Neighborhoods and Health

Edited by

ICHIRO KAWACHI
LISA F. BERKMAN

OXFORD
UNIVERSITY PRESS
2003

7

Geocoding and Measurement of Neighborhood Socioeconomic Position: A U.S. Perspective

NANCY KRIEGER
SALLY ZIERLER
JOSEPH W. HOGAN
PAMELA WATERMAN
JARVIS CHEN
KERRY LEMIEUX
ANNIE GJELSVIK

Where do you live? Take a moment to visualize your neighborhood. Ask yourself: What are its class composition and boundaries? What kinds of data would you use to characterize neighborhood conditions, let alone explore their connections to health?

Ask any of these questions, and you immediately engage with core issues of geocoding and measurement of neighborhood socioeconomic position for public health research and practice. In this chapter we focus on conceptual and methodologic issues germane to public health monitoring in the United States. Many of the concepts, methods, and questions, however, are likely to be relevant to other countries (Gordon, 1995; Lee et al., 1995; Hankins et al., 1998; Cadum et al., 1999; da Silva et al., 1999; Mustard et al., 1999; Pringle et al., 1999) as well as to etiologic studies of neighborhood effects on health and other aspects of well-being (Brooks-Gunn and Duncan, 1997; Diez-Roux, 1998; MacIntyre and Ellaway, 2000).

GEOCODING AND MONITORING SOCIAL INEQUALITIES IN HEALTH: OPPORTUNITIES AND OBSTACLES—A U.S. PERSPECTIVE

Public health interest in delimiting the boundaries and socioeconomic composition of the neighborhoods in which people live dates back nearly two centuries. At issue is the extent to which neighborhood data are informative about—and relevant to the health of—their inhabitants. In 1826, for example, Louis René Villermé (1782–1863), one of the leading public health professionals in France, used area-based economic data and mortality rates to demonstrate empirically, for the first time, that social conditions—and not just the "natural environment"—played a decisive role in patterning mortality among neighborhoods of Paris (Villermé, 1828; Coleman, 1982). Specifically, he found that death rates were highest in areas whose residents paid the least in "untaxed rents," a type of tax paid only by the wealthy. By contrast, no correlations existed between neighborhood mortality rates and altitude, distance from the Seine, temperature, and other environmental features. Subsequently, in 1845, Friedrich Engels cited data in his classic text, *The Condition of the Working Class in England*, indicating that mortality rates depended on both "class of street" as well as "class of house" (Engels, 1958 (1845); Davey Smith, 1997). Thus, among people who lived in worse homes, mortality was higher for those residing in poorer compared to more affluent neighborhoods. More recently, a revived concern about social inequalities in health in the United States and other countries (Townsend et al., 1990; Evans et al., 1994; Braveman, 1996; Krieger et al., 1997a; Pamuk et al., 1998; Mann et al., 1999; Berkman and Kawachi, 2000; Kim et al., 2000), coupled with new advances in geographic information system (GIS) technologies (Moore and Carpenter, 1999; Richards et al., 1999; Yasnoff and Sondik, 1999), have led to new interest in analyzing population health in relation to neighborhood characteristics, the theme of this book.

In the United States, however, still another factor spurs current interest in using area-based socioeconomic measures: the relative paucity, if not absence, of socioeconomic data in most U.S. public health databases, as reviewed below (Krieger and Fee, 1996; Krieger et al., 1997b). From a public health standpoint, augmenting U.S. vital statistics and other public health surveillance systems with socioeconomic data is especially important because these data provide the fundamental basis for routinely monitoring population burdens of disease and death, along with births and their attendant joys and sometimes sorrow (Institute of Medicine, 1988, 1997). Absent adequate socioeconomic data, we cannot monitor—or gauge progress or setbacks in rectifying—socioeconomic disparities in health, let alone assess the contribution of socioeconomic position to equally troubling racial/ethnic inequalities in health (U.S. Department of Health and Human Services, 1991; Krieger et al., 1993; Williams and Collins, 1995; Krieger et al., 1997a; Freeman, 1998). Beyond this, routine public health data documenting social inequalities in health are vital to shaping the analyses, aspirations, and actions of social movements and governments in their efforts to create healthy public policies, at the local, national, and global levels, to reduce social disparities in health (Fox, 1989; Townsend et al., 1990; Braveman, 1996; Krieger et al., 1997b; Mann et al., 1999; Shaw et al., 2000).

The Problem: Paucity or Absence of Socioeconomic Data in U.S. Vital Statistics

Despite more than two centuries of international vital statistics documenting that people subjected to poorer living and working conditions live shorter, less healthy lives (Antonovsky, 1967; Fox, 1989; Townsend et al., 1990), it was only in 1989 that reporting of birth and death data stratified by educational level became routine in the United States (Tolson et al., 1991)—60 years after the last American attempt, in the 1930s, to produce national mortality data stratified by social class (Whitney, 1934; Krieger and Fee, 1996). Even so, as revealed by a 1997 survey of all U.S. state health departments, although 100% of states now collect information on education of mother and father and decedent, only 7% of cancer registries, 4% of tuberculosis (TB) registries, and no AIDS registries collect data on education (Krieger at al., 1997b). Moreover, although all states collect data on occupation of decedent, and 80% do so for their cancer registries, only 12% and 9%, respectively, break out these data by occupation. Other sources fare worse: only half the states collect—and none publish—birth data in relation to occupation of mother or father, and occupational data in state TB and AIDS registries distinguish only between health care and non–health care occupations. None of these public health databases obtain information on income or other relevant socioeconomic data.

These gaps in data at the state level are reflected at the national level. Before 1998, the annual federal report, *Health, United States*, contained little socioeconomic data (National Center for Health Statistics, 1996; Krieger et al., 1997a; National Center for Health Statistics, 1997). In 1998, however, this report included its first-ever chartbook on "Socioeconomic Status and Health," drawing on data from birth and death records plus the National Health Interview Survey (Pamuk et al., 1998). An absence of socioeconomic data in other databases, however, meant critical public health problems, such as cancer, TB, and HIV/AIDS, could not be included. The 1999 issue unfortunately reverted to earlier practice, with socioeconomic data once again appearing in only ten of the report's seventy-three tables on "Health Status and Determinants" (National Center for Health Statistics, 1999). Moreover, fully 70% of the 467 U.S. public health objectives for the year 2010 lack quantitative targets for reducing socioeconomic disparities in health, given a lack of baseline data (U.S. Department of Health and Human Services, 2000), nor will this problem be rectified by better national surveys. First, as acknowledged by the National Center for Health Statistics, such surveys offer sparse or no data at the local level, the principal locale for most public health planning and programs (Meriwether 1996; Pollack and Rice, 1997; U.S. Department of Health and Human Services, 2000). Second, most surveys are not designed to provide routine data on diverse and relatively small racial/ethnic populations (e.g., subpopulations of Asians and Pacific Islanders, Hispanics, and American Indians) (U.S. Department of Health and Human Services, 1991; Williams and Collins, 1995; Nolan et al., 1996; Krieger et al., 1997b).

Geocoding and Area-Based Socioeconomic Measures: A Potential Solution

Fortunately, the problem is not intractable. In fact, tracts are likely to be part of the solution. This is because we can use census-derived, area-based data to characterize persons in both public health databases and the total population, thereby permitting ascertainment of population-based incidence or prevalence rates stratified by area-based socioeconomic position (Krieger et al., 1992, 1997a). The key to this strategy is the methodology of geocoding.

Geocoding refers to the process of identifying an address's latitude, longitude, and assigned geographic codes, which in the United States include its census-defined state, county, census tract, and block-group codes, along with codes for political jurisdictions (e.g., congressional districts), economic regions, plus post office–defined ZIP codes (Kaplan and Van Valey 1980; U.S. Department of Commerce 1990; U.S. Bureau of the Census, 1991). As shown in Figure 7–1, the basic geographic building blocks in the United States are literally the census blocks, with an average population of 85, nested within the census block-group (average population of 1,000), in turn nested within the census tract (average population of 4,000) (Kaplan and Van Valey, 1980; U.S. Department of Commerce, 1990). Census tracts were first demarcated in New York City in 1910 and were assigned to several major U.S. cities by 1940; as of 1990 block-group codes were assigned nationwide (Shevky and Bell, 1955;

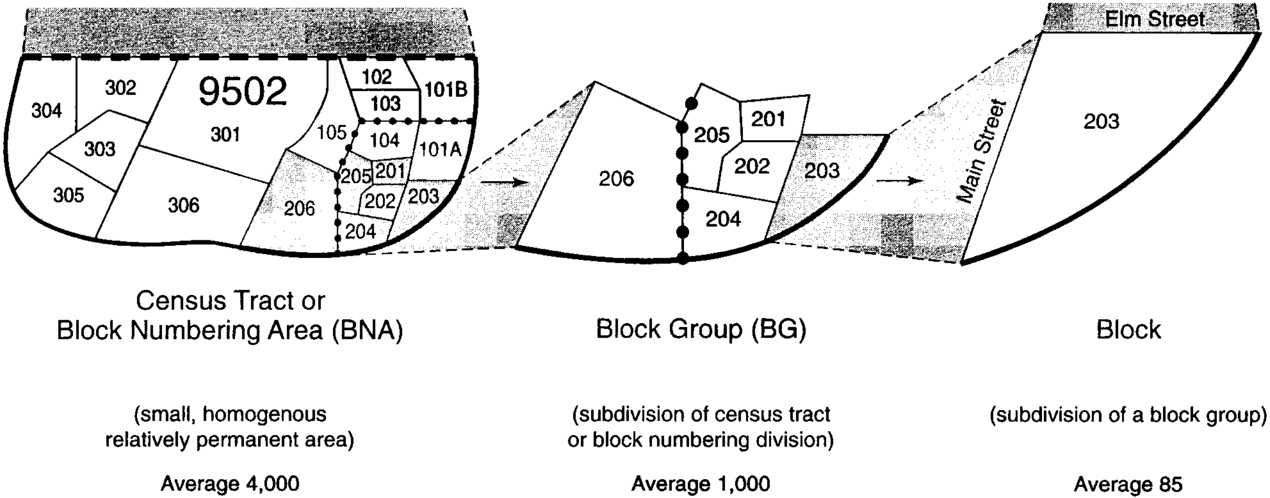

FIGURE 7–1. 1990 census geography: census tracts, block groups, and blocks (U.S. Department of Commerce, 1994). *Source:* US Bureau of the Census. Geographical areas reference manual. Washington, D.C.: US Department of Commerce, 1994.

Kaplan and Van Valey, 1980). These areas are intended to contain populations reasonably homogeneous with regard to socioeconomic composition. Block-groups constitute the smallest unit for which area-based socioeconomic measures are feasible, because, to protect confidentiality, no socioeconomic data are released at the block level (U.S. Department of Commerce, 1990; U.S. Bureau of the Census, 1991). By contrast, ZIP codes typically contain upwards of 30,000 people, are intended to facilitate delivery of mail, often cut across census tract and block-group boundaries, and are not designed to be economically homogeneous (Kaplan and Van Valey, 1980; U.S. Bureau of the Census, 1991; U.S. Post Office, 2000).

Three strengths of using area-based socioeconomic measures for monitoring social inequalities in health are that they (*1*) can be appended to any database with addresses, as is the case for key vital statistics and other public health surveillance databases (Krieger et al., 1997a, 1997b; Yasnoff and Sondik, 1999; Thrall, 1999a), (*2*) provide data for determining contextual as well as compositional neighborhood effects on health, above and beyond effects that are due to individual-level socioeconomic position (Diez-Roux, 1998; MacIntyre and Ellaway, 2000), and (*3*) can be applied equally to all persons, regardless of age, gender, and employment status (Krieger et al., 1997a; Krieger, 1992; Carstairs and Morris 1989). This methodology thereby avoids well-known problems associated with individual-level education and occupation data, that is, how to classify people who have not completed their education or who are not in the paid-labor force (children, housewives, househusbands, unemployed, or retired persons) (Carstairs and Morris, 1989; Sorensen, 1994; Krieger et al., 1997a, 1999). Three drawbacks are (*1*) they can be misconstrued as only a "proxy" for individual-level socioeconomic data (rather than seen as complementary data that, in fact, can be analyzed together with individual-level data in multilevel models), (*2*) they reflect socioeconomic context at the time of case ascertainment, not necessarily during the relevant etiologic period, and (*3*) they can be outdated, given the decennial nature of the U.S. census. Rendering this last objection moot, however, the U.S. is already shifting toward implementing the more frequent American Community Survey, which will generate data at the tract and block-group level (U.S. Census Bureau, 1999).

Given a wide array of census data, a variety of geographic levels (such as the census tract, block-group, and ZIP code), and a plethora of geocoding and other GIS software (Moore and Carpenter, 1999; Richards et al., 1999; Thrall, 1999b), the question then becomes: are there any standard, well-accepted area-based socioeconomic measures that have been validated for public health monitoring in the United States? Briefly stated, "No" (Krieger et al., 1997a; Lynch and Kaplan, 2000).

As of mid-2000, only seven U.S. public health studies have explored the validity of using area-based socioeconomic measures (Krieger, 1991, 1992; Cherkin, 1992; Greenwald et al., 1994; Krieger et al., 1996; Geronimus et al., 1996; Geronimus and Bound, 1998). Unfortunately, all used different census-derived socioeconomic variables and examined only a handful of health outcomes. All compared socioeconomic gradients in health detected by area-based and individual-level socioeconomic measures to assess the validity of using area-based measures to monitor socioeconomic gradients in health. Yielding results similar to the handful of comparable methodologic studies from the United Kingdom (Carr-Hill and Rice, 1995), Canada (Mustard et al., 1999) and Australia (Hyndman et al., 1995), five of the U.S. studies found that estimates of socioeconomic gradients in health produced with block-group or census tract socioeconomic measures were similar to, but often slightly less than, those produced with individual-level socioeconomic data (Krieger, 1991, 1992; Cherkin, 1992; Greenwald, 1994; Krieger et al., 1996); results were less consistent for the two studies that employed ZIP-code data (Geronimus et al., 1996; Geronimus and Bound, 1998; see also Krieger and Gordon, 1999; Davey Smith and Hart, 1999). Among these seven studies, four used differently categorized measures of neighborhood social class composition, educational and poverty level, and unemployment rate (Krieger, 1991, 1992; Krieger et al., 1996; Geronimus and Bound, 1998), two used measures of average annual family income (Greenwald et al., 1994; Cherkin et al., 1992), and two used data on median family income and educational level (Geronimus and Bound, 1996; Geronimus and Bound, 1998).

A profusion of measures is likewise apparent in the small but growing U.S. public health literature with demonstrated interest in using area-based socioeconomic measures. Employing an extremely eclectic array of census-derived variables at markedly different levels of geography, these investigations have focused on such diverse outcomes as births (Collins and David, 1997; O'Campo et al., 1997; Roberts, 1997; Wasserman et al., 1998), deaths (Kitagawa and Hauser, 1973; Hann et al., 1987; Davey Smith et al., 1996a, 1996b), cancer (Devessa and Diamond, 1980; Krieger, 1990; Greenwald et al., 1994; Liu et al., 1998; Prehn and West, 1998; Krieger et al., 1999; Morris, 1999; Arbes, 1999), cardiovascular disease (Casper et al., 1999; Iwashyna et al., 1999; Sayegh et al., 1999), AIDS and other infectious diseases (Zierler et al., 2000, Morse et al., 1991; Fife et al., 1992; Hu et al., 1994; Ellen et al., 1995; Simon et al., 1995; Chen et al., 1998), asthma (Wissow et al., 1988), mental health (Goldsmith et al., 1982), and violence and injuries (Collins and David, 1997; Harries, 1997; Anderson et al., 1998; Grisso et al., 1999; Hinton et al., 1999; Powell and Tanz, 1999). However, despite a common concern about socioeconomic inequalities in health, estimates of effect for socioeconomic position cannot readily be compared across these studies because of the lack of consistency in area-based socioeconomic measures. The sociologic literature is similarly rife with a heterogeneous assortment of measures developed over the past several decades, none systematically assessed for their relevance to health (Shevky and Bell, 1955; Rossi and Gilmartin, 1980; White, 1987; Brooks-Gunn et al., 1997). Moreover, with regard to government standards, the only federally designated demarcation is for "poverty area," defined as an area where 20% or more of the population lives below the poverty line (U.S. Bureau of Census, 1985; Jargowsky, 1997). By contrast, health research and government reports in the United Kingdom employ a small set of validated and well-established area-based measures of deprivation, with key ones—such as the Townsend index—explicitly developed for public health use (Townsend et al., 1988; Townsend, 1993; Carstairs, 1995; Gordon, 1995; Lee et al., 1995; see also Chapter 8).

In summary, in the United States, public health researchers and practitioners confront uncertainty about three important aspects of area-based socioeconomic measures: (*1*) validity, (*2*) content—including both which measures are included and how they are constructed, and (*3*) geographic level. No standard validated measure exists, whether for use by public health departments or by individual researchers.

GENERATING SOLUTIONS: A PROJECT TO DEVELOP VALID AREA-BASED SOCIOECONOMIC MEASURES FOR U.S. PUBLIC HEALTH DATABASES

To address this gap in U.S. public health research and practice, we are engaged in a project whose goal is to generate a valid, robust, easy to interpret, and easy to construct area-based socioeconomic measure that can readily be used by health departments anywhere in the United States, for any health outcome, from birth to death, whether for women or men, young or old, among any racial/ethnic group (Krieger et al., NIH Grant 1 R01 HD36865-01). Throughout, our guiding principle is to generate area-based socioeconomic measures that are grounded in sound statistical methodology yet whose meaning is readily evident to both data users and the public at large.

In presenting preliminary results of our project, five lessons stand out that are relevant to any project engaged with geocoding and/or using census-derived socioeconomic data. First, it is critical to ascertain the accuracy of geocoding software. Second, it is essential to develop protocols that protect confidentiality of the databases that are being geocoded. Third, it is important to evaluate the extent to which different databases can be geocoded to specified geographic levels. Fourth, it is necessary to be aware of how boundaries for different geographic regions are defined and may cross-cut one another. Fifth, it is imperative that analyses be conceptually grounded rather than algorithm-driven empirical exercises in data reduction.

Scope of Project

To evaluate the validity and feasibility of monitoring social inequalities in health using area-based socioeconomic measures, our project is using public health surveillance data from two states, Massachusetts and Rhode Island, along with 1990 census data from three geographic levels: census block-group (BG), census tract (CT), and ZIP code (ZC). The study's five main tasks are to (*1*) geocode the public health databases to the BG, CT, and ZC level, (2) at each geographic level generate the same set of area-based socioeconomic indicators, (3) within and across geographic levels examine patterns of associations between these variables and, where feasible, with individual-level socioeconomic data, (4) link the area-based socioeconomic measures to the geocoded health databases, and (5) estimate and compare socioeconomic gradients in health detected for each health outcome, using the same set of diverse area-based socioeconomic measures at each geographic level.

Public Health Databases: Eligibility and Confidentiality

For the purpose of our project, we defined as eligible all public health surveillance system databases in the states of Massachusetts and Rhode Island that included age, gender, race/ethnicity, and street address at the time of case ascertainment. Staff from the Massachusetts Department of Public Health and the Rhode Island Department of Health provided information on the eligibility of their surveillance system data for our project. Of the more than fifty surveillance systems surveyed, nine met eligibility criteria: birth, death, cancer, tuberculosis, sexually transmitted diseases, nonfatal gun and stab wounds, domestic violence, childhood lead screening, and the Health Interview Survey. Although HIV/AIDS surveillance systems also were eligible, we excluded them because the Massachusetts data were already being used in a pilot study conducted by two of our team's investigators (Zierler et al., 2000). In total, our project has health outcome data for 650,012 residents of Massachusetts and 324,078 residents of Rhode Island for the proposed years of monitoring surrounding the census year 1990, thus nearly 1 million records combined.

Confidentiality protocols were set by the respective health department's internal review boards, with input from the directors of each database and from project staff. To preserve confidentiality, for each database we created files for geocoding containing only each record's address, a dummy variable to identify the data source, and a newly generated unique identifier which the health department could then use to link the geocoded record back to the original database. Working with staff at each health department, we then created databases for analysis containing solely the relevant demographic and health data plus newly assigned geocodes. Thus, no street addresses are in any of our project's databases, except for the public domain death certificate data.

Selecting Geocoding Tools: Options and Accuracy

Recognizing that numerous companies and software products claim to geocode accurately, we decided to put the claims to the test (see Krieger et al., 2001, on p. 170). Although we located a few articles evaluating the capacity and user-friendliness of several commercial geocoding software products (Thrall, 1999b; Richards et al., 1999; Moore and Carpenter, 1999), we found none that explicitly compared the cost, timeliness, and accuracy of geocoding firms and software.

After searching the Internet to determine the range of services offered by commercial geocoding firms and ascertaining which software products were licensed to our workplace, we selected four firms and one software program. To each we submitted (*1*) information on the number of records we wanted to geocode (six submissions of 250,000 records each) and (2) seventy test addresses with known census tracts, consisting of fifty "incorrect" but geocodable addresses provided by a geocoding specialist at the Massachusetts Cancer Registry, which we supplemented with twenty telephone book addresses randomly selected from the local telephone book. The "incorrect" addresses contained various common errors, including out-of-range street addresses, abbreviated or misspelled street names, wrong street types (e.g., "avenue" instead of "street"), and correct towns but wrong zip codes. Census tracts of the fifty "incorrect" addresses were earlier verified by staff at the cancer registry; tracts of the twenty "telephone book" addresses were verified by project staff using the U.S. Bureau of the Census geocoding web site (U.S. Bureau of the Census, 2000).

Notably, the four firms and the software program varied dramatically in cost, timeliness, attitude, and, most importantly, accuracy. Estimated costs among the firms varied by a factor of two, ranging from $8,800 to $15,800. Project costs for the software program were lower ($1,460), but this price included only the cost of training two project staff members, not the cost of their labor for geocoding addresses nor the price of the software (which had been licensed to our university). If, however, we purchased the necessary software, costs would have increased by an additional $1,690 at the time of ascertaining this information for the test file. (At the time of preparing this chapter, however, the cost would have risen by an additional $11,190 due to increased expenses associated with upgrading or updating the software.)

Representatives for each firm and program estimated it would take one to two weeks to geocode the test file; actual turnaround time ranged from five hours to three weeks. Company A, the least expensive, friendliest, and suitably prompt geocoding firm did best, correctly geocoding 84% of the test addresses (80% of the "incorrect" addresses, 95% of the telephone book addresses). Company D, by contrast, despite costing twice as much, was the worst: it took 4 times as long, provided poor customer service, and correctly geocoded only 44% of the test addresses (36% of the "incorrect" addresses, 65% of the telephone book addresses).

Based on these results, we selected company A and sent them our files. Then, to evaluate the "real world" accuracy of the firm, we randomly selected 150 geocoded addresses from the public domain death certificate data. Our strategy was to compare company A's results to (*1*) the gold standard, defined as the census block-group maps generated by the U.S. Census Bureau, using actual maps and guides with street address ranges, (*2*) the U.S. Census Bureau Website (which permits geocoding one manually-entered record at a time, but only to the tract level) (U.S. Bureau of the Census, 2000) , and (*3*) the software program, which we nevertheless decided to use for its map-making abilities. We also drove by several of the selected addresses to ascertain their exact location. Company A continued to do an impressively accurate job, with the overall percentage of addresses geocoded correctly to the tract level equaling 96%, the same as that of the U.S. Census Bureau web site (95%). In doing this exercise, however, we also encountered one additional disturbing result: on several occasions the software program could accurately geocode an address but then mapped it to the wrong part of town.

The marked variability in accuracy of geocoding products accordingly raises several important issues. Notably, none of the public health studies we have seen using geocoded data have included their rationale for selecting a particular geocoding firm or software product, nor have any presented data on the accuracy of the selected geocoding methodology. Our results raise questions about the validity of the results of previously published studies using geocoding methodologies of unverified accuracy.

Selecting Census Data: Options and Boundaries

To obtain the original, uncompressed raw census data at the block-group and census tract level, we downloaded data from the U.S. Census Bureau

Summary Tape File (STF) 3A; the ZIP code data are in STF 3B (U.S. Bureau of the Census, 1991). We also obtained data from a menu-driven commercial product, which, although intended to be user-friendly, provided little documentation (Geolytics, 2000). Reassuringly, both sources provided identical data for a series of randomly selected block-groups, tracts, and ZIP codes.

In working with the ZIP code data, however, we learned that their boundaries stretch the geographic imagination and can pose interesting problems for geocoding projects (see Krieger et al., 2002, on pp. 170–171). First, unlike block-groups and census tracts, which are physically delimited geographic areas with official and public boundaries demarcated by the U.S. Census Bureau (U.S. Bureau of the Census, 1991; Kaplan and Van Valey, 1980), ZIP code boundaries are far more fluid. As described by staff at the U.S. Census Bureau, these boundaries "resemble spaghetti and follow delivery routes" (Stuber, 1999) and can also be "point locations (as in rural post office buildings)" (Nichols, 1999). Consequently, "most companies that make maps of ZIP codes interpolate where a boundary is by talking to local postmasters or using street and address databases to come up with some approximation" (Nichols, 1999), and "the imputed boundaries are not recognized or used by the U.S. Postal Service" (Stuber, 1999). Thus, without publicly available official ZIP code boundaries, ZIP code borders may vary by geocoding firm and software product (U.S. Post Office, 2000). Second, we also learned that because cross-county ZIP codes are not assigned state codes in the STF 3B files, selecting on ZIP codes by state can give misleading impressions of missing data, because neither cross-county nor cross-state ZIP codes will be selected.

Census-Derived Socioeconomic Measures: Conceptual and Operational Definitions

Before analyzing the actual census socioeconomic data, we mapped out six conceptual domains relevant to characterizing socioeconomic position (Krieger et al., 1997; Council of Economic Advisors for the President's Initiative on Race, 1998). Our intent was to think through which types of measures might be pertinent before using statistical methods to characterize relationships between these measures. We then operationalized each domain in terms of available census variables, as shown in Table 7–1. These domains and the relevant variables are

1. *Occupational class*, defined in terms of (*a*) percent of employed persons age 16 and older in working-class jobs, made up chiefly of nonsuper-

TABLE 7–1. Selected U.S. Census-Derived Area-Based Socioeconomic Measures: Conceptual and Operational Definitions

Variable: conceptual definition	*Census Text Definition: numerator ÷ denominator*	*1990 Census Code Definition: numerator ÷ denominator*
Working class: nonsupervisory employees	Σ (persons employed in the 8 working class census occupational categories) ÷ Σ (all 13 census occupational categories)	(P0780004 + . . . + P0780006 + P0780008 + P0780010 + . . . + P0780013) ÷ (P0780001 + . . . + P0780013)
Unemployment: persons actively seeking employment (as defined by the census)	# persons age 16+ in the labor force and unemployed ÷ # persons age 16+ in the labor force	(P0710003 + P0710007 + P0710011 + P0710015 + P0710019 + P0710023 + P0710027 + P0710031 + P0710035 + P0710039) ÷ (P0710002 + P0710003 + P0710006 + P0710007 + P0710010 + P0710011 + P0710014 + P0710015 + P0710018 + P0710019 + P0710022 + P0710023 + P0710026 + P0710027 + P0710030 + P0710031 + P0710034 + P0710035 + P0710038 + P0710039)
Low income: relative measure of poverty, set at half the national median income (European Union definition of poverty)	# households with income <$15,000 ÷ # households	(P0800001 + . . . + P0800005) ÷ (P0800001 + . . . + P0800025)
Median household income: income of households where half of households have income below this level and half have income above it	median household income	P080A
High income: income at least 5 times higher than the median income	# households with income $150,000 + ÷ # households	P0800025 ÷ (P0800001 + . . . + P0800025)

(continued)

TABLE 7–1. Selected U.S. Census-Derived Area-Based Socioeconomic Measures: Conceptual and Operational Definitions (Ccontinued)

Variable: conceptual definition	*Census Text Definition: numerator ÷ denominator*	*1990 Census Code Definition: numerator ÷ denominator*
Presence of poverty: persons or households whose household income is below the threshold to cover basic needs, definedin relation to food (US census definition)	# persons for whom ratio of (1989 household_income/U.S. poverty line) is <0.5 ÷ # persons for whom poverty status is determined	P1210001 ÷ (P1210001 + . . . + P1210009)
	# persons with 1989 household income <U.S. poverty line ÷ # persons for whom poverty status is determined	(P1170013 + . . . + P1170024) ÷ (P0170001 + . . . + P0170024)
	# households with 1989 income <U.S. poverty line ÷ # households	(P1270016 + . . . + P1270030) ÷ (P1270001 + . . . + P1270030)
Absence of poverty: persons whose household income is two times more than the poverty line	# persons for whom ratio of (1989 household_income/US poverty line) is ≥2 ÷ # persons for whom poverty status is determined	P1210009 ÷ (P1210001 + . . . + P1210009)
Low assets: assets sufficient to own only housing stock of very low value	# housing units valued at <$50,000 ÷ # specified owner-occupied housing units	(H0610001 + . . . + H0610008) ÷ (H0610001 + . . . + H0610020)
High assets: sufficient assets to own a home at about 4× the median U.S. value of owner-occupied homes	# housing units valued at $300,000+ ÷ # specified owner-occupied housing units	(H0610018 + H0610019 + H0610020) ÷ (H0610001 + . . . + H0610020)
Less than high school: adults with low educational attainment	# persons aged 25+ with no high school diploma ÷ # persons aged 25+	(P0570001 + P0570002) ÷ (P0570001 + . . . + P0570007)
College: adults with high educational attainment	# persons aged 25+ with 4+ years of college ÷ # persons aged 25+	(P0570006 + P0570007) ÷ (P0570001 + . . . + P0570007)
Crowding: at household level, more people than rooms	# occupied housing units with >1 person per room ÷ # occupied housing units	(H0690002 + H0690003 + H0690005 + H069006 + H0690008 + H0690009 + H0690011 + H0690012) ÷ (H0690001 + . . . + H0690012)

visory employees, represented by eight of the thirteen census-defined occupational categories (Krieger et al., 1997a; Wright, 1997; Wright et al., 1982) and (*b*) percent of persons age 16 and older who are unemployed;
2. *Income*, defined in terms of (*a*) median annual household income, (*b*) low income, referring to the percentage of households with annual income below half the annual median income, a measure of poverty commonly used in the European Union (Gordon and Spicker, 1999) and equivalent to an annual household income below $15,000 in 1989, and (*c*) high income, referring to the percentage of households with an annual income of $150,000 or higher, the highest income category reported by the U.S. Census;
3. *Poverty*, defined in terms of (*a*) the percentage of persons with household incomes below the U.S. federal poverty line, a threshold that varies by size and age composition of the household, and on average equaled $12,647 for a family of four in 1989 (U.S. Bureau of the Census, 1991), (*b*) extreme poverty, referring to the percentage of persons in households below 50% of the poverty line, and (*c*) not impoverished, referring to the percentage of persons with household incomes at least 200% higher than the poverty line;
4. *Wealth*, defined in terms of (*a*) high assets, referring to the percentage of owner-occupied homes valued at $300,000 or more and (*b*) low assets, referring to the percentage of owner-occupied homes valued at less than $50,000;
5. *Education*, defined in terms of (*a*) low education, referring to the percentage of adults age 25 and older who have not completed high school and (*b*) high education, referring to the percentage of adults age 25 and older who have completed at least 4 years of college education; and
6. *Crowding*, defined in terms of (*a*) household crowding, referring to the percentage of households with more than one person per room and (*b*) population density, a contextual variable.

We also considered two additional variables—lack of car ownership and percentage of rented households—because they are used as indicators of deprivation in the United Kingdom (Lee et al., 1995). Emphasizing the importance of context, however, we found neither variable to be consistently associated with either poverty or median household income, given the much higher rates of car and home ownership in the U.S.

Additionally, to explore characteristics of selected socioeconomic indexes employed in other studies, we created several measures incorporating data on one or more of the single-variable census measures. These included

1. *U.S. versions of two U.K. area-based deprivation indexes*, the Townsend and Carstairs indexes, which combine data on such characteristics as unemployment, lack of car ownership, household crowding, rented households, and low social class (Lee et al., 1995; Carstairs, 1995; Townsend, 1993);
2. *Two measures of income inequality*: (*a*) the Gini coefficient, and (*b*) the ratio of the median income of the top fifth to the bottom fifth of the income distribution (Bernstein et al., 2000);
3. *The Index of Local Economic Resources* employed in the recent Centers for Disease Control (CDC) atlas on cardiovascular mortality, which combines data on the percentage of persons unemployed, in white collar jobs, and median family income (Casper et al., 1999).

We also combined several of the single census indicators into several theoretically driven *a priori composite indicators*, that is, a measure classifying areas simultaneously in terms of occupational class, wealth, and poverty levels. The final step was then to generate all of the specified variables at each level of geography (BG, CT, and ZC). Once accomplished, we then did a visual "reality check" on our constructs by visiting several Boston block-groups represented in the Massachusetts death certificate data and identified by our data as being poor, middle-income, or affluent.

PRELIMINARY RESULTS: GEOCODING AND PATTERNS OF ASSOCIATION AMONG SOCIOECONOMIC DATA

Importantly, from the standpoint of the feasibility of using area-based socioeconomic measures for monitoring social inequalities in health, the success rate of geocoding the administrative surveillance databases was suitably high. As shown in Table 7–2, of the nearly 1 million records we submitted for geocoding, fully 92% were geocoded successfully to the census block-group level. Another 6% could not be geocoded to the block-group level but were successfully geocoded to the census tract level, whereas only 0.2% could be geocoded only to the ZIP code level. Among all the addresses submitted, only 1.8% could not be geocoded. Although these percentages varied somewhat among the thirteen surveillance systems and one survey database, the median percentage of addresses geocoded to at least the census tract level equaled 97.2%.

Our data analysis strategy is comprised of three parts: data exploration, model building, and confirmatory analyses. In this section we provide detailed results of our exploratory analysis of census variables. Data exploration focuses on understanding important features of variable distributions at each aggregation level from both the census and surveillance

TABLE 7–2. Feasibility of Geocoding Public Health Surveillance System Databases

		% Geocoded to:			
Database	*N*	*BG*	*CT only*	*ZIP only*	*% Not Geocoded*
Total	**970,086**	**92%**	**6%**	**0.2%**	**1.8%**
Massachusetts					
Birth	267,724	95	5	0	0
Death	162,071	92	8	0	1
Cancer	170,004	92	8	0	0
Gun/stab wounds	7,723	83	10	0	7
TB + STDs	42,490	75	23	0	2
Rhode Island					
Birth	106,443	94	3	1	2
Death	28,942	90	5	0	5
Cancer	21,392	92	8	0	0
Domestic violence	15,222	94	5	0	0
TB	576	91	3	0	7
STDs	9,353	88	1	1	10
Lead	135,567	91	1	0	9
HIS	2,579	75	0	7	18

*Years vary for each database, with years selected: (a) to be temporally proximal to the 1990 census, and (b) to ensure adequate sample size for calculation of stable rates.

TB, tuberculosis; STDs, sexually transmitted diseases; HIS, health interview survey; BG, block-group; CT, census tract; ZC, zip code.

data. For the census data we review analyses that involve both numerical and graphic summaries, stratified by level of aggregation and by state. These include univariate summaries, bivariate associations, and associations between individual census variables and the selected existing area-based socioeconomic indexes. Univariate summaries indicate whether census variables cluster around a particular value or set of values, identify possibly outlying observations and the extent of missing data, and provide information useful for identifying potentially meaningful cutpoints or functional transformations (e.g., log or square root) needed for statistical modeling. They also permit comparisons of variability and distribution across levels of aggregation.

Figure 7–2 shows the distribution of several selected census variables using box and whisker plots. Heavy right skewness is evident is several variables, such as those related to poverty and wealth. The majority of

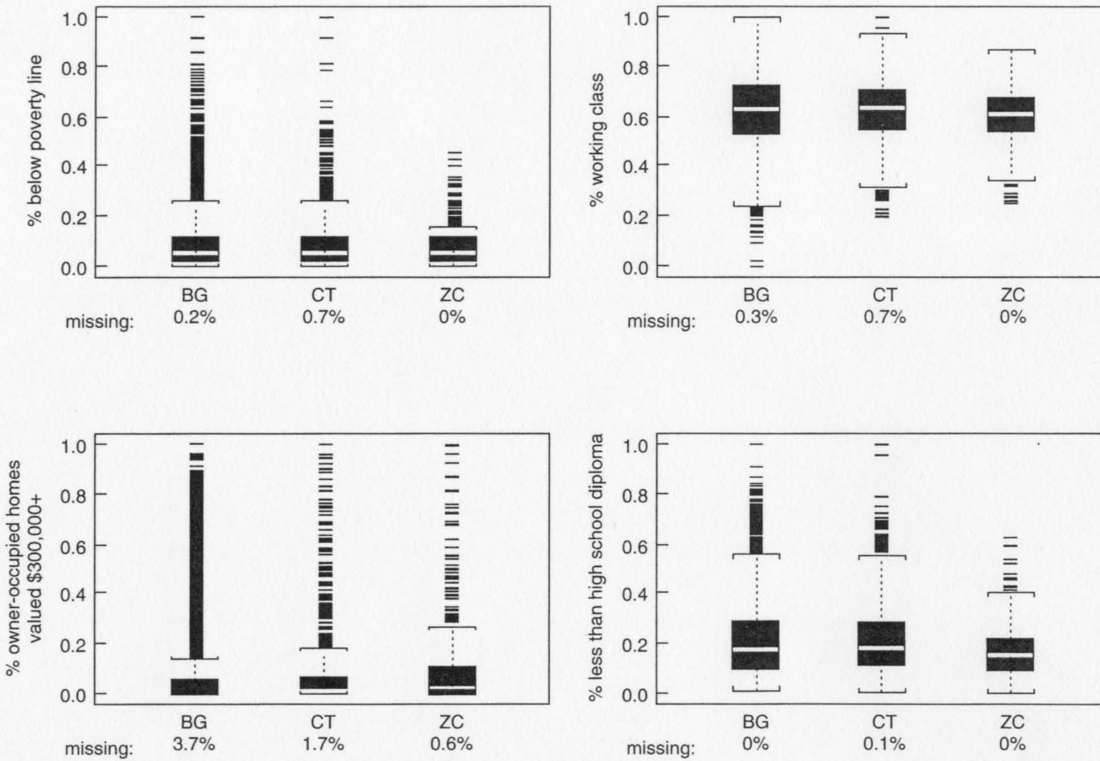

FIGURE 7–2. Boxplots showing univariate distributions and percent missing of selected socioeconomic U.S. Census variables (1990) for Massachusetts block groups, census tracts, and ZIP codes.

variables show little appreciable difference in either shape or central tendency across levels of aggregation. Specifically, quantiles defining the middle 50% of the distribution remain relatively constant across aggregation unit. As the aggregation units get larger, however, some "smoothing" is evident, reflected by attenuation of the minimum and maximum values in larger aggregation units. This phenomenon occurs because larger aggregation units encompass the smaller ones and average their values. It is interesting, however, to compare the observed range of values in poverty and wealth: although many block-groups and census tracts have poverty rates in excess of 50%, the maximum poverty rate in ZIP codes is 46%. By contrast, the full range of values for wealth appears largely to be preserved across levels of aggregation. Thus, aggregation by ZIP code preserves the full range of values for indicating extreme wealth, but extreme poverty appears to be "smoothed out."

We used Spearman's rank correlations and two-way scatterplots to summarize associations between individual census variables at each aggregation level. (Rank correlations are preferred due to skewed distributions.) The goals of the association analysis are to understand whether patterns of association are consistent across level of aggregation and whether groups of variables tend to cluster together to delineate underlying, latent area-level characteristics. The second goal can and will be addressed more formally using factor analysis (Bartholomew and Knott, 1999, Chapter 3).

Table 7–3 lists pairwise correlations by level of aggregation; the boldface is used to indicate clusters of variables with high pairwise correlation (e.g., greater than 0.7 in absolute value). Although correlation is relatively constant across aggregation level, the clusters tend to emerge with greater definition at higher levels of aggregation. The first cluster characterizes economic resources in relation to income (or lack thereof) and includes all the selected income and poverty measures. A second cluster, slightly more difficult to discern, captures aspects of wealth and social class and includes variables pertaining to class composition, assets, and educational level. Importantly, for both clusters the direction of association remains the same across levels of aggregation, even as the degree of association sometimes changes.

The third component of our exploratory analysis examines the nature of association between individual census variables and the selected existing area-based socioeconomic indexes. This analysis is undertaken to help understand the extent to which existing indexes reflect variation in important census indicators. As an illustration, Figure 7–3 shows scatterplots of poverty rate versus three area-based socioeconomic indexes: the Townsend index, the Gini coefficient of income inequality, and the CDC's

TABLE 7–3. Spearman Rank Correlation Matrix of Selected Socioeconomic U.S. Census Variables (1990) for Massachusetts Census Tracts, Showing Two Clusters of Highly Correlated Variables*

	Working Class (%)	*Unemployed (%)*	*Median Household (HH) Income*	*<50% Median HH Income*	*HH Income ≥$150,000*	*Persons < Poverty (%)*	*Persons ≥200% Poverty (%)*	*Owner-Occupied Homes ≥$300,000 (%)*	*Adults < High School (%)*	*Adults ≥ 4 yrs College (%)*	*Crowded HH (%)*
% unemployed	***0.614***										
Median household (HH) income	**−0.686**	**−0.650**									
<50% median HH income	**0.622**	**0.638**	**−0.938**								
HH income ≥ $150,000	**−0.711**	−0.576	**0.691**	**−0.61 5**							
% persons < poverty	0.522	**0.617**	**−0.857**	**0.882**	−0.557						
% persons ≥ 200% poverty	**−0.625**	**−0.654**	**0.936**	**−0.919**	**0.643**	**−0.921**					
% owner-occupied homes ≥ $300,000	***−0.684***	−0.487	**0.602**	−0.5 23	**0.702**	−0.470	0.545				
% adults < high school	*0.774*	*0.693*	*−0.768*	*0.755*	*−0.682*	*0.670*	*−0.742*	*−0.658*			
% adults ≥ 4 yrs college	*−0.906*	*−0.670*	*0.668*	*−0.608*	*0.707*	−0.503	*0.605*	*0.702*	*−0.826*		
% crowded HH	0.463	0.531	−0.608	0.579	−0.482	**0.668**	**−0.674**	−0.416	0.545	−0.438	
Population per square mile	0.216	0.347	−0.490	0.519	−0.345	0.593	−0.549	−0.293	0.395	−0.203	0.523

*Correlations >0.6 highlighted in bold identify a cluster of varibales pertaining to economic resources, in relation to poverty and income.

Correlations >0.6 highlighted in italics identify a cluster of variables related to social class, education, and wealth.

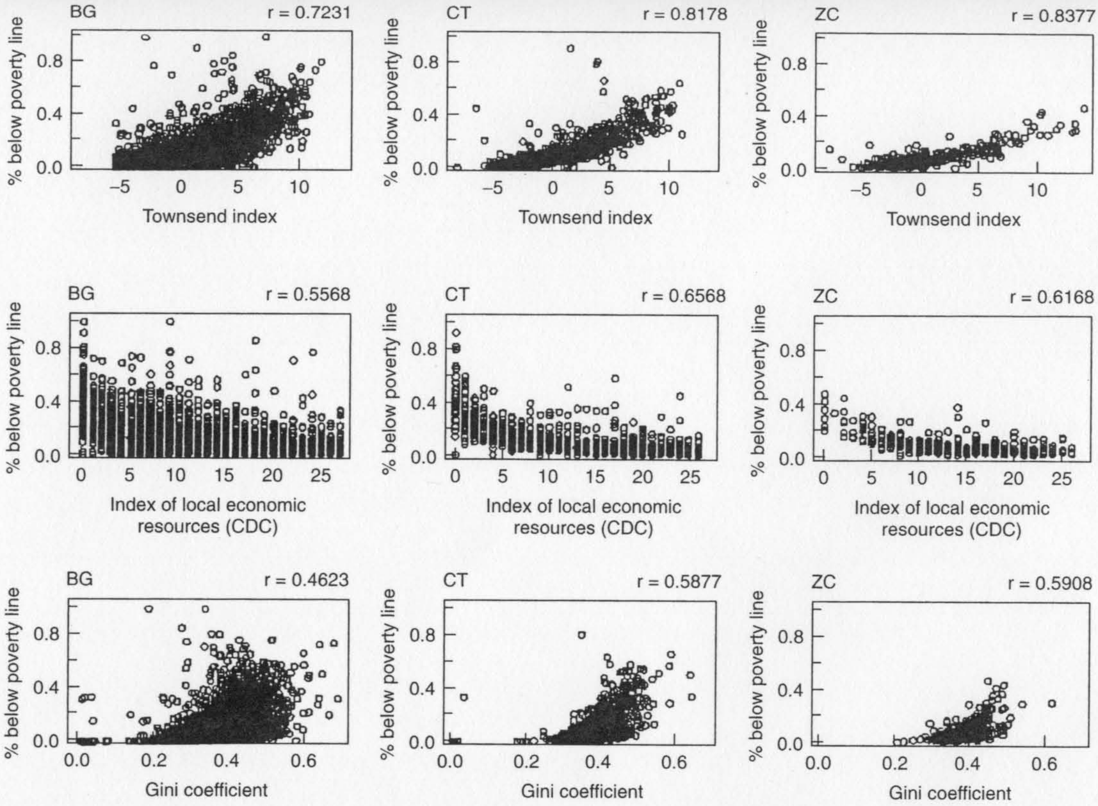

FIGURE 7–3. Scatterplots of poverty rate vs. Townsend index, Gini coefficient, and the CDC Index of Local Economic Resources, for Massachusetts block groups, census tracts, and ZIP codes using 1990 U.S. Census data.

Index of Local Economic Resources. Two features are readily apparent. First, as expected, the variability of each index at a fixed percentage of poverty (i.e., vertical variation in the plot) decreases as the area of aggregation increases. Second, at any given level of poverty, the variability in each index is quite substantial, even at the extremes, and especially for the CDC index. The high variation in the CDC index across fixed levels of poverty may indicate that it is not aptly named. The Gini coefficient also exhibits considerable variability at fixed poverty levels, but this is to be expected because the Gini coefficient measures economic disparity within an area and not per se (i.e., an area with uniform poverty can have the same Gini coefficient as does one with uniform wealth). This may be obvious but points to why a given measure may have different utility at different levels of aggregation: given economic segregation, the Gini coefficient and other measures of economic inequality are likely to be informative only at higher levels of aggregation (Jargowsky, 1997; Soobader and LeClere, 1999).

Our next steps involve both model building and confirmatory analyses. Building on our correlation analyses, we use factor analysis to identify subsets of variables that cluster in particular domains and use information on latent factors in our eventual construction of the index (Bartholomew and Knott, 1999, pp. 42 ff.). We also address two important issues that make our analysis slightly more complex than simply applying factor analysis to the census variables. First, census variables for adjacent geographic areas are likely to be spatially correlated, and incorporating this correlation is likely to improve our estimation of area-specific factors (Clayton and Kaldor, 1987; Moore and Carpenter, 1999). Second, especially at the ZIP code level, census data are derived from differently sized populations, making it necessary to incorporate appropriate weights.

Having obtained factor scores in addition to our single-variable socioeconomic measures and other existing indexes, we turn to the question of socioeconomic gradients in health outcomes. A standard approach for addressing this question is Poisson regression, appropriately generalized to deal with potential spatial correlation (e.g., Breslow and Clayton, 1993), in which the dependent variable is the number of events in an area (standardized by the population size) and the independent variables include various area-based socioeconomic measures. This modeling approach permits evaluating the utility of our various socioeconomic measures, both within and across levels of geography and across the different health outcomes on the bases of predictive accuracy, relevance to public health monitoring, transparency of meaning, and ease of use by government agencies that eventually will use these measures.

CONCLUSION

In conclusion, we believe our project affords an example of why it is useful to remember the basic fact that we each live in neighborhoods and that data on socioeconomic characteristics of these neighborhoods—whether compositional or contextual—can be useful for describing and monitoring social inequalities in health. This insight is not new: the first peer-reviewed U.S. public health studies using census tract socioeconomic data to augment vital statistics appeared shortly after World War II, more than a half-century ago (Terris, 1948; Cohart, 1954; Ellis, 1968). The seemingly obvious utility of area-based socioeconomic measures to overcome the absence of socioeconomic data in U.S. public health databases, however, is belied by the still scant research on what measures to use at which level of geography. Perhaps now, with the end the cold war; with renewed discussion of economic inequality, social class, and health; and with new challenges to individualistic frameworks arising from reinvigorated social movements for social justice and human rights (Copenhagen, 1995; Braveman, 1996; Krieger and Birn, 1998; Mann et al., 1999; Shaw et al., 1999; Berkman and Kawachi, 2000; Kim et al., 2000), plus new technologies like GIS (Thrall, 1999a, 1999b; Richards et al., 1999b; Yasnoff and Sondik, 1999; Moore and Carpenter, 1999), it will finally be possible in the United States, to put social inequalities in health on the proverbial map.

Recent Findings from the Public Health Disparities Geocoding Project

Since preparation of this chapter, empirical and methodological analyses of the described study—since named the *Public Health Disparities Geocoding Project*—have been published in scientific journals. Highlights and citations are provided below.

Analytic findings

Our central finding is that census tract (CT) and block group (BG) area-based socioeconomic measures (ABSMs) performed similarly, detecting gradients expected based on the extant literature, whereas ZIP code (ZC) measures are more problematic, at times generating effect estimates substantially smaller than, larger than, and sometimes in even the opposite direction of, those generated with BG and CT measures. Additionally, measures of economic deprivation (e.g., percent of persons below poverty) consistently detected the expected gradients, whereas measures pertaining to occupation, education, and wealth often were less sensitive. One implication is that, from a monitoring perspective, measures of economic deprivation may be best for public health surveillance systems; from an etiologic perspective, however, it may be useful to employ diverse ABSMs to elucidate relevant pathways by which socioeconomic conditions affect the specified health outcomes. For further discussion, see the following

- Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV, and Carson R (2002). Geocoding and monitoring US socioeconomic inequalities in mortality and cancer incidence: Does choice of area-based measure and geographic level matter?—*The Public Health Disparities Geocoding Project. Am. J. Epidemiol.* 156: 471–482.
- Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV, and Carson R (in press). Choosing area-based socioeconomic measures to monitor social inequalities in low birthweight and childhood lead poisoning—*The Public Health DIsparities Geocoding Project* (US). *J. Epidemiol. Community Health.*
- Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV (in press). Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based soecioconomic measures—*The Public Health Disparities Geocoding Project* (US). *Public Health Reports.*

Methodologic studies

These studies: (*1*) documented the need for ensuring accuracy of geocoding for public health research (since accuracy across several geocoding firms ranked from as low as 44% to up to 96%), and (2) provided empirical evidence of bias introduced by the potential mismatch between census-defined and ZIP code-defined geographic areas (e.g., for colon cancer, the socioeconomic gradient observed using ZC ABSMs was in the opposite direction of the expected gradient that was observed with the BG and CT ABSMs). For further discussion, see the following:

- Krieger N, Waterman PD, Lemieux K, Zierler S, Hogan JW (2001). On the wrong side of the tracts? Evaluating accuracy of geocoding for public health research. *Am J Public Health* 91:1114–1116.
- Krieger N, Waterman PD, Chen JT, Soobader M-J, Subramanian SV, and Carson R (2002). ZIP code caveat: bias due to spatiotemporal mis-

matches between ZIP codes and US census-defined areas—*The Public Health Disparities Geocoding Project. Am J Public Health* 92:1100–1102.

ACKNOWLEDGMENTS

We gratefully acknowledge our collaborators at the Massachusetts Department of Public Health and the Rhode Island Department of Health, listed below, along with the helpful guidance of our project's initial advisory board: Peter Townsend and David Gordon (University of Bristol), and Michael White and Andrew Foster (Brown University).

Massachusetts Department of Public Health: Daniel J. Friedman, PhD, Assistant Commissioner; Alice Mroszczyk, Program Coordinator for 24A/B/111B Review Committee; *Weapon Related Injury Surveillance System (WRISS):* Victoria Ozonoff, PhD, Project Director; Beth C. Hume, MPH, Data Manager/Analyst; Laurie Janelli, Site Coordinator; Patrice R. Cummings, MPH, Epidemiologist; *Cancer Registry:* Susan Gershman, PhD, Director; Mary Mroszczyk, Geocoding/Special Projects Coordinator; Ann R. MacMillan, Data Analyst; *Bureau of Communicable Disease Control:* Alfred DeMaria, Jr., MD, Assistant Commissioner; Yuren Tang, MD, MPH, Chief, Surveillance Program; *Registry of Vital Records and Statistics:* Elaine Trudeau, Registrar of Vital Records; Charlene Zion, Public Information Office.

Rhode Island Department of Health: Jay Buechner, PhD, Chief; *Health Interview Survey:* Janice Fontes, MA, Principal Systems Analyst; *Vital Statistics:* Roberta Chevoya, State Registrar of Vital Records; *Division of Disease Prevention and Control:* John Fulton, PhD, Associate Director; Ted Donnelly, RN, MPH, Senior Public Health Epidemiologist; *Violence Against Women Prevention Program:* Wendy Verhoek-Oftedahl, PhD, Epidemiologist; Joyce Babcock, MAT, Assistant Epidemiologist; *Lead:* Bob Vanderslice, PhD, Chief; Susan Feeley, MPH, Epidemiologist.

We also thank Ichiro Kawachi for sharing the program that calculates Gini coefficients based on census-defined income categories. Research for this chapter was supported by NIH grant #1 R01 HD36865-01 ("Area-based socioeconomic measures for health data," Principle Investigator, Nancy Krieger).

REFERENCES

Anderson CL, Agran PF, Winn DG, Tran C (1998). Demographic risk factors for injury among Hispanic and non-Hispanic white children: An ecologic analysis. *Injury Prevention* 4: 33–38.

Antonovsky A (1967). Social class, life expectancy and overall mortality. *Millbank Mem Fund Q* 45: 31–73.

Arbes SJ, Jr, Olshan AF, Caplan DJ, Schoenbach VJ, Slade GD, Symons MJ (1999). Factors contributing to the poorer survival of black Americans diagnosed with oral cancer (United States). *Cancer Causes Control* 10: 513–523.

Barholomew DJ and Knott M (1999). *Latent Variable Models and Factor Analysis*, 2nd ed. New York: Oxford University Press.

Berkman L and Kawachi I, eds. (2000). *Social Epidemiology*. New York: Oxford University Press.

Bernstein J, McNichol EC, Mishel L, Zahradnik R (2000). *Pulling Apart: A State-by-State Analysis of Income Trends*. Washington, DC: Center on Budget and Policy Priorities, Economic Policy Institute.

Braverman P (1996). *Equity in Health and Health Care: A WHO/SIDA Initiative*. Geneva: World Health Organization.

Brooks-Gunn J, Duncan GJ, Lawrence Arber J, eds. (1997). *Neighborhood Poverty, Volume II: Policy Implications in Studying Neighborhoods.* New York: Russell Sage Foundation Press.

Brooks-Gunn J, and Duncan GJ (1997). *Consequences of Growing Up Poor.* New York: Russell Sage Foundation Press.

Breslow NE and Clayton DG (1993). Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 88: 9–25.

Cadum E, Costa G, Biggeri A, Martuzzi M (1999). Deprivation and mortality: A deprivation index suitable for geographical analysis of inequalities. *Epidemiologia e Prevenzione* 23: 175–187.

Carr-Hill R and Rice N (1995). Is enumeration district level an improvement on ward level analysis in studies of deprivation and health? *J Epidemiol Community Health* 49(suppl 2): S28–S29.

Carstairs V (1995). Deprivation indices: Their interpretation and use in relation to health. *J Epidemiol Community Health* 49(suppl 2): S3–S8.

Carstairs V and Morris R (1989). Deprivation and mortality: An alternative to social class? *Communitiy Med* 11: 210–219.

Casper ML, Barnett E, Halvorson JA, Elmes GA, Brahan VE, Majeed ZA, Bloom AS, Stanley S (1999). *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.

Chen FM, Breiman RF, Farley M, Plikaytis B, Deaver K, Cetron MS (1998). Geocoding and linking data from population-based surveillance and the U.S. Census to evaluate the impact of median household income on the epidemiology of invasive *Streptococcus pneumoniae* infections. *Am J Epidemiol* 148: 1212–1218.

Cherkin DC, Grothaus L, Wagner EH (1992). Is magnitude of co-payment effect related to income? Using census data for health services research. *Soc Sci Med* 34: 33–41.

Clayton DG and Kaldor J (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43: 671–681.

Cohart EM (1954). Socioeconomic distribution of stomach cancer in New Haven. *Cancer* 7: 455–461.

Coleman W (1982). *Death Is a Social Disease: Public Health and Political Economy in Early Industrial France.* Madison, Wis: University Wisconsin Press, pp. 149–180.

Collins JW, Jr, and David RJ (1997). Urban violence and African-American pregnancy outcome: An ecologic study. *Ethnicity Disease* 7: 184–190.

Copenhagen Declaration on Social Development (1995). World Summit for Social Development, 6–12 March 1995. URL: http://www.icsw.org/copenhagen_imp...ground_information/copenhagen.html

Council of Economic Advisors for the President's Initiative on Race (1998). *Changing America: Indicators of Social and Economic Well-Being by Race and Hispanic Origin.* URL: http://www.whitehouse.gov/WH/EOP/CEA/html/publications.htm

da Silva LM, Paim JS, Cost M da C (1999). Inequalities in mortality, space and social strata. *Revista de Saude Publica* 33: 187–197.

Davey Smith G (1997). Down at heart—The meaning and implications of social inequalities in cardiovascular disease. *J R Coll Physicians Lond* 31: 414–424.

Davey Smith G and Hart C (1999). Use of census-based aggregate variables to proxy for socioeconomic group: Evidence from national samples (letter). *Am J Epidemiol* 150: 996–997.

Davey Smith, G, Neaton JD, Wentworth D, Stamler R, Stamler J (1996a). Socioeconomic differentials in mortality among men screened for the Multiple Risk Factor Intervention Trial: I. White men. *Am J Public Health* 86: 486–496.

Davey Smith G, Wentworth D, Neaton JD, Stamler R, Stamler J (1996b). Socioeconomic differentials among men screened for the Multiple Risk Factor Intervention Trial: II. Black men. *Am J Public Health* 86: 497–504.

Devesa SS and Diamond EL (1980). Association of breast cancer and cervical cancer incidence with income and education among whites and blacks. *J Natl Cancer Inst* 65: 515–528.

Diez-Roux AV (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *Am J Public Health* 88: 216–222.

Ellen JM, Kohn RP, Bolan GA, Shiboski S, Krieger N (1995). Socioeconomic differences in sexually transmitted disease rates among black and white adolescents, San Francisco, 1990 to 1992. *Am J Public Health* 85: 1546–1548.

Ellis JM (1968). Socioeconomic differentials in mortality from chronic diseases. In Jaco EG, ed.: *Patients, Physicians, Illness.* Glencoe, Ill: Free Press, pp. 30–37.

Engels F (1845 [1958]). *The Condition of the Working Class in England,* transl Henderson WO and Chaloner WH). Stanford, Calif: Stanford University Press, pp. 120–121.

Evans RG, Barer ML, Marmor TR, eds. (1994). *Why Are Some People Healthy and Others Not? The Determinants of Health of Populations.* New York: Aldine de Gruyter.

Fife D and Mode C (1992). AIDS incidence and income. *J Acquir Immune Defic Syndr Hum Retrovirol* 5: 1105–1110.

Fox J, ed. (1989). *Health Inequalities in European Countries.* Aldershot, UK: Gower.

Freeman HP (1998). The meaning of race in science—Considerations for cancer research: concerns of special populations in the National Cancer Program. *Cancer* 82: 219–225.

Geolytics Inc. (2000). *Census CD+.* East Brunswick, NJ. URL: http://www.geolytics.com

Geronimus AT and Bound J (1998). Use of census-based aggregate variables to proxy for socioeconomic group: Evidence from national samples. *Am J Epidemiol* 148: 475–486.

Geronimus AT, Bound J, Neidert LJ (1996). On the validity of using census geocode data to proxy individual socioeconomic characteristics. *J Am Statist Assoc* 91: 529–537.

Goldsmith HF, Lee AS, Rosen BM (1982). *Small Area Social Indicators.* National Institutes of Mental Health, Mental Health Service System Reports Series BN, No. 3, DHHS Pub. No. (ADM) 82-1189. Washington, DC: U.S. Government Printing Office.

Goldstein H (1995). *Multilevel Statistical Models,* 2nd ed. New York: Oxford University Press.

Gordon D (1995). Census based deprivation indices: Their weighting and validation. *J Epidemiol Community Health* 49(suppl 2): S39–S44.

Gordon D and Spicker P, eds. (1999). *The International Glossary on Poverty.* London: Zed Books.

Greenwald JP, Polissar NL, Borgatta EF, McCorkle R (1994). Detecting survival effects of socioeconomic status: Problems in the use of aggregate data. *J Clin Epidemiol* 47: 903–909.

Grisco JA, Schwarz DF, Hirschinger N, Sammel M, Brensinger C, Santanna J, Lowe

RA, Anderson E, Shaw LM, Bethel CA, et al. (1999). Violent injuries among women in an urban area. *N Engl J Med* 341: 1899–1905.

Haan M, Kaplan GA, Camacho T (1987). Poverty and health: Prospective evidence from the Alameda County Study. *Am J Epidemiol* 125: 989–998.

Hankins C, Tran T, Hum L, Laberge C, Lapointe N, Lepine D, Montpetit M, O'Shaughnessy MV (1998). Socioeconomic geographical links to human immunodeficiency virus seroprevalence among childbearing women in Montreal, 1989–1993. *Int J Epidemiol* 27: 691–697.

Harries K (1997). Social stress and trauma: Synthesis and spatial analysis. *Soc Sci Med* 45: 1251–1264.

Hinton RY, Lincoln A, Crockett MM, Sponseller P, Smith G, (1999). Fractures of the femoral shaft in children: Incidence, mechanisms, and sociodemographic risk factors. *J Bone Joint Surg* 81: 500–509.

Hu DJ, Frey R, Costa SJ, et al. (1994). Geographical AIDS rates and sociodemographic variables in Newark, New Jersey, metropolitan area. *AIDS Public Policy J* 9: 20–25.

Hyndman JCT, D'Arcy D, Holman J, Hockey RL, Donovan RJ, Corti B, Rivera J (1995). Misclassification of social disadvantage based on geographical areas: Comparison of postcode and collector's districts analyses. *Int J Epidemiol* 24: 165–176.

Institute of Medicine Committee for the Study of the Future of Public Health (1988). *The Future of Public Health.* Washington, DC: National Academy Press.

Institute of Medicine Committee on Using Performance Monitoring to Improve Community Health (1997). *Improving Health in the Community: A Role for Performance Monitoring.* Washington, DC: National Academy Press.

Iwashyna TJ, Christakis NA, Becker LB (1999). Neighborhoods matter: A population-based study of provision of cardiopulmonary resuscitation. *Ann Emergency Med* 34: 459–468.

Jargowsky PA (1997). *Poverty and Place: Ghettos, Barrios, and the American City.* New York: Russell Sage.

Kaplan CP and Van Valey TL (1980). *Census '80: Continuing the Fact Finding Tradition.* Washington, DC: U.S. Government Printing Office.

Kim JY, Millen JV, Irwin A, Gershman J (2000). *Dying for Growth: Global Inequality and the Health of the Poor.* Monroe, Maine: Common Courage Press.

Kitagawa EM and Hauser PM (1973). *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology.* Cambridge, Mass: Harvard University Press.

Krieger N (1992). Overcoming the absence of socioeconomic data and medical records: Validation and application of a census-based methodology. *Am J Public Health* 82: 703–710.

Krieger N (1990). Social class and the black/white crossover in the age-specific incidence of breast cancer: A study linking census-derived data to population-based registry records. *Am J Epidemiol* 131: 804–814.

Krieger N (1991). 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* 45: 35–42.

Krieger N and Birn AE (1998). A vision of social justice as the foundation of public health: Commemorating 150 years of the spirit of 1848. *Am J Public Health* 88: 1603–1606.

Krieger N, Chen JT, Ebel G (1997b). Can we monitor socioeconomic inequalities in health? A survey of U.S. health departments' data collection and reporting practices. *Public Health Reports* 112: 481–491.

Krieger N, Chen JT, Selby JV (1999a). Comparing individual-based and household-based measures of social class to assess class inequalities in women's health: A methodological study of 684 U.S. women. *J Epidemiol Community Health* 53: 612–623.

Krieger N and Fee E (1996). Measuring social inequalities in health in the United States: An historical review, 1900–1950. *Int J Health Serv* 26: 391–418.

Krieger N and Gordon D (1999). Re: "Use of census-based aggregate variables to proxy for socioeconomic group: Evidence from national samples" (letter). *Am J Epidemiol* 150: 892–896.

Krieger N, Quesenberry, Jr, C, Peng T, Horn-Ross P, Stewart S, Brown S, Swallen K, Guillermo T, Suh D, Alvarez-Martinez L, Ward F (1999b). 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, 1988–92 (United States). *Cancer Causes Control* 10: 525–537.

Krieger N, Rowley D, Hermann AA, Avery B, Phillips MT (1993). Racism, sexism, and social class: Implications for studies of health, disease, and well-being. *Am J Prev Med* 9(suppl): 82–122.

Krieger J, Song L, Solet D (1996). Use of social status measures in community assessment. Paper presented at the American Public Health Association 124th annual meeting, New York , 18–21 November 1996.

Krieger N, Williams DR, Moss NE (1997a). Measuring social class in U.S. public health research: Concepts, methodologies, and guidelines. *Annu Rev Public Health* 18: 341–378.

Lee P, Murie A, Gordon D (1995). *Area Measures of Deprivation: A Study of Current Methods and Best Practices in the Identification of Poor Areas in Great Britain*. Birmingham, UK: Centre for Urban and Regional Studies, University of Birmingham.

Liu L, Deapen D, Bernstein L (1998). Socioeconomic status and cancers of the femal breast and reproductive organs: A comparison across racial/ethnic populations in Los Angeles County, California (United States). *Cancer Causes Control* 9: 369–380.

Lynch J and Kaplan G (2000). Socioeconomic position. In Berkman L and Kawachi I, eds.: *Social Epidemiology*. New York: Oxford University Press, pp. 13–35.

MacIntyre S and Ellaway A (2000). Ecological approaches: Rediscovering the role of the physical and social environment. In Berkman L and Kawachi I, eds.: *Social Epidemiology*. New York: Oxford University Press, pp. 332–348.

Mann JM, Gruskin S, Grodin MA, Annas GJ, eds. (1999). *Health and Human Rights*. New York: Routledge.

Merriwether RA (1996). Blueprint for a National Public Health Surveillance System for the 21st century. *J Public Health Management Practice* 2: 16–23.

Moore DA and Carpenter TE (1999). Spatial analytic methods and geographic information systems: Use in health research and epidemiology. *Epidemiol Rev* 21: 143–161.

Morse DL, Lessner L, Medvesky MG, Glebatis DM, Novick LF (1991). Geographic distribution of newborn HIV seroprevalence in relation to four sociodemographic variables. *Am J Public Health* 81(suppl): 25–29.

Mustard DL, Derksen S, Berthelot JM, Wolfson M (1999). Assessing ecologic proxies for household income: A comparison of household and neighbourhood level income measures in the study of population health status. *Health Place* 5: 157–171.

National Center for Health Statistics (1996). *Health, United States 1995.* Hyattsville, Md: Public Health Service.

National Center for Health Statistics (1997). *Health, United States, 1996–97, and Injury Chartbook.* Hyattsville, Md: U.S. Department of Health and Human Services.

National Center for Health Statistics (1999). *Health, United States, 1999, with Health and Aging Chartbook.* Hyattsville, Md: U.S. Department of Health and Human Services.

Nichols M (1999). Personal communication via e-mail. Products and Service Staff, Geography Division, U.S. Census Bureau, 13 December 1999.

Nolan LJ, Freeman WL, D'Angelo AJ (1996). Local research: Needed guidance for the Indian Health Service's urban mission. *Public Health Reports* 320: 111.

O'Campo P, Xue X, Wang MC, Caughy M (1997). Neighborhood risk factors for low birthweights in Baltimore: A multilevel analysis. *Am J Public Health* 87: 1113–1118.

Pamuk E, Makuc D, Heck K, Reuben C, Lochner K (1998). *Socioeconomic Status and Health Chartbook.* Health, United States, 1988. Hyattsville, Md: National Center for Health Statistics.

Pollack AM and Rice DP (1997). Monitoring health care in the United States—A challenging task. *Public Health Reports* 112: 108–113.

Powell EC and Tanz RR (1999). Child and adolescent injury and death from urban firearm assaults: Association with age, race and poverty. *Injury Prevention* 5: 41–47.

Prehn AW and West DW (1998). Evaluating local differences in breast cancer incidence rates: A census-based methodology (United States). *Cancer Causes Control* 9: 511–517.

Pringle DG, Walsh J, Hennessy M, eds. (1999). *Poor People, Poor Places: A Geography of Poverty and Deprivation in Ireland.* Dublin: Oak Tree Press.

Richards TB, Croner CM, Rushton G, Brown CK, Fowler L (1999). Geographic information systems and public health: Mapping the future. *Public Health Reports* 114: 359–373.

Roberts EM (1997). Neighborhood social environments and the distribution of low birthweight in Chicago. *Am J Public Health* 87: 597–603.

Rossi RJ and Gilmartin KJ (1980). *The Handbook of Social Indicators: Sources, Characteristics, and Analysis.* New York: Garland STPM Press.

Sayegh AJ, Swor R, Chu KH, Jackson R, Gitlin J, Domeier RM, Basse E, Smith D, Fales W. Does race or socioeconomic status predict adverse outcome after out of hospital cardiac arrest: A multi-center study. *Resuscitation* 40: 141–146.

Shaw M, Dorling D, Gordon D, Davey-Smith G (1999). *The Widening Gap: Health Inequalities and Policy in Britain.* Bristol, UK: University of Bristol Press.

Shevsky E and Bell W (1955). *Social Area Analysis: Theory, Illustrative Application, and Computational Procedures.* Stanford, Calif: Stanford University Press.

Simon PA, Hu DJ, Diaz T, Kerndt PR (1995). Income and AIDS rates in Los Angeles County. *AIDS* 9: 281–284.

Soobader M-J and LeClere FB (1999). Aggregation and the measurement of income inequality: Effects on morbidity. *Soc Sci Med* 48: 733–744.

Sorensen A (1994). Women, family, and class. *Annual Review of Sociology* 20: 27–47.
Stuber C (1999). Personal communication via e-mail. U.S. Bureau of Census, 13 December 1999.
Terris M (1948). Relation of economic status to tuberculosis mortality by age and sex. *Am J Public Health* 38: 1061–1070.
Thrall GI (1999a). The future of GIS in public health management and practice. *J Public Health Management Practice* 5: 75–82.
Thrall SE (1999b). Geographic information system (GIS) hardware and software. *J Public Health Management Practice* 5: 82–90.
Tolson GC, Barnes JM, Gay GA, Kowaleski JL (1991). The 1989 revision of the U.S. standard certificates and reports. National Center for Health Statistics. DHHS Pub. No. PHS 91-1465. *Vital Health Stat* 4 (28).
Townsend P (1993). *The International Analysis of Poverty.* New York: Harvester/ Wheatsheaf.
Townsend P, Davidson N, Whitehead M (1990). *Inequalities in Health: The Black Report and the Health Divide.* London: Penguin Books.
Townsend P, Phillimore P, Beattie A (1988). *Health and Deprivation: Inequality and the North.* London: Croom Helm.
U.S. Bureau of the Census (2000). The Census Tract Street Locator. URL: http://tier2.census.gov/ctsl/ctsl.htm
U.S. Bureau of the Census (1991). *Census of Population and Housing, 1990: Summary Tape File 3 Technical Documentation.* Washington, DC: Bureau of the Census.
U.S. Bureau of the Census (1985). *Poverty Areas in Large Cities, PC80-2-8D: 1980 Census of Population, Vol. 2, Subject Reports.* Washington, DC: U.S. Government Printing Office.
U.S. Bureau of the Census (1999). *American Community Survey: Questions and Answers.* Washington, DC: U.S. Department of Commerce. URL: http://www.census.gov/acs/www
U.S. Department of Commerce (1990). *Census '90 Basics.* Washington, DC: U.S. Government Printing Office.
U.S. Department of Health and Human Services (1991). *Health of Minorities and Low-Income Groups,* 3rd ed. Washington, DC: U.S. Government Printing Office.
U.S. Department of Health and Human Services (2000). *Healthy People 2010 (Conference Edition, in Two Volumes).* Washington, DC. URL: http://www.health.gov/healthypeople
U.S. Post Office (2000). URL: http://www.framed.usps.com/ncsc/lookups/zip_faqs.htm
Villermé LR (1828). Mémoire sur la mortalité en France dans la classe aisée et dans la classe indigente. *Mémoires de l'Académie royale de medicine* 1: 51–98.
Wasserman CR, Shaw GM, Selvin S, Gould JB, Syme SL (1998). Socioeconomic status, neighborhood social conditions, and neural tube defects. *Am J Public Health* 88: 1674–1680.
White MJ (1987). *American Neighborhoods and Residential Differentiation.* New York: Russell Sage Foundation.
Whitney JS, ed. (1934). *Death Rates by Occupation Based on Data of the U.S. Census Bureau 1930.* New York: National Tuberculosis Association.
Williams DR and Collins C (1995). U.S. socioeconomic and racial differences in health: Patterns and explanations. *Ann Rev Sociol* 21: 329–386.
Wissow LS, Gittelsohn AM, Szklo M, Starfield B, Mussman M (1988). Poverty, race, and hospitalization for childhood asthma. *Am J Public Health* 78: 777–782.

Wright EO (1997). *Class Counts: Comparative Studies in Class Analysis.* Cambridge: Cambridge University Press.

Wright EO, Costello C, Hachen D, Sprague J (1982). The American class structure. *Am Sociol Rev* 47: 709–726.

Yasnoff WA and Sondik EJ (1999). Geographic information systems (GIS) in public health practice and the new millennium. *J Public Health Management Practice* 5(4): viii–xi.

Zierler S, Krieger N, Tang Y, Coady W, Siegfried E, DeMaria A, Auerbach J (2000). Economic deprivation and AIDS incidence. *Am J Public Health* 90:1064–1073.