Identifying the Impact of Social Determinants of Health on Disease Rates Using Correlation Analysis of Area-Based Summary Information

Ruiguang Song, PhD[a]
H. Irene Hall, PhD[a]
Kathleen McDavid Harrison, PhD, MPH[b]
Tanya Telfair Sharpe, PhD[b]
Lillian S. Lin, PhD[a]
Hazel D. Dean, ScD, MPH[c]

ABSTRACT

Objectives. We developed a statistical tool that brings together standard, accessible, and well-understood analytic approaches and uses area-based information and other publicly available data to identify social determinants of health (SDH) that significantly affect the morbidity of a specific disease.

Methods. We specified AIDS as the disease of interest and used data from the American Community Survey and the National HIV Surveillance System. Morbidity and socioeconomic variables in the two data systems were linked through geographic areas that can be identified in both systems. Correlation and partial correlation coefficients were used to measure the impact of socioeconomic factors on AIDS diagnosis rates in certain geographic areas.

Results. We developed an easily explained approach that can be used by a data analyst with access to publicly available datasets and standard statistical software to identify the impact of SDH. We found that the AIDS diagnosis rate was highly correlated with the distribution of race/ethnicity, population density, and marital status in an area. The impact of poverty, education level, and unemployment depended on other SDH variables.

Conclusions. Area-based measures of socioeconomic variables can be used to identify risk factors associated with a disease of interest. When correlation analysis is used to identify risk factors, potential confounding from other variables must be taken into account.

[a]Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of HIV/AIDS Prevention, Atlanta, GA

[b]Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Office of Health Equity, Atlanta, GA

[c]Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Office of the Director, Atlanta, GA

Address correspondence to: Ruiguang Song, PhD, Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of HIV/AIDS Prevention, 1600 Clifton Rd. NE, MS E-48, Atlanta, GA 30333; tel. 404-639-4801; fax 404-639-8642; e-mail <RSong@cdc.gov>.

In recent years, scientists have come to understand that genetic influence and personal behavior do not fully explain disparities in infectious diseases.[1–3] Complex, integrated, and overlapping social structures and economic systems, collectively referred to as social determinants of health (SDH), are now thought to affect disease morbidity and mortality.[4] The structural inequities in societal resources that contribute to better health outcomes for some people are apparent. However, inequities in societal resources are not always detectable by traditional methods for measuring disease burden to establish causal links. Improving the measurement of social determinants and connecting them with disease burden will provide evidence to support policy development and action.

In public health, we are interested not only in reducing the morbidity and mortality of diseases in the entire population, but also in achieving equity of health outcomes among subpopulations, particularly those with socioeconomic disadvantages. To accomplish this goal, we first need a system for measuring and monitoring the status of health in the population. For infectious diseases, we have population-based surveillance systems to monitor morbidity and mortality rates. Although information on demographic and geographic variables at the individual level is collected, little information on social environment and health service variables is typically collected in those data systems.

Each person's health is affected by the person's behavior, which, in turn, is associated with his or her social or economic status (e.g., income, education, and marital status) and the corresponding environmental conditions (e.g., the proportion of people in a neighborhood who live below the federal poverty level [FPL] or who do not have a high school education). Although SDH variables at the individual level are important for evaluating the equity of health among groups, SDH variables at the group level are also important because people do not live in isolation and some infectious diseases are transmitted through physical contact. Another advantage of using SDH variables at the group level is that they are available from many data sources that cannot be linked with morbidity and mortality data at the individual level. For example, the National Health Interview Survey[5] and the Behavioral Risk Factor Surveillance System[6] collect data not only on health conditions, but also on health-related risk and health-care services. The decennial U.S. Census data and the American Community Survey (ACS) data are particularly useful because summary information is available for small geographic areas and that information can be linked to surveillance data systems.

SDH typically represent social and physical environmental factors that cannot be controlled by the individual but that have significant impact on the individual's health.[7] Public health entities such as the Centers for Disease Control and Prevention (CDC), the U.S. Department of Health and Human Services, and the World Health Organization recognize that addressing SDH can contribute to health equity.[4,8,9] SDH may explain overlapping risk factors that are common among groups that bear a disproportionate burden of some diseases. Studying the SDH of infectious diseases is challenging because of the limited amounts and types of SDH data available in population-based surveillance systems in the U.S.[10–17]

In this article, we introduce a quantitative method for identifying the SDH variables that influence morbidity of a specific disease—acquired immunodeficiency syndrome (AIDS). We present the results of linking U.S. AIDS surveillance data with SDH data from ACS and determining the significant SDH variables by correlation analysis of summary statistics at a local level.

METHODS

A disease of interest in a population can be measured by the incidence, or diagnosis rate, of the disease in the population during a specific period. This rate may or may not depend on a person's geographic area of residence. Differences in rates in geographic areas may reflect differences in the environmental conditions or the socioeconomic structures in these areas.

Data sources

American Community Survey. In the United States, a census is conducted every 10 years to get a snapshot of the entire population.[18] For a dynamic and timelier picture of the nation, the U.S. Census Bureau also conducts the ACS.[19] The ACS is a nationwide survey that collects essentially the same information on people and housing status that was collected on the long-form questionnaire used in past decennial censuses. In 2010 and thereafter, the ACS replaces the long-form questionnaire by collecting long-form–type information throughout the decade rather than only once every 10 years. The ACS collects not only housing and demographic information, but also socioeconomic variables, including marital status, education, language spoken at home, employment, occupation, and family income. Each year, the ACS sample comprises about 1% of the total population.

Geographic area for analysis of SDH. Individual-level data from the ACS are available for public use. To maintain confidentiality, the geographic location of each

individual can be identified only by public-use microdata area (PUMA), defined for the 2000 Census as a geographic area with a population of at least 100,000. Because of the confidentiality requirement, we cannot link the ACS data with disease case surveillance data at the individual level, but we can link data from ACS geographic areas with surveillance data.

To identify the impact of SDH variables on the morbidity of AIDS, we compared the population by geographic area of residence. The areas can be as large as (multistate) census region, as small as census tract, or something in between (e.g., state or county). If the areas are too large, we have fewer data points with which to assess the association. In many situations, county is the most appropriate area. However, if the areas are too sparsely populated, the data points may vary so much that the association cannot be determined. The PUMA, as defined in ACS, avoids this problem. On the other hand, the PUMA may be small in terms of square miles in population-dense areas. Ideally, the areas for SDH studies should be defined so that most of the people live and conduct daily activities in the same area. Taking these concerns into account, we defined an area, called county-PUMA overlap area (CPOA), as either a county or a PUMA (whichever was larger) so that the area could be identified in both the ACS data and the case surveillance data.

AIDS surveillance data. Since 1982, all 50 U.S. states and the District of Columbia have reported AIDS cases to CDC in a uniform format. All cases are reported to CDC without identifying information. We used data on people with a diagnosis of AIDS (all ages) reported to CDC through June 2009 to calculate AIDS diagnosis rates. Rates per 100,000 person-years were calculated for CPOAs.

Data analysis

First, we estimated the diagnosis rate for each CPOA. Depending on the study objectives and the frequency of the disease, this rate may be based on a single year or multiple years of data. If too many CPOAs had no cases in a calendar year, then the rate was estimated for a three- or five-year period. When too many CPOAs have no cases, the data will be less likely to differentiate rates according to the characteristics associated with CPOAs.

Next, we used the ACS data to measure SDH variables in each CPOA during the period for which the AIDS diagnosis rate was based. Except for the population density variable, which is the population divided by the land area in the CPOA, we measured all demographic and SDH variables as proportions. For example, we measured poverty in a CPOA as the proportion of people living below the FPL. There are often two ways to define a proportion. For example, an analysis of the effect of gender in a population can be based on the proportion of males or the proportion of females. If one of the two defined variables is positively correlated with a variable, then the other will be negatively correlated with the variable, and vice versa. For ease of comparison of correlations among demographic and SDH variables, in each such instance, we chose the variable that was positively correlated with the AIDS diagnosis rate.

Using ACS data, we examined the following demographic and socioeconomic variables (variable names used in our analyses are in parentheses; *q* indicates a log-transformed proportion):

1. Population density (log_dens)
2. Proportion female (p_female)
3. Proportion aged ≤30 years (p_young)
4. Proportion Hispanic (q_hisp)
5. Proportion non-Hispanic black (q_black)
6. Proportion of minority race/ethnicity (q_xwhite)
7. Proportion not currently married (p_single)
8. Proportion below the FPL (p_pov)
9. Proportion with less than a high school education (p_hsch)
10. Proportion unemployed (p_unemp)
11. Proportion moved in the past 12 months (p_moved)
12. Proportion foreign-born (q_foreign)

After determining the diagnosis rate and the demographic and SDH variables at the CPOA level, we estimated the Pearson product moment correlations among these variables. Correlation measures the linear relationship between variables, but many relationships are not linear. Whether the relationship is linear can be assessed visually by using scatter plots. When the relationship is not linear, transformations can be performed on variables to make the relationship close to linear so that the relationship can be detected through the measure of correlation. Also, outliers and the distribution of a variable can affect the measure of the correlation between this variable and other variables. In other words, transformation should be considered to make either the distribution resemble a normal distribution or the relationship with other variables close to linear. In our study, the following variables were log-transformed: AIDS diagnosis rate, population density, proportion of foreign-born people, and proportions of

subpopulations—Hispanic, non-Hispanic black, and all racial/ethnic minority groups (including both Hispanic and non-Hispanic black people).

Demographic and SDH variables are often correlated. However, correlations between these variables and the diagnosis rate could be caused by, or confounded with, other demographic or SDH variables. To control, or adjust, for the confounding, we estimated partial correlations[20]—the correlation between two variables, with an adjustment for a third variable. If the correlation of a variable (e.g., X) with the AIDS diagnosis rate does not change after adjustment for a third variable, then the third variable is considered to have no impact on the effect of variable X on the AIDS diagnosis rate.

An indirect effect of an SDH variable (X) on the AIDS diagnosis rate through a third variable (Y) can be defined as a product of two correlations: the correlation between X and Y and the correlation between Y and the AIDS diagnosis rate. By contrast, the partial correlation can be considered a direct effect. The crude, or unadjusted, correlation between an SDH variable and the AIDS diagnosis rate can be roughly decomposed into two parts: the direct or partial correlation and the indirect correlation through a third variable.

To visualize the relationships among the AIDS diagnosis rate, demographic variables, and SDH variables and the strength of correlation, we applied multidimensional scaling.[21] This technique projects all variables of interest onto a two-dimensional plane so that each variable is represented by a point on the plane, and the relative strength of the correlation between two variables is reflected by the distance between the two points. Each pair of highly correlated variables is connected with a line. Arrows are added at the end(s) of each line to denote potential causal directions. The health outcome variable (e.g., the AIDS diagnosis rate) is connected only by unidirectional lines with arrows pointed to the health outcome variable. A line with arrows on both ends indicates that the relationship is posited to be mutually causal.

RESULTS

The ACS data showed 3,141 counties and 1,153 PUMAs in the 50 states and the District of Columbia (DC). After combining counties within a PUMA and merging PUMAs within a county, we identified 949 CPOAs in the 50 states and DC. The median population among CPOAs was 163,848 from 2006 through 2008 (range: 93,125 to 9,831,675). In 2008, more than 3% of CPOAs (32 of 949) had no AIDS diagnosis. When we included all three years (2006, 2007, and 2008), the number of CPOAs with no AIDS diagnoses was four (0.4%), so we selected 2006 through 2008 as our study period.

During these three years (2006, 2007, and 2008), the AIDS diagnosis rate in all CPOAs (combined) was about 12 cases per 100,000 person-years. However, AIDS diagnosis rates in the 949 CPOAs were not similar (range: zero to 122 cases per 100,000 person-years; standard deviation [SD] = 14.3). This variation was not just due to random variation. If the AIDS diagnosis rate was constant in all CPOAs and the rate was 12 cases per 100,000 person-years, then the SD of AIDS diagnosis rates in the 949 CPOAs would be 1.98 based on binomial variation in each CPOA, with a possible range of three to 25 cases per 100,000 person-years.

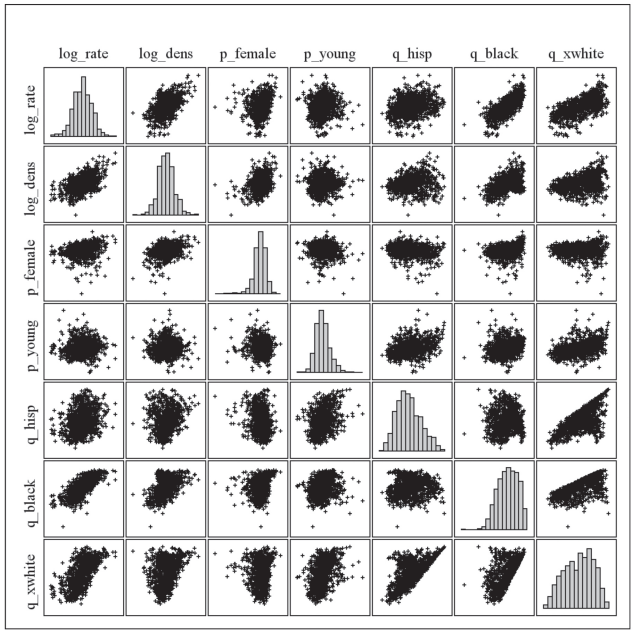
To answer the question of what caused, or was associated with, the difference in rate among CPOAs, we examined 12 demographic and socioeconomic variables derived from the ACS data. Histograms and scatter plots of the log AIDS diagnosis rate and the demographic and socioeconomic variables (or their log transformations) are shown in Figure 1.

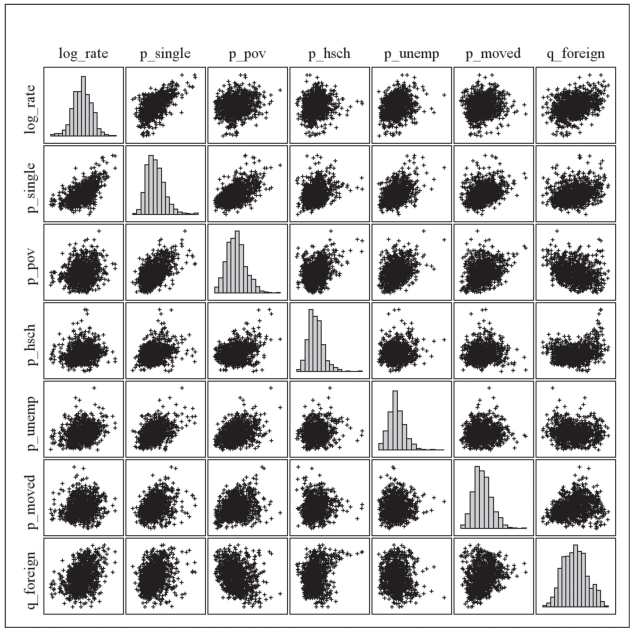
Using the transformed variables, we estimated the correlation of each demographic or SDH variable with the AIDS diagnosis rate (Table 1). The correlations with the AIDS diagnosis rate ranged from 0.06 to 0.74. All correlations were significantly different from zero. The demographic variables most strongly correlated with AIDS diagnosis rates were the proportion of the black population (correlation coefficient [ρ] = 0.74) and the proportion of people of minority race/ethnicity (ρ=0.65). The SDH variables most strongly correlated with AIDS diagnosis rates were the proportion of unmarried people (ρ=0.59) and population density (ρ=0.52).

Demographic and SDH variables were often correlated. For example, a higher proportion of foreign-born population in a CPOA was highly associated with a higher proportion of Hispanic population (ρ=0.84); a higher proportion of people of minority race/ethnicity in a CPOA was associated with a higher proportion of unmarried people (ρ=0.63); and a higher proportion of young people in a CPOA was associated with a higher proportion of people who had moved during the past 12 months (ρ=0.57).

We calculated the partial correlation between each demographic or SDH variable and the AIDS diagnosis rate, adjusting for each of the other demographic or SDH variables, one at a time (Table 2). The strong correlation between the proportion of black people and the AIDS diagnosis rate was not significantly affected by other demographic or SDH variables, although there was some minor impact from population density, the proportion of people of minority races/ethnicities,

Figure 1. Histograms and scatter plots of the AIDS diagnosis rate and demographic and SDH variables: National HIV Surveillance System and American Community Survey data, 2006–2008[a]





continued on p. 75

and the proportion of unmarried people. On the other hand, the effects of many SDH variables on the AIDS diagnosis rate disappeared after adjusting for the proportion of black people. For example, because a disproportionate number of black people lived below the FPL, the effect of poverty on the AIDS diagnosis rate was reduced from a correlation of 0.17 to 0.02.

To understand these relationships better, we used the partial correlations to examine the interactions between SDH variables and their effects on the AIDS diagnosis rate. For example, poverty was highly correlated with marital status, and both were correlated with AIDS diagnosis rates (Table 1). Adjusting for poverty did not change the effect of marital status on AIDS diagnosis rates (Table 2). The unadjusted correlation (ρ=0.59) was almost the same as the correlation adjusted for poverty (ρ=0.60). However, adjusting for marital status changed the effect of poverty on AIDS diagnosis rates from positively correlated (ρ=0.17 unadjusted) to negatively correlated (ρ=–0.19 adjusted).

Similarly, adjusting for education level did not change the effect of the minority race/ethnicity proportion on the AIDS diagnosis rate. The unadjusted correlation (ρ=0.65) was almost the same as the correlation adjusted for education level (ρ=0.66). However, adjusting for minority race/ethnicity proportion changed the effect of education level on the AIDS diagnosis rate from positively correlated (ρ=0.15 unadjusted) to negatively correlated (ρ=–0.26 adjusted).

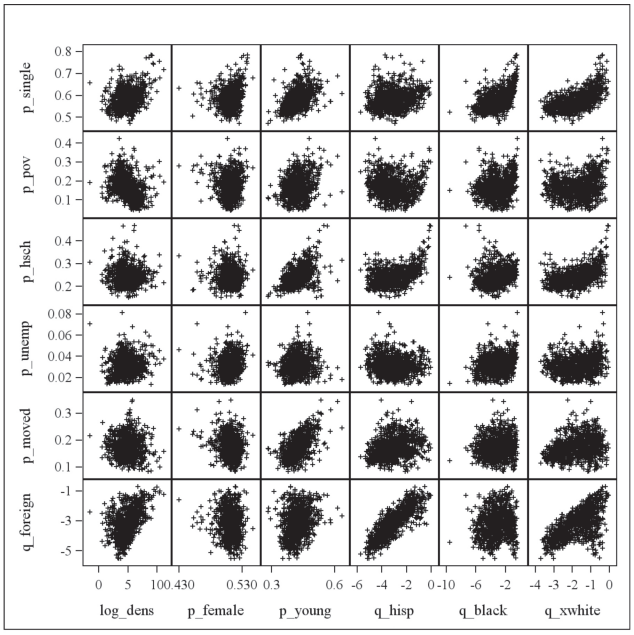
The proportion of young people in a CPOA was not highly associated with AIDS diagnosis rates (Table 1). However, the proportion of young people was highly correlated with the proportion of people of minority race/ethnicity and the proportion of unmarried people. Because the proportion of people of minority race/ethnicity and the proportion of unmarried people in a CPOA were highly correlated with AIDS diagnosis rates, the proportion of young people in a CPOA had a significant indirect and positive impact on AIDS diagnosis rates (Table 3). On the other hand, given the proportion of people of minority race/ethnicity or the proportion of unmarried people in a CPOA, the proportion of young people had a significant direct, but negative, impact on AIDS diagnosis rates (Table 2).

We calculated the indirect effect of each demographic or SDH variable on the AIDS diagnosis rate through each of the other demographic or SDH variables. A comparison of the direct correlations in Table 2 with the indirect correlations in Table 3 shows that demographic and SDH variables with high unadjusted correlations with AIDS diagnosis rates had mostly direct effects on AIDS diagnosis rates. Indirect effects of other demographic and SDH variables were mostly through these variables.

On the basis of the estimated correlations in Table 1, the 12 demographic and SDH variables and the AIDS diagnosis rate were projected on a plane by using mul-

tidimensional scaling (Figure 2). Each pair of variables with a correlation coefficient of >0.4 was connected with a line. Racial/ethnic group variables were connected with a two-headed arrow line, indicating that an increase in one of the two variables resulted in a decrease in the other variable. Also, lines connecting a racial/ethnic group variable with another variable were unidirectional and always pointed from the racial/ethnic group variable to the other variable.

Figure 1 (continued). Histograms and scatter plots of the AIDS diagnosis rate and demographic and SDH variables: National HIV Surveillance System and American Community Survey data, 2006–2008[a]



[a]Log transformations have been applied to AIDS diagnosis rate, population density, and proportions of racial/ethnic and foreign-born population groups.

AIDS = acquired immunodeficiency syndrome
SDH = social determinants of health
log_rate = AIDS diagnosis rate
log_dens = population density
p_female = proportion female
p_young = proportion aged ≤30 years
q_hisp = proportion Hispanic
q_black = proportion non-Hispanic black
q_xwhite = proportion of minority race/ethnicity
p_single = proportion not currently married
p_pov = proportion below the federal poverty level
p_hsch = proportion with less than a high school education
p_unemp = proportion unemployed
p_moved = proportion moved in the past 12 months
q_foreign = proportion foreign-born

DISCUSSION

Our study uncovered the complexity of measuring the exact contribution of SDH to AIDS diagnoses. When correlation analysis is applied to area-based measures of socioeconomic variables, potential confounding effects from other variables must be taken into account. As demonstrated in our study, complicated interactive relationships between race/ethnicity and other variables such as poverty and education exist. Black people are more likely to live in densely populated areas and to experience higher rates of poverty, lower levels of educational attainment, and lower marriage rates compared with people of other racial/ethnic groups.[22,23]

Results show that the correlation of a specific SDH variable with the morbidity or mortality rate of a disease could be superficial. Part or most of the correlation could be caused by other SDH variables. For example, the effect of poverty on the AIDS diagnosis rate was reduced from a correlation of 0.17 to 0.02 after adjusting for the proportion of black people in the population in a given area. Moreover, the interaction between SDH variables can completely change the direction of the correlation of an SDH variable with the morbidity or mortality rate. For example, poverty was positively correlated with the AIDS diagnosis rate (ρ=0.17), but the correlation became negative (ρ=–0.19) after adjusting for marital status. This finding suggests that SDH variables should not be examined in isolation. Interactions between SDH variables must be considered in studying the impact of SDH variables on a specific disease.

The partial correlation separates the direct correlation from the indirect correlation through other variables. Higher orders of partial correlations (correlations adjusted for, or controlled by, more than one variable) can be considered, but the results will be difficult to interpret (similar to interpreting interactions in which more than two variables are involved).

A correlation between two variables may or may not be due to a causal relationship. Some variables may be correlated with AIDS diagnosis rates through other variables (observed or unobserved). From a statistical point of view, these variables could be confounding, mediation, or effect-modification variables. The stronger the correlation, the more likely that the variable is causal because a strong correlation means that it is less likely that a variable with an even stronger correlation is lurking in the background.

To uncover possible causal relationships, we can use partial correlation. A partial correlation measures the strength of a relationship between two variables while controlling the effect of other variables. Whether the partial correlation indicates a causal relationship

Table 1. Correlations between the AIDS diagnosis rate and demographic and SDH variables based on data at the CPOA level: National HIV Surveillance System and American Community Survey data, 2006–2008

		Demographic/SDH variables											
		F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*	*F9*	*F10*	*F11*	*F12*
Demographic/ SDH variables	*Log_rate*	*0.52*	*0.22*	*0.06*	*0.26*	*0.74*	*0.65*	*0.59*	*0.17*	*0.15*	*0.22*	*0.11*	*0.37*
log_dens (F1)	0.52		0.35	0.04	0.19	0.46	0.28	0.33	–0.34	–0.05	0.08	–0.02	0.49
p_female (F2)	0.22	0.35		–0.17	–0.22	0.29	0.05	0.18	–0.01	–0.02	0.12	–0.18	–0.12
p_young (F3)	0.06	0.04	–0.17		0.33	0.15	0.41	0.46	0.28	0.51	0.07	0.57	0.26
q_hisp (F4)	0.26	0.19	–0.22	0.33		0.04	0.66	0.21	–0.09	0.43	–0.07	0.27	0.84
q_black (F5)	0.74	0.46	0.29	0.15	0.04		0.62	0.53	0.22	0.18	0.25	0.11	0.14
q_xwhite (F6)	0.65	0.28	0.05	0.41	0.66	0.62		0.63	0.21	0.50	0.18	0.27	0.61
p_single (F7)	0.59	0.33	0.18	0.46	0.21	0.53	0.63		0.51	0.28	0.43	0.33	0.26
p_pov (F8)	0.17	–0.34	–0.01	0.28	–0.09	0.22	0.21	0.51		0.34	0.36	0.27	–0.28
p_hsch (F9)	0.15	–0.05	–0.02	0.51	0.43	0.18	0.50	0.28	0.34		0.12	0.07	0.22
p_unemp (F10)	0.22	0.08	0.12	0.07	–0.07	0.25	0.18	0.43	0.36	0.12		0.05	–0.09
p_moved (F11)	0.11	–0.02	–0.18	0.57	0.27	0.11	0.27	0.33	0.27	0.07	0.05		0.23
q_foreign (F12)	0.37	0.49	–0.12	0.26	0.84	0.14	0.61	0.26	–0.28	0.22	–0.09	0.23	

AIDS = acquired immunodeficiency syndrome
SDH = social determinants of health
CPOA = county-PUMA [public-use microdata area] overlap area
log_rate = AIDS diagnosis rate
log_dens = population density (F1)
p_female = proportion female (F2)
p_young = proportion aged ≤30 years (F3)
q_hisp = proportion Hispanic (F4)
q_black = proportion non-Hispanic black (F5)
q_xwhite = proportion of minority race/ethnicity (F6)
p_single = proportion not currently married (F7)
p_pov = proportion below the federal poverty level (F8)
p_hsch = proportion with less than a high school education (F9)
p_unemp = proportion unemployed (F10)
p_moved = proportion moved in the past 12 months (F11)
q_foreign = proportion foreign-born (F12)

Table 2. Partial correlations between SDH variable X and the AIDS diagnosis rate, adjusted for SDH variable Y, based on data at the CPOA level: National HIV Surveillance System and American Community Survey data, 2006–2008

		Y											
		F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*	*F9*	*F10*	*F11*	*F12*
X	*Log_rate*	*0.52*	*0.22*	*0.06*	*0.26*	*0.74*	*0.65*	*0.59*	*0.17*	*0.15*	*0.22*	*0.11*	*0.37*
log_dens (F1)	0.52		0.48	0.52	0.49	0.30	0.46	0.42	0.62	0.53	0.51	0.52	0.42
p_female (F2)	0.22	0.05		0.24	0.30	0.01	0.25	0.15	0.23	0.23	0.20	0.25	0.29
p_young (F3)	0.06	0.05	0.10		–0.03	–0.08	–0.29	–0.30	0.01	–0.02	0.05	0.00	–0.04
q_hisp (F4)	0.26	0.19	0.32	0.26		0.34	–0.28	0.17	0.28	0.22	0.28	0.24	–0.10
q_black (F5)	0.74	0.67	0.73	0.75	0.76		0.57	0.63	0.74	0.74	0.73	0.74	0.75
q_xwhite (F6)	0.65	0.61	0.65	0.68	0.65	0.35		0.44	0.63	0.66	0.63	0.64	0.57
p_single (F7)	0.59	0.52	0.58	0.64	0.57	0.35	0.32		0.60	0.58	0.57	0.60	0.56
p_pov (F8)	0.17	0.43	0.18	0.16	0.20	0.02	0.04	–0.19		0.13	0.10	0.15	0.31
p_hsch (F9)	0.15	0.21	0.16	0.14	0.04	0.03	–0.26	–0.02	0.10		0.13	0.15	0.08
p_unemp (F10)	0.22	0.21	0.20	0.22	0.25	0.05	0.14	–0.05	0.17	0.21		0.22	0.27
p_moved (F11)	0.11	0.14	0.15	0.09	0.04	0.04	–0.09	–0.12	0.06	0.10	0.10		0.02
q_foreign (F12)	0.37	0.16	0.41	0.37	0.29	0.40	–0.04	0.28	0.44	0.35	0.40	0.36	

SDH = social determinants of health

AIDS = acquired immunodeficiency syndrome

CPOA = county-PUMA [public-use microdata area] overlap area

log_rate = AIDS diagnosis rate

log_dens = population density (F1)

p_female = proportion female (F2)

p_young = proportion aged ≤30 years (F3)

q_hisp = proportion Hispanic (F4)

q_black = proportion non-Hispanic black (F5)

q_xwhite = proportion of minority race/ethnicity (F6)

p_single = proportion not currently married (F7)

p_pov = proportion below the federal poverty level (F8)

p_hsch = proportion with less than a high school education (F9)

p_unemp = proportion unemployed (F10)

p_moved = proportion moved in the past 12 months (F11)

q_foreign = proportion foreign-born (F12)

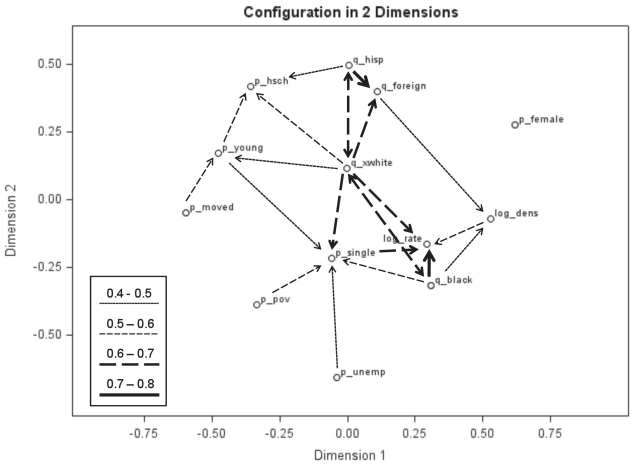
Table 3. Indirect effect of SDH variable X on the AIDS diagnosis rate, through SDH variable Y, based on data at the CPOA level: National HIV Surveillance System and American Community Survey data, 2006–2008

		Y											
		F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*	*F9*	*F10*	*F11*	*F12*
X	*Log_rate*	*0.52*	*0.22*	*0.06*	*0.26*	*0.74*	*0.65*	*0.59*	*0.17*	*0.15*	*0.22*	*0.11*	*0.37*
log_dens (F1)	0.52		0.08	0.00	0.05	0.34	0.18	0.19	–0.06	–0.01	0.02	0.00	0.18
p_female (F2)	0.22	0.18		–0.01	–0.06	0.21	0.03	0.11	0.00	0.00	0.03	–0.02	–0.04
p_young (F3)	0.06	0.02	–0.04		0.09	0.11	0.27	0.27	0.05	0.08	0.02	0.06	0.10
q_hisp (F4)	0.26	0.10	–0.05	0.02		0.03	0.43	0.12	–0.02	0.06	–0.02	0.03	0.31
q_black (F5)	0.74	0.24	0.06	0.01	0.01		0.40	0.31	0.04	0.03	0.06	0.01	0.05
q_xwhite (F6)	0.65	0.15	0.01	0.02	0.17	0.46		0.37	0.04	0.08	0.04	0.03	0.23
p_single (F7)	0.59	0.17	0.04	0.03	0.05	0.39	0.41		0.09	0.04	0.09	0.04	0.10
p_pov (F8)	0.17	–0.18	0.00	0.02	–0.02	0.16	0.14	0.30		0.05	0.08	0.03	–0.10
p_hsch (F9)	0.15	–0.03	0.00	0.03	0.11	0.13	0.33	0.17	0.06		0.03	0.01	0.08
p_unemp (F10)	0.22	0.04	0.03	0.00	–0.02	0.19	0.12	0.25	0.06	0.02		0.01	–0.03
p_moved (F11)	0.11	–0.01	–0.04	0.03	0.07	0.08	0.18	0.19	0.05	0.01	0.01		0.09
q_foreign (F12)	0.37	0.25	–0.03	0.02	0.22	0.10	0.40	0.15	–0.05	0.03	–0.02	0.03	

SDH = social determinants of health
AIDS = acquired immunodeficiency syndrome
CPOA = county-PUMA [public-use microdata area] overlap area
log_rate = AIDS diagnosis rate
log_dens = population density (F1)
p_female = proportion female (F2)
p_young = proportion aged ≤30 years (F3)
q_hisp = proportion Hispanic (F4)
q_black = proportion non-Hispanic black (F5)
q_xwhite = proportion of minority race/ethnicity (F6)
p_single = proportion not currently married (F7)
p_pov = proportion below the federal poverty level (F8)
p_hsch = proportion with less than a high school education (F9)
p_unemp = proportion unemployed (F10)
p_moved = proportion moved in the past 12 months (F11)
q_foreign = proportion foreign-born (F12)

requires knowledge beyond statistics. With additional knowledge of causal relationships, we can apply path analysis and causal modeling using structural equations.[24] In structural equation models, as opposed to functional models, all variables can be considered random rather than fixed. Structural equations express relationships among several variables that can be either directly observed variables (manifest variables) or unobserved hypothetical variables (latent variables).

Figure 2. Strength of correlations and potential causal relationships among demographic and SDH variables and the AIDS diagnosis rate using multidimensional scaling: National HIV Surveillance System and American Community Survey data, 2006–2008[a]



[a]Each pair of variables with a correlation coefficient of >0.4 is connected with a line. Racial/ethnic group variables are connected with a two-headed arrow line, indicating that an increase in one of the two variables results in a decrease in the other variable. Lines connecting a racial/ethnic group variable with another variable are unidirectional and always point from the racial/ethnic group variable to the other variable.

SDH = social determinants of health

AIDS = acquired immunodeficiency syndrome

log_rate = AIDS diagnosis rate

log_dens = population density

p_female = proportion female

p_young = proportion aged ≤30 years

q_hisp = proportion Hispanic

q_black = proportion non-Hispanic black

q_xwhite = proportion of minority race/ethnicity

p_single = proportion not currently married

p_pov = proportion below the federal poverty level

p_hsch = proportion with less than a high school education

p_unemp = proportion unemployed

p_moved = proportion moved in the past 12 months

q_foreign = proportion foreign-born

The approach we have described can be used to assess the SDH of other diseases by linking ACS data to other surveillance systems—for example, sexually transmitted diseases and tuberculosis surveillance systems. The value of this approach is that ACS data collection is ongoing, and the data are accurate and readily available. Also, this approach does not require specialized software or sophisticated statistical methods.

SDH variables can be measured at various geographic levels. The choice of geographic level may affect the analysis results for area-based measures. For example, the health inequality measures at the geographic ZIP-code level can be different from the measures at U.S. census block group or census-tract level.[25,26] From a statisticial point of view, if the geographic area is too small, the area-based SDH measures may be too variable to be meaningful. On the other hand, if the area is too large, important geographic differences may become unidentifiable. The smaller the geographic area, the closer the area-based measures are related to the individual's socioeconomic position. However, if the area is too small, then the area-based measures do not necessarily reflect the socioeconomic conditions that would affect the individual's social behavior. For instance, very few people would live, work, and have social activities only in an area defined by census block or tract. We think the area should be large enough to cover the normal activities of most of the people who live in the area. Our proposed area, the CPOA, is a combination of county and PUMA, and is probably the geographic area that best meets these criteria. The CPOA is clearly defined, rarely changes over time, and, most importantly, is identifiable in many data sources. Of cause, if data are available to identify smaller geographic areas, one can follow the method proposed in this article to identify and compare the impacts of SDH variables measured at different geographic levels.

Although information at the group level is very useful for identifying SDH factors, results based on group-level SDH information may not apply to individuals. For example, we found that a higher proportion of unmarried people in a CPOA was associated with a higher rate of AIDS diagnoses. However, this association does not necessarily mean that the rate of AIDS diagnoses in an area was higher for unmarried people. To draw conclusions at the individual level, one needs SDH information at the individual level.

CONCLUSIONS

Area-based measures of socioeconomic variables can be used to identify risk factors associated with a disease of interest. However, the health effects of demographic

and socioeconomic variables are complex. The impact can be direct or indirect through other variables. The magnitude or even the direction of the impact can be changed due to interactions between variables. When correlation analysis is used to identify risk factors, potential confounding from other variables must be taken into account. The complexities of measuring SDH contributions to disease morbidity, as illustrated in this article, call for careful consideration when developing interventions, programs, and policies to reduce disease transmission and provide access to care.

The authors thank Timothy A. Green, PhD, and Marie Morgan of the Centers for Disease Control and Prevention (CDC) for their helpful comments and editorial suggestions, which led to a significant improvement of the original article.

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of CDC.

REFERENCES

1. Institute of Medicine. Unequal treatment: confronting racial and ethnic disparities in health care. Washington: National Academies Press; 2003.
2. Marmot M. Social determinants of health inequalities. Lancet 2005;365:1099-104.
3. Raphael D. Introduction to the social determinants of health. In: Rafael D, editor. Social determinants of health: Canadian perspectives. Toronto: Canadian Scholar's Press Inc.; 2004. p. 1-18.
4. World Health Organization, Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. Geneva: WHO; 2008. Also available from: URL: http://www.who.int/social_determinants/thecommission/finalreport/en/ [cited 2010 Jun 23].
5. National Center for Health Statistics (US). Design and estimation for the National Health Interview Survey, 1995–2004. Vital Health Stat 2 2000(130).
6. Chowdhury P, Balluz L, Town M, Chowdhury FM, Bartolis W, Garvin W, et al. Surveillance of certain health behaviors and conditions among states and selected local areas—Behavioral Risk Factor Surveillance System, United States, 2007. MMWR Surveill Summ 2010;59(1):1-220.
7. Tarlov AR. Public policy frameworks for improving population health. Ann N Y Acad Sci 1999;896:281-93.
8. Centers for Disease Control and Prevention (US), National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention. Strategic plan, 2010–2015. Atlanta: CDC; 2010. Also available from: URL: http://www.cdc.gov/nchhstp [cited 2010 Jun 22].
9. Department of Health and Human Services (US). Healthy people 2020 [cited 2011 May 16]. Available from: URL: http://www.healthypeople.gov/2020/about/default.aspx
10. Organisation for Economic Co-operation and Development. OECD factbook 2009: economic, environmental and social statistics. Paris: OECD; 2009.
11. Braveman PA. Monitoring equity in health and healthcare: a conceptual framework. J Health Popul Nutr 2003;21:181-92.
12. Braveman P, Gruskin S. Defining equity in health. J Epidemiol Community Health 2003;57:254-8.
13. Kindig D, Day P, Fox DM, Gibson M, Knickman J, Lomas J, et al. What new knowledge would help policymakers better balance investments for optimal health outcomes? Health Serv Res 2003;38(6 Pt 2):1923-37.
14. Satcher D. Ethnic disparities in health: the public's role in working for equality. PLoS Med 2006;3:e405.
15. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. Am J Public Health 2005;95:312-23.
16. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures—the Public Health Disparities Geocoding Project. Am J Public Health 2003;93:1655-71.
17. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health 1997;18:341-78.
18. Census Bureau (US). Measuring America: the decennial censuses from 1790 to 2000. Washington: Census Bureau; 2002. Also available from: URL: http://www.census.gov/prod/2002pubs/pol02marv.pdf [cited 2010 Jun 23].
19. Census Bureau (US). A compass for understanding and using American Community Survey data: what PUMS data users need to know. Washington: Government Printing Office (US); 2009.
20. Rummel RJ. Understanding correlation. Honolulu: University of Hawaii, Department of Political Science; 1976.
21. Schiffman SS, Reynolds ML, Young FW. Introduction to multidimensional scaling: theory, methods, and appplications. New York: Academic Press Inc.; 1981.
22. Taylor EM, Adimora AA, Schoenbach VJ. Marital status and sexually transmitted infections among African Americans. J Fam Issues 2010;31:1147-65.
23. Koball HL, Moiduddin E, Henderson J, Goesling B, Besculides M. What do we know about the link between marriage and health? J Fam Issues 2010;31:1019-40.
24. Bollen KA. Structural equations with latent variables. New York: John Wiley & Sons; 1989.
25. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. Am J Epidemiol 2002;156:471-82.
26. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. 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 2003;57:186-99.