Selecting and Applying a Standard Area-based Socioeconomic Status Measure for Public Health Data: Analysis for New York City

To describe and monitor disparities in health and conditions of public health concern, a standardized and meaningful socioeconomic status (SES)-specific measure that uses readily available data is needed. In 2011, the New York City (NYC) Department of Health and Mental Hygiene adopted neighborhood-level poverty as a standard SES measure for public health data analysis, following the recommendations of a workgroup of Health Department epidemiologists. This report describes the history of and rationale for the neighborhood-level poverty measure and presents analyses of data from several existing surveillance systems to demonstrate how the measure can be used.

Introduction

It is well-established that people with low socioeconomic status (SES) bear a disproportionate burden of poor health and premature death regardless of the city or state in which they live. They are more likely to be exposed to environmental hazards, including illicit drugs and violence. They also are more likely to reside in low-quality, often-crowded housing in neighborhoods with few opportunities for physical activity and limited access to healthy food. Poor health is concentrated in New York City Department of Health and Mental Hygiene May 2013

Key Findings:

- The income gap in New York City (NYC) is among the highest in the United States. A standardized socioeconomic measure that is meaningful and easy to use would help to document and monitor progress in addressing health disparities resulting from SES differences.
- Although many public health data systems do not routinely collect individual socioeconomic indicators such as income and education level, area-level socioeconomic indicators can be determined for any dataset that includes individual-level geographic data, such as residential address.
- The National Institutes of Health-funded Public Health Disparities Geocoding Project (PHDGP) recommends the use of census tract poverty (the percentage of individuals living below the federal poverty threshold within a census tract) as a standard measure.
- The NYC Health Department adapted the PHDGP indicator to NYC and recommended that all analyses of routinely collected surveillance data with geographic information include a neighborhood-level poverty variable.
- While age-adjusted mortality rates in NYC are higher in neighborhoods with greater poverty, this gap narrowed from 1990 to 2000.

“Quality data on health disparities can guide policy discourse and focus attention on the interplay between neighborhood, race, poverty, inequality and health.”

certain neighborhoods in part because health resources and opportunities are unequally distributed.4,5

Reduction of health disparities in NYC would save thousands of lives each year.6 To achieve that goal, disparities must be measured in a way that is meaningful, easy-to-understand and actionable. Quality data on health disparities can guide policy discourse and focus attention on the interplay between neighborhood, race, poverty, inequality and health. Without quality data, public health conversations may exclude socioeconomic factors or frame them in ways that limit rather than promote action.

Rationale for Area-Based Poverty Measures

Most public health surveillance systems do not routinely measure individual SES variables such as income and level of education. Instead, they often use race/ethnicity as a proxy measure of SES.7 This substitution is problematic because racial/ethnic categories combine the effects of culture-specific practices, social differences, economic status and discrimination.8 In addition, categorizing an individual’s race or ethnicity can be difficult because race/ethnicity is inconsistently defined, data are often missing9 and many people self-identify as multiracial or multiethnic. In the 2010 Census, 3% of the United States (U.S.) population, or nine million people, and 4% of New Yorkers reported more than one race.10

Using race/ethnicity as the sole measure of socioeconomic status also obscures the fact that widely differing SES strata are grouped together within racial/ethnic categories. For example, the National Health and Nutrition Examination Survey (NHANES) estimates that the prevalence of diabetes among non-Hispanic black adults in the U.S. is twice that for non-Hispanic whites.11 But an analysis of NHANES data demonstrated that the prevalence among non-Hispanic black adults with household incomes below the federal poverty threshold (FPT) was 1.8 times higher than that among blacks earning over 400% of FPT (See Box 1 for more on the federal poverty threshold). Similarly, diabetes prevalence among non-Hispanic whites living below FPT was 2.1 times higher than that among whites living over 400% of FPT.9 Presenting associations solely by race/ethnicity can deflect attention from equally important

10 United States Census website. Available at: census.gov/hhes/www/poverty/
economic, environmental and neighborhood factors. Finally, the number of new public health data resources that do not regularly capture race/ethnicity (e.g., laboratory reporting, insurance claims data, pharmacy databases) is rising, providing further impetus to agree upon a direct SES measure.

In contrast to individual-level SES measures, SES measures related to neighborhood of residence are widely available. Most public health surveillance systems routinely capture neighborhood of residence (e.g., street address, ZIP code), and neighborhood SES indicators are easily obtained from U.S. Census data.

Neighborhood-level SES, often used as a proxy for individual SES when individual data are unavailable, has been shown to predict a wide range of health outcomes such as mortality, low birth weight, hypertension and attitudes toward smoking, alcohol use and dietary fat intake, independent of individual-level SES. In addition, the magnitude of health disparities between higher and lower SES neighborhoods is often as great as or greater than that among racial/ethnic groups.

Further, a step-wise gradient of health outcomes (or dose-response) is often observed across levels of neighborhood SES. Finally, a standard measure would enable comparisons across health outcomes and geographic areas, and would facilitate the monitoring of disparities based on SES over time.


The Public Health Disparities Geocoding Project

The Public Health Disparities Geocoding Project (PHDGP) is a National Institutes for Health-funded project based at the Harvard School of Public Health. From 2000 to 2005, PHDGP used 1990 Census data and public health surveillance data from Massachusetts and Rhode Island to examine various neighborhood SES measures to identify a readily available and meaningful direct measure of neighborhood-level SES. \(^{13}\) The measures included percentage living below FPT, percentage with less than a high school education, percentage living in housing with at least one person per room, median household income and the Townsend index, a composite measure that uses data on crowding, unemployment, car ownership and home ownership. The PHDGP examined how well each of these measures functioned for three area sizes: census block group, census tract and ZIP code.

The percentage of persons earning below FPT within a census tract, hereafter referred to as census tract poverty, predicted disparities across an array of health outcomes more consistently than the other SES measures and area sizes examined. \(^{14}\)

Adapting a Neighborhood Poverty Measure to New York City

The NYC Health Department formed the Poverty Measures Workgroup in April 2010 to adapt the work of the PHDGP to NYC. The Workgroup, which included staff from six different program areas, met regularly between May and December 2010. The members concluded that a standard measure of SES was needed to allow for consistency when presenting SES-related findings to enable comparisons between jurisdictions and the national level, and to evaluate trends and effects of interventions over time.

Several important issues needed to be considered in adapting the suggested neighborhood poverty measure and cut-points to NYC. First, the distribution of the NYC population differs in important ways from the population in which the PHDGP SES indicator was developed. The cut-points proposed by the PHDGP resulted in four nearly equally sized groups in Massachusetts and Rhode Island but very different sized groups in NYC.

\(^{14}\) Public Health Disparities Geocoding Project Monograph. Available at: [hsph.harvard.edu/thegeocodingproject/](http://hsph.harvard.edu/thegeocodingproject/)
In 2010, only 10% of the NYC population resided in census tracts with <5% neighborhood-level poverty but 43% of the population resided in census tracts with ≥20% neighborhood-level poverty. The Workgroup ultimately arrived at a six-category expansion of the original PHDGP-recommended four categories: <5%, 5% to <10%, 10% to <20%, 20% to <30%, 30 to <40% and ≥40% residents living below FPT. This six-category expansion provides the flexibility to present data in categories meaningful to NYC while retaining comparability to data analyzed using the PHDGP-recommended categories. The six categories can be collapsed to four categories that comprise more equally sized groups of the NYC population as follows: <10%, 10% to <20%, 20% to <30% and ≥30%. Alternatively, the six categories can be collapsed to the four PHDGP-recommended categories when comparison across jurisdictions is desired.

Second, there are key data sources in NYC that do not contain street address data necessary for census tract-level analysis. For example, the annual Community Health Survey, which provides representative data on population health indicators such as disease prevalence, behavioral risk factors and environmental exposures in NYC, does not collect street address. The smallest possible geographic unit for analysis of this data is ZIP code. In addition, many Health Department publications have used even larger United Hospital Fund (UHF) areas as the geographic unit for area-based analysis and display (See Box 2 for more on census tracts, ZIP codes and UHF areas).

---

**Box 2: Census tract, ZIP code and UHF area**

**Census tract**
- Census tracts were created to provide a stable set of geographic units for presentation of decennial Census data. They are relatively permanent, having been delineated with the intention of enabling comparisons between Census periods. However, Census tracts are occasionally divided due to population growth or combined as a result of population decline.
- Census tract boundaries were drawn to encompass populations homogenous in relation to economic status and living conditions.
- Census tracts are used by federal, state and local agencies to determine eligibility and resource allocation for programs such as urban empowerment zones, medically underserved areas and income-dependent tax credits.
- Census tracts generally contain between 1,500 and 8,000 people, with an average size of 4,000 people.

**ZIP (Zone Improvement Plan) code**
- The ZIP code is a postal coding system launched in 1963 in response to increasing volumes of mail.
- The first number in the code represents a general geographic region, from “0” in the East, moving to “9” in the West. The next two numbers represent regional areas and the final two identify specific post offices.
- ZIP codes generally include 30,000 or more people. They are typically not socio-demographically homogeneous and do not conform to census statistical area boundaries.
- Because ZIP codes do not correspond to census geographical entities, the Census Bureau produces ZIP Code Tabulation Areas (ZCTAs) to approximate ZIP code service areas. A description of how ZCTAs are created is available on the Census Bureau website: [census.gov/geo/reference/zctas.html](http://census.gov/geo/reference/zctas.html).

**United Hospital Fund (UHF) area**
- UHF areas were defined by United Hospital Fund staff by combining ZIP code areas to approximate the boundaries of community planning districts. The current community planning district boundaries were established by local law in 1975. Each of the 59 districts is represented by a Community Board which serves as an advisor to city government on matters relating to its district.
- There are 42 UHF neighborhoods, each containing an average of 200,000 individuals.
- UHF area boundaries do not exactly correspond to community planning district boundaries because ZIP code areas and community planning district boundaries overlap.

---

15 United States Postal Service website: about.usps.com/welcome.htm
16 United Hospital Fund. New York City Community Health Atlas, 2002. Available at: uhfnyc.org/publications/99007
17 New York City Department of City Planning website: nyc.gov/html/dcp/
Using different neighborhood sizes to calculate neighborhood-level SES measures requires an understanding of the effect of neighborhood size on the description of disparities. The Workgroup examined the effect of neighborhood size on the proportion of the population in each neighborhood poverty-level group using 2000 Census data. As neighborhood size increased from census tract (2,217 geographic units) to ZIP code (176) to UHF area (42), the percentage of New Yorkers in the highest and lowest neighborhood poverty categories decreased (Figure 1).

In 2000, only 7% of New Yorkers resided in census tracts where <5% of the population lives below FPT. Two percent of New Yorkers resided in ZIP codes where <5% of the population lived below FPT and there were no UHF areas where <5% of the population lived below FPT. At the other end of the spectrum, 12% of New Yorkers resided in census tracts where ≥40% of the population lived below FPT, compared with 8% in ZIP codes and 6% in UHF areas.

Using four categories created more equally-sized groups (Figure 2): 26% of New Yorkers resided in census tracts, 19% in ZIP codes, and 11% in UHF areas where <10% of the population lived below FPT. For the highest poverty category, 25% of New Yorkers resided in census tracts, 25% in ZIP codes, and 26% in UHF areas where ≥30% of the population lived below FPT.
To further evaluate the applicability of the neighborhood-level poverty measure to NYC data, the Workgroup applied it to two previously geocoded Health Department datasets: all-cause mortality and tuberculosis. Both outcomes are known to be associated with SES. Rates of each health outcome were determined for each poverty category using permutations of each neighborhood size (census tract, ZIP code and UHF area) and both the six and four-level poverty categories. In addition, for the six categories of census tract poverty, rates were determined within each of four major racial/ethnic groups and for two time periods (1990 and 2000 for mortality; 2000 and 2008 for tuberculosis). Population denominators and neighborhood poverty classification for all analyses were based on 2000 Census data except for the 1990 mortality data, which used 1990 Census data. Chi-square tests for trend were performed and were significant at the p<0.05 level or below for each comparison presented here.

**All-Cause Mortality**

Crude all-cause mortality rates did not show clear disparities by neighborhood-level poverty, but when the rates were age-adjusted so that each poverty group had the same age structure, a clear and stepwise gradient of increasing mortality with increasing neighborhood poverty emerged (Figure 3). Age-adjusted mortality ranged from 5.9 per 1,000 for census tracts with <5% of the population in poverty to 10.2 per 1,000 for census tracts with ≥40% poverty.

The findings and the degree of disparities were similar whether the census tract, ZIP code or UHF area neighborhood size was used (Figure 4).

---

**Figure 3** Crude and age-adjusted mortality rates by census tract poverty, NYC, 2000

![Crude and age-adjusted mortality rates by census tract poverty, NYC, 2000](image)

Source: U.S. Census 2000; NYC Health Department, Bureau of Vital Statistics

**Figure 4** Age-adjusted mortality rates by census tract, ZIP code and UHF area poverty, NYC, 2000

![Age-adjusted mortality rates by census tract, ZIP code and UHF area poverty, NYC, 2000](image)

Source: NYC Health Department, Bureau of Vital Statistics

---

For mortality, a neighborhood SES gradient was present within each major racial/ethnic group (Figure 5). For example, among non-Hispanic blacks, mortality rates ranged from 7.3 per 1,000 for areas with <5% poverty to 12.8 per 1,000 for areas with ≥40% poverty.

To examine use of census tract poverty over time, mortality rates in two time periods were analyzed. From 1990 to 2000 mortality rates declined in each neighborhood SES category (Figure 6). Among census tracts with <5% of the population in poverty, the mortality rate decreased from 7.4 to 5.9 per 1,000 from 1990 to 2000, a 21% reduction. Among census tracts with ≥40% poverty, mortality dropped from 13.4 to 10.2 per 1,000 in the same time period, a 24% reduction. The difference in mortality rates between the highest and the lowest SES groups narrowed from 6.0 to 4.3 per 1,000.

**Tuberculosis**

Analysis of tuberculosis (TB) data showed increasing incidence of TB with higher neighborhood poverty for each neighborhood size and for both the six- and four-category poverty classifications (data not shown). A neighborhood SES gradient was present within each major racial/ethnic group (Figure 7). For example, among non-Hispanic blacks, TB incidence ranged from 9.99 per 100,000 in...
To examine trends over time, data from 2008 (the most recent year for which geocoded tuberculosis data were available) were compared with data from 2000. There was a decrease in TB incidence in all neighborhood-level poverty groups, regardless of neighborhood size. Analysis at the census tract level showed that the absolute magnitude of the decreases in TB incidence was higher in poorer neighborhoods (Figure 8). In areas with <5% poverty, TB incidence decreased from 4.9 to 4.1 per 100,000 whereas in areas with ≥40% poverty, TB incidence decreased from 26.4 to 15.6 per 100,000 in the same time period.

Conclusions and Recommendations

This report demonstrates the rationale, feasibility and usefulness of an area-based SES measure. The NYC Health Department has adopted neighborhood-level poverty, or the percent of population in an area living below FPT, as an agency standard variable. All routinely collected surveillance data that include geographic information (i.e., street address) should be geocoded, and analyzed and presented with neighborhood-level poverty as a standard variable. The Health Department additionally has
recommended that surveillance datasets for which individual-level SES is collected should also be analyzed using neighborhood-level poverty as a standard variable. Neighborhood-level poverty should be measured, analyzed, and displayed in a way that enables its use as a measure by other jurisdictions and allows comparisons across jurisdictions, particularly urban ones.

The neighborhood unit used for analysis and presentation should be the smallest geographic area that the data support and release policy allows, with census tract preferable to ZIP code preferable to UHF area. Importantly, the adoption of this measure does not preclude using other area-based or individual SES measures. It is possible that in the future, different SES measures may be found to better describe disparities and suggest interventions for some outcomes.

Six categories of neighborhood-level poverty should be used for data analysis, with the potential to collapse to four categories for data display. The standard cut-points for defining categories of neighborhood-level poverty in NYC should be: <5%, 5% to <10%, 10% to <20%, 20% to <30%, 30% to <40%, and ≥40% residents living below FPT. Collapsed categories should be: <10%, 10% to <20%, 20% to <30%, and ≥30%, corresponding to low-, medium-, high- and very high-poverty categories.

To enable all NYC Health Department programs to conduct these analyses, a repository of standard census tract, ZIP code, and UHF denominators for 1990, 2000, and 2010 is available, along with intercensal population estimates, where possible. Standard denominators are stratified by sex, age, race/ethnicity and neighborhood-level poverty in the recommended six and four categories, when possible. Neighborhood-level poverty for the years 1990 and 2000 was determined by using U.S. Census data collected in those years. As of 2010, the decennial U.S. Census does not collect the data necessary to determine neighborhood-level poverty. Those data are now collected by the American Community Survey (ACS) administered by the Census Bureau. The ACS samples a small percentage of the population every year and generates census tract poverty estimates by pooling the most recent five years of data.¹⁹ For data from 2005 forward, the area-based poverty classification of each census tract will change annually to reflect updated estimates from the ACS. For data from 2007 forward, ZIP Code and UHF poverty classifications will be updated annually. In addition, the Health Department has provided Statistical Analysis System (SAS) datasets and methodology documentation to aid its analysts in conducting standard neighborhood-

level SES analyses, including trend analyses. The new methodology includes guidance related to the preferred poverty definition and denominators to use for given years of data.

Health disparities in NYC and the U.S. are longstanding and pervasive. To continue to make headway in reducing these disparities, we must measure them in a way that is meaningful and actionable. In the past health disparities have generally been presented solely in terms of racial inequality. This manner of analyzing disparities by itself is not ideal because race comprises a combination of economic, social and cultural factors. The neighborhood-level poverty measure is a readily available tool to measure economic disparities in a way that is easy to understand. Its adoption as a standard variable for analysis and presentation should contribute to a greater understanding of the effects of neighborhood, race and poverty on health outcomes.

**Technical Notes**

**Data Sources:** Mortality data from 1990 and 2000 were collected and analyzed by the Health Department’s Bureau of Vital Statistics. Tuberculosis data from 2000 and 2008 were collected and analyzed by the Health Department’s Bureau of Tuberculosis Control.

Population denominators from the U.S. Census 2000 were used for the calculation of all rates, except for 1990 mortality rates, which used U.S. Census 1990 data. All age-adjusted rates were adjusted to the U.S. 2000 Standard Population. Neighborhood poverty classifications were based on data from the U.S. Census 2000, except for 1990 calculations, which used the U.S. Census 1990.

**Data Limitations:** This report has the following limitations. The datasets analyzed were selected based on availability of geocoded data and Poverty Measures Workgroup members’ interest. They do not represent all data types that the Health Department routinely analyzes. The analyses also exclude individuals for whom no street address is available, such as homeless people.

**Interpretation of Data Findings:** The statistical analyses presented in this report were conducted to demonstrate the relationship between area-based poverty and various health outcomes. The mechanisms underlying that relationship are heterogeneous and complex. We do not attempt to explain them in this report.