

Income-Related Health Disparity and Its Determinants in New York State: Racial/Ethnic and Geographical Comparisons¹

Kajal Lahiri and Zulkarnain Pulungan

Summary:

Using self-assessed health status together with several indicators of individual morbidity and socio-demographic characteristics, we study the quality of health and income related health disparity in five racial/ethnic groups as well as across 17 geographic areas of New York State. The American Indian/Alaskan Natives and Hispanics are found to do the worst, whereas, geographically, the North Country in Upstate New York and Bronx County in Downstate score the worst on both counts. Three major contributing factors to income related health disparity are found to be household income, employment status, and education. However, the contribution of each of these determinants varies significantly among racial/ethnic groups as well as across geographic areas, suggesting targeted public policy initiatives to eliminate health disparity between rich and poor.

Keywords: BRFSS data; Self-assessed health; Ordered Probit; Income related health inequality; Concentration index; Concentration curve; Decomposition analysis

¹ This study has been supported by the Center for the Elimination of Minority Health Disparities, University at Albany, SUNY, and has benefited from presentations at the First Annual Health Disparities Conference, Columbia University (March 10–11, 2006), NIH Health Disparity Conference (October 23–24, 2006), and at a roundtable discussion at the Center for the Elimination of Minority Health Disparities, University at Albany. Comments and help from Richard Alba, Bill Jenkins, Richard Keller, Catherine Lawson, Louis McNutt, Thad Mirer, Jim Scanlan, Larry Schell, and Barbara Wallace are gratefully acknowledged. We, however, are responsible for all remaining errors and omissions.

INTRODUCTION

Persistent health inequality in the United States along multiple dimensions has been put at the forefront by one of the challenging goals of *Healthy People 2010* (USDHHS, 2000): to eliminate health disparities among all segments of the population, including differences that occur by gender, race, or ethnicity, education or income, disability, geographic location, and sexual orientation. In this paper we study the quality of health and health inequality among racial/ethnic groups as well as across geographic areas of the State of New York. Even though certain aggregate indicators of health (e.g., life expectancy at birth, mortality rate, etc.) in New York have improved during last few decades (see NCHS, 2006), health disparities among racial/ethnic groups and among regions continue to exist. For example, as we will show below, the prevalence of diabetes is almost twice as high among Blacks compared to that among Whites; on the other hand, many New York City neighborhoods and unsuspecting areas of Upstate New York are characterized by extreme poor health. This is the first study to look at the health status and its disparity among New Yorkers along these dimensions.

A question arises as to what causes poor health and health disparities. A large number of studies have reported that socioeconomic status (SES) is a key factor affecting quality of health and health disparity (see for example, Adler & Newman, 2002; Cutler & Lleras-Muney, 2006; Adams, Hurd, McFadden, Merrill, & Ribeiro, 2003; Cutler, Deaton, & Lleras-Muney, 2006; Deaton, 2006). There are four broad pathways—health care, environmental exposure, health behavior, and chronic stress—through which SES affects health (Adler & Ostrove, 1999). Because SES is an important mediator for quality of health, studying health disparity cannot be separated from studying disparity in SES.

Rawls' First Principle of Justice (1971) requires that all individuals should have the same opportunity to achieve their potential health levels; see Bommier and Stecklov (2002). An egalitarian viewpoint of social justice requires that people in equal need of health care be treated equally, irrespective of characteristics such as income, place of residence, race, and so forth. Since discrimination in access to health care is likely to be based on income, we focus on income-related health inequality in this paper. There seems to be broad consensus among health policy analysts that socioeconomic inequality in health is indeed inequitable and unjust, and is consistent with the Institute of Medicine (2002) definition of health disparity—any difference in health after adjusting for health care needs. This definition recognizes that factors such as income may be mediators of disparity in health care.²

Numerous studies on measuring quality of health and health distributions have focused on mortality rates, prevalence of diseases/risk factors, psychological morbidity, quality of or access to health care services, and health care utilization rates.³ In addition to looking at many of these factors, in this study we focus on a measure of health more generally, and calculate an index of health and health inequality based on self-assessed health (SAH) status. SAH is defined as the response to the survey question “Would you say that in general your health is: excellent, very good, good, fair, or poor?” (Centers for Disease Control and Prevention (CDC), 1999-2004).

SAH has been shown to be a good measure of overall health status. In their review, Idler and Benyamini (1997) show that SAH has strong predictive validity for

² We found that between 40% and 50% of the total health inequality in our sample is due to income-related health inequality—an estimate that is much higher than 25% reported by Wagstaff and van Doorslaer (2004) for Canada.

³ See, for instance, Williams and Collins (1995), Ayanian, Weissman, Chasan-Taber, and Epstein (1999), and Shishehbor, Litaker, Pothier, and Lauer (2006).

mortality. Sickles and Taubman (1997) compiled results from worldwide studies on the association between self-assessed health and mortality, and reported that a lower level of SAH has higher mortality odds. Manor, Mathews, and Power (2001) found that SAH has a strong association with longstanding illness. Furthermore, Lahiri, Vaughn, and Wixon (1995) show that SAH is a useful predictor of the severity of diseases and disability. Humphries and van Doorslaer (2000) found that health inequality calculated on the basis of SAH status gives similar results to the results calculated based on a more objective health indicator (viz. McMaster Health Utility Index). More recently, Safaei (2006) finds SAH to be statistically more reliable than the binary chronic conditions as a measure of overall health.

In this paper we generate a continuous measure of health by modeling the five-category SAH as an Ordered Probit Model (McKelvey & Zavoina, 1975) conditioned by several objective determinants including different diseases, behavioral risk factors, and socio-demographic characteristics. The estimated values from this model are used as a measure of individual health and income related health inequality using concentration index and concentration curve (Kakwani, Wagstaff, & van Doorslaer, 1997). Furthermore, to be useful for policy purposes, the income related health inequality is decomposed into its determinants (Wagstaff, van Doorslaer, & Watanabe, 2003) for the whole sample and specific sub-samples.

The paper is organized as follows: The estimation procedures—the methods to calculate quality of health, income related health inequality and their determinants—are described; the data used in the empirical analysis is documented; results are then presented; and finally we summarize our conclusions.

METHODS

In modeling SAH we follow the same procedures as Cutler and Richardson (1997, 1998) and Groot (2000). In the empirical modeling of the quality of health, three related concepts are distinguished: a true quality of health denoted as h^* , a vector of objective measures of health denoted as \mathbf{h}^o , and a subjective measure of health denoted as h^s . The true quality of health is a latent variable, which is unobservable. What we observe is a vector of objective indicators and a subjective measure of health. The true unobserved quality of health h^* is assumed to be a function of the vector of observed and objective measures of health, and a vector of individual characteristics denoted by \mathbf{x} . The subjective measure of health is measured on an ordinal scale with m self-assessed response categories. For the purpose of measuring health and health inequality we transform these ordinal responses into a cardinal measure. In this paper we used an ordered response model to transform the order scale variable into a cardinal variable. To control for possible heterogeneity in self-assessed health, we estimate an Ordered Probit model with heteroskedasticity in errors. The model is formulated as follows:

$$h_i^* = \mathbf{h}_i^o \boldsymbol{\gamma} + \mathbf{x}_i \boldsymbol{\beta} + s(\mathbf{z}_i, \boldsymbol{\eta}) \varepsilon_i \quad (1)$$

$$h_i^s = j \Leftrightarrow \mu_j \leq h_i^* \leq \mu_{j+1} \text{ for } j = 0, 1, \dots, m-1$$

$$\mu_0 = -\infty \text{ and } \mu_m = +\infty$$

$$i = 1, 2, \dots, n$$

where $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, $\boldsymbol{\eta}$ are vectors of coefficients, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{m-1})$ is an unknown vector of thresholds to be estimated together with the vectors of coefficients, ε_i is the error term and is assumed to be normally distributed, $s(\mathbf{z}_i, \boldsymbol{\eta}) = \sigma \sqrt{(1 + \exp(\mathbf{z}_i \boldsymbol{\eta}))}$ is a scale function

to control for heteroskedasticity, and n is the number of observations. \mathbf{z}_i is a vector of observed variables that affect the variance of the error term.⁴

The model is estimated using the maximum likelihood approach. The predicted quality of health, $\hat{h}_i^* = \mathbf{h}_i^0 \hat{\boldsymbol{\gamma}} + \mathbf{x}_i \hat{\boldsymbol{\beta}}$, is used as a measure of individual health. The predicted health from the estimated Ordered Probit model will purge at least some part of the variation in SAH that is due to subjective idiosyncrasies of the respondents, not supported by objective health measures. Following van Doorslaer and Jones (2003), we rescale this prediction to be in the $[0, 1]$ interval as $h_i = (\hat{h}_i^* - \hat{h}_{\min}^*) / (\hat{h}_{\max}^* - \hat{h}_{\min}^*)$, where \hat{h}_{\max}^* and \hat{h}_{\min}^* are the maximum and the minimum of the predicted quality of health, respectively.

Using the estimated quality of health h_i , we measure income related health inequality using concentration curves and health concentration index (Wagstaff, Paci, & van Doorslaer, 1991).⁵ A concentration curve plots the cumulative proportion of health $L(s)$ against the cumulative proportion of population s (starting with the lowest socioeconomic status and ending with the highest socioeconomic status), as shown in Figure 1. If the concentration curve $L(s)$ coincides with the diagonal, health is equally distributed over socioeconomic status. This means that there is no inequality in health across socioeconomic hierarchies in the population. The farther the concentration curve from the diagonal the larger is the degree of inequality. The area between concentration curve and the diagonal provides a measure of inequality. The concentration index is defined as twice the area between the concentration curve and the diagonal. The coefficient ranges from 0 (when across income everybody enjoys exactly the same

⁴ van Doorslaer and Jones (2003) have shown that this heteroskedastic model accommodates possible individual-specific heterogeneity in the subjective thresholds $\boldsymbol{\mu}$.

⁵ See also Lecluyse and Cleemput (2005) and Clarke and Ryan (2006).

health) to 1 (when all population's health is concentrated in the hands of the richest individual).

<<Fig. 1 approx here>>

The concentration index can be calculated using equation (see Kakwani et al., 1997):

$$\hat{C} = \frac{2}{n\mu} \sum_{i=1}^n h_i R_i - 1 \quad (2)$$

where R_i is the i th individual fraction rank in socioeconomic status and μ is the mean of quality of health. The variance is estimated using the Huber–White procedure. The disadvantage of the concentration index is its lack of straightforward interpretation in a natural unit, while its advantage is that it takes into account both coefficient variation of health and correlation between health and income rank (Milanovic, 1997). Koolman and van Doorslaer (2004) present a lucid interpretation of the concentration index.

Furthermore, to be more meaningful for policy purposes, income related health inequality is decomposed into its determinants as demonstrated by Wagstaff et al. (2003). Define a vector of explanatory variables as $\mathbf{w} = (\mathbf{h}^0 \quad \mathbf{x})$. Given the relationship between health and explanatory variables as in equation (1), the concentration index can be written as

$$\hat{C} = \sum_{k=1}^K (\hat{\beta}_k \bar{w}_k / \bar{h}) C_k \quad (3)$$

where \bar{h} is the mean of h , \bar{w}_k is the mean of variable w_k from the vector of explanatory variables \mathbf{w} , and C_k is the concentration index of variable w_k .

DATA, DESCRIPTIVE STATISTICS, AND IMPUTATION

Data and Descriptive Statistics

The data used in this study are obtained from the New York State sample of the *Behavioral Risk Factor Surveillance System (BRFSS)* over 1999–2004, with a total of 22,083 sample observations.⁶ Every year health departments of all states, with technical and methodological assistance from CDC, conduct monthly telephone interviews on randomly selected noninstitutional adults aged 18 years or older. The surveys are developed and conducted to monitor major behavioral risks among adults associated with premature morbidity and mortality. The number of observations is not the same for all variables. The differences can be attributed to (i) the absence of some questions in some years—for example, coronary heart disease was asked only in the interviews for the years of 1999, 2001, and 2003; and (ii) missing values due to “do not know,” “not sure” responses, and refusals to answer. In addition, population socioeconomic characteristics are obtained from 2000 census information (U.S. Census Bureau, 2000).

Based on Census 2000 (U.S. Census Bureau, 2000), New York State population is 18,976,457 persons, which is 6.7% of the U.S. population. Sixty two percent of the population is non-Hispanic White, 15.9% is Black, 15.1% is Hispanic, 5.5% is Asian, and 0.4% is American Indian and Alaskan Native. As a comparison, the U.S. population consists of 69.1% non-Hispanic White, 12.3% Black, 12.5% Hispanic, 3.6% Asian, and 0.9% American Indian and Alaskan Native. So the percentage of minority population of New York State is higher than that of the United States.

In this paper, we divide New York State into 17 geographic areas, which consist of 9 counties of Downstate and 8 economic development regions of Upstate. Upstate

⁶Sehili, Elbasha, Moriarty, and Zack (2005) have used these data source to study health inequality in the United States in terms of physically healthy days.

New York was divided into economic development regions due to small samples in individual counties. Racial/ethnic groups included in the comparison are non-Hispanic White (White), non-Hispanic Black (Black), Hispanic, Asian/Pacific Islander (Asian), and American Indian and Alaskan Native (AIAN).

<<Fig. 2 approx. here>>

<<Fig. 3 approx here>>

The population characteristics vary across New York State geographic areas, especially between Downstate and Upstate of New York. Sixty three percent of the New York State residents live in Downstate. Percentage of individuals who live in poverty in Downstate varies from 5% (Nassau County) to 25% (Bronx County), while in Upstate it varies from 4% (Putnam County) to 15% (Allegany County) as presented in Figure 2.⁷ Income inequality, measured by Gini coefficient⁸, varies across the areas from 0.33 (Nassau County) to 0.45 (Bronx County) for Downstate; and for Upstate, it varies from 0.36 (Hudson Valley) to 0.42 (North Country) (see Figure 3). In addition, the racial/ethnic composition of the population also varies considerably across the areas, where Downstate population is more diverse compared to Upstate population. For example 21% of Downstate population is Black compared to 7% of Upstate population. Figure 4 presents the percent Black population across the areas.

<<Fig. 4 approx here>>

⁷All maps in this paper were created using ArcView GIS (Environmental Systems Research Institute [ESRI], 2002). The breakpoints between classes are determined using a statistical formula (Jenks optimization) that minimizes the variance within each class.

⁸Lorenz curve plots the cumulative proportion of income against the cumulative proportion of individuals ranked by income. The Gini coefficient is defined as twice the area between the Lorenz curve and the diagonal, and ranges from 0 to 1.

We follow the BRFSS guideline that the minimum number of observations to be meaningful for interpretation is 50. Specifically, the description of variables used in this study is as follows.

Socio-demographic variables. The average age of respondents in the sample is 45 years. Comparing racial/ethnic groups, White has the highest average age (48 years) and Asian has the lowest (39 years). Since age is an important determinant of health, we need to adjust for differing age distributions when comparing the prevalence of the diseases among racial/ethnic groups or areas. Across the geographic areas of New York State, the average age varies from 43 years (Queens County) to 49 years (Mohawk). Sixty percent of White respondents are married, compared to 37% of Black respondents. The percentage of married respondents in New York County is the lowest (35%) compared to other areas.

Education level varies considerably among racial/ethnic groups as well as across the areas. Only 13% of AIAN respondents have 4 years or more of college, compared to 61% of Asian respondents; across the areas, the percentage varies from 19% in Bronx County to 52% in New York County. The percentage of respondents who were unable to work varies from 1.6% (Asian) to 14% (AIAN), while the percentage across the areas varies from 1.3% (Westchester County) to 11.3% (North Country). Annual household income also varies considerably among racial/ethnic groups as well as across the areas. AIAN average annual household income is \$34,390, compared to \$62,470 of Asian. Across the geographic areas, the average varies considerably from \$44,579 (Kings County) to \$75,831 (Rockland County).

Eighty-six percent of the respondents have health insurance plans, but only 68% among Hispanics have health plans compared to 91% among Whites. Across the areas, the percentage varies from 77% (Queens County) to 92% (Western New York). The percentage of respondents who could not afford to see a doctor in the past 12 months also varies considerably among racial/ethnic groups, ranging from 8% (White) to 21% (Hispanic). Twenty percent of Rockland County respondents could not afford to see a doctor at least once, while in the Capital Region only 7% of the respondents had that experience.

Bad health habits. The U.S. Surgeon General’s Report (US-DHHS, 2004) has concluded that smoking is a source of many kinds of diseases and harms every organ of the body. Thus, smoking status could be a good explanatory variable to be included in equation (1). The percentage of smokers among racial/ethnic groups varies noticeably, ranging from 13% (Asian) to 34% (AIAN); the percentage of smokers across the state varies from 15% (Westchester County) to 28% (North Country) (see Figure 5). In addition to smoking, not exercising is categorized as a bad health habit that can result in several kinds of diseases. The percentage of respondents who participated in any sort of exercise in the past 30 days is 74%. Sixty-two percent of Hispanic respondents participated in any exercise in the past 30 days, compared to 79% for White respondents; across the areas, Bronx County has the lowest percentage (66%) and North Country has the highest (80%).

<<Fig. 5 approx here>>

Self-assessed health status. Figure 6 presents the distribution of SAH by racial/ethnic groups. Twenty-two percent of the respondents considered their health as “excellent”,

while only 4% considered their health as “poor”. The distribution varies considerably among racial/ethnic groups. The percentage of “excellent” and “very good” health is significantly higher for Whites than for Blacks, Hispanics, or AIANs. In addition, the percentage of “fair” and “poor” health is lower for Whites than for Blacks, Hispanics, or AIANs. The percentage of “excellent” and “very good” health varies from 43% (Bronx County) to 62% (Capital District), and the percentage of “fair” and “poor” health varies noticeably from 10% (Finger Lakes) to 24% (Bronx County). Based on both criteria, Bronx County has the lowest quality of health compared to the other areas. It is noteworthy that different cut points (criteria) can give different classifications. For example, if we use the percentage of “excellent” health as the criterion of quality of health then Westchester does the best.

<<Fig. 6 here>>

Figure 7 presents the distribution of SAH by income groups. The figure indicates that as income increases the percentage of “excellent” health increases and the percentage of “poor” health decreases. This pattern indicates a strong association between income and quality of health—the so-called negative income health gradient.

<<Fig. 7 here>>

Number of days physical and mental health not good in the past 30 days. The pattern of number of days where physical health was not good is the same as the number of days where mental health was not good. Comparing racial/ethnic groups, Asian has the lowest average on both measures and AIAN has the highest. Across the areas the patterns of these two measures are different (see Figures 8 and 9). The lowest average number of days where physical health was not good is 2.6 days (Rockland County) and the highest

average is 4.3 days (Mohawk). While the lowest average number of days where mental health was not good is 1.3 days (North Country), and the highest is 4.2 days (Bronx County).

<<Fig. 8 approx here>>

<<Fig. 9 approx here>>

Limited activities due to health problems. The overall prevalence of limited activities due to health problem is 16%. The prevalence varies substantially among racial/ethnic groups ranging from 6% (Asian) to 31% (AIAN). Across the areas, the prevalence varies from 12% (Queens County) to 22% (Rockland County).

Body mass index (BMI). Obesity—defined as BMI greater than 30 kg/m²—is the second leading cause of preventable death after smoking in the United States and is as a major cause of morbidity and disability (Must et al., 1999; Mokdad, Karks, Stroup, & Gerberding, 2004). Hence this variable is a good predictor of quality of health to be included in equation (1). Comparing racial/ethnic groups, the average BMI varies from 24 (Asian) to 28 (Black); across the areas it varies from 25 (New York County) to 28 (North Country). Figure 10 presents the prevalence of obesity across the geographic areas.

<<Fig. 10 approx here>>

Asthma. The prevalence of asthma among respondents is 11.8%. Comparing racial/ethnic groups, the prevalence varies substantially from 5.9% (Asian) to 17.7% (AIAN). Across the areas, the prevalence varies from 8.5% (Westchester County) to 17.4% (North Country). Figure 11 presents the prevalence of asthma across the areas.

<<Fig. 11 approx here>>

Hypertension (high blood pressure). The prevalence of high blood pressure among respondents is 24.3%. It is only 12.2% for Asians compared to 31.2% for AIANs. Across the areas, the prevalence varies from 20.9% (New York County) to 33.5% (Mohawk). Figure 12 presents the prevalence of hypertension across the areas.

<<Fig. 12 approx here>>

Coronary heart disease. Coronary heart disease prevalence is 4.2% in the pooled sample. The prevalence varies from 2% (Asian) to 7% (AIAN). Blacks, Hispanics, and Asians have lower prevalence than Whites. Across the regions, the prevalence varies from 3.3% (Richmond County) to 5.2% (Bronx County and Nassau County). North Country and Rockland County are excluded from the comparison for lack of sufficient observations.

Myocardial infarction. The prevalence of myocardial infarction is 3.5%. Asians have the lowest prevalence (less than 0.1%) and AIANs have the highest (7.6%). Across the areas, the prevalence varies from 2.2% (Westchester County) to 5.3% (Central New York). North Country is excluded from the comparison for inadequate number of observations.

Stroke. The prevalence of stroke is 1.9%. Among racial/ethnic groups the prevalence varies considerably, ranging from 0.4% (Asian) to 6% (AIAN). The prevalence varies considerably across the areas, ranging from 1.5% (Hudson Valley) to 4.4% (Southern Tier). North Country and Rockland County are excluded from the comparison.

Diabetes. The overall prevalence of diabetes is 6.4%. It varies from 4.7% for Asians to 11.6% for AIANs. Across the areas, the prevalence varies from 5.0% (Suffolk County) to 14.1% (North Country). Figure 13 presents the prevalence across the areas.

<<Fig. 13 approx here>>

Arthritis. The prevalence of arthritis is 25.1% among the respondents. The prevalence varies substantially among racial/ethnic groups, ranging from 8.1% (Asian) to 35.1% (AIAN). Across the State, the prevalence varies from 18.2% (Bronx County) to 34.7% (North Country).

Pain, aching, stiffness, and swelling in or around a joint. The prevalence of this medical condition is 38.3%. The prevalence varies considerably, ranging from 23.7% for Asians to 52.5% for AIANs. Across the areas, the prevalence varies from 29.0% (Kings County) to 46.7% (Western New York).

High blood cholesterol. The prevalence of high blood cholesterol is 31%. The prevalence varies from 24.5% for Hispanics to 48.3% for AIANs. Across the areas the prevalence varies from 25.5% (Southern Tier) to 41.6% (Rockland County).

Multiple Imputation

Information on some diseases and risk factors was not collected in some survey years. Table 1 presents the pattern of missing values attributed to the absence of questions in the survey questionnaires. For example, during 1999–2004, information on high blood pressure was collected only in 1999, 2001, and 2003 and not in 2000, 2002, and 2004.

<<Table 1 approx here>>

In order to include all important diseases and risk factors as covariates in equation (1), we needed to fill in the missing values in our pooled sample. Otherwise, an omitted variable bias would result in the coefficient estimates of included variables. A currently accepted procedure to impute missing values is the multiple-imputation method of Rubin

(1987) and Schafer (1997). The basic idea of multiple imputations is to create two or more completed datasets using the correlation structure of the available covariates, and then do analysis on each completed dataset. Subsequently, we make inferences based on both within and between variability of the estimates obtained from the completed datasets.

In this method, the missing values are filled in by drawing random samples from the conditional distribution of missing values given the observed values. Assuming the joint distribution of the variables is multivariate normal, and using the Markov Chain Monte Carlo (MCMC) method to obtain simulation-based estimates of the posterior parameters of the distribution, values from the conditional distribution for the missing values are drawn randomly given the observed values. It is noteworthy that most of our missing values are binary, rather than multivariate normal. However, Horton, Lipsitz, and Parzen (2003) show that the parameter estimates from the imputed dataset are unbiased as long as the imputed values are not rounded to binary (0, 1) values.

The performance of the multiple-imputation method can be seen in our case by comparing the descriptive statistics of the imputed variables before and after imputation, as presented in Table 2. The table shows that the mean and standard deviation of each variable before and after imputation are almost the same. Since the “missingness” does not depend on any variables in the dataset, the missing values are considered to be *missing completely at random* (MCAR). The MCAR characteristic of the missing values implies that the statistics obtained from incomplete data are unbiased. Since the statistics obtained from the imputed datasets are almost the same as those obtained from the

incomplete (original) dataset, the statistics obtained from the imputed datasets are also unbiased.⁹

<<Table 2 approx here>>

RESULTS

Coefficient Estimates

Table 3 presents the coefficient estimates of equation (1). Since this study is based on pooled cross-sectional observational data without controlling for endogeneity, the coefficient estimates do not necessarily suggest any causality relationship—they merely reflect a measure of association between quality of health and the explanatory variables. So it is possible that the association reflects reverse causality. For example, good health may have a positive effect on income. However, the higher the coefficient’s absolute value the stronger is the association between quality of health and the corresponding explanatory variable.

<<Table 3 approx here>>

As the SAH ranges from “poor” (= 1) to “excellent” (= 5), a positive coefficient of an explanatory variable indicates that a higher value of the variable is associated with a higher quality of health, while a negative coefficient indicates that a higher value of the variable is associated with a lower quality of health. From Table 3, we can see that health status declines steadily as age increases. The negative coefficient estimate for gender indicates that females are healthier than males on average. All racial/ethnic dummies have negative coefficient estimates, implying that after controlling for objective health measures, the self-reported health of minority populations are lower than that of the

⁹In this study, we use SAS® to perform the multiple-imputation procedure and also all other calculations.

White population. It may mean that there are omitted covariates in the regression (e.g. severity of diseases and risk factors, neighborhood effects, discrimination, etc.) that systematically affect the health of the minorities. Kobetz, Daniel, and Earp (2003) found that neighborhood poverty is associated with a greater likelihood of poor SAH.¹⁰

The negative coefficient of body mass index indicates that higher body mass index is associated with lower quality of health. With the dummy for elementary school or lower as the base, the coefficient estimates of all education levels are positive. These estimates tell us that higher education is associated with a better quality of health. The negative coefficient estimate of the dummy for living in New York City indicates that the conditional mean of quality of health of New York City population is lower than that of the rest of New York State population. It is noteworthy that the dummies for other cities such as Utica, Syracuse, Buffalo, Rochester, and Albany were not statistically significant and therefore were excluded from the equation. Respondents having a health insurance plan have better quality of health than respondents without a health plan, as expected. The coefficient estimate of annual household income is positive indicating that higher income is associated with better quality of health.

The coefficient estimate of smoking is negative which indicates smokers have lower quality of health than nonsmokers. Participating in physical activities or exercise has a positive association with the quality of health. Consuming more fruits and vegetables is associated with a better quality of health. This finding is consistent with the belief that dietary differences in fruits and vegetables contribute to differences in

¹⁰It may also be due to relatively different thresholds used by White while reporting SAH; see Banks, Marmot, Oldfield, and Smith (2006). However, this explanation is less likely in our case because we allow for heteroskedastic errors where the race/ethnicity variables are statistically significant. See footnote 4.

morbidity for chronic diseases (James & Nelson, 1997). A number of researchers have found that poor neighborhoods tend to have poor diets; certain aspects of disadvantaged neighborhoods act to hinder the procurement of healthy food; see Ecob and MacIntyre (2000) and Diez-Roux et al. (1999). Thus, the fruit & vegetable variable in our regression may be capturing certain omitted other neighborhood characteristics that affect health adversely.

All coefficient estimates of health variables (diseases and risk factors) are negative as expected, and almost all of them are statistically significant at the 5% level of significance. The relative magnitudes of the coefficient estimates are quite sensible. The diseases or risk factors generally considered serious such as diabetes, coronary heart disease, myocardial infarction, and stroke have relatively high coefficient estimates in absolute value. While the diseases or risk factors considered less serious have relatively low coefficient estimates in absolute value. These findings based on the New York State population are broadly consistent to the results obtained by Cutler and Richardson (1997, 1998) and Groot (2000) based on the U.S. population.

In many studies, it has been debated whether higher income inequality in a society is associated with poor average health. Van Ourti, van Doorslear, and Koolman (2006) show that when the relationship between income and health is concave, proportional income growth increases average health, and rising income inequality reduces average health. Wilkinson and Pickett (2006) compile results from 155 published peer review papers about the relationship between income inequality and population health. About seventy percent of the results suggest that health status is lower in societies where income is more unequal. The proponents of the association between income inequality and health

are, for example, Wilkinson (1992), Kennedy, Kawachi, Glass, and Prothrow-Smith (1998), Soobader and LeClere (1999), and Subramanian and Kawachi (2003, 2004, 2006).

Studies on the relationship between income inequality and health have been conducted using various levels of data, from census tract level to national level, and based on cross section and time series data. The measure of health outcome also varies including self-assessed health status, mortality rate, or life expectancy. For example, Subramanian and Kawachi (2003) test the association between income inequality and individual poor self-assessed health states in the United States.

Deaton and Lubotsky (2003) have, however, found that after controlling for the racial composition of population in a city, the effect of income inequality on health disappears. They argue that the higher is the percentage of minorities (e.g., Blacks) the higher the income inequality in the city.

In addition to the estimation results reported in Table 3, we also estimated equation (1) with three additional variables: county Gini coefficient (as a measure of income inequality), percent Black, and percent Hispanic. We found that the coefficients of all these variables were insignificant when the dummy for New York City area was included. Without the New York City dummy, however, the Gini coefficient was significant in this multilevel regression even when we controlled for percent Black and percent Hispanic, which is inconsistent to those found by Deaton and Lubotsky (2003). In our case, it can be explained by the fact that the patterns of income inequality and percent blacks across regions are quite different and, hence, are not collinear (see Figures 3 and 4). Since the New York City dummy is picking up the effect of the three variables and the

effect of income inequality is weak, we decided to use the specification in Table 3 in subsequent analysis.

Table 4 presents the coefficient estimates of the scale function. These coefficient estimates indicate that the error in equation (1) is heteroskedastic and is a function of gender, age, race/ethnicity, annual household income, having health plan, and education. However, d’Uva, van Doorslaer, Lindeboom, O’Donnell, and Chatterji (2006) found that reporting heterogeneity of health status does not have a large quantitative impact on the measures of health inequality.

<<Table 4 approx here>>

Estimates of Health Index

Quality of health by racial/ethnic groups. Table 5 presents the average estimated quality of health and health adjusted life expectancy (HALE). Comparing racial/ethnic groups, Asian followed by White has the highest average estimated quality of health, while AIAN followed by Hispanic and Black have the lowest. The average age varies considerably among racial/ethnic groups from 38.6 years through 47.6 years. In addition, the average estimated quality of health of a group depends on age distribution in the group. A group with a higher proportion of young individuals, ceteris paribus, will have a better quality of health relative to groups with a lower proportion of young individuals. Comparing the average quality of health between groups in a population with different age distributions could be misleading.

<<Table 5 approx here>>

Several methods can be used to control for the effects of age distribution. The simplest method is to compare the average estimated quality of health between groups of

population by age groups. Another method is to incorporate the quality of health into the life table of the group. In other words, we combine morbidity and mortality data to obtain the estimates of Health Adjusted Life Expectancy (HALE) (see Molla, Madans, Wagener, & Crimmins, 2003). The HALE measures the expected life (years) in perfect health condition. This measure is also called Healthy Life Expectancy (HLE). Since dependable life tables for different racial/ethnic groups are not available, in this study HALE is calculated based on the general U.S. population life table of 2002 (Arias, 2004). Thus, HALE estimated in this paper is used to compare the quality of health among groups of populations by eliminating the effect of age distribution without differentiating the mortality rates among the groups. HALE for each racial/ethnic group by age groups is presented in Table 5.

The table shows that Whites in the youngest age group (20–24) have the highest HALE followed by Asian, and Hispanic has the lowest followed by AIAN. A 20-year old White individual is expected to live for 44.2 years in perfect health condition, while a Hispanic with the same age is expected to live for 36.8 years in perfect health condition. Thus, at age 20, a White individual is expected to live almost 7.5 years in perfect health longer than a Hispanic individual. It is clear from these results that by eliminating the effect of age distribution, Whites do better than Asian, while Hispanics do worse than AIAN. This is a remarkable result. Also note that if HALE for each racial/ethnic group is calculated based on its own life table, the disparity across racial/ethnic groups could be higher as quality of health is correlated with life expectancy (Mullahy, 2001).

Regional disparity in health. In this part, we do not compute HALE for each area for two reasons. First, the distributions of age across the areas are very similar so the

effect of age distribution is negligible. Second, not all areas have enough observations required to compute HALE. The average quality of health by areas is presented in Figure 14. Nassau, Suffolk, Rockland, and Westchester Counties are the brightest areas reflecting the highest quality of health. In contrast, Bronx County is in the darkest area followed by Richmond County, North Country, Kings County, Queens County, and Southern Tier. None of the Upstate areas is in the brightest group, while Downstate areas vary from the brightest to the darkest, indicating that health disparity across Downstate areas is higher than that across Upstate areas. This finding is consistent with the socioeconomic variations across geographic areas presented in Figures 2, 3, and 4.

<<Fig. 14 approx here>>

It is very common that quality of health is measured using dichotomized SAH; cf. CDC. For example, quality of health of a group may be defined as a percentage of individuals in “very good” and “excellent” health condition (e.g., Keppel, Percy, & Klein, 2004); or it may be defined as the complement of the percentage of individuals in “poor” and “fair” health condition. Unfortunately, this means that the health rank of a group depends on the chosen cut-off point in dichotomizing the SAH. The procedure used in this paper circumvents this problem of arbitrariness by utilizing all five categories of SAH.

Inequality

Similar to the case of the quality of health, we also compare income-related health inequality among racial/ethnic groups as well as across geographic areas of New York State. In addition, this section also presents the decomposition results for each racial/ethnic group and for different geographic areas.

Income-related health inequality by race/ethnicity. Figure 15 presents the concentration index by racial/ethnic groups with corresponding 95%-confidence intervals. The standard errors of the concentration index are calculated using the Huber–White robust estimator. All indices are significantly greater than zero, indicating that health disparities between the rich and the poor exist in all groups. There is, however, substantial variation in the coefficients among groups. The highest inequality index is found within the AIAN group followed by Hispanics. The lowest inequality is found within the Asian group followed by Whites.

<<Fig. 15 approx here>>

Another way to compare income-related health inequality between groups of populations is by comparing their concentration curves. Figure 16 presents the concentration curves expressed as the deviation of the concentration curve from the diagonal in order to amplify the differences between racial/ethnic groups. The figure provides more obvious evidence of the differences of income-related health inequalities between racial/ethnic groups. The Asian curve strictly dominates others, while the AIAN curve is dominated by others. These indicate that AIAN is the most unequal at all percentiles, while Asian is the least. Therefore, differences between racial/ethnic groups are not only in terms of the average quality of health but also in terms of health distribution itself over income among individuals within each group.

<<Fig. 16 approx here>>

Regional variation in income related health inequality. Concentration indices across the geographic areas are presented in Figure 17. North Country and Mohawk are in the darkest areas indicating the highest income related health inequality, while the

brightest areas are represented by—from the lowest to the highest—Nassau, Westchester, Queens, and Suffolk County. Figures 14 and 17 show that areas with lower quality of health tend to have higher income related health inequality. These values have been plotted in Figure 18 for the 8 Upstate regions and 9 Downstate counties. The simple correlation coefficient between them is -0.67 and is statistically significant. The worst two regions in terms of both health inequality and health are North Country (4) and Bronx County (9). It is interesting that both regions 4 and 9 report very high levels of medical risk factors like diabetes, obesity, and asthma, but the racial compositions of the two areas are diametrically opposite. On the other hand, two best counties are Nassau (11) and Westchester (17) where a very high average level of health is achieved with very low health inequality. Interestingly, these two counties have a rather high percentage of African Americans in the population.

<<Fig. 17 approx here>>

<<Fig. 18 approx here>>

For more detailed information about the magnitude and significance of the concentration index across the areas, the corresponding 95%-confidence intervals by areas are presented in Figure 19. All coefficients are significantly different from zero, indicating that income-related health inequalities exist in all areas. The statistical significance of the difference in the concentration indices between geographic areas can be seen by comparing the confidence intervals. For instance, New York County has a significantly higher concentration index than those of Nassau, Westchester, Queens, and Suffolk Counties. But the concentration index of New York County is not significantly different from those of Kings and Bronx Counties.

<<Fig. 19 approx here>>

An important public policy question is: what are the main factors that are contributing to the inequality? This can be answered by decomposing the inequality into its determinants, as presented in the next section.

Decomposition of income related health inequality. Decomposition analysis demonstrates the relative contribution of the different components of inequalities for different racial/ethnic groups as well as for different geographic areas. The analysis offers policy makers a starting point for geographical targeting to eliminate health disparity between the rich and the poor. The analysis does not provide unequivocal causal pathways between the determinants of health and income related health inequality. However, this method describes health inequality across income levels and provides explanations for the observed patterns for different groups of population.

In this paper, we are interested in analyzing income-related health inequality attributable to socio-demographic factors including age, sex, race/ethnicity, marital status, education, employment status, health insurance, smoking status, and access to a doctor. Therefore, in the decomposition analysis, the explanatory variables included in equation (1) are only socio-demographic variables. The percent contribution of each component is calculated by racial/ethnic groups and also by geographic areas. Among all these factors, the contributions of the main variables by racial/ethnic groups are presented in Table 6.

<<Table 6 approx here>>

For the overall New York State population, among the socio-demographic variables, three major factors contributing to income-related health inequality are

household income, employment status, and education. Each of these three factors contributes at least 18% to the inequality. Our estimate of the contribution of income is similar to that in van Doorslaer and Jones (2003) who found that in Canada income contributes between 41% to 47% of income-related health inequality. However, we find that income is relatively less important for the disadvantaged minority groups (*viz.*, Black, Hispanic, and AIAN). For Asian the corresponding percentage is very high (63%). If health status were distributed equally across income levels, employment status, and education levels, then the disparity between the rich and the poor in New York State population would be 79% lower.

After controlling for other factors, race/ethnicity contributes only 6% to the inequality. Race/ethnicity is highly intertwined with household income, employment status, and education—Blacks, Hispanics, and AIANs have lower household income, lower education levels, and lower employment rates compared to those of Whites. That is another reason why separate analysis for each group is necessary.

The pattern of the contributions of the socio-demographic variables varies considerably between racial/ethnic groups. Income has the largest contribution to the inequalities for all groups (except AIAN), indicating that income redistribution policy is the most effective policy to eliminate health disparity between the poor and the rich within these ethnic/racial groups. For AIANs, the most effective public policy is to ensure employment opportunities to all individuals. For Whites, the second largest contributor to income-related health inequality is employment status (18%) followed by education (16%); for Blacks the second largest is employment status (28%) followed by education (20%); for Hispanics the second largest is employment status (29%) followed

by education (24%); for Asians the second largest is education (17%) followed by inability to see a doctor (12%). The differences in the patterns of the contributions indicate that different groups of population need different policies to eliminate health disparity between the rich and the poor.

Comparing across the geographic areas, the contribution of each factor to the inequality varies noticeably, too (see Table 7). Income is the largest contributor to the inequality for all areas, ranging from 30% (Bronx County) to 48% (Rockland County). In addition to income, employment status and education also contribute substantially to the inequality in all areas. The contribution of employment status varies considerably ranging from 13% (Westchester County) to 29% (Bronx County), while the contribution of education ranges from 12% (Richmond County) to 23% (Bronx County). Thus, in addition to income redistribution policy, ensuring employment opportunities and good educational access are also effective ways to eliminate health disparity between the rich and the poor in these areas. For Queens and New York Counties, race/ethnicity contributes to the inequality relatively high—10% and 10% respectively—compared to other areas. This indicates that providing good access to health care for minorities in these two counties is also an effective way to eliminate the disparity.

<<Table 7 approx here>>

CONCLUSIONS

Following recent developments in measuring quality of health and health inequality, we use self-assessed health status conditioned by several objective determinants as a comprehensive measure of individual health. Among racial/ethnic groups, AIANs followed by Hispanics have the lowest average quality of health, while

after adjusting for age distribution, Hispanics have the lowest average quality of health. Asians have the highest average followed by Whites, while after adjusting for age distribution, Whites have the best quality of health. These results highlight that when comparing quality of health between groups of populations, one needs to consider the age distribution within each group.

In Upstate New York, North Country and Southern Tier have the lowest average quality of health, whereas in Downstate Bronx, Richmond, Kings, and Queens Counties have the lowest. Comparing all regions of New York State, Nassau, Suffolk, Rockland, and Westchester Counties have very high levels of average health. However, quality of health is more unequal across Downstate areas than Upstate, as one would possibly expect.

We find statistically significant income-related health inequality within each racial/ethnic group and each geographic area of New York State. The highest inequality is found within the AIAN group followed by Hispanics, while the lowest inequality is found within the Asian group, followed by Whites. Across the 17 geographic areas, the highest income-related health inequality is found in North Country followed by Mohawk, while the lowest inequality is found in Nassau County followed by Westchester County. Areas with lower average quality of health have larger health disparity between the rich and the poor; the correlation is -0.67 and is statistically significant.

Decomposition analysis of income related health inequality presented in this paper offers New York health policy makers certain guidelines to eliminate health disparity in the population. Three major factors generating the disparity are household income, employment status, and education. Contribution of each of these factors varies

considerably between racial/ethnic groups as well as across the geographic areas suggesting different pathways from income to health. Income has the largest contribution to the disparity for all groups but AIANs and for all areas. This indicates that income redistribution policy is the most effective policy to eliminate the health disparity within each group (except AIAN), as well as within each area. The relative health status of AIAN in almost all recorded dimensions is disturbingly bad, and the most effective policy for this group was found to be the generation of employment opportunities. For Blacks and Hispanics, public policy that ensures employment opportunities is also an effective way to eliminate disparity.

Our results underscore the need for different public policy initiatives for different racial/ethnic groups and different geographic areas to eliminate health disparity between the rich and the poor. In the long run, policies that can ensure equality in income (e.g., income redistribution), employment opportunities, and educational access will have a substantial impact on improving the average quality of health and in eliminating health disparity.

REFERENCES

- Adams, P., Hurd, M., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal between health and socioeconomic status. *Journal of Econometrics*, *112*, 3–56.
- Adler, N. E., & Newman, K. (2002). Socioeconomic disparities in health: Pathways and policies. *Health Affairs*, *21*, 60–76.
- Adler, N. E., & Ostrove, J. M. (1999). SES and health: What we know and what we don't. *Annals of the New York Academy of Sciences*, *896*, 3–5.

- Arias, E. (2004). *United States life tables, 2002*. National Vital Statistics Reports 53. Hyattsville, MD: National Center for Health Statistics.
- Ayanian, J. Z., Weissman, J. S., Chasan-Taber, S., & Epstein, A. M. (1999). Quality of care by race and gender for congestive heart failure and pneumonia. *Medical Care*, 37, 1260–1269.
- Banks, J. B., Marmot, M., Oldfield, Z., & Smith, J. P. (2006). The SES health gradient on both sides of the Atlantic. National Bureau of Economic Research working paper 12674. Cambridge, MA.
- Bommier, A., & Stecklov, G. (2002). Defining health inequality: Why Rawls succeeds where social welfare theory fails. *Journal of Health Economics*, 21, 497–513.
- Centers for Disease Control and Prevention (CDC) (1999-2004). Behavioral Risk Factor Surveillance System Survey Questionnaire. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.
- Centers for Disease Control and Prevention (CDC) (1999-2004). *Behavioral Risk Factor Surveillance System Survey Data*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.
- Clarke, P. M., & Ryan, C. (2006). Self-reported health: Reliability and consequences for health inequality measurement. *Health Economics*, 15, 645–352.
- Cutler, D. M., Deaton, A., & Lleras-Muney, A. (2006). The determinants of mortality. *Journal of Economic Perspectives*, 20, 97–120.
- Cutler, D. M., & Lleras-Muney, A. (2006). Education and health: Evaluating theories and evidence. National Bureau of Economic Research working paper 12352. Cambridge, MA.

- Cutler, D. M., & Richardson, E. (1997). Measuring the health of the U.S. population. *Brooking papers on economic activity. Microeconomics 1997*, 217–271.
- Cutler, D. M., & Richardson, E. (1998). The value of health: 1970–1990. *American Economic Review*, 88, 97–100.
- Deaton, A. (2006). Global pattern of income and health: Facts, interpretation, and policies. *National Bureau of Economic Research Working Paper 12735*. Cambridge, MA.
- Deaton, A., & Lubotsky, D. (2003). Mortality, inequality and race in American cities and states. *Social Science & Medicine*, 56, 1139–1153.
- Diez-Roux, A. V., Nieto, F. J., Caulfield, L., Tyroler, H. A., Watson, R. L., & Szklo, M. (1999). Neighborhood differences in diet: The Atherosclerosis Risk in Communities (ARIC) study. *Journal of Epidemiology and Community Health*, 53, 55–63.
- d’Uva, T. B., van Doorslaer, E., Lindeboom, M., O’Donnell, O., & Chatterji, S. (2006). Does reporting heterogeneity bias the measure of health disparity? Timbergen Institute discussion paper, TI 2006-033/3.
- Ecob, R., & MacIntyre, S. (2000). Small area variations in health related behaviors: Do these depend on the behavior itself, its measurement, or on personal characteristics? *Health and Place*, 6, 261–274.
- Environmental Systems Research Institute (ESRI). (2002). *ArcView GIS*. Version 3.3. Redlands, CA: Author.
- Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, 19, 403–420.

- Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A potential for bias when rounding in multiple imputation. *American Statistician*, *57*, 229–232.
- Humphries, K. H., & van Doorslaer, E. (2000). Income-related health inequality in Canada. *Social Science Medicine*, *50*, 663–671.
- Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*, 21–37.
- Institute of Medicine. (2002). *Unequal treatment: Controlling racial and ethnic disparities in health care*. Washington, DC: National Academies Press.
- James, W., & Nelson, M. (1997). The contribution of nutrition to inequalities in health. *British Medical Journal*, *314*, 1545–1550.
- Kakwani, N., Wagstaff, A., & van Doorslaer, E. (1997). Socioeconomic inequalities in health: Measurement, computation, and statistical inference. *Journal of Econometrics*, *77*, 59–103.
- Kennedy, B. P., Kawachi, I., Glass, R., & Prothrow-Stith, D. (1998). Income distribution socioeconomic status, self-rated health in the United States: Multilevel analysis. *BMJ*, *317*, 917–921.
- Keppel, K. G., Percy, J. N., & Klein, R. J. (2004). Measuring progress in healthy people 2010. Statistical notes 25. Hyattsville, MD: National Center for Health Statistics.
- Kobetz, E., Daniel, M., & Earp, J. A. (2003). Neighborhood poverty and self-reported health among low-income, rural women, 50 years and older. *Health & Place*, *9*, 263–271.
- Koolman, X., & van Doorslaer, E. (2004). On the interpretation of a concentration index of inequality. *Health Economics*, *13*, 649–56.

- Lahiri, K., Vaughan, D. R., & Wixon, B. (1995). Modeling SSA's sequential disability determination process using matched SIPP data. *Social Security Bulletin*, 58, 1–41.
- Lecluyse, A., & Cleemput, I. (2005). Making health continuous: Implications of different methods on the measurement of inequality. *Health Economics*, 15, 99–104.
- Manor, O., Matthews, S., & Power, C. (2001). Self-rated health and limiting longstanding illness: Inter-relationships with morbidity in early adulthood. *International Journal of Epidemiology*, 30, 600–607.
- McKelvey, R., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4, 103–120.
- Milanovic, B. (1997). A simple way to calculate the Gini coefficient, and some implications. *Economics Letters*, 56, 45–49.
- Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2004). Actual causes of death in the United States, 2000. *Journal of American Medical Association*, 291, 1238–1245.
- Molla, M. T., Madans, J. H., Wagener, W. K., & Crimmins, E. M. (2003). *Summary measures of population health: Report of findings on methodologic and data issues*. Hyattsville, MD: National Center for Health Statistics.
- Mullahy, J. (2001). Live long, live well: Quantifying the health of heterogeneous population. *Health Economics*, 10, 429–440.
- Must, A., Spadano, J., Coakley, E. H., Field, A.E., Colditz, G., & Dietz, W. H. (1999). The disease burden associated with overweight and obesity. *Journal of American Medical Association*, 282, 1523–1529.

- National Center for Health Statistics (NCHS). (2006). *Health, United States, 2006: With chartbook on trends in the health of Americans*. Hyattsville, MD: Author.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley.
- Safaei, J. (2006). Income and health inequality across Canadian provinces. *Health & Place, 9*, 3.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman and Hall.
- Sehili, S., Elbasha, E. H., Moriarty, D. G., & Zack, M. M. (2005). Inequalities in self-reported physical health in the United States, 1993–1999. *Health Economics, 14*, 377–389.
- Shishehbor, M. H., Litaker, D., Pothier, C. E., & Lauer, M. S. (2006). Association of socioeconomic status with functional capacity, heart rate recovery, and all-cause mortality. *Journal of American Medical Association, 295*, 784–792.
- Sickles, R. C., & Taubman, P. (1997). Mortality and morbidity among adults and elderly. In M. R. Rozenzweig & O. Stark (Eds.), *Handbook of population and family economics*. Amsterdam: North-Holland.
- Soobader, M.-J., & LeClere, F. B. (1999). Aggregation and the measurement of income inequality: Effects on morbidity. *Social Science & Medicine, 48*, 733–744.
- Subramanian, S. V., & Kawachi, I. (2003). The association between state income inequality and worse health is not confounded by race. *International Journal of Epidemiology, 32*, 1022–1028.

- Subramanian, S. V., & Kawachi, I. (2004). Income inequality and health: What have we learned so far? *Epidemiologic Reviews*, 26, 78–91.
- Subramanian, S. V., & Kawachi, I. (2006). Whose health is affected by income inequality? A multilevel interaction analysis of contemporaneous and lagged effects of state income inequality on individual self-rated health in the United States. *Health & Place*, 12, 141–156.
- U.S. Census Bureau. (2000). *American FactFinder fact sheet*. Retrieved January 10, 2007, from <http://factfinder.census.gov/>.
- U.S. Department of Health and Human Services (US-DHHS). (2000). *Healthy People 2010: Understanding and improving health* (2nd ed.). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Health and Human Services (US-DHHS). (2004). *The Health Consequences of Smoking: A Report of Surgeon General*. US Government Printing Office, Washington, DC.
- Van Doorslaer, E., & Jones, A. M. (2003). Inequalities in self-reported health: Validation of a new approach to measurement. *Journal of Health Economics*, 22, 61–87.
- Van Ourti, T., van Doorslaer, E., & Koolman, X. (2006). *The effect of growth and inequality in incomes on health inequality: Theory and empirical evidence from the European Panel*. Tinbergen Institute discussion paper, TI 2006-108/3. Rotterdam: Erasmus University.
- Wagstaff, A., Paci, P., & van Doorslaer, E. (1991). On the measurement of inequalities in health. *Social Science & Medicine*, 33, 545–557.

- Wagstaff, A., & van Doorslaer, E. (2004). Overall versus socioeconomic health inequality: A measurement framework and two empirical illustrations. *Health Economics, 13*, 297–301.
- Wagstaff, A., van Doorslaer, E., & Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics, 112*, 207–223.
- Wilkinson, R. G. (1992). Income distribution and life expectancy. *British Medical Journal, 304*, 165–168.
- Wilkinson, R. G., & Pickett, K. E. (2006). Income inequality and population health: A review and explanation of the evidence. *Social Science & Medicine, 62*, 1768–1784.
- Williams, D. R., & Collins, C. (1995). US socioeconomic and racial differences in health: Patterns and explanations. *Annual Review of Sociology, 21*, 349–386.

Table 1. Missing Data Pattern in New York State BRFSS Sample

Variable	Year					
	1999	2000	2001	2002	2003	2004
Could not afford to see a doctor	√	√	.	.	√	√
Participate in any physical activities or exercises	.	√	√	√	√	√
Fruit and vegetable servings per day	.	√	.	√	√	.
Heavy drinking	√	.	√	√	√	√
Activities limited due to health problem	.	√	√	.	√	√
Ever had asthma	.	√	√	√	√	√
Ever told blood pressure high	√	.	√	.	√	.
Ever told had coronary heart disease	√	.	√	.	√	.
Ever told had myocardial infarction	√	.	√	.	√	.
Ever told had stroke	√	.	√	.	√	.
Ever told had arthritis	.	√	√	√	√	√
Ever told blood cholesterol high	√	.	√	.	√	.
Had pain, aching, stiffness, and swelling	.	√	√	.	.	.
Participate in phys. activities or exercises	.	√	√	√	√	√
Fruit and vegetable servings per day	.	√	.	√	√	.

Note: √ means the information was collected.

Table 2. Mean and Standard Deviation based on Original and Imputed Datasets

Variable	Original dataset		Imputed dataset		Ratio	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Number of days physical health not good	3.562	8.002	3.578	8.011	1.004	1.001
Number of days mental health not good	3.317	7.501	3.322	7.501	1.001	1.000
Ever told had diabetes	0.070	0.256	0.070	0.256	1.000	1.000
Annual household Income (\$1,000)	50.243	37.202	49.361	37.298	0.982	1.003
Could not afford to see doctor	0.113	0.316	0.113	0.316	1.000	1.000
Heavy drinking	0.133	0.340	0.134	0.340	1.005	1.000
Activities limited due to health problem	0.189	0.391	0.185	0.390	0.979	0.996
Ever had asthma	0.119	0.324	0.119	0.324	0.998	1.000
Ever told blood pressure high	0.276	0.447	0.280	0.447	1.013	1.001
Ever told had coronary heart disease	0.049	0.217	0.048	0.216	0.968	0.997
Ever told had myocardial infarction	0.043	0.204	0.042	0.203	0.959	0.997
Ever told had stroke	0.026	0.158	0.024	0.158	0.931	0.999
Ever told had arthritis	0.280	0.449	0.281	0.449	1.004	1.001
Ever told blood cholesterol high	0.324	0.468	0.304	0.469	0.937	1.001
Had pain, aching, stiffness or swelling	0.420	0.494	0.443	0.498	1.056	1.008
Participate in phys. activities or exercises	0.754	0.431	0.754	0.430	1.000	1.000
Fruit and vegetable servings per day	3.855	2.192	3.862	2.191	1.002	1.000

Table 3. Coefficient Estimate of the Ordered Probit Model

Variable	Coefficient Standard P-value		
	Estimate	Error	
Intercept	4.0997	0.1629	0.0000
Age 25–29	0.1848	0.0567	0.0011
Age 30–34	0.1662	0.0532	0.0018
Age 35–39	0.1028	0.0545	0.0593
Age 40–44	0.0531	0.0529	0.3148
Age 45–49	0.0814	0.0552	0.1404
Age 50–54	0.0088	0.0576	0.8784
Age 55–59	0.0532	0.0608	0.3811
Age 60–64	0.0130	0.0680	0.8478
Age 65–69	–0.1666	0.0739	0.0243
Age 70–74	–0.1162	0.0755	0.1236
Age 75–79	–0.3492	0.0842	0.0000
Age 80–84	–0.3112	0.0939	0.0009
Age >=85	–0.5571	0.1275	0.0000
Sex (male=1)	–0.0373	0.0239	0.1182
Black	–0.1423	0.0418	0.0007
Hispanic	–0.4037	0.0447	0.0000
Asian	–0.4191	0.0693	0.0000
AIAN	–0.1585	0.1451	0.2748
Other	–0.2822	0.0857	0.0010
Marital status	–0.0501	0.0240	0.0370
Body mass index/27	–0.6948	0.0614	0.0000
Grades 9–11 (Some high school)	0.3806	0.0843	0.0000
Grade 12 or GED (High school graduate)	0.5138	0.0762	0.0000
College 1 year to 3 years (Some college or technical school)	0.6270	0.0778	0.0000
College 4 years or more (College graduate)	0.7991	0.0793	0.0000
Self-employed	0.2432	0.0396	0.0000
Out of work	0.0186	0.0508	0.7136
A homemaker	–0.0158	0.0463	0.7335
A student	0.1424	0.0658	0.0306
Retired	–0.0979	0.0463	0.0343
Unable to work	–0.4573	0.0652	0.0000
Having health plan	0.1101	0.0399	0.0058
Annual Household Income (\$1,000)	0.0048	0.0004	0.0000
Smoking	–0.2418	0.0275	0.0000
Participating in any physical activities or exercises	0.3076	0.0289	0.0000
Fruit and vegetable servings per day	0.0424	0.0072	0.0000
Number of days physical health not good	–0.0595	0.0025	0.0000
Number of days mental health not good	–0.0157	0.0016	0.0000
Ever told had diabetes	–0.7772	0.0543	0.0000
Could not afford to see doctor	–0.3063	0.0469	0.0000

Heavy drinking	0.0439	0.0338	0.1957
Activities limited due to health problem	-0.6144	0.0392	0.0000
Ever had asthma	-0.2065	0.0348	0.0000
Ever told blood pressure high	-0.3967	0.0299	0.0000
Ever told had coronary heart disease	-0.4685	0.0721	0.0000
Ever told had myocardial infarction	-0.4392	0.0894	0.0001
Ever told had stroke	-0.3093	0.0838	0.0004
Ever told had arthritis	-0.1240	0.0319	0.0002
Ever told blood cholesterol high	-0.2090	0.0271	0.0000
Had pain, aching, stiffness or swelling in or around a joint	-0.2228	0.0386	0.0001
Dummy for NY City	-0.1934	0.0266	0.0000
Threshold 2	1.7705	0.0618	0.0000
Threshold 3	3.6140	0.1100	0.0000
Threshold 4	5.2001	0.1531	0.0000

McKelvey-Zavoina $R^2 = 0.60$

Note: Reference for Age group dummies is 18–24; for Education it is grade 8 or less; and for Employment status it is employed for wage.

Table 4. Coefficient Estimate of the Heteroskedasticity Scale Function

	Coefficient Estimate	Standard Error	P-value
Sex (male=1)	0.2084	0.0544	0.0002
Age 25–29	0.1109	0.1276	0.3851
Age 30–34	-0.0662	0.1386	0.6336
Age 35–39	-0.0365	0.1342	0.7861
Age 40–44	-0.0735	0.1342	0.5843
Age 45–49	0.0519	0.1251	0.6784
Age 50–54	0.1943	0.1236	0.1161
Age 55–59	0.2911	0.1230	0.0180
Age 60–64	0.4317	0.1208	0.0004
Age 65–69	0.2825	0.1471	0.0573
Age 70–74	0.2632	0.1429	0.0660
Age 75–79	0.4133	0.1461	0.0048
Age 80–84	0.2888	0.1809	0.1116
Age >=85	0.8155	0.1737	0.0000
Black	0.3716	0.0818	0.0000
Hispanic	0.4014	0.0810	0.0000
Asian	0.3521	0.1505	0.0198
AIAN	0.6897	0.2313	0.0029
Annual household income (\$1,000)	-0.0024	0.0009	0.0067
Having health plan	-0.2058	0.0794	0.0096
Education higher than high school	-0.1035	0.0574	0.0717
Sex (male=1)	0.2084	0.0544	0.0002

Table 5. Average Estimated Quality of Health and Health-Adjusted Life Expectancy (HALE)

		All	White	Black	Hispanic	Asian	AIAN
Average quality of health		0.750	0.765	0.715	0.678	0.778	0.665
Age	Life expectancy	HALE (in year)					
20–24	58.23	43.05	44.24	40.23	36.81	43.75	37.52
25–29	53.50	39.35	40.44	36.62	33.40	40.05	33.74
30–34	48.74	35.52	36.48	32.90	29.87	36.19	30.33
35–39	44.00	31.68	32.51	29.22	26.35	32.33	26.73
40–44	39.33	27.97	28.69	25.66	23.02	28.65	23.28
45–49	34.78	24.41	25.02	22.30	19.90	25.11	20.23
50–54	30.36	20.99	21.48	19.15	16.99	21.68	17.82
55–59	26.09	17.76	18.16	16.17	14.28	18.48	15.27
60–64	22.01	14.74	15.05	13.47	11.85	15.30	12.77
65–69	18.19	11.96	12.16	10.95	9.68	12.58	10.17
70–74	14.69	9.56	9.68	8.77	7.87	10.61	8.11
75–79	11.54	7.39	7.48	6.75	6.07	8.38	6.13
80–84	8.79	5.63	5.70	5.22	4.69	6.34	4.86

Table 6. Decomposition of Health Inequality by Racial/Ethnic Groups

Race/Ethnicity	Percentage contribution of each factor to health inequality						
	Age	Race/Ethnicity	Education	Employment	Income	Smoking	Could not afford to see doctor
All	7.00	5.71	18.63	18.99	41.19	2.48	7.34
White	13.07	-	16.86	18.17	45.40	2.99	6.31
Black	4.71	-	20.73	28.26	35.79	4.64	7.66
Hispanic	6.69	-	24.16	28.47	32.12	0.31	9.16
Asian	-1.96	-	17.45	8.71	63.19	1.35	12.31
AIAN	8.44	-	16.24	42.96	30.67	2.83	1.50

Note: Small contributors are not presented in this table.

Table 7. Decomposition of Health Inequality by Geographic Areas

Geographic Area	Percentage contribution of each factor to health inequality (%)						
	Age	Race/Ethnicity	Education	Employment	Income	Smoking	Could not afford to see doctor
H Valley	11.723	2.898	14.400	20.680	43.801	1.540	7.715
C Region	11.425	2.729	17.185	18.509	41.877	3.617	6.735
Mohawk	11.509	1.980	17.181	22.284	35.310	3.728	10.032
N Country	2.343	-0.423	15.372	26.392	41.477	4.242	12.748
C NY	12.493	1.875	13.862	21.981	41.164	3.272	7.961
S Tier	8.061	2.046	17.799	26.207	34.958	4.643	8.578
W NY	12.135	3.085	15.937	23.205	39.463	2.891	6.031
F Lakes	7.440	4.098	18.908	15.137	44.668	4.314	7.844
Bronx	4.601	6.943	22.959	28.562	30.380	1.221	7.394
Kings	5.322	8.382	21.643	21.876	35.676	1.549	7.726
Nassau	12.651	3.603	16.421	14.401	47.478	1.680	5.509
New York	4.861	9.910	19.671	17.994	40.805	1.522	7.281
Queens	2.677	10.454	19.950	13.675	44.474	0.793	10.956
Richmond	10.606	3.828	11.995	23.479	41.043	1.586	8.990
Rockland	1.940	4.599	14.693	19.430	48.030	0.057	12.672
Suffolk	13.244	4.008	14.982	15.427	45.839	2.931	5.636
Westchester	7.353	6.875	19.880	13.252	46.472	1.854	6.365

Note: Small contributors are not presented in this table.

Figure 1. Concentration Curve

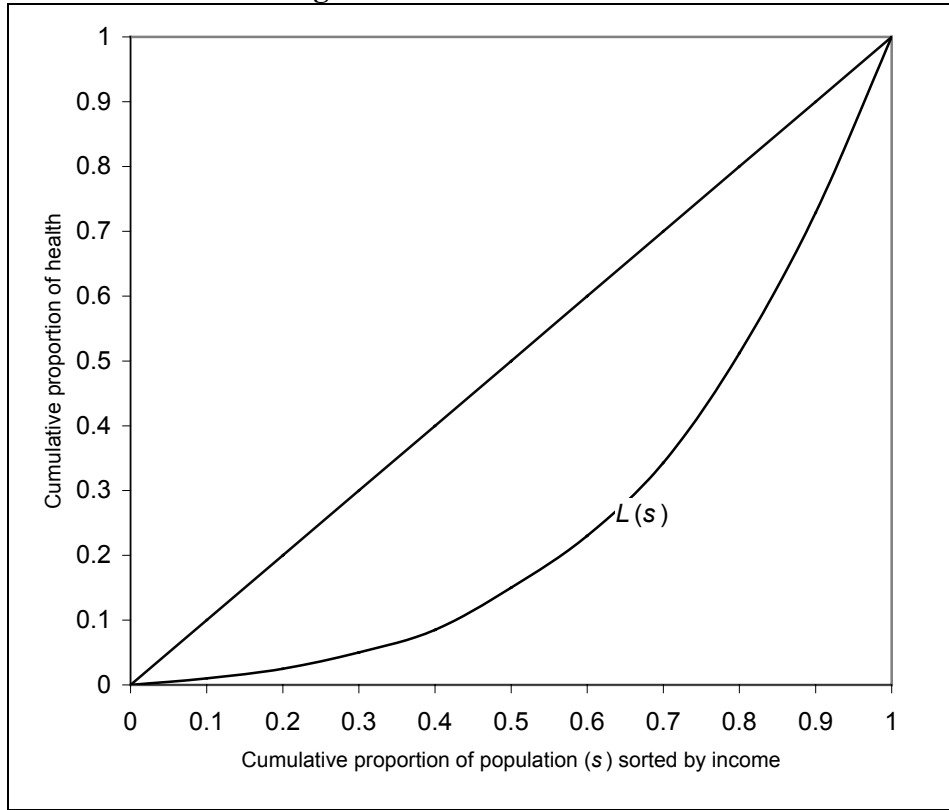


Figure 2. Percent Population in Poverty

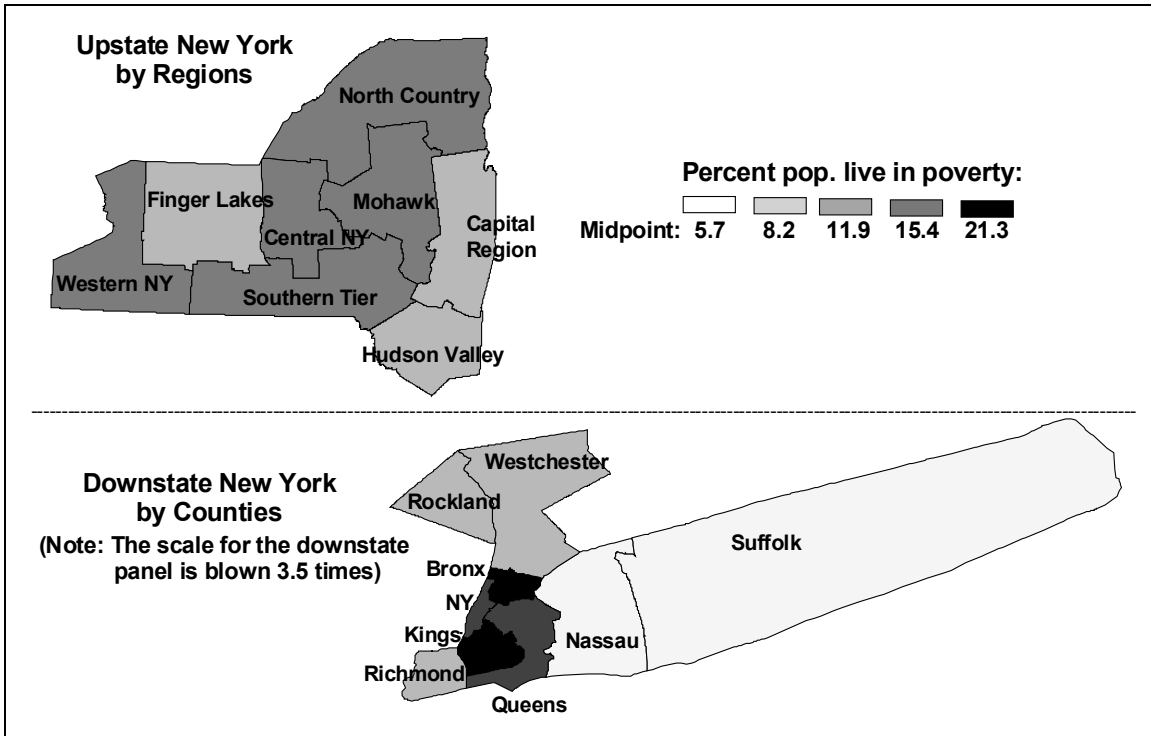


Figure 3. Income Inequality

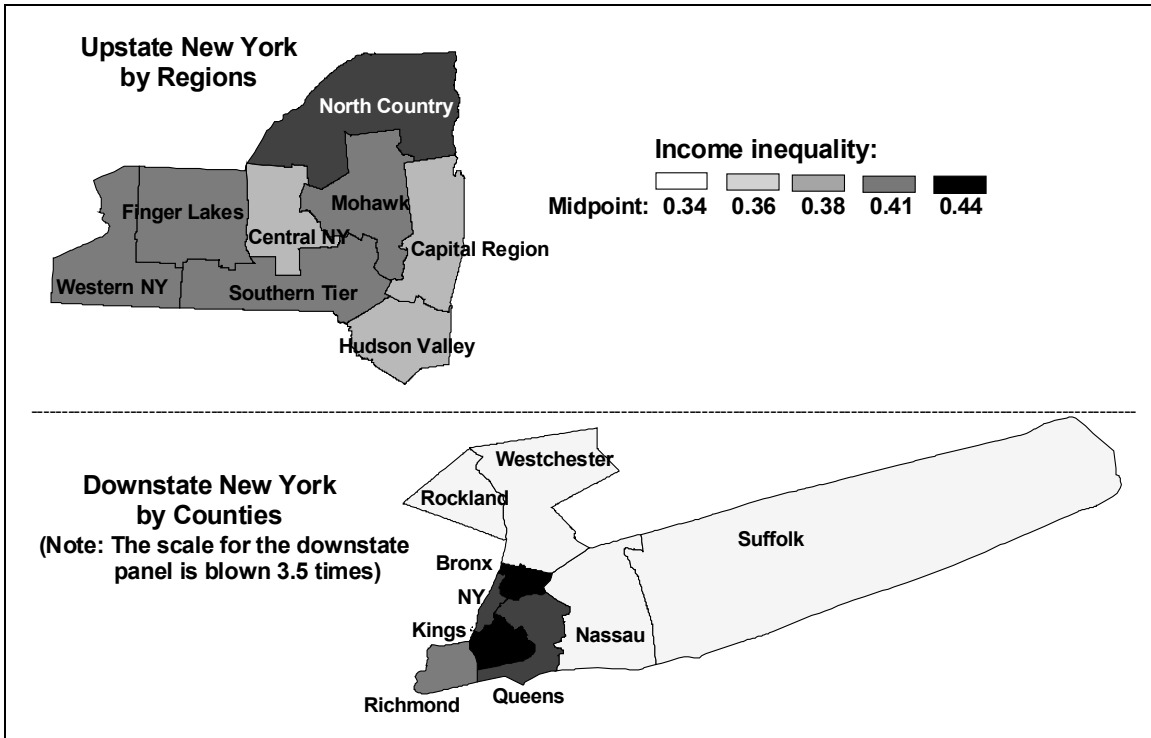


Figure 4. Percent Black Population

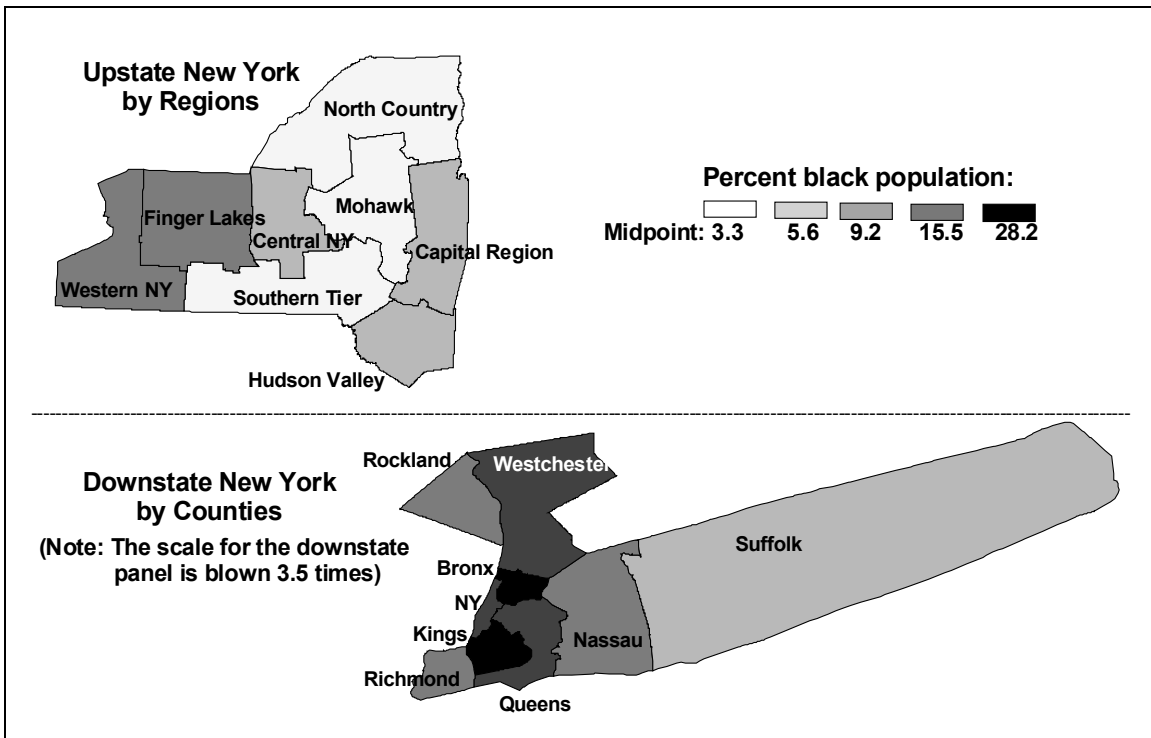


Figure 5. Prevalence of Smoking

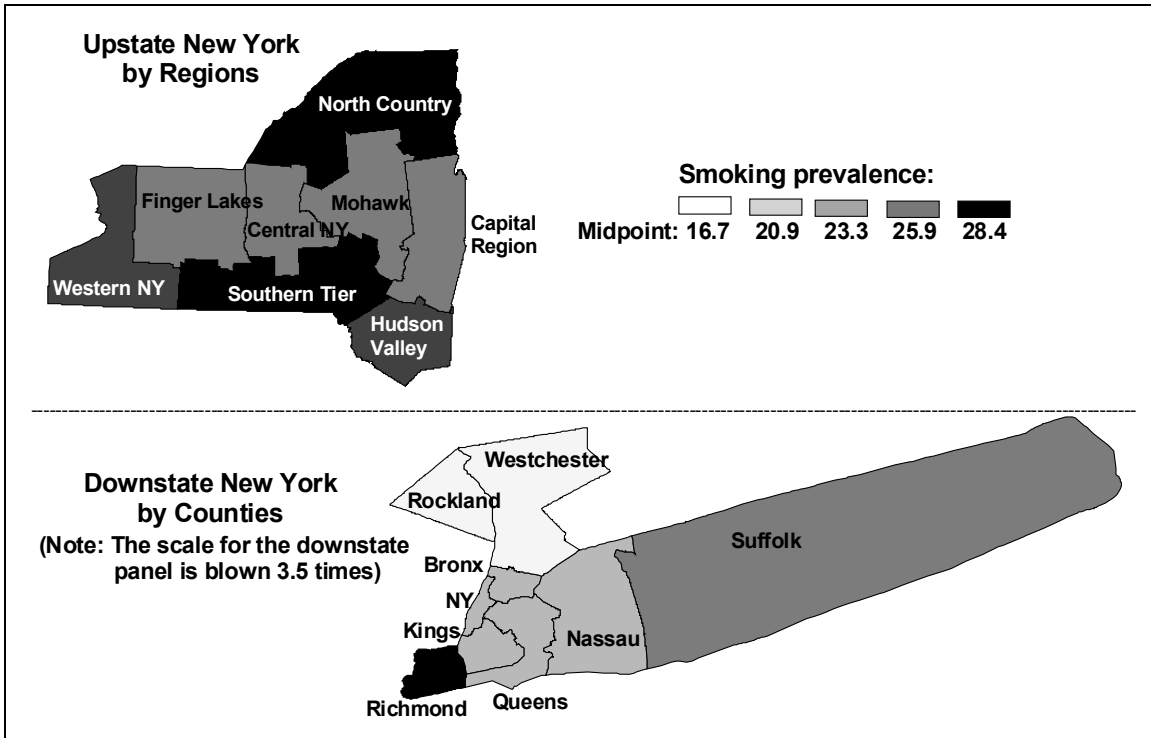


Figure 6. Distribution of Self-Assessed Health by Racial/Ethnic Groups

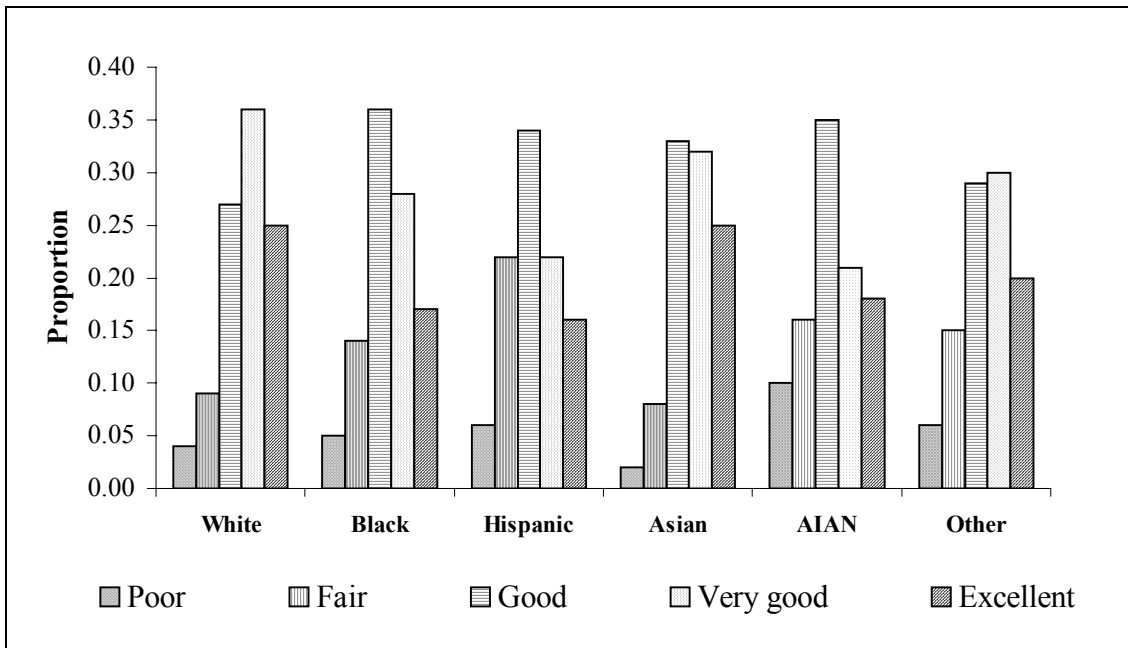


Figure 7. Distribution of Self-Assessed Health by Annual Household Income

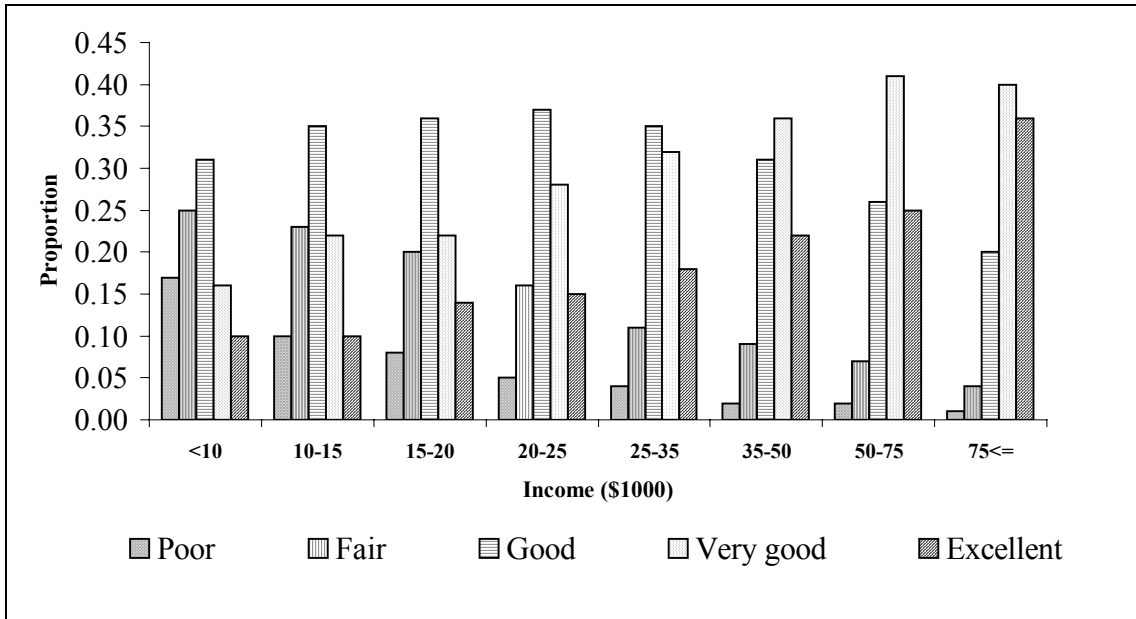


Figure 8. Number of Days Physical Health not Good

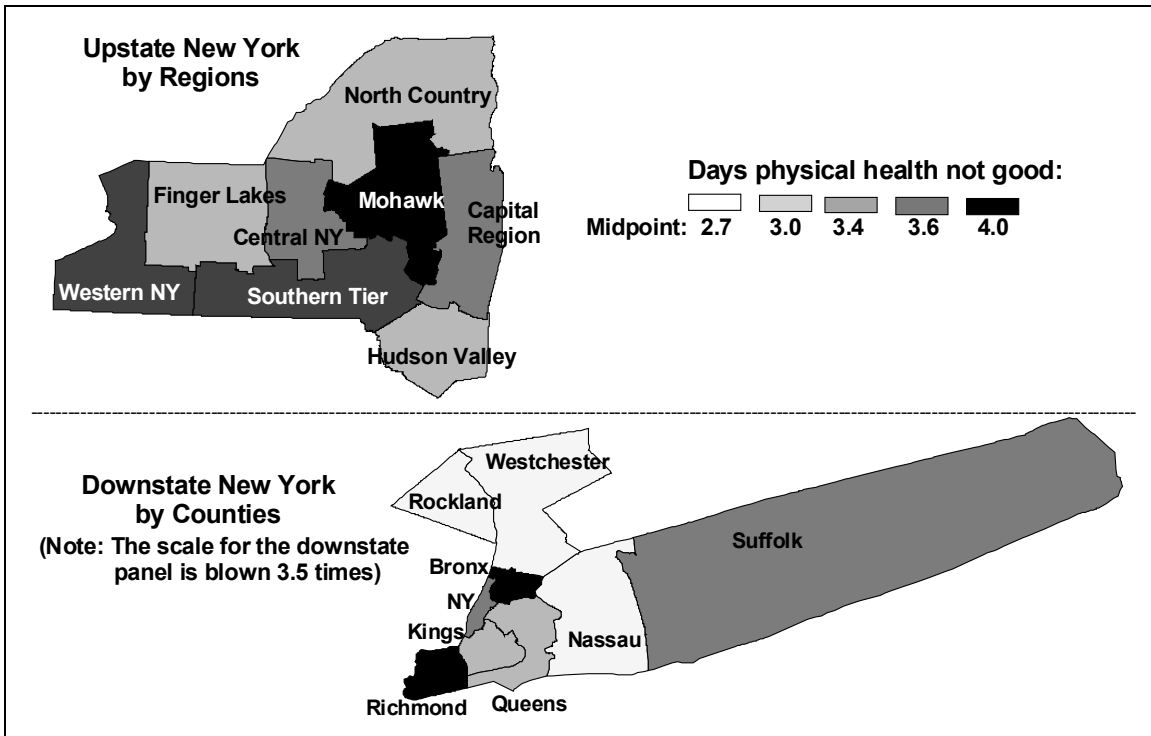


Figure 9. Number of Days Mental Health not Good

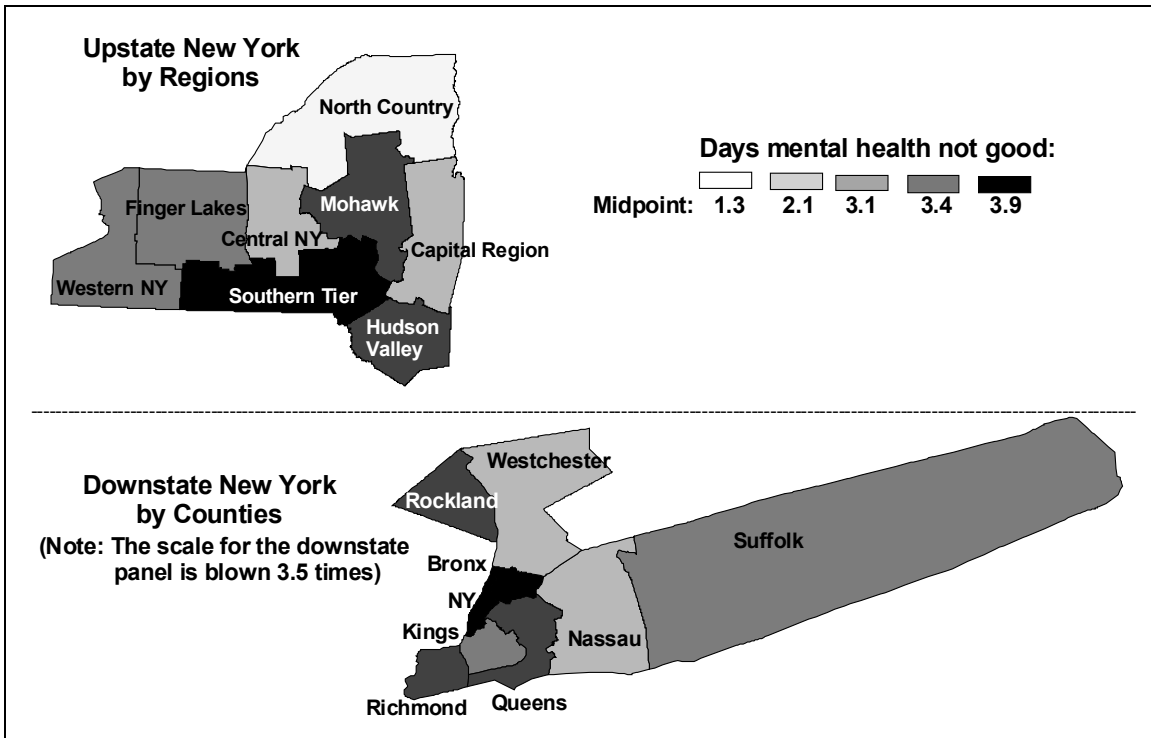


Figure 10. Prevalence of Obesity

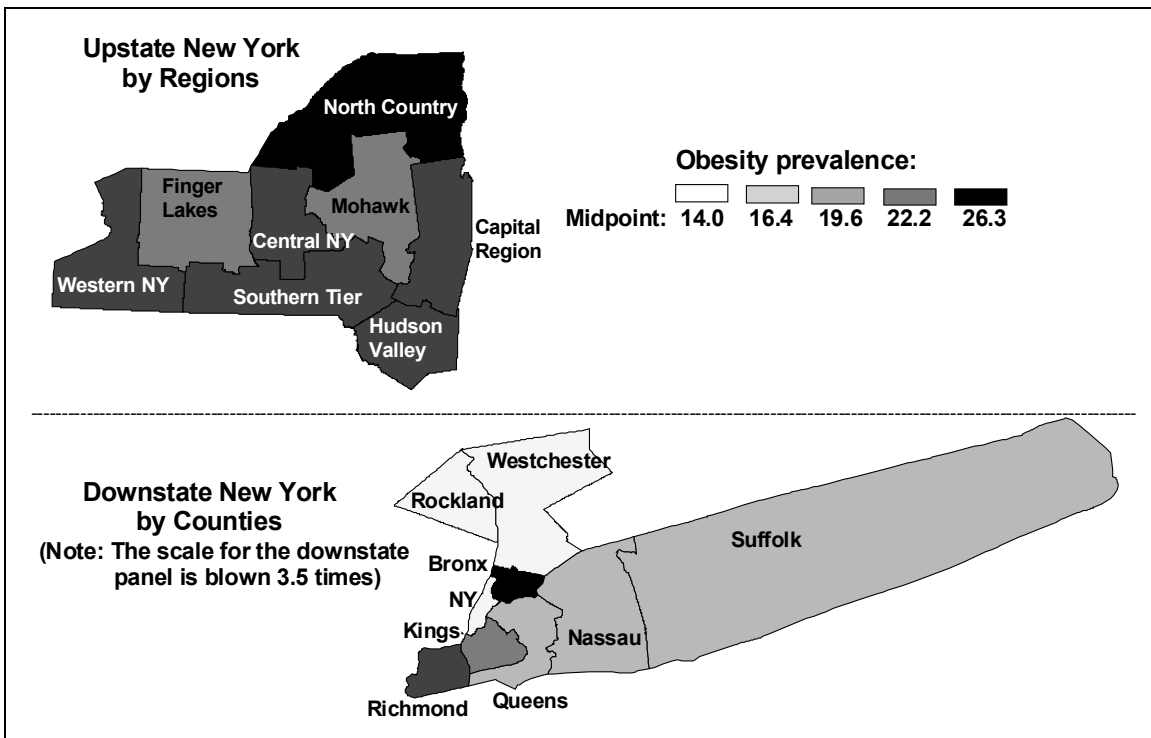


Figure 11. Prevalence of Asthma

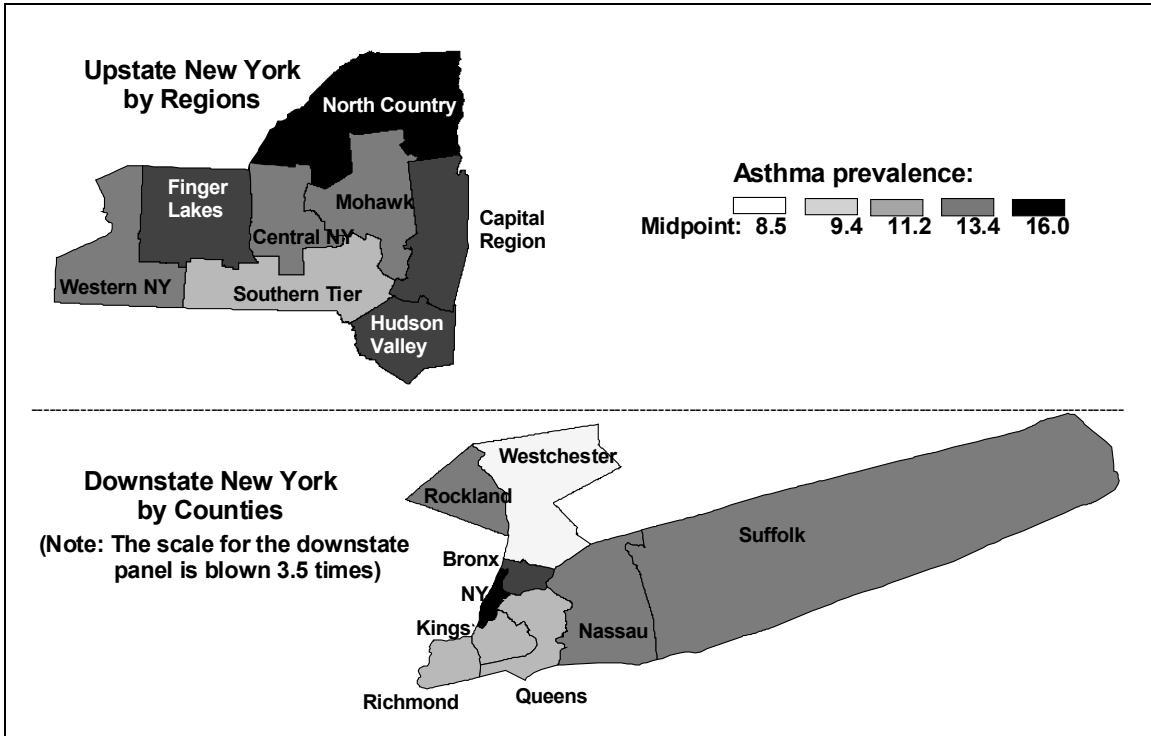


Figure 12. Prevalence of Hypertension

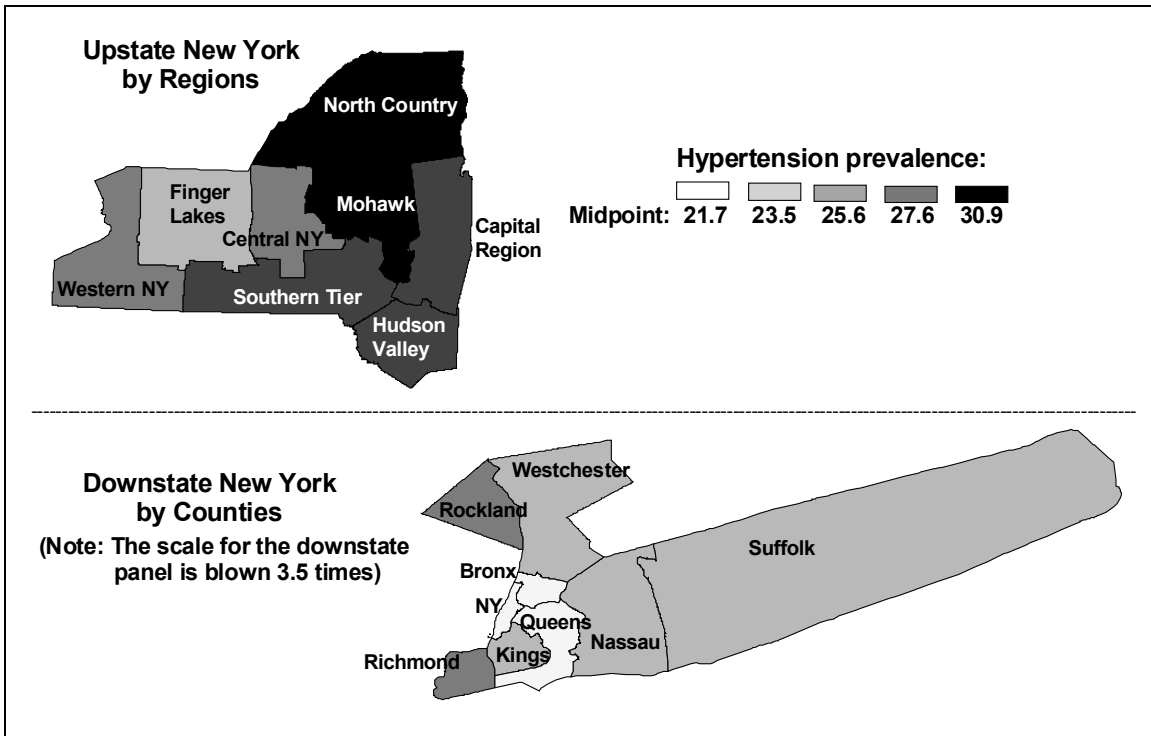


Figure 13. Prevalence of Diabetes

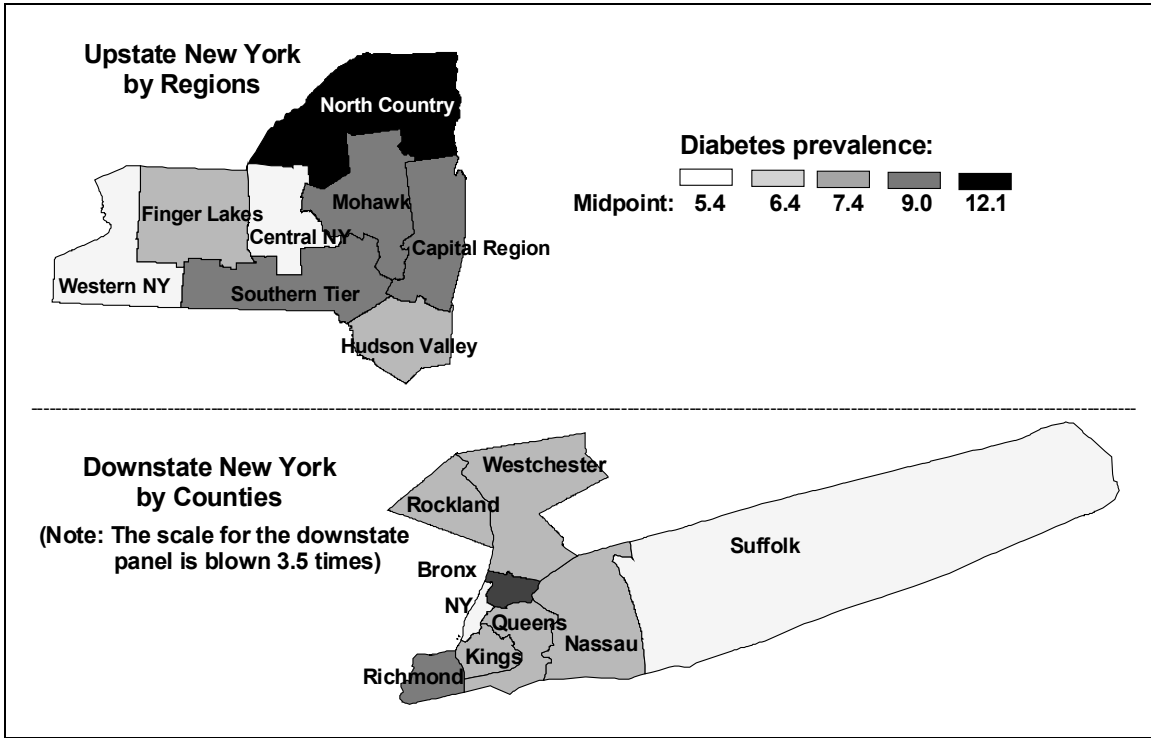


Figure 14. Average Estimated Quality of Health by Geographic Areas

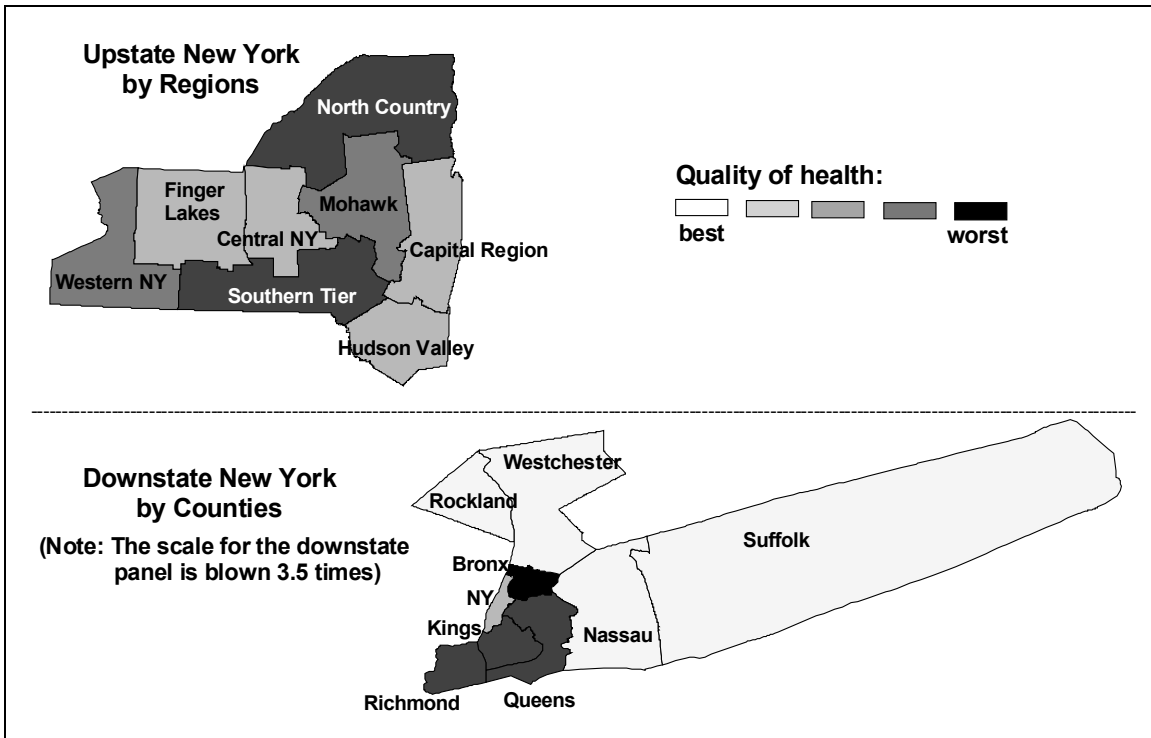


Figure 15. Concentration Indices with 95%-Confidence Intervals by Racial/Ethnic Groups

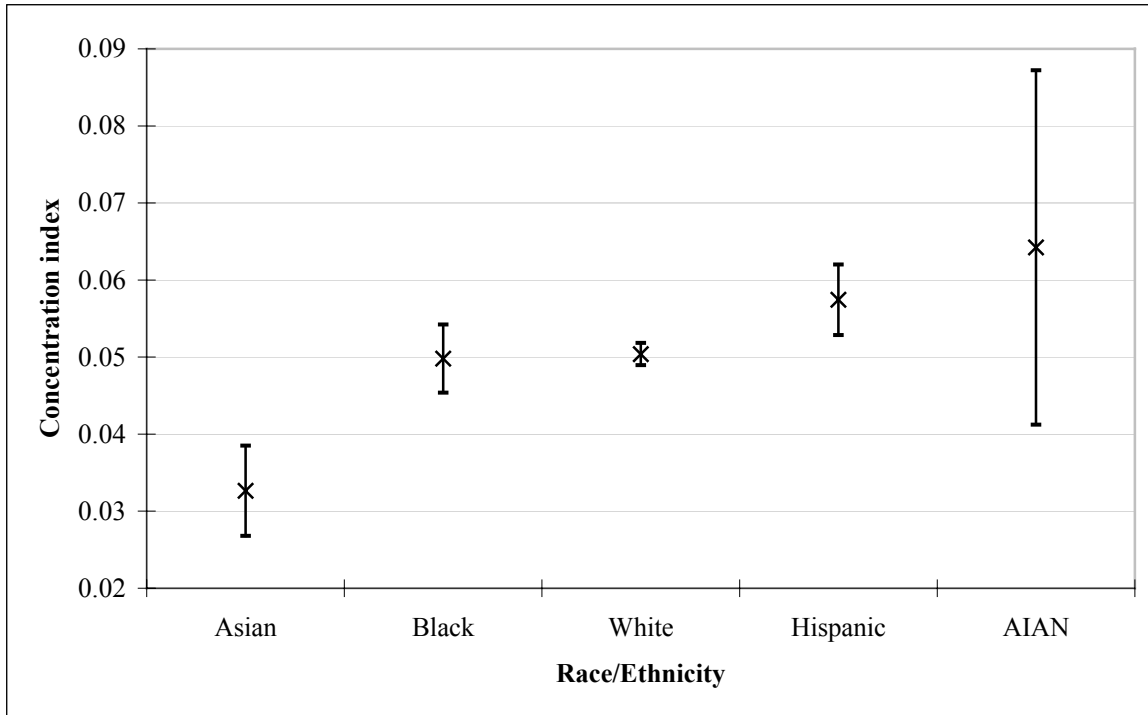


Figure 16. Concentration Curve by Racial/Ethnic Groups

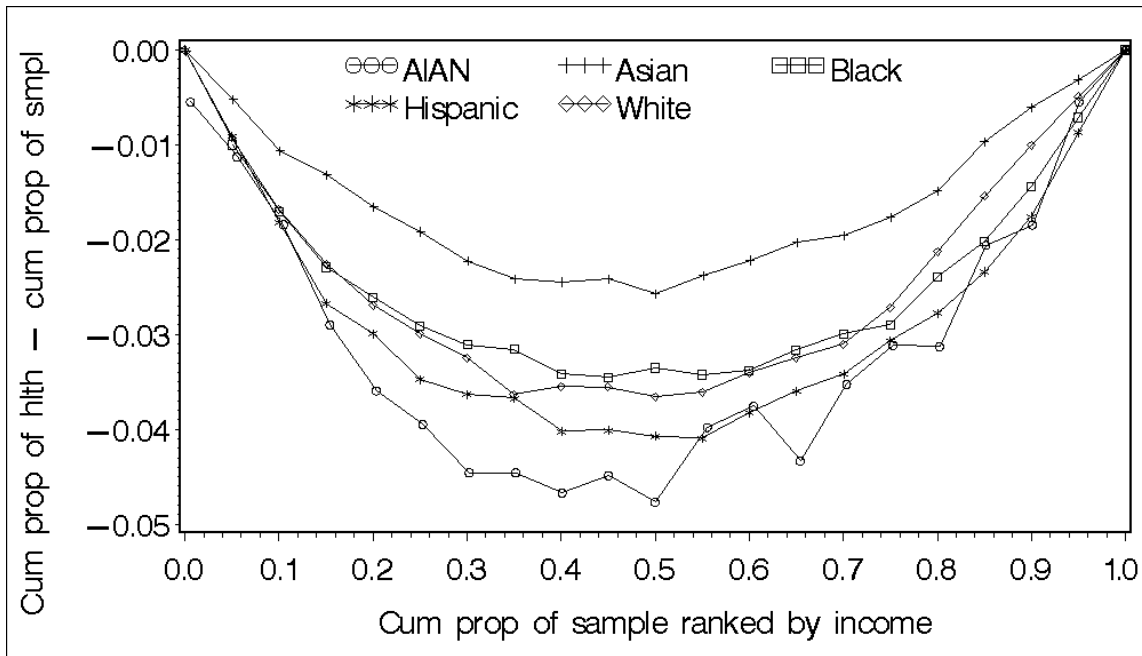


Figure 17. Income Related Health Inequality by Geographic Areas

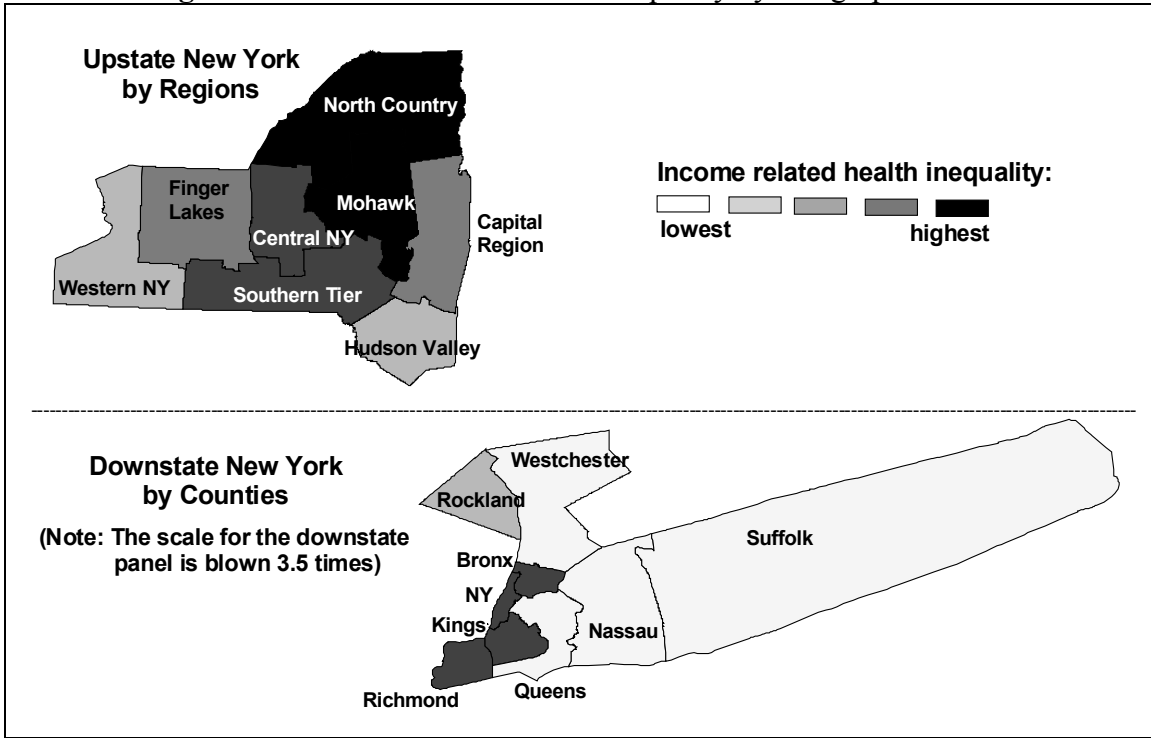


Figure 18. Average Health vs. Income Related Health Inequality



Note: 1 - Hudson Valley; 2 - Capital Region; 3 - Mohawk; 4 - North Country; 5 - Central New York; 6 - Southern Tier; 7 - Western New York; 8 - Finger Lakes; 9 - Bronx County; 10 - Kings County; 11 - Nassau County; 12 - New York County; 13 - Queens County; 14 - Richmond County; 15 - Rockland County; 16 - Suffolk County; 17 - Westchester County.

Figure 19. Regional Concentration Indices with 95%-Confidence Intervals

