

Data and Methods

Selecting an Area Socioeconomic Measure—The Poverty Rate

A variety of socioeconomic variables at the ecological level, such as poverty rate, median family income, percentage of population with at least a high school education, percentage employed in white collar occupations, unemployment rate, housing tenure, household crowding, and automobile ownership, as well as composite indices that combine some or all of these variables, have been used to analyze area socioeconomic inequalities in cancer outcomes (16,17,21–23,25,26,28,30,34,36,39–42). These single and composite area measures are generally taken to represent important aspects of the social environment—such as economic deprivation, social inequality, resource availability, opportunity structure, or living conditions (6,16,17,34,60).

Although all of the above variables are useful in describing socioeconomic inequalities in cancer, the poverty rate (the percentage of population below the poverty level) was chosen as the preferred area measure for a variety of reasons. In the United States, the poverty rate refers to the percentage of families or individuals classified as being below the official poverty threshold. The poverty thresholds are updated annually by the U.S. Bureau of the Census to reflect changes in the Consumer Price Index. In the 1990 decennial census, the official

poverty threshold for a family of four was \$12,674, and 12.8% of the U.S. population was below this poverty threshold (61,62). The poverty rate is a measure of economic deprivation and an uneven distribution of economic resources in a given population. It also correlates highly with other measures of socioeconomic position and deprivation, such as educational attainment, unemployment rate, and occupational composition. For example, for the U.S. as a whole, the weighted correlation coefficients of the 1990 county poverty rate with other socioeconomic variables were as follows: percentage of population with at least a high school education (–0.73), median family income (–0.77), percentage of population in white collar occupations (–0.43), and unemployment rate (0.78). The corresponding national correlations at the census tract level were –0.67, –0.66, –0.51, and 0.72. Similar county- and census tract-level correlations were also observed for the combined 11 SEER registration areas. Moreover, poverty rate had one of the largest relative weights in generating a composite area socioeconomic index for the U.S. (16,17).

Yet another advantage for choosing poverty rate as an area measure is that a priori cutpoints may be specified based on prior empirical research and policy relevance (21). The following cutpoints for poverty rate were selected: < 10%, 10%–19.99%, ≥ 20%. Areas

with a poverty rate of 20% or higher are often considered to be distressed or severely disadvantaged areas. For other area variables, quintiles or quartiles of a distribution that classify an equal number of areas or an equal amount of population into the given categories are measures generally used to denote cutpoints. Although quintiles and quartiles are useful statistical categories, they are intrinsically less meaningful from a programmatic and policy standpoint than the aforementioned poverty categories.

In all of the analyses, the poverty rate is measured at either the county or census tract level and is derived from the 1990 decennial census. Figures 2.1–2.3, pages 20–22, show respectively the county-level distribution of poverty rate, median family income, and percentage of population with at least a high school diploma in 1990, indicating a fairly similar geographic distribution of socioeconomic disadvantage in the U.S. The socioeconomic classification of counties based on the 1990 poverty rate is also temporally stable for the study period 1975–1999 in that the 1990 poverty rate is highly correlated with the 1980 rate ($r = 0.91$ for the U.S. and 0.90 for the SEER regions). Similar associations were observed when the three-category poverty variable was compared in 1990 and 1980 ($\gamma = 0.94$ for the U.S. and 0.91 for the SEER regions; Figures 2.1 and 2.4, pages 20 and 23).

Socioeconomic and Demographic Characteristics of Area Poverty Groups

Table 2.1, page 24, shows the distribution of selected sociodemographic characteristics across the three area poverty groups in 1990. For the U.S. as a whole, 13% of the population lived in counties with poverty rates of 20% or more and 31% of the population lived in counties with a poverty rate of less than 10%. However, the population distribution varied substantially by race/ethnicity. Whereas 27% of blacks, 31% of American Indians/Alaska natives, and 21% of the Hispanic population lived in counties with poverty rates of 20% or more, only 7% of Asians and Pacific Islanders and 10% of non-Hispanic whites were represented in the highest poverty county group. The three poverty groups also varied substantially in terms of median family income, concentration of high school and college graduates, white collar employment, and unemployment rate. Also, the high poverty county group had a higher proportion of rural population than the low poverty county group.

Taken together, the 11 SEER registration areas were somewhat more well off than the total U.S. population. Eight percent of the SEER population lived in high poverty counties and 43% of the SEER population lived in low poverty counties, as compared with 13% and 31% respectively of the U.S. population. Moreover, the SEER county poverty groups did not differ much in their urbanization levels.

The area poverty groups for the 11 SEER registries based on census tracts differed from those based on counties in that a higher

proportion of the population was concentrated in the high poverty census tract group than in the high poverty county group (18% vs. 8%). Moreover, more than 40% of the black, American Indian/Alaska native, and Hispanic population lived in census tracts with a poverty rate of 20% or higher, as compared with 23%, 32%, and 6% respectively of the black, American Indian/Alaska native, and Hispanic populations living in counties with a poverty rate of 20% or higher. Income, education, occupation, and unemployment differentials between area poverty groups were also greater for census tracts than for counties.

Computing Incidence and Mortality Rates for Area Poverty Groups

To compute cancer incidence and mortality rates, three categories of area poverty rate were used to classify all U.S. counties, SEER counties, and SEER census tracts into three population groups, which ranged from being the least disadvantaged (richest) to the most disadvantaged (poorest) area group. While the county geocode in the national mortality database refers to the residence of the decedent at the time of cancer death (57), the county or census tract geocode in SEER relates to the place of residence at the time of cancer diagnosis (63).

The analysis of census tract-level socioeconomic patterns in incidence were based on 379,070 men and 347,245 women newly diagnosed with invasive cancers between January 1, 1988, and December 31, 1992, in 11 population-based SEER cancer registries (55). The analysis of census tract-level socioeconomic

patterns in stage of disease were based on 942,839 men and 881,216 women diagnosed with invasive cancers between January 1, 1988, and December 31, 1999. The 11 SEER registries cover about 14% of the total U.S. population and include the states of Connecticut, Hawaii, Iowa, New Mexico, and Utah, and the metropolitan areas of Atlanta, Detroit, Los Angeles, San Francisco and Oakland, San Jose and Monterey, and Seattle. For the analysis of temporal county-level socioeconomic patterns in cancer incidence, data from 9 SEER registries (excluding Los Angeles and San Jose/Monterey from the above list) were used, which consisted of 1,210,279 male and 1,153,028 female incidence cases during the 1975–1999 time period. All analyses of SEER data were based on the November 2001 data submission (55). Demographic and medical variables from the SEER database used in this report included age at diagnosis, sex, race/ethnicity, county and census tract of residence, stage of disease at diagnosis, therapy/treatment (cancer surgery), survival time (months) and vital status, and year of diagnosis (63).

The analysis of mortality trends was based on 6,277,958 male and 5,516,968 female cancer deaths drawn from the annual national mortality data files from 1975 through 1999. Age-sex-race-county-specific population estimates from 1975 to 1999 served as denominators for computing cancer mortality rates and county-based poverty-specific incidence rates over time (64). Age-sex-race-census tract-specific decennial census populations in 1990, multiplied by 5, served as denominators for computing average annual

cancer incidence rates for the 1988–1992 time period (65). Of the 11.8 million U.S. cancer deaths and 2.36 million newly diagnosed cancers cases from 9 SEER registries between 1975 and 1999, 12 cancer deaths and 231 cancer incidence cases could not be matched with the county poverty rate because of missing or invalid county geocode. Of the 1.82 million invasive cancer cases diagnosed in 11 SEER registries during 1988–1999, 2.57% (46,904) had missing census tract information and hence could not be linked to the area poverty rate. The proportion of cancer cases with missing census tract data was somewhat higher among men than among women (2.81% vs. 2.32%) and among American Indians/Alaska natives (6.37%), Asian and Pacific Islanders (4.76%), Hispanics (3.17%), and non-Hispanic whites (2.17%) than among blacks (1.32%). The higher proportion of American Indian/Alaska native and API cases with missing data was due to a higher proportion of cases with unknown census tracts in the New Mexico, San Jose/Monterey, and Hawaii registries, which have a substantially higher concentration of the American Indian and API populations respectively. However, the cancer cases with and without census tract codes were similar in their age distribution, nativity and marital status composition, stage, and year of diagnosis.

Incidence and mortality rates for each area poverty group were age-adjusted by the direct method using the age composition of the 2000 U.S. standard population and five-year age-specific cancer incidence and mortality rates (55). While trends in mortality rates for the U.S.

are presented on an annual basis, trends in SEER incidence and mortality are analyzed as three-year moving average rates. The moving average smoothing technique allows the identification of a trend more clearly by reducing some of the variability associated with the small populations represented in the SEER annual rates.

Computing Five-Year Cause-Specific Survival Rates for Area Poverty Groups

The analysis of patient survival included 442,415 men and 398,147 women (with known census tract poverty rates) in 11 SEER registries who were diagnosed with primary invasive cancers during 1988–1994 and were followed for vital status through December 31, 1999. Five-year cause-specific survival rates were computed for each poverty group. Survival times were measured in months. The patients lost to follow-up, those alive at the end of the five-year follow-up, and those dying of causes other than the underlying cancer during the follow-up period were treated as censored observations (66). Cancer patients who died of unknown causes, those whose initial diagnosis was found on the death certificate or at autopsy, and those who were not being actively followed were excluded from the analysis.

Statistical Significance and Suppression of Rates and Counts

Socioeconomic gradients were generally described in terms of relative incidence and mortality rate ratios. Rate ratios and differences in rates were tested for statistical significance at

the 0.05 level (54,55). In all analyses, the lowest area poverty group was selected as the reference category. When the number of incidence cases or deaths used to compute incidence or mortality rates are small, those rates tend to have poor reliability. Therefore, to discourage misinterpretation and misuse of rates or counts that are unstable and to protect confidentiality of cancer patients and decedents because of the small numbers of cases or deaths, incidence and mortality rates as well as case and death counts are not shown in tables and figures if the case or death counts are fewer than 16. A case or a death count of less than approximately 16 results in a standard error of the incidence or mortality rate that is approximately 25% or more as large as the rate itself. Equivalently, a case or death count of less than approximately 16 results in the width of the 95% confidence interval around the rate being at least as large as the rate itself. These relationships were derived under the assumption of a Poisson process and with the standard population age distribution close to the observed population age distribution (67). Because of the small numbers of cases and the above criteria for rate reliability, stage, treatment, and survival analyses for American Indians/Alaska natives are deemed unreliable and are generally not presented in this report.

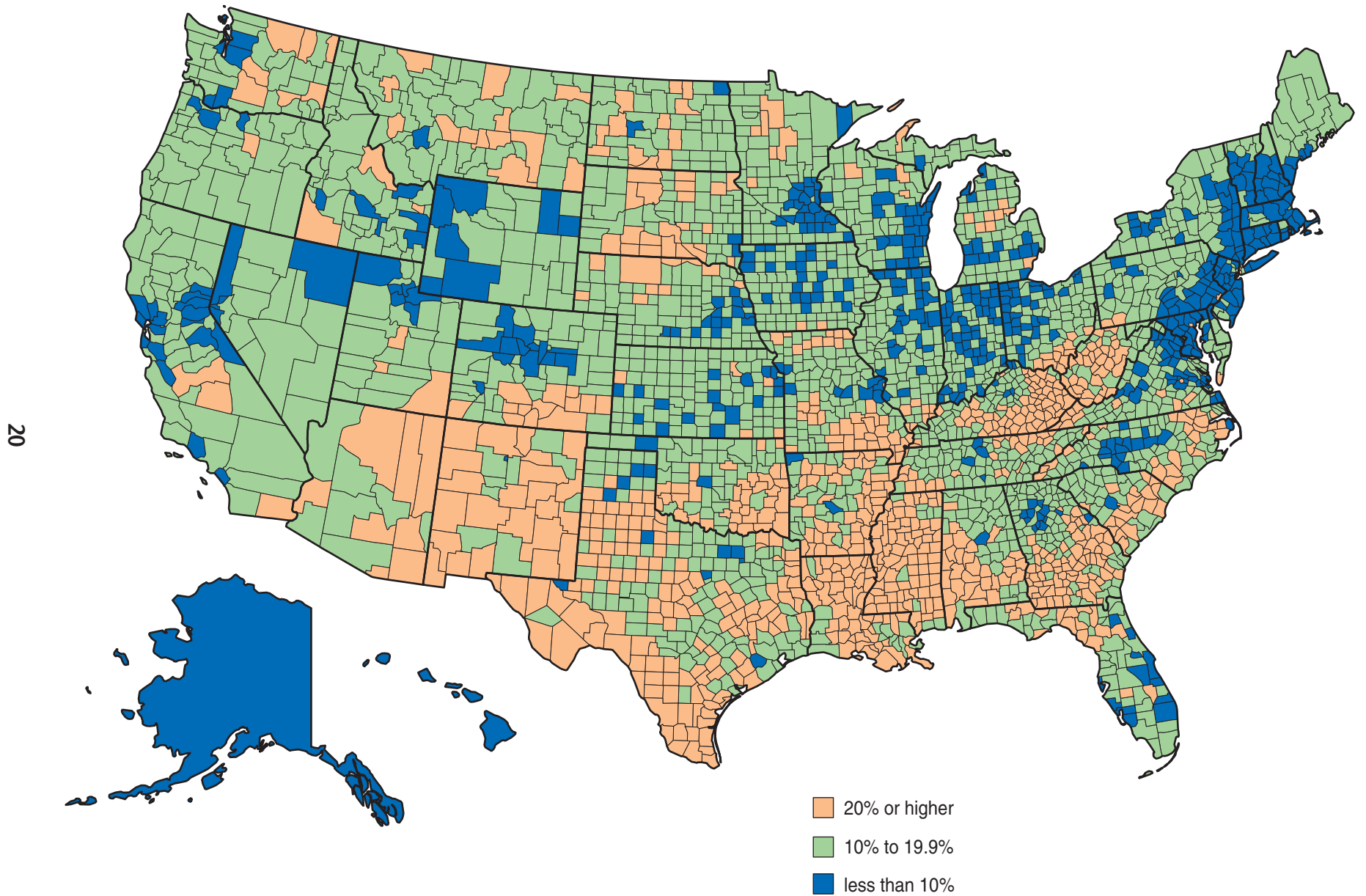
Use of County Versus Census Tract Poverty Rates for U.S. Mortality and SEER Databases

Counties are far more heterogeneous in their socioeconomic and demographic composition than smaller geographic areas such as census tracts, zip codes, or block groups (16). As such,

census tracts are preferable to counties for the purposes of classifying areas into socioeconomic groups and for examining area socioeconomic patterns in health outcomes, especially over a relatively short time horizon. For confidentiality protection of individual information on death certificates, however, the national mortality database does not include data for geographic areas smaller than counties. Therefore, for all analyses of temporal and cross-sectional mortality data, the county-level poverty rate was used.

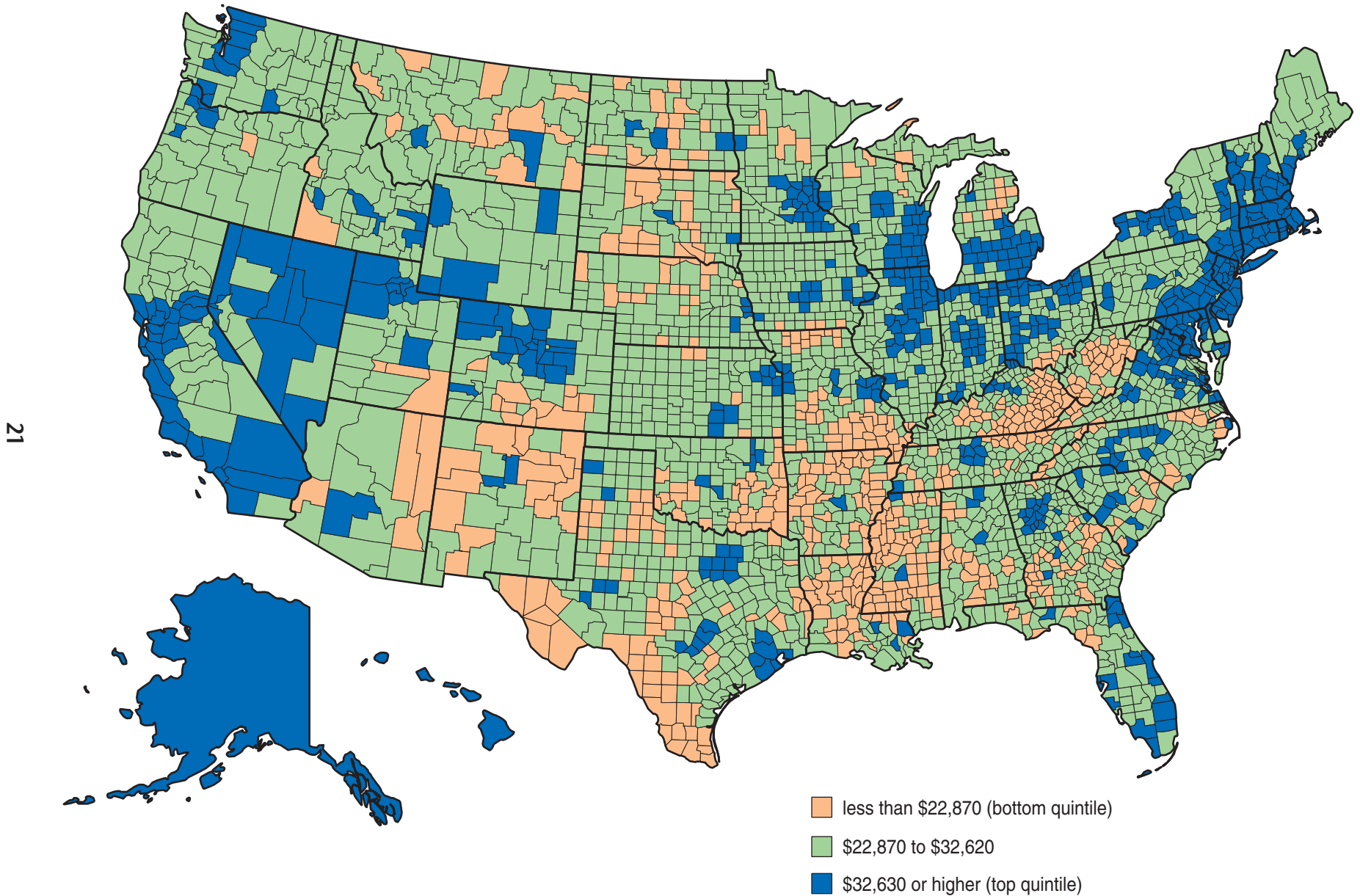
The SEER database, on the other hand, contained the county geocode from 1975 to 1999 and the census tract geocode from 1988 to 1999. Therefore, for the SEER incidence trend analyses from 1975 through 1999, the county poverty rate was used. However, for cross-sectional racial/ethnic and socioeconomic patterns in cancer incidence for the 1988–1992 period, the census tract poverty rate was used. Socioeconomic patterns in SEER cancer incidence, using the census tract poverty rate, could not be assessed for a more recent time period or in a temporal fashion because of a lack of relevant population denominator data at the census tract level. All stage, treatment, and survival analyses in this report involved the use of numerator-based SEER data; hence, for both temporal and cross-sectional analyses, the census tract poverty rate was used.

Figure 2.1. Percentage of County Population Below Poverty Level, United States, 1990



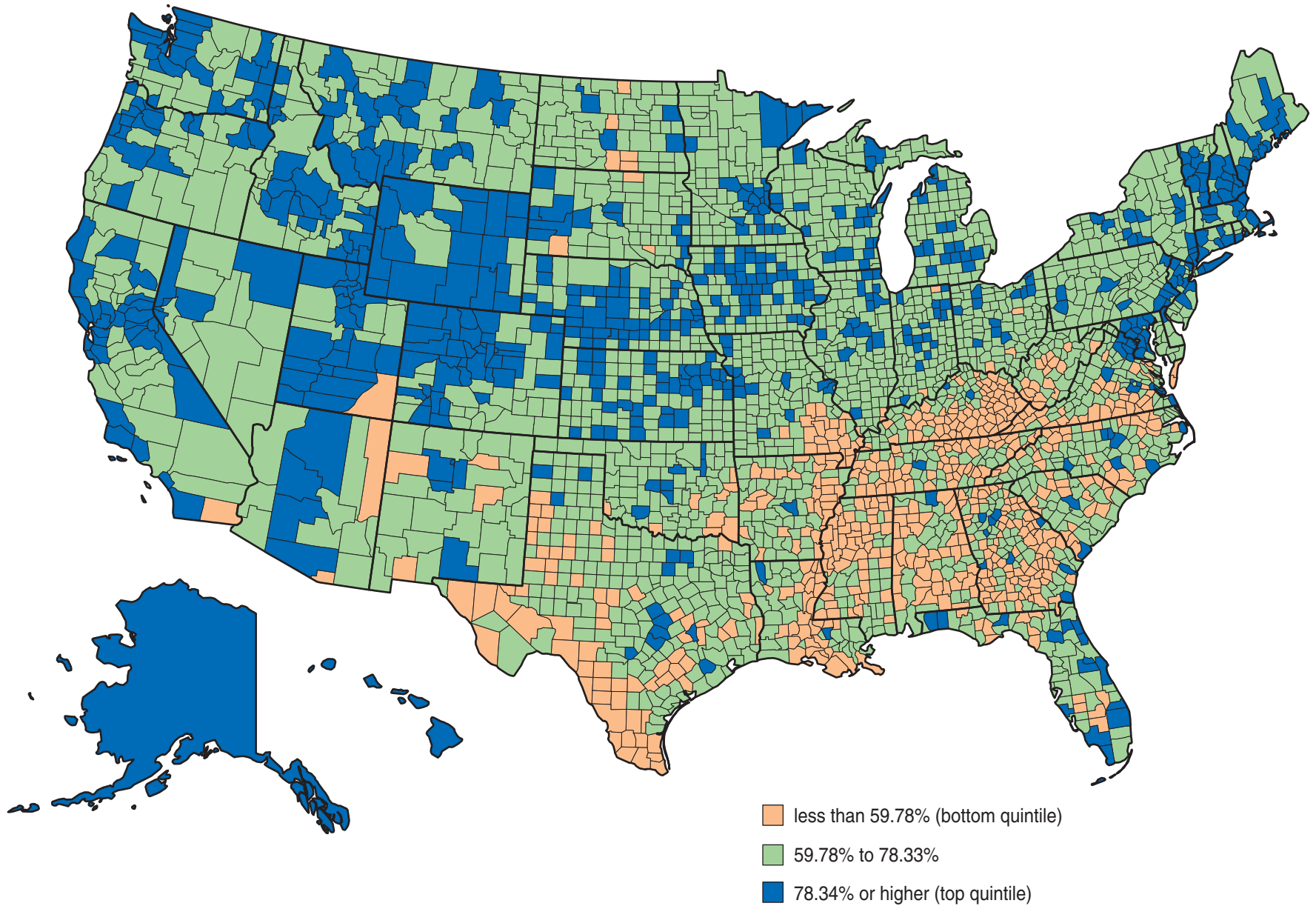
Source: U.S. Bureau of the Census. Data for Alaska and Hawaii are shown at the state level.

Figure 2.2. Median Family Income by County, United States, 1990



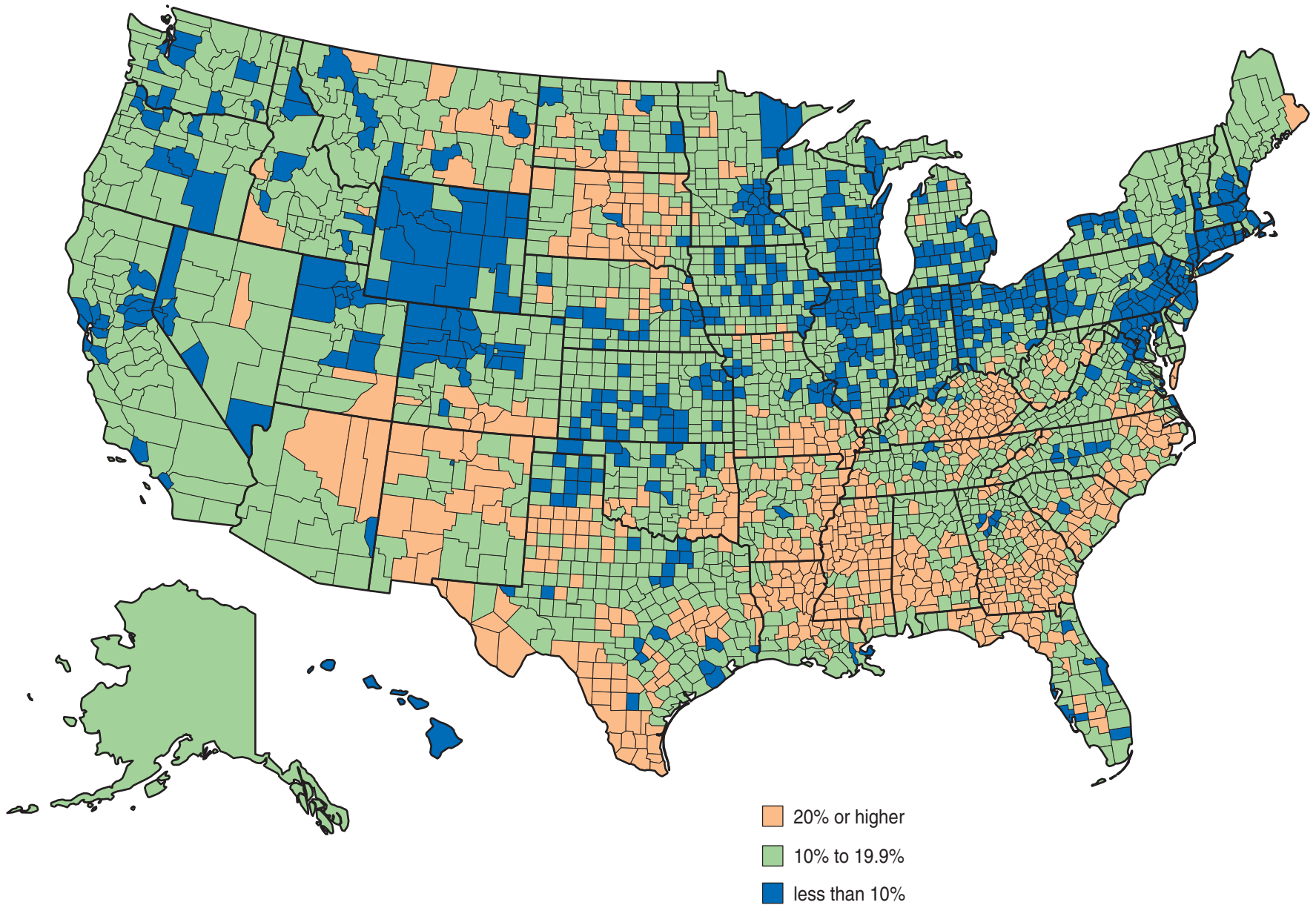
Source: U.S. Bureau of the Census. Data for Alaska and Hawaii are shown at the state level.

Figure 2.3. Percentage of Population With at Least a High School Diploma, United States, 1990



Source: U.S. Bureau of the Census. Data for Alaska and Hawaii are shown at the state level.

Figure 2.4. Percentage of County Population Below Poverty Level, United States, 1980



Source: U.S. Bureau of the Census. Data for Alaska and Hawaii are shown at the state level.

Table 2.1. Selected Socioeconomic and Demographic Characteristics of Area (Census Tract and County) Poverty Groups, 1990: United States and 11 SEER Registration Areas

Area Poverty Rate (% Population Below Poverty Level)	% Total Population	% Non- Hispanic White Population	% Black Population	% American Indian Population	% Asian & Pacific Islander Population	% Hispanic Population	% Population Age Group < 25 Years	% Population Age Group 65+ Years	% High School Diploma or More	% College Graduates	% White Collar Occupation	Median Family Income (\$)	Unemployment Rate (%)	% Urban Population
United States														
County Poverty Rate														
Less than 10%	31.17	35.08	16.72	18.69	40.82	17.54	34.99	11.87	81.42	24.90	62.41	44,555	4.63	78.39
10% to 19.99%	55.76	54.70	56.15	49.94	51.86	61.80	36.51	12.84	74.44	18.94	56.10	33,765	6.56	76.04
20% or higher	13.06	10.22	27.13	31.37	7.32	20.66	38.74	13.04	63.43	14.92	51.41	25,447	9.71	64.07
Total	100.00	100.00	100.00	100.00	100.00	100.00	36.12	12.54	75.20	20.30	58.10	35,338	6.30	75.20
11 SEER Registration Areas														
County Poverty Rate														
Less than 10%	43.44	52.14	27.83	23.55	49.81	18.72	35.09	11.01	83.06	27.42	64.91	46,058	4.70	88.50
10% to 19.99%	48.30	40.97	49.00	44.40	49.36	75.37	37.75	10.93	74.78	22.90	59.18	39,035	6.66	90.20
20% or higher	8.27	6.89	23.17	32.25	0.83	5.91	38.73	12.17	69.39	13.91	53.76	34,099	11.75	87.39
Total	100.00	100.00	100.00	100.00	100.00	100.00	36.99	11.04	78.00	24.30	61.70	39,035	6.20	89.20
11 SEER Registration Areas														
Census Tract Poverty Rate														
Less than 10%	59.28	72.55	24.96	29.42	60.22	28.31	33.72	11.42	85.03	29.32	67.13	48,077	4.27	86.65
10% to 19.99%	22.90	20.41	25.58	23.03	23.75	31.06	38.46	11.48	73.53	17.39	52.50	32,041	7.01	80.82
20% or higher	17.82	7.04	49.46	47.54	16.03	40.66	45.69	9.02	56.31	11.05	43.50	21,458	14.85	91.20
Total	100.00	100.00	100.00	100.00	100.00	100.00	36.99	11.04	78.00	24.30	61.70	40,112	6.20	89.20

Notes: The 11 SEER registries include the states of Connecticut, Hawaii, Iowa, New Mexico, and Utah, and the metropolitan areas of Atlanta, Detroit, Los Angeles, San Francisco and Oakland, San Jose and Monterey, and Seattle. Median family income for each area group is the weighted median of the county or census tract median family income values.