block group, census tract, and zip code) for persons residing in areas with the least and the most socioeconomic resources, Massachusetts, 1989–1991

Area-based socioeconomic measure	Rate for areas with the least resources			Rate for areas with the most resources			Incidence rate ratio for least versus most					
							BG		CT		ZC	
	BG‡	CT‡	ZC‡	BG	CT	ZC	IRR‡	95% CI‡	IRR	95% CI	IRR	95% CI
Working class (categories)	929.7	966.6	900.3	718.9	749.8	647.1	1.29	1.23, 1.36	1.29	1.22, 1.36	1.39	1.30, 1.49
Median household income (quintiles)	954.9	1,006.7	927.0	747.9	781.1	698.9	1.28	1.22, 1.34	1.29	1.23, 1.35	1.33	1.26, 1.39
Poverty (categories)	1,030.7	1,060.4	1,070.5	763.3	800.1	766.8	1.35	1.29, 1.42	1.33	1.26, 1.39	1.40	1.32, 1.47
Gini coefficient (quintiles)	865.5	937.1	884.3	840.2	854.9	822.7	1.03	0.98, 1.08	1.10	1.04, 1.15	1.07	1.01, 1.14
Wealth (categories)	834.3	886.1	880.5	703.7	751.1	665.9	1.19	1.13, 1.24	1.18	1.13, 1.23	1.32	1.26, 1.39
Crowding (categories)	1,119.4	1,024.6	944.7	782.7	837.6	803.5	1.43	1.23, 1.67	1.22	1.00, 1.5	1.18	0.69, 2.00
Low education (categories)	962.4	986.6	960.8	752.3	780.4	734.9	1.28	1.22, 1.34	1.26	1.20, 1.33	1.31	1.23, 1.39
Townsend index (quintiles)	1,001.9	1,049.9	938.2	743.2	777.8	733.3	1.35	1.28, 1.42	1.35	1.28, 1.42	1.28	1.21, 1.35
Index of Local Economic Resources (quintiles)	952.5	1,005.9	953.3	726.7	769.8	681.5	1.31	1.25, 1.37	1.31	1.25, 1.37	1.40	1.34, 1.46
SEP1‡ (categories)	1,025.6	1,036.3	1,043.9	687.4	741.7	646.2	1.49	1.38, 1.61	1.40	1.30, 1.51	1.62	1.43, 1.82
SEP index (quintiles)	934.8	1,004.2	934.4	712.1	754.5	672.1	1.31	1.25, 1.38	1.33	1.27, 1.4	1.39	1.33, 1.46
Median value	*954.9*	*1,005.9*	*938.2*	*743.2*	*777.8*	*698.9*	*1.31*		*1.29*		*1.33*	

* Average annual age-standardized† rates (per 100,000) and age-adjusted incidence rate ratios. Cutpoints for the measures shown are provided in the Appendix (aje.oupjournals.org).

† Age-standardized to the year 2000 standard million (57).

‡ BG, block group; CT, census tract; ZC, zip code; IRR, incidence rate ratio; CI, confidence interval; SEP, socioeconomic position.

with a weaker gradient), and diabetes (with a stronger gradient). By contrast, for mortality due to human immunodeficiency virus and to homicide and legal intervention, measures intended to reflect poverty detected notably larger gradients. For human immunodeficiency virus, estimates ranged from no effect (wealth) to a >20-fold effect (RII for poverty, Townsend index, and crowding), with a median IRR between 3 and 4 and a median RII between 5 and 7. For homicide and legal intervention, estimates ranged from a twofold effect (IRR for wealth) to a >30-fold effect (RII for poverty, crowding, and Townsend index), with a median IRR between 9 and 11 and a median RII between 22 and 24. For all outcomes, the precision of the effect estimates was greater for the RII than for the IRR.

Analysis of the Rhode Island mortality data (table 4, section c) yielded similar patterns, except that somewhat stronger socioeconomic gradients were apparent both for median household income and for all outcomes except homicide and legal intervention. For both states, analyses of mortality and individual-level education data comparing persons with a high school education or less with persons with more than a high school education showed gradients pointing in the same direction (data not shown; available upon request).

Alternatively, for cancer incidence, level of geography mattered for several of the sites but not all (table 4, sections b and d). For example, census block and tract-level measures detected expected socioeconomic gradients (27, 28, 66) for three cancer sites not captured by zip code measures (breast and prostate cancer in Massachusetts and lung cancer in Rhode Island). In Massachusetts, gradients detected using zip code data were in the direction opposite that observed using block group and tract socioeconomic data for colon cancer (IRR and RII) and for all sites combined (RII only).

Visually summarizing key results, figure 2 depicts socioeconomic gradients in all-cause mortality for Massachusetts by employing the three block group measures that most consistently detected socioeconomic gradients in health while differently delimiting the population at risk: poverty (single-variable, categorical), SEP1 (composite, categorical), and the Townsend index (composite, quintile).

DISCUSSION

Findings

This study—which to our knowledge was the first systematic US investigation of area-based socioeconomic measures suitable for monitoring population health and the first that simultaneously compared diverse area-based socioeconomic measures within and across levels of geography—provided empirical evidence that both choice of measure and level of geography matter. Specifically, examining mortality and cancer incidence for two New England states during the period around 1990 in conjunction with 1990 US Census data, we obtained three findings. First, measures designed to detect economic deprivation were most robust, consistently detecting socioeconomic gradients not only for the leading causes of death and cancer, as did the other measures, but also for deaths due to human immunodeficiency virus and homicide and legal intervention, whose gradients were detected less well or missed by measures of education and wealth. Second, census block group and census tract

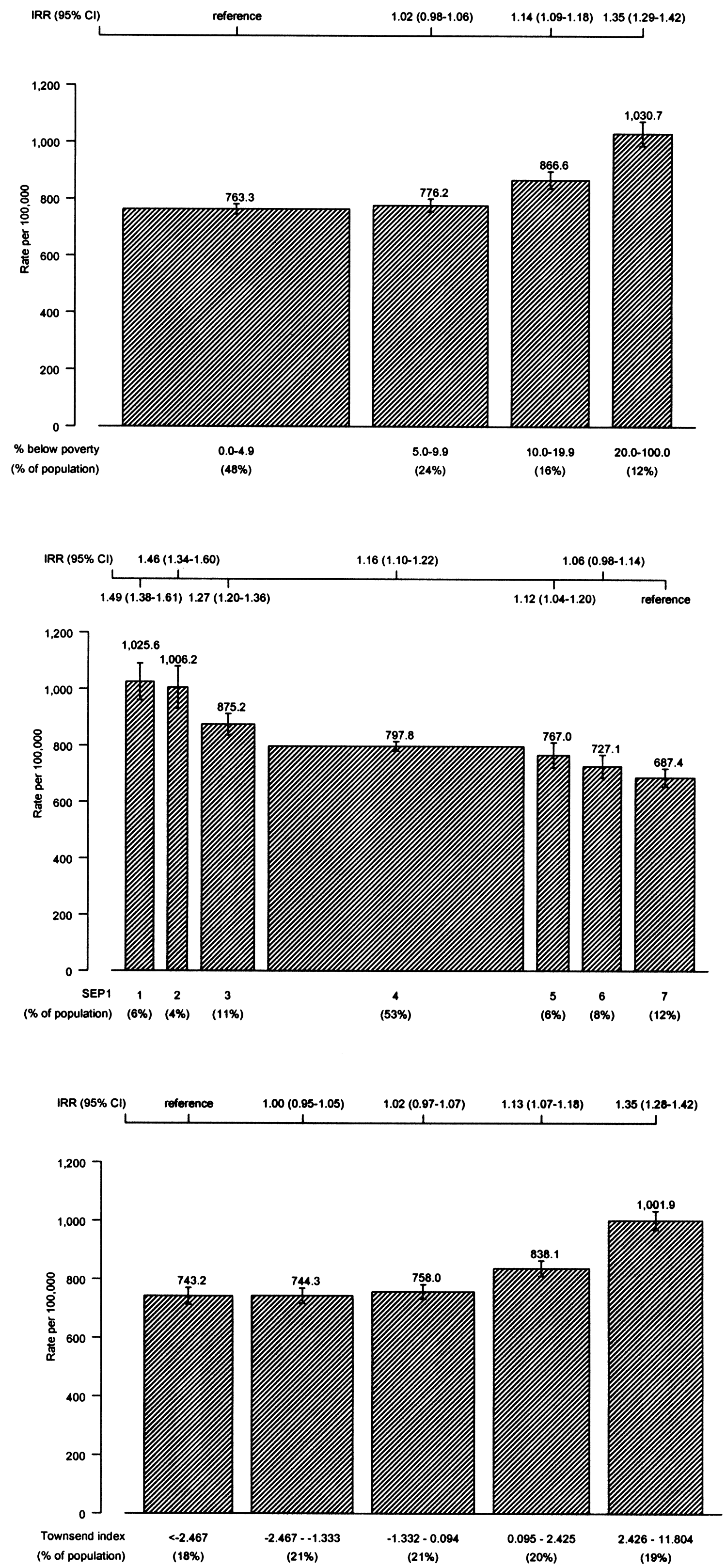

FIGURE 2. Age-adjusted all-cause mortality rates per 100,000 person-years (*y*-axes) and incidence rate ratios for mortality (*x*-axis above each section) according to three socioeconomic measures (top, percent below US poverty line; middle, SEP1; bottom, Townsend index) at the US Census block group level, Massachusetts, 1989–1991. For detailed definitions of measures, see table 2. The width of the bars is proportional to the percentage of the population they contain. IRR, incidence rate ratio; CI, confidence interval; SEP, socioeconomic position.

measures performed similarly for virtually all outcomes; zip code measures, however, in some cases failed to detect gradients or detected gradients contrary to those observed with the block group and tract measures. Third, categories based on quintiles and a priori cutpoints detected similar socioeconomic gradients, but only the latter could be uniformly applied across levels of geography within and across states.

Study limitations

Several sources of error and bias could have affected our findings. If, for example, underregistration or misclassification of cases were either nondifferential with respect to poverty or increased with respect to poverty (66), the net effect would be to underestimate socioeconomic gradients in the specified outcomes. A conservative bias would also have occurred if persons subject to socioeconomic deprivation were less likely to have a geocodable address (12); table 3 suggests that our results were unlikely to have been affected by this problem. Were such biases operative, however, they would have equally affected analyses at each geographic level and thus would not invalidate comparisons of socioeconomic gradients across socioeconomic measures and across levels of geography. Adding further credence to our findings, the proportion of areas without data on the area-based socioeconomic measures was so low as to render negligible the impact of these missing data, and we minimized geocoding error by using a commercial firm whose accuracy we validated with records from the study's death and birth databases (50).

Additional concerns pertain to the construction of the area-based socioeconomic measures. One controversy centers on the benefits and drawbacks of using single-variable indicators versus composite indicators—a topic as relevant to individual-level socioeconomic data as to area-based socioeconomic measures (6, 26, 39–42). A related controversy pertains to establishing categorical cutpoints for socioeconomic data (6, 26, 39–42). To address these issues empirically, we employed a variety of single-variable and composite socioeconomic measures, using cutpoints based on both percentile distributions and a priori considerations. It is notable that several of the single-variable measures, especially those intended to measure poverty, detected the same magnitude of socioeconomic inequality in health as the composite measures, and categorical variables based on a priori cutpoints and quintiles detected gradients of the same magnitude. However, while the a priori cutpoints could be uniformly applied to each level of geography in each state, the data-dependent cutpoints differed by level within and across states, rendering comparison of findings across regions and geographic levels more problematic.

Other caveats pertain to temporal and spatial scale. From an etiologic perspective, misclassification of SEP may occur if SEP at the time of disease diagnosis or death differs from that at the time of exposure to conditions causing the outcome (6, 11). From a monitoring standpoint, however, use of temporally congruent socioeconomic data is appropriate for delimiting population distributions of the specified outcomes. It is also notable that all of our study's area-based socioeconomic measures can be meaningfully compared across decennial censuses, a necessary attribute for monitoring socioeconomic trends over time (67).

Analyses conducted for this first phase of our project did not take into account either spatial correlation of geographic areas (e.g., nesting of block groups within tracts) or issues of adjacency (e.g., effects of living in a poor block group adjacent to chiefly poor block groups versus more affluent block groups). Although use of multilevel models to take into account geographic nesting would have improved the precision of our effect estimates, existing literature suggests that it would not have substantially changed the estimates themselves or the patterns of associations we observed (68–70). Had the analyses taken into account issues of adjacency, however, different and additional effect estimates might have been obtained (68–70). The type of aggregation bias typically referred to in epidemiologic literature as the "ecologic fallacy" (71–75) is not germane to the present study design, since individuals constituted the unit of observation for both the dependent variables (health outcomes) and the independent variables (living in an area with certain sociodemographic characteristics). Instead, the validity of using area-based socioeconomic measures depends on the extent to which areas constitute meaningful geographic units (12, 76)—a different question from whether they are "proxies" for individual-level socioeconomic data.

Interpretation and implications

The patterning of socioeconomic gradients in health detected by the selected area-based socioeconomic measures employed in this study, within and across levels of geography and across health outcomes, is likely to reflect both the different meanings of the areas investigated and the different pathways by which diverse aspects of SEP influence health (6, 39, 40, 77). It is notable that almost all measures detected gradients across the full socioeconomic spectrum in the direction expected on the basis of extant literature (1, 3, 66). The fact that patterns at the block group and tract levels were largely similar for Massachusetts and Rhode Island but patterns at the zip code level differed within and across these states is perhaps not surprising, given that census tract and block groups would, by design, be expected to contain more homogenous populations than zip codes (6, 25). Epidemiologic studies that have investigated the use of individual-based socioeconomic measures versus area-based measures have reported similar performance for the block group and tract measures (or their equivalents) (12, 13, 32, 35, 38, 78–83) and inconsistent results for zip code data (32, 35). Together, these results underscore the conclusion that additional effort expended to geocode health data to the tract and block group level is likely to offset the greater ease of obtaining potentially less informative zip code data.

However, the novel finding that results based on zip codes versus tracts and block groups differed chiefly for cancer incidence, not mortality, cannot simply be attributed to level of geography per se. One speculative explanation is that exclusion of persons geocoded to zip codes not included in the 1990 Census introduced more of a selection bias in relation to SEP for cancer incidence (9.4 percent) than for mortality (3.2 percent). The consistent lack of association between the health outcomes and the area-based Gini coefficient, in turn, is probably due to the relatively small size of the geographic areas studied, since, given the realities of economic residential segregation in the United States, meaningful measurement of income inequality requires analysis of larger regions (61, 70). Perhaps most importantly, the finding that the single-variable and composite measures explicitly capturing aspects of economic impoverishment consistently detected the sharpest socioeconomic gradients in health across different specific causes of death and types of cancer underscores the profound impact of material deprivation on health and the fundamental necessity of evaluating area-based socioeconomic measures across more than just one or two outcomes.

In conclusion, drawing on our a priori criteria pertaining to external validity, robustness, completeness, and user-friendliness, along with Rossi and Gilmartin's (85) criteria for valid and useful social indicators—that they be 1) conceptually based; 2) constructed from valid, reliable, and accessible data using appropriate statistical techniques; 3) comparable over time and across population groups; and 4) readily understandable, with normative value relevant to timely policy making—we offer a tentative recommendation, pending our analyses of additional data from public health surveillance systems. Specifically, our findings suggest that efforts to monitor US socioeconomic inequalities in health using area-based socioeconomic measures will be best served by those tract or block group measures that are 1) most attuned to capturing economic deprivation, 2) meaningful across regions and over time, and 3) easily understood, and hence based on readily interpretable variables with a priori categorical cutpoints. One likely candidate meeting all of these criteria is the measure "percentage of persons living below the US poverty line."

ACKNOWLEDGMENTS

This work was funded by grant 1 R01 HD36865-01 from the National Institute of Child Health and Human Development (Principal Investigator, Dr. Nancy Krieger).

The authors thank Drs. Daniel Friedman (Bureau of Health Statistics, Research and Evaluation, Massachusetts Department of Health) and Jay Buechner (Office of Health Statistics, Rhode Island Department of Health) for facilitating the use of health department data in the conduct of this study and for their helpful comments. The authors likewise thank the following health department staff for supplying mortality and cancer registry data: Massachusetts Department of Public Health—Alice Mroszczyk; *Cancer Registry*: Dr. Susan Gershman, Mary Mroszczyk, and Ann R. MacMillan; *Registry of Vital Records and Statistics*: Elaine Trudeau and Charlene Zion; Rhode Island Department of Health—*Vital Statistics*: Roberta Chevoya; *Division of Disease Prevention and Control*: Dr. John Fulton.

REFERENCES

1. National Center for Health Statistics. Health, United States, 1998 with socioeconomic status and health chartbook. Hyattsville, MD: National Center for Health Statistics, 1998. (DHHS publication no. (PHS) 98-1232).
2. Evans T, Whitehead M, Diderichsen F, et al. Challenging inequities in health: from ethics to action. New York, NY: Oxford University Press, 2001.
3. Leon D, Walt G, eds. Poverty, inequality, and health: an international perspective. New York, NY: Oxford University Press, 2001.
4. Berkman L, Kawachi I, eds. Social epidemiology. New York, NY: Oxford University Press, 2000.
5. Krieger N, Chen JT, Ebel G. Can we monitor socioeconomic inequalities in health? A survey of U.S. health departments' data collection and reporting practices. Public Health Rep 1997;112:481–91.
6. Krieger N, Williams D, Moss N. Measuring social class in US public health research: concepts, methodologies and guidelines. Annu Rev Public Health 1997;18:341–78.
7. Tolson GC, Barnes JM, Gay GA, et al. The 1989 revision of the U.S. standard certificates and reports. Hyattsville, MD: National Center for Health Statistics, 1991. (Vital and health statistics, series 1, no. 28) (DHHS publication no. PHS 91-1465).
8. Whitney JS, ed. Death rates by occupation based on data of the U.S. Census Bureau 1930. New York, NY: National Tuberculosis Association, 1934.
9. Krieger N, Fee E. Measuring social inequalities in health in the United States: an historical review, 1900–1950. Int J Health Serv 1996;26:391–418.
10. Public Health Service, US Department of Health and Human Services. Healthy People 2010 (conference edition, in two volumes). Washington, DC: US GPO, 2000. (http://www.health.gov/healthypeople). [Last accessed June 10, 2001].
11. Krieger N, Zierler S, Hogan JW, et al. Geocoding and measurement of neighborhood socioeconomic position. In: Kawachi I, Berkman LF, eds. Neighborhoods and health. New York, NY: Oxford University Press (in press).
12. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. Am J Public Health 1992;82:703–10.
13. 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 Community Health 1991;45:35–42.
14. Carstairs V, Morris R. Deprivation and mortality: an alternative to social class? Community Med 1989;11:210–19.
15. Sorensen A. Women, family, and class. Annu Rev Sociol 1994; 20:27–47.
16. Watkins RJ. Introduction. In: Watkins RJ, Swift AL Jr, Green HW, et al. Golden anniversary of census tracts, 1956. Washington, DC: American Statistical Association, 1956:1–2.
17. Nathan WB. Health conditions in North Harlem 1923–1927. New York, NY: National Tuberculosis Association, 1932.
18. Green HW. Tuberculosis and economic strata, Cleveland's Five-City Area, 1928–1931. Cleveland, OH: Anti-Tuberculosis League, 1932.
19. Green HW. Infant mortality and economic status, Cleveland Five-City Area. Cleveland, OH: Cleveland Health Council, 1939.
20. Terris M. Relation of economic status to tuberculosis mortality by age and sex. Am J Public Health 1948;38:1061–70.
21. Cohart EM. Socioeconomic distribution of stomach cancer in

New Haven. Cancer 1954;8:455–61.
22. Sheps C, Watkins JH. Mortality in the socio-economic districts of New Haven. Yale J Biol Med 1947-48;20:51–80.
23. Richards TB, Croner CM, Rushton G, et al. Geographic information systems and public health: mapping the future. Public Health Rep 1999;114:359–73.
24. Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. Epidemiol Rev 1999;21:143–61.
25. Bureau of the Census, US Department of Commerce. Geographical areas reference manual. Washington, DC: Bureau of the Census, 1994. (http://www.census.gov/geo/www/garm.html). [Last accessed June 28, 2001].
26. Lynch J, Kaplan G. Socioeconomic position. In: Berkman L, Kawachi I, eds. Social epidemiology. New York, NY: Oxford University Press, 2000:13–35.
27. Krieger N, Quesenberry C Jr, Peng T, et al. Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, black, Hispanic, and white residents of the San Francisco Bay Area, 1988–1992 (USA). Cancer Causes Control 1999;10:525–37.
28. Liu L, Deapen D, Bernstein L. Socioeconomic status and cancers of the female breast and reproductive organs: a comparison across racial/ethnic populations in Los Angeles County, California (United States). Cancer Causes Control 1998;9:369–80.
29. Ellen JM, Kohn RP, Bolan GA, et al. Socioeconomic differences in sexually transmitted disease rates among black and white adolescents, San Francisco, 1990 to 1992. Am J Public Health 1995;85:1546–8.
30. Chen FM, Breiman RF, Farley M, et al. Geocoding and linking data from population-based surveillance and the US Census to evaluate the impact of median household income on the epidemiology of invasive *Streptococcus pneumoniae* infections. Am J Epidemiol 1998;148:1212–18.
31. Wissow LS, Gittelsohn AM, Szklo M, et al. Poverty, race, and hospitalization for childhood asthma. Am J Public Health 1988;78:777–82.
32. Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. Am J Epidemiol 1998;48:475–86.
33. Smith GD, Neaton JD, Wentworth D, et al. Socioeconomic differentials in mortality among men screened for the Multiple Risk Factor Intervention Trial: I. white men. Am J Public Health 1996;86:486–96.
34. Smith GD, Wentworth D, Neaton JD, et al. Socioeconomic differentials among men screened for the Multiple Risk Factor Intervention Trial: II. black men. Am J Public Health 1996;86:497–504.
35. Geronimus AT, Bound J, Neidert LJ. On the validity of using census geocode data to proxy individual socioeconomic characteristics. J Am Stat Assoc 1996;91:529–37.
36. Roberts EM. Neighborhood social environments and the distribution of low birthweight in Chicago. Am J Public Health 1997;87:597–603.
37. Diez-Roux AV, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. N Engl J Med 2001;345:99–106.
38. Cherkin DC, Grothaus L, Wagner EH. Is magnitude of co-payment effect related to income? Using census data for health services research. Soc Sci Med 1992;34:33–41.
39. 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, United Kingdom: Centre for Urban and Regional Studies, University of Birmingham, 1995.
40. Gordon D. Census based deprivation indices: their weighting and validation. J Epidemiol Community Health 1995;49(suppl 2):S39–44.
41. Townsend P, Phillimore P, Beattie A. Health and deprivation: inequality and the North. London, United Kingdom: Croom Helm, 1988.
42. Carstairs V. Deprivation indices: their interpretation and use in relation to health. J Epidemiol Community Health 1995;49(suppl 2):S3–8.
43. Krieger N. Theories for social epidemiology for the 21st century: an ecosocial perspective. Int J Epidemiol 2001;30:668–77.
44. Bureau of the Census, US Department of Commerce. American fact finder. Massachusetts. Washington, DC: Bureau of the Census, 2001. (http://factfinder.census.gov/bf/_lang=en_vt_name=DEC_2000_SF1_U_DP1_geo_id=04000US25.html). [Last accessed August 30, 2001].
45. Bureau of the Census, US Department of Commerce. American fact finder. Rhode Island. Washington, DC: Bureau of the Census, 2001. (http://factfinder.census.gov/servlet/BasicFactsTable?_lang=enand_vt_name=DEC_1990_STF1_DP1and_geo_id=04000US44). [Last accessed August 30, 2001].
46. Health Care Financing Administration, US Department of Health and Human Services. International classification of diseases, ninth revision, clinical modification (ICD-9-CM). (Computer file). 6th ed. Washington, DC: Health Care Financing Administration, 2000.
47. National Cancer Institute. SEER cancer statistics review, 1973–1999. Bethesda, MD: National Cancer Institute, 2001. (http://www.seer.cancer.gov/csr/1973_1999). [Last accessed July 11, 2001].
48. Bureau of the Census, US Department of Commerce. Census of population and housing, 1990. Summary Tape File 3 technical documentation. (http://www.census.gov/td/stf3/contents/html). Washington, DC: Bureau of the Census, 1991.
49. Bureau of the Census, US Department of Commerce. Geographic changes for Census 2000 + glossary. Washington, DC: Bureau of the Census, 2001. (http://www.census.gov/geo/www/tiger/glossary.html). [Last accessed July 3, 2001].
50. Krieger N, Waterman P, Lemieux K, et al. On the wrong side of the tracts? Evaluating accuracy of geocoding for public health research. Am J Public Health 2001;91:1114–16.
51. Council of Economic Advisors for the President's Initiative on Race. Changing America: indicators of social and economic well-being by race and Hispanic origin. Washington, DC: US GPO, 1998. (http://www.access.gpo.gov/eop/ca/index.html). [Last accessed May 15, 2000].
52. Casper ML, Barnett E, Halverson JA, et al. Women and heart disease: an atlas of racial and ethnic disparities in mortality. Morgantown, WV: Office for Social Environment and Health Research, West Virginia University, 1999.
53. Lawley DN, Maxwell AE. Factor analysis as a statistical method. New York, NY: Macmillan Publishing Company, 1971.
54. Bartholomew DJ. Latent variable models and factor analysis. London, United Kingdom: Charles Griffin and Company, 1987.
55. Breslow NE, Day NE, eds. Statistical methods in cancer research. Vol 2. The design and analysis of cohort studies. (IARC Scientific Publication no. 82). Lyon, France: International Agency for Research on Cancer, 1987.
56. Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia, PA: Lippincott-Raven, 1998.
57. Anderson RN, Rosenberg HM. Age standardization of death rates: implementation of the year 2000 standard. Hyattsville, MD: National Center for Health Statistics, 1998. (National vital statistics reports, vol 37, no. 3).
58. Breslow NE, Day NE. Statistical methods in cancer research. Vol 1. The analysis of case-control studies. (IARC Scientific

Publication no. 32). Lyon, France: International Agency for Research on Cancer, 1980:47.
59. Boyle P, Parkin DM. Statistical methods for registries. In: Jensen OM, Parkin DM, MacLennan R, et al, eds. Cancer registration: principles and methods. (IARC Scientific Publication no. 95). Lyon, France: International Agency for Research on Cancer, 1991:126–58.
60. Bureau of the Census, US Department of Commerce. Poverty areas. Washington, DC: Bureau of the Census, 2001. (http://www.census.gov/population/socdemo/statbriefs/povarea.html). [Last accessed October 8, 2001].
61. Jargowsky PA. Poverty and place: ghettos, barrios, and the American city. New York, NY: Russell Sage Foundation, 1997.
62. Pamuk ER. Social class inequality in mortality from 1921 to 1972 in England and Wales. Popul Stud 1985;39:17–31.
63. Wagstaff A, Paci P, van Doorslaer E. On the measurement of inequalities in health. Soc Sci Med 1991;33:545–57.
64. Davey Smith G, Hart C, Hole D, et al. Education and occupational social class: which is the more important indicator of mortality risk? J Epidemiol Community Health 1998;52:153–60.
65. SAS Institute, Inc. SAS/STAT user's guide, version 6. 4th ed. Vol 1. Cary, NC: SAS Institute, Inc, 1989.
66. Kogevinas M, Pearce N, Susser M, et al, eds. Social inequalities in cancer. Lyon, France: International Agency for Research on Cancer, 1997. (IARC Scientific Publication no. 136).
67. Gordon D, Spicker P, eds. The international glossary of poverty. (CROP International Series on Poverty). London, United Kingdom: Zed Books, 1999.
68. Subramanian SV, Duncan C, Jones K. Multilevel perspectives on modeling census data. Environment and Planning A 2001; 33:399–417.
69. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. Soc Sci Med 1998;46:97–117.
70. Soobader M, LeClere FB. Aggregation and the measurement of income inequality, effects on morbidity. Soc Sci Med 1999;48: 733–44.
71. Alker HR Jr. A typology of ecologic fallacies. In: Doggan M, Rokkan S, eds. Social ecology. Cambridge, MA: MIT Press, 1969:69–86.
72. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. Am J Public Health 1998;88:216–22.
73. Macintyre S, Ellaway A. Ecological approaches: rediscovering the role of the physical and social environment. In: Berkman L, Kawachi I, eds. Social epidemiology. New York, NY: Oxford University Press, 2000:332–48.
74. Robinson WS. Ecological correlations and the behavior of individuals. Am Sociol Rev 1950;15:351–7.
75. Langbein LI, Lichtman AJ. Ecological inference. Beverly Hills, CA: Sage Publications, 1978:27–8.
76. Oppenshaw S, Taylor PJ. The modifiable areal unit problem. In: Wrigley N, Bennett RJ, eds. Quantitative geography. London, United Kingdom: Routledge and Kegan Paul, 1981:60–9.
77. Smith GD, Whitley E, Dorling D, et al. Area based measures of social and economic circumstances: cause specific mortality patterns depend on the choice of index. J Epidemiol Community Health 2001;55:149–50.
78. Greenwald JP, Polissar NL, Borgatta EF, et al. Detecting survival effects of socioeconomic status: problems in the use of aggregate data. J Clin Epidemiol 1994;47:903–9.
79. Carr-Hill R, Rice N. Is enumeration district level an improvement on ward level analysis in studies of deprivation and health? J Epidemiol Community Health 1995;49(suppl 2): S28–9.
80. Mustard CA, Derksen S, Berthelot JM, et al. Assessing ecologic proxies for household income: a comparison of household and neighbourhood level income measures in the study of population health status. Health Place 1999;5:157–71.
81. Reijneveld SA, Verheij RA, de Bakker DH. The impact of area deprivation on differences in health: does the choice of the geographical classification matter? J Epidemiol Community Health 2000;54:306–13.
82. Hyndman JC, D'Arcy D, Holman J, et al. Misclassification of social disadvantage based on geographical areas: comparison of postcode and collector's districts analyses. Int J Epidemiol 1995;24:165–76.
83. Diez-Roux AV, Kiefe CI, Jacobs DR Jr, et al. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. Ann Epidemiol 2001;11:395–405.
84. Rossi RJ, Gilmartin KJ. The handbook of social indicators: sources, characteristics, and analysis. New York, NY: Garland STPM Press, 1980.
85. Bureau of the Census, US Department of Commerce. Census geography. Washington, DC: Bureau of the Census, 2001. (http://factfinder.census.gov/home/en/epss/census_geography.html). [Last accessed May 29, 2001].
86. Kennedy BP, Kawachi I, Prothrow-Stith D. Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States. BMJ 1996;312:1004–7.
87. Cowell FA. Measuring inequality. 2nd ed. (LSE Handbooks in Economics series). London, United Kingdom: Prentice Hall, 1995.