Health Inequality and Its Determinants in New York[1]

by

Kajal Lahiri and Zulkarnain Pulungan
Department of Economics
University at Albany – SUNY
Albany, NY 12222
Email: klahiri@albany.edu

[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 a NIH Health Disparity Conference (October 23-24, 2006), the First Annual Health Disparities Conference, Columbia University (March 10-11, 2006), and at a roundtable discussion at the Center for the Elimination of Minority Health Disparities, University at Albany. Comments from Richard Alba, Bill Jenkins, Richard Keller, Catherine Lawson, Louis McNutt, Thad Mirer, Jim Scanlan, and Larry Schell are gratefully acknowledged. We, however, are responsible for all remaining errors and omissions.

Health Inequality and Its Determinants in New York

Summary: Self-assessed health status conditioned by several objective measures of health and socio-demographic characteristics are used to measure health inequality. We compare the quality of health and health inequality among different racial/ethnic groups as well as across 10 economic development regions in New York State. In terms of average health and health inequality, American Indian/Alaskan Natives and Hispanics are found to be the worst, and North Country and Southern Tier regions lag behind the rest of the State. Three major contributing factors to health inequality are found to be employment status, education, and income. However, the contribution of each of these determinants varies significantly among racial/ethnic groups as well as across regions, suggesting targeted public health initiatives for vulnerable populations to eliminate overall health disparity.

Keywords: BRFSS data; Self-assessed health; Ordered Probit; Health inequality; Gini coefficient; Lorenz curve; Decomposition analysis

1. Introduction

The goals of Healthy People 2010 - the national statement on health objectives - are twofold: first, to help individuals of all ages to increase life expectancy and to improve their quality of life; and second, to eliminate health disparities among segments of the population, including differences that occur by gender, race or ethnicity, education or income, disability, geographic location, and sexual orientation (US-DHHS, 2000).

Achieving the Healthy People 2010 goals needs effective public policies that require a precise and consistent measure of quality of health and health inequality.[2] Different groups of the population have different health and socioeconomic characteristics, which vary considerably over regions. In addition, the causes of health disparity may also be different for different groups. So in order to improve the quality of health and to eliminate health disparity, policy makers need to identify the main sources of disparity in different groups. While the quality of health of a particular group may be improved more effectively through education, for another group better health insurance or employment initiatives may be more effective.

Numerous studies on measuring the quality of health and health disparity 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 this study we focus on measuring the quality of health more generally, and calculate the quality of health and its 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:

[2] We will use 'inequality' and 'disparity' interchangeably in this paper to mean differences in health status within and between groups of people.
[3] See, for instance, Williams and Collins (1995), Ayanian *et al*. (1999), Shishehbor *et al*. (2006), and Safaei (2006).

excellent, very good, good, fair, or poor?" (*Centers for Disease Control and Prevention*, CDC).

SAH has been shown to be a good measure of overall health conditions. In their review, Idler and Benyamini (1997) show that SAH has strong predictive validity for mortality. Sickles and Taubman (1997) compiled results from worldwide studies on the association between self-assessed health and mortality, and reported that lower level of SAH has higher mortality odds. Manor *et al*. (2001) found that SAH has a strong association with longstanding illness. Furthermore, Hu *et al*. (2001) show that SAH is a useful predictor of the severity of diseases. 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.

Dichotomized SAH status—defined as the percentage of 'excellent' and 'very good' health, or 'fair' and 'poor' health—has been used as a measure of health of a population (*e.g.,* Keppel *et al*. 2004). In this paper SAH is modeled using an Ordered Probit Model (McKelvey and Zavoina, 1975). The estimated values from this model -- which is conditioned by several objective determinants including different diseases or risk factors, and socio-demographic characteristics -- are used as a measure of individual health. These predicted values are utilized to measure health inequality using Gini coefficient and Lorenz curve (Kakwani *et al*. 1997). Furthermore, to be useful for policy purposes, the health inequality is decomposed into its determinants (Wagstaff *et al.* 2003).

The primary goal of this paper is to measure health inequality between and within racial/ethnic groups as well as across regions of New York State. We also decompose health inequality into its determinants and characterize the sources of inequality for different population groups. This is the first study to look at the health status of New Yorkers along these dimensions.

The paper is organized as follows. Section 2 describes the estimation procedures of SAH - the methods to calculate quality of health, health inequality and their determinants. The data used in the empirical analysis are described in Section 3. The results are presented in section 4. Finally, section 5 summarizes our conclusions.

2. 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^*, an objective measure 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 $\mathbf{x}_i$. 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:

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

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

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

$$i = 1,2,...,n$$

where **γ, β, η** are vectors of coefficients, **μ** = (μ_1,…,μ_{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 maximum likelihood estimation. The predicted quality of health, $\hat{h}_i^* = \mathbf{h}_i^o \hat{\boldsymbol{\gamma}} + \mathbf{x}_i \hat{\boldsymbol{\beta}}$, 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 re-scale 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.

[4] van Doorslaer and Jones (2003) have shown that this heteroskedastic model accommodates possible individual-specific heterogeneity in the subjective thresholds **μ**.

Using the estimated quality of health h_i, we measure health inequality using pseudo-Lorenz curves and health Gini coefficient (Wagstaff *et al*. 1991).[5] A Lorenz curve plots the cumulative proportion of health $L(s)$ against the cumulative proportion of population s (starting with the lowest health and ending with the highest health), as shown in Figure 1. If the Lorenz curve $L(s)$ coincides with the diagonal, health is equally distributed. This means that there is no health inequality in the population. The farther the Lorenz curve from the diagonal the larger is the degree of inequality. The area between Lorenz curve and the diagonal provides a measure of inequality. The Gini coefficient is defined as twice the area between the Lorenz curve and the diagonal. The coefficient ranges from 0 (when everybody enjoys exactly the same health) to 1 (when all population's health is concentrated in the hands of one person).

The Gini coefficient can be calculated using equation (see Kakwani *et al*., 1997):

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

where R_i is the i^{th} individual fraction rank in health and μ is the mean of quality of health. The variance is estimated using the Huber-White procedure. The disadvantage of the Gini coefficient 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 health rank (Milanovic, 1997).

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

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

$$\hat{G} = \sum_{k=1}^{K} \left(\hat{\beta}_k \overline{w}_k / \overline{h} \right) G_k \tag{3}$$

where $\overline{h}$ is the mean of h, $\overline{w}_k$ is the mean of variable w_k from the vector of explanatory variables **w,** and G_k is the Gini coefficient ranked by health for variable w_k.

3. Data, descriptive statistics, and imputation

3.1. 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 22,083 sample observations.[6] Every year health departments of all states, with technical and methodological assistance from the Centers for Disease Control and Prevention (CDC), conduct monthly telephone interviews on randomly selected non-institutional 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.

Racial/ethnic groups included in the comparison are Non-Hispanic White (White), Non-Hispanic Black (Black), Hispanic, Asian/Pacific Islander (Asian), and American Indian Alaskan Native (AIAN). Regions included in the comparison are the economic development regions of New York State, which consist of 10 groups of counties (see

[6] Sehili *et al.* (2005) have used this data source to study health inequality in the U.S. in terms of physically healthy days.

Table A1 in the Appendix). Descriptive statistics of all variables by racial/ethnic groups are presented in Table A2, while descriptive statistics for all variables by regions are presented in Table A3. For some variables, the descriptive statistics for Asian or AIAN are not reported because the number of observations is small. In this case, we follow the BRFSS guideline that the minimum number of observations to be meaningful for interpretation is 50. Specifically, the variables used in this study are 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 different age distributions when comparing the prevalence of the diseases among racial/ethnic groups or regions. Comparing across regions of New York State, average age varies from 43 (New York City) to 49 years (Mohawk). Sixty percent of White respondents are married, compared to 37 percent of Black respondents. The percentage of married respondents in New York City is the lowest (43%) compared to other regions.

Education level varies considerably among racial/ethnic groups as well as across the regions. Only 13 percent of AIAN respondents have 4 years or more of college, compared to 61 percent of Asian respondents; across the regions, the percentage varies from 26 percent in North Country to 42 percent in Hudson Valley. The percentage of respondents who were unable to work varies from 1.6 percent (Asian) to 14 percent (AIAN), while the percentage across regions varies from 2.3 percent (Capital Region) to 11.3 percent (North Country). Annual household income also varies considerably among racial/ethnic groups as well as across regions. AIAN average annual household income is

$34,390, compared to $62,470 of Asian. Across the regions, the average varies considerably from $46,864 (North Country) to $71,848 (Long Island).

Eighty-six percent of the respondents have health plans, but only 68 percent among Hispanics have health plan compared to 91 percent among Whites. Across regions, the percentage varies from 80 percent (New York City) to 92 percent (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 percent (White) to 21 percent (Hispanic). Fourteen percent of New York City respondents could not afford to see a doctor at least once, while in Capital District only 7 percent of the respondents had that experience.

The U.S. Surgeon General Report has concluded that smoking is a source of many kinds of diseases and it 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 percent (Asian) to 34 percent (AIAN); the percentage of smokers across regions varies from 19 percent (Hudson Valley) to 28 percent (North Country). In addition to smoking, drinking alcohol, and not exercising are categorized as bad health habits that can result in several kinds of diseases. The percentage of heavy drinkers among racial/ethnic groups varies from 10.8 percent (Black) to 17.9 percent (AIAN), and the percentage across regions ranges from 1.4 percent (Hudson Valley) to 23.1 percent (North Country). The percentage of respondents who participated in any sort of exercise in the past 30 days is 74 percent. Sixty-two percent of Hispanic respondents participated in any exercise in the past 30

days, compared to 79 percent for White respondents; across regions, New York City has the lowest percentage (71%) and North Country has the highest (80%).

- Self assessed health status:

Figure 2 presents the distribution of SAH by racial/ethnic groups. Twenty-two percent of the respondents considered their health as 'excellent', while only 4 percent 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 White than for Black, Hispanic, or AIAN. In addition, the percentage of 'fair' and 'poor' health is lower for White than for Black, Hispanic, or AIAN. The percentage of 'excellent' and 'very good' health varies from 49 percent (New York City) to 62 percent (Capital Region), and the percentage of 'fair' and 'poor' health varies noticeably from 10 percent (Finger Lakes) to 19 percent (New York City) (see Table A3). Based on both criteria, New York City has the lowest quality of health compared to the other regions. 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 North Country is the worst.

Figure 3 presents the distribution of SAH by income groups. The figure indicates that as income increases the percentage of 'excellent' health increases. In contrast, the figure shows that as income increases the percentage of 'poor' health decreases. This pattern indicates a strong association between income and quality of health.

- 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 regions the patterns of these two measures are different. The lowest average number of days where physical health was not good is found in Hudson Valley and the highest average is found in Mohawk. While the lowest average number of days where mental health was not good is found in North Country, and the highest is found in Southern Tier.

- Limited activities due to health problems:

The overall prevalence of limited activities due to health problem is 16 percent. The prevalence varies substantially among racial/ethnic groups ranging from 6 percent (Asian) to 31 percent (AIAN). The prevalence varies from 15 percent (New York City) to 21 percent (Mohawk).

- Body mass index (BMI):

Obesity - defined as BMI greater than 30 kg/m^3 – is the second leading cause of preventable death after smoking in the U.S. and is as a major cause of morbidity and disability (Must *et al*. 1999; Mokdad *et al*., 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 regions it varies from 26 (Long Island) to 28 (North Country).

- Asthma:

The prevalence of Asthma among respondents is 11.8 percent. Comparing racial/ethnic groups, the prevalence varies substantially from 5.9 percent (Asian) to 17.7 percent (AIAN). Regionally, the prevalence varies from 9.8 percent (Southern Tier) to 13.8 percent (Finger Lakes).

- High blood pressure:

The prevalence of high blood pressure among respondents is 24.3 percent. It is only 12.2 percent for Asian compared to 31.2 percent for AIAN. Across regions the prevalence varies from 22.2 percent (New York City) to 33.5 percent (Mohawk).

- Coronary heart disease:

Coronary heart disease prevalence is 4.2 percent in the pooled sample. The prevalence varies from 2 percent (Asian) to 7 percent (AIAN). Black, Hispanic, and Asian have lower prevalence than White. Across regions, the prevalence varies from 2.4 percent (North Country) to 4.8 percent (Southern Tier).

- Myocardial Infarction:

The prevalence of myocardial infarction is 3.5 percent. Asian has the lowest prevalence (less than 0.1%) and AIAN has the highest (7.6%). Across regions prevalence varies from 2.4 percent (Capital district) to 4.5 percent (Mohawk). It is noteworthy that North Country is excluded from the comparison because the sample size is too small (sample size <50).

- Stroke:

The incidence of stroke is 1.9 percent. Comparing racial/ethnic groups the prevalence varies considerably, ranging from 0.4 percent (Asian) to 6 percent (AIAN). The prevalence varies considerably across the regions, ranging from 1.5 percent (Hudson Valley) to 4.4 percent (Southern Tier).

- Diabetes:

The overall prevalence of diabetes is 6.4 percent. It varies from 4.7 percent for Asian to 11.6 percent for AIAN. Across regions the prevalence varies from 5.5 percent (Long Island) to 14.1 percent (North Country).

- Arthritis:

The prevalence of Arthritis is 25.1 percent in the whole population, but varies substantially among racial/ethnic groups, ranging from 8.1 percent (Asian) to 35.1 percent (AIAN). Across regions the prevalence varies from 20.2 percent (New York City) to 34.7 percent (North Country).

- Pain, aching, stiffness, and swelling in or around a joint:

The prevalence of this medical condition is 38.3 percent. The prevalence varies considerably, ranging from 23.7 percent for Asian to 52.5 percent for AIAN. Across regions the prevalence varies from 34.2 percent (New York City) to 46.7 percent (Western New York).

- High blood cholesterol:

The prevalence of high blood cholesterol is 31 percent. It varies from 24.5 percent for Hispanic to 48.3 percent for AIAN. Across regions the prevalence varies from 27.4 percent (Mohawk) to 34.1 percent (Long Island).

3.2. 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.

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 doing 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 *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 *et al.* (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.[7]

4. Results

4.1. Coefficient estimates

Table 3 presents the coefficient estimates of equation (1). Since this study is based on pooled cross-section 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 a coefficient's absolute value the stronger is the association between quality of health and the corresponding explanatory variable.

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 from age group 25-39. 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 even after controlling for objective health measures, the self-reported health of minority populations are lower than that of the White population. It may mean that there are omitted covariates in the

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

regression (*e.g.,* severity of diseases and risk factors, neighborhood effects, discrimination, etc.) that systematically affect the health of the minorities. Kobetz *et al.* (2003) found that neighborhood poverty is associated with a greater likelihood of poor SAH.[8]

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 are not statistically significant and therefore are 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 non-smokers. 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 morbidity for chronic diseases (James and Nelson 1997). A number of researchers have

[8] It may also be due to relatively different thresholds used by White while reporting SAH, see Banks *et al.* (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 fn. 4.

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 neighborhood characteristics too 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.

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 *et al*. (2006) found that reporting heterogeneity of health status does not have a large quantitative impact on the measures of health inequality.

4.2. Quality of health

4.2.1. 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 has the lowest. The average age varies considerably among racial/ethnic groups from 38.6 years through 47.6 years (see Table A2 in appendix). 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 quality of health between groups in a population with different age distributions could be misleading.

Several methods can be used to control for the effects of the age distribution. The simplest method is by comparing the average estimated quality of health between groups of population by age groups. Another method is by incorporating 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 *et al*., 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 that eliminates the effect of age distribution without differentiating the mortality rates among the groups. HALE for each racial/ethnic group by age groups are presented in Table 5.

The table shows that White in the youngest age group (20-24) has 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 44.24 years in perfect health condition, while a Hispanic with the same age is expected to live for 36.81 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 White does better than Asian, while Hispanic does 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).

4.2.2. Geography of health

In this part, we do not compute HALE for each region for two reasons. First, the distributions of age across the regions are very similar so the effect of age distribution is negligible. Second, not all regions have enough observations required to compute HALE. The average quality of health by region is presented in Figure 4. Hudson Valley and Long Island are in the brightest areas reflecting the highest quality of health. In contrast, North Country and Southern Tier are in the darkest areas reflecting the lowest quality of health.

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; 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. Figures A1 and A2 in the Appendix present the patterns of quality of health using two different cut points. From both figures, it is obvious that two different cut points used in dichotomizing the SAH give two somewhat different patterns of quality of health. The procedure used in this paper circumvents this problem of arbitrariness.

4.3. Health inequality

Similar to quality of health, we also compare health inequality among racial/ethnic groups as well as across regions. In addition, this section also presents the decomposition results for each racial/ethnic group and for different regions.

4.3.1. Health inequality by racial/ethnic groups

Gini coefficients by racial/ethnic groups with corresponding 95%-confidence intervals are presented graphically in Figure 5. The standard errors of the Gini coefficients are calculated using the Huber-White robust estimator. All coefficients are significantly greater than zero, indicating that health inequalities exist in all groups. There is, however, substantial variation in the coefficients among groups. The highest inequality is found within AIAN group followed by Hispanic. The lowest inequality is found within the Asian group followed by White. The figures also show that the differences in Gini coefficients between groups are statistically significant.

Another way to compare inequalities between groups of populations is by comparing their Lorenz curves. Figure 6 presents the Lorenz curves expressed as the deviation of the Lorenz curve from the diagonal in order to amplify the differences

between racial/ethnic groups. The figure provides more obvious evidence of the differences of health inequalities among racial/ethnic groups. The Asian curve strictly dominates others, while AIAN curve is strictly dominated by others. These indicate that AIAN is the most unequal at all percentiles, while Asian is the least. Therefore, differences among racial/ethnic groups are not only in terms of the average quality of health but also in terms of health distribution itself among individuals within each group.

4.3.2. Geography of health inequality

Figure 7 presents health Gini coefficients for each region. North Country represents the darkest area indicating the highest health inequality, while the brightest areas are represented by - from the lowest to the highest - Long Island, Hudson Valley, Finger Lakes, and Capital Region. Comparing Figures 5 and 7, it is clear that regions in the dark areas in Figure 5 are also in the dark areas in Figure 7. In the other words, regions with lower quality of health tend to have higher health inequality. It is noteworthy that the dichotomized SAH measure cannot be used to calculate a Gini coefficient. This is another advantage of the continuous measure of health we use in this work.

For more detailed information about the magnitude and significance of the Gini coefficients across regions, the corresponding 95%-confidence intervals by regions are presented in Figure 8. All coefficients are significantly different from zero, indicating health inequalities exist in all regions. The statistical significance of the difference in the Gini coefficients between regions can be seen by comparing the confidence intervals. For instance, New York City has significantly higher Gini coefficient than those of Long

Island, Hudson Valley, and Capital Region. But the Gini coefficient of New York City is not significantly different from those of Mohawk and Central New York.

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.

4.3.3. Decomposition of health inequality

The decomposition analysis demonstrates differences in the components of inequalities for different racial/ethnic groups as well as for different regions. We are interested in analyzing health inequality attributable to socio-demographic factors including age, sex, race/ethnicity, marital status, education, employment status, health insurance, smoking status, access a doctor, and living in New York City. The percentage contribution of each component is calculated by racial/ethnic groups and also by regions. Among all these factors, the contributions of the main variables by racial/ethnic groups are presented in Table 6. The contributions by regions are presented in Table 7.

For the overall New York State population, among the socio-demographic variables, three major factors contributing to health inequality are employment status, annual household income, and age. Each of these three factors contributes more than 20 percent to health inequality. Our estimate of the effect of annual household income is similar to that in Wagstaff and van Doorslear (2004) who found that income inequality in Canada is around 25% of overall health inequality. However, we find that income is relatively less important for the disadvantaged minority groups (*viz*., Black, Hispanic and AIAN). For instance, the corresponding percentages are 31.56 and 11.01 for Asian and

AIAN respectively. A similar differential effect was found in Subramanian and Kawachi (2006) between state income inequality and average health for the U.S. The observed association between income inequality and health inequality on the one hand, and between health inequality and average health on the other found in this paper implies a 'pollution effect' of income inequality on average health across regions, *cf*. Subramanian and Kawachi (2004). This correlation in our sample of New York regions was found to be –0.84.

If health status were distributed equally across different employment status, household incomes, and age groups, health inequality attributable to socio-demographic variables in New York State population would be 66 percent lower. After controlling for other factors, race/ethnicity contributes only 4.5 percent to health inequality. As can be seen from Table A2 in the Appendix, race/ethnicity is highly intertwined with employment status, income, and education - Black, Hispanic, and AIAN have lower education levels, lower employment rates, and lower household incomes compared to those of White. That is why separate analysis for each group is necessary.

Comparing racial/ethnic groups, the pattern of contribution of the socio-demographic variables varies considerably. The largest contributor to health inequality within the White population is Age (28.46%) followed by employment status (24.51%); for Black it is employment status (32.98%) followed by age (19.94%); for Hispanic it is employment status (31.58%) followed by education (22.73%); for Asian is annual household income (31.56%) followed by age (18.04%); and for AIAN it is employment status (49.66%) followed by age (18.51%). Since inequality in employment status has the highest contribution to health inequality within Black, Hispanic, and AIAN, the most

effective public policy initiative to eliminate health inequality is to ensure employment opportunities to all in these minority groups, particularly the AIAN. Education is another important factor, but it is more important for Hispanic. Interestingly, for Asian, income inequality is the most important (32%) contributor to its health inequality.

Comparing across regions, the contribution of each factor to health inequality varies noticeably (see Table 7). For example, the contribution of education to health inequality ranges from 21.08% (Capital Region) to 43.94 % (Southern Tier). In New York City employment status has the highest contribution to health inequality. Thus, the most effective policy to eliminate health inequality in New York City is to provide employment opportunities to minority populations. The second most effective policy is to ensure good educational access to all. Moreover, the contribution of race/ethnicity to health inequality in New York City is 7.31%, while in Southern Tier it is only 0.78 percent. This is an obvious result given the diversity of New York City population.

5. Conclusions

Following recent developments in socio-demographic and spatial research on 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 distributions, Hispanics have the lowest quality of health. Asians have the highest average followed by Whites, while after adjusting for age distributions, Whites have the best quality of health. This result highlights that when comparing quality of health between groups of populations, one needs to consider the age distribution within

each group. Across the 10 economic development regions of New York State, Southern Tier followed by New York City has the lowest average quality of health, and Hudson Valley followed by Long Island has the highest. These differences are mostly statistically significant.

We find statistically different health inequality, both spatially and across racial/ethnic groups. The highest health inequality is found within the AIAN group followed by Hispanic, while the lowest health inequality is found within the Asian group, followed by White. Across the 10 regions, the highest health inequality is found in the North Country followed by Southern Tier, while the lowest health inequality is found in Long Island followed by Hudson Valley. Groups with lower average quality of health tend to have higher health inequality.

The statistical decomposition analysis shows the contribution of several socio-economic and demographic factors to health inequality for different racial/ethnic groups as well as for different regions. After controlling for age distribution, the three major factors generating health inequality are employment, education, and household income -- each contributing around 20 percent to the total variation in health inequality. For the disadvantaged minorities (Black, Hispanic, and AIAN), employment status is the most important factor -- alone responsible for more than 30% of health inequality. The contribution of these three factors varies across regions, but employment status is again found to be relatively more important. For instance, in the North Country, nearly 44% of its health disparity is explained by employment.

Our results underscore the need for different public health policy initiatives for different racial/ethnic groups and different regions to eliminate overall health disparity.

In general, policies that can ensure equality in employment opportunities, educational access, and income will have a substantial impact on improving the average quality of health and in reducing health inequality. Unfortunately, there is no quick fix.

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 Estimate	Standard Error	P-value
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

Variable		Race/ethnicity					
		All	**White**	**Black**	**Hisp.**	**Asian**	**AIAN**
Age	(%)	21.63	28.46	19.94	16.91	18.04	18.51
Race/ethnicity	(%)	4.53	-	-	-	-	-
Education	(%)	17.13	15.74	17.32	22.73	17.24	12.47
Employment status	(%)	24.51	24.11	32.98	31.58	15.23	49.66
Annual household Income ($1,000)	(%)	22.44	23.79	17.26	14.92	31.56	11.01
Smoking	(%)	3.16	3.51	5.01	2.73	2.94	4.45
Could not afford to see doctor	(%)	6.42	5.21	7.11	10.10	13.78	4.71

Note: Variables with small contributions are not presented in this table.

Table 7. Decomposition of Health Inequality by Regions

Variable		Region*									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Age	(%)	17.83	26.04	22.15	27.34	25.50	18.25	26.88	25.16	27.34	25.46
Race/ethnicity	(%)	7.31	2.69	3.58	1.84	1.64	0.78	1.35	1.33	1.74	2.51
Education	(%)	19.75	15.59	16.91	16.75	14.44	11.76	14.12	14.39	15.58	17.07
Employment status	(%)	25.10	21.52	21.83	21.08	29.90	43.94	28.32	30.65	27.85	21.45
Annual household income ($1,000)	(%)	21.20	25.98	26.38	23.88	18.97	15.36	21.02	16.96	20.05	23.58
Smoking	(%)	2.39	3.35	3.16	4.01	3.55	4.26	3.80	4.83	3.37	4.25
Could not afford to see doctor	(%)	7.17	5.40	6.87	5.78	6.34	6.39	5.35	7.52	5.07	6.37

Note:* 1 is New York City; 2 is Long Island; 3 is Hudson Valley; 4 is Capital Region; 5 is Mohawk; 6 is North Country; 7 is Central New York; 8 is Southern Tier; 9 is Western New York; and 10 is Finger Lakes.

Figure 1. Lorenz Curve

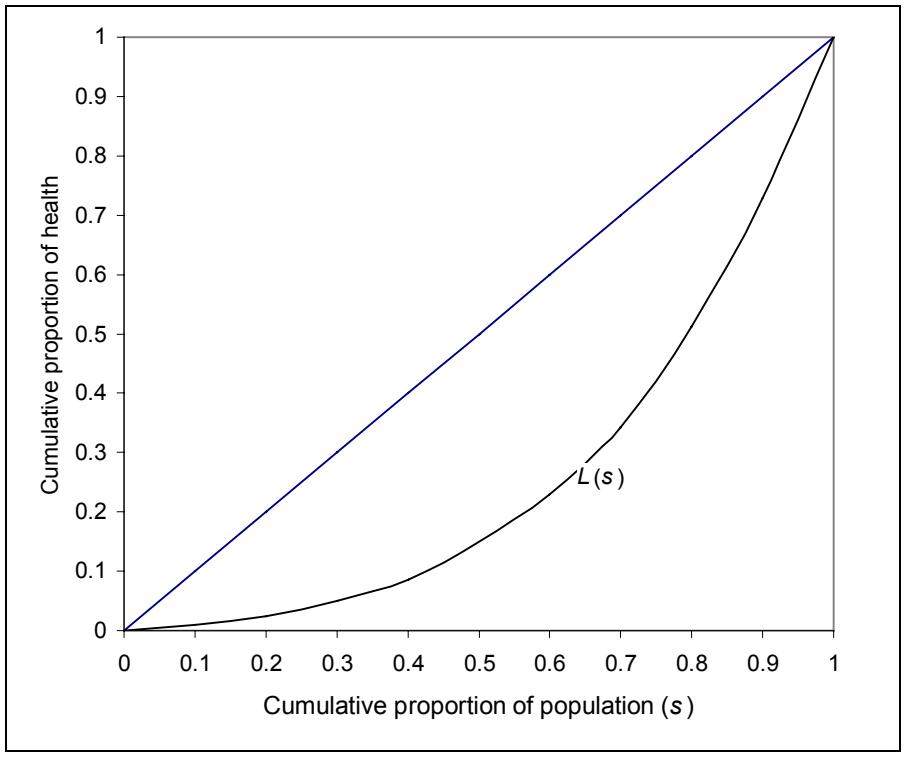


Figure 2. Distribution of SAH by Racial/Ethnic Groups

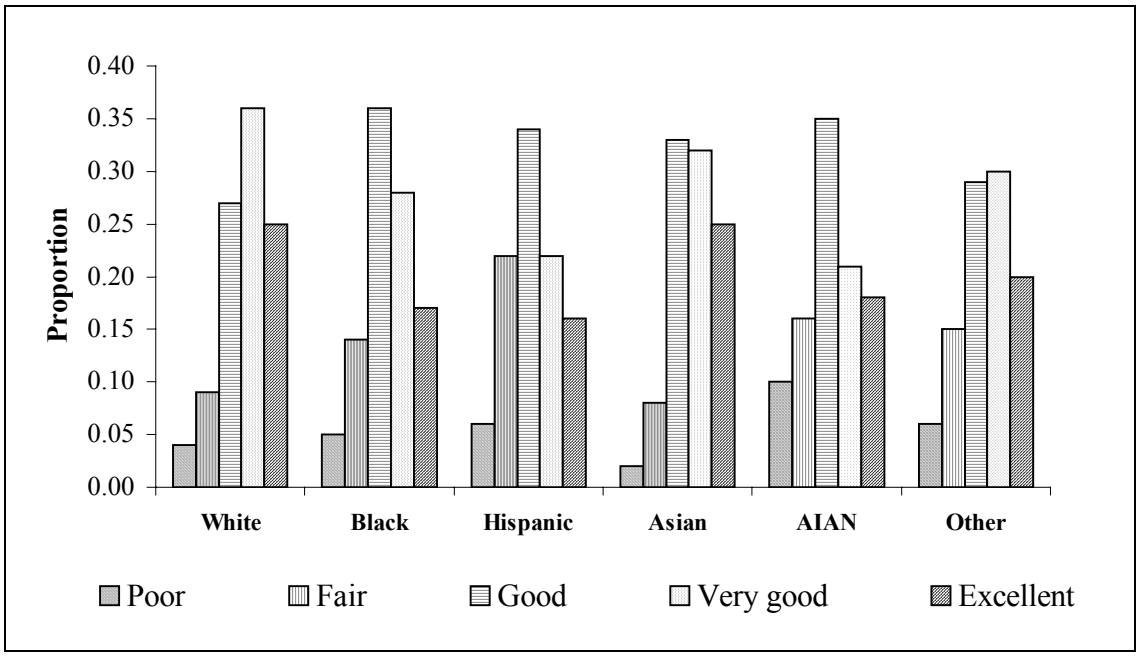


Figure 3. Distribution of SAH by Annual Household Income Groups

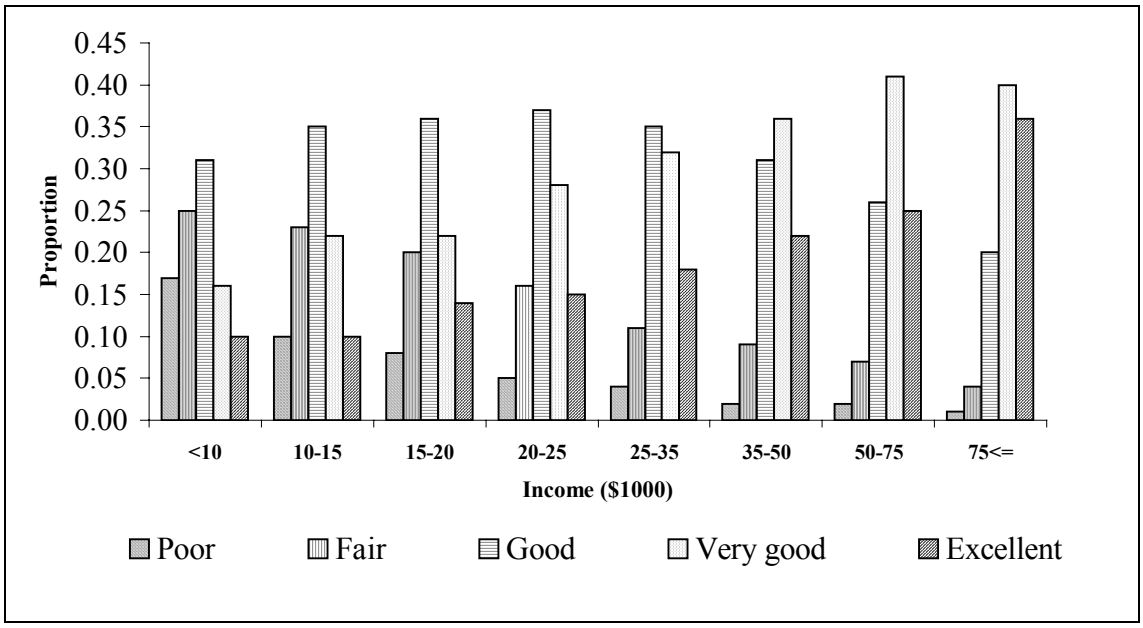


Figure 4. Average Estimated Quality of Health by Regions

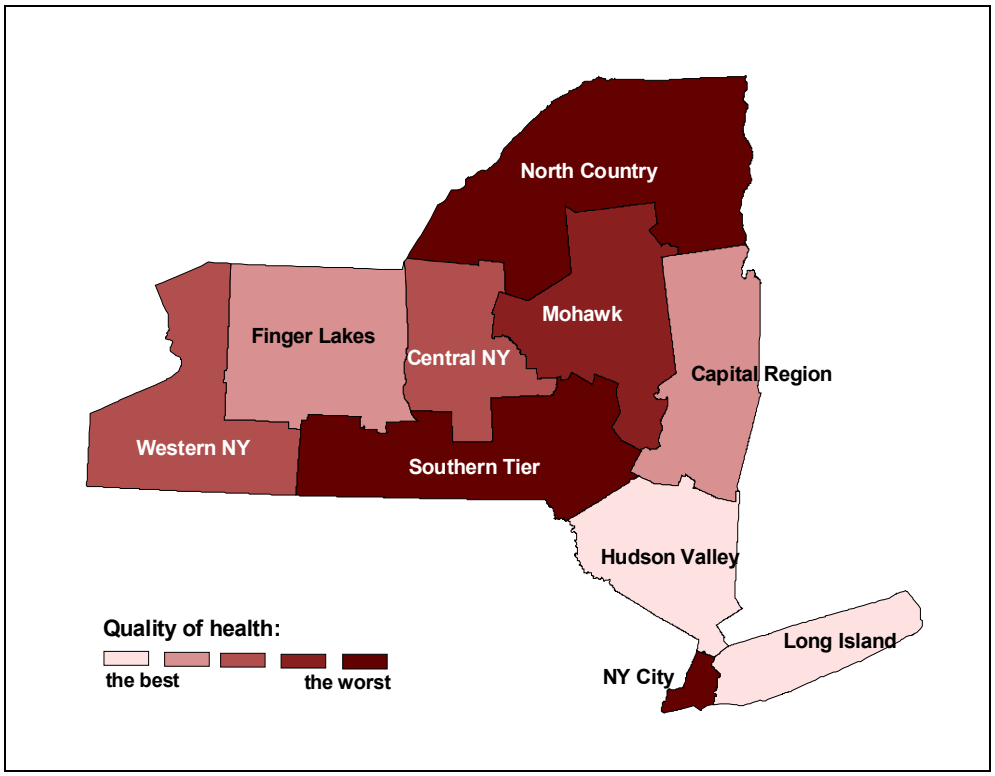


Figure 5. Gini Coefficient with Corresponding 95%-Confidence Interval by Racial/Ethnic Groups

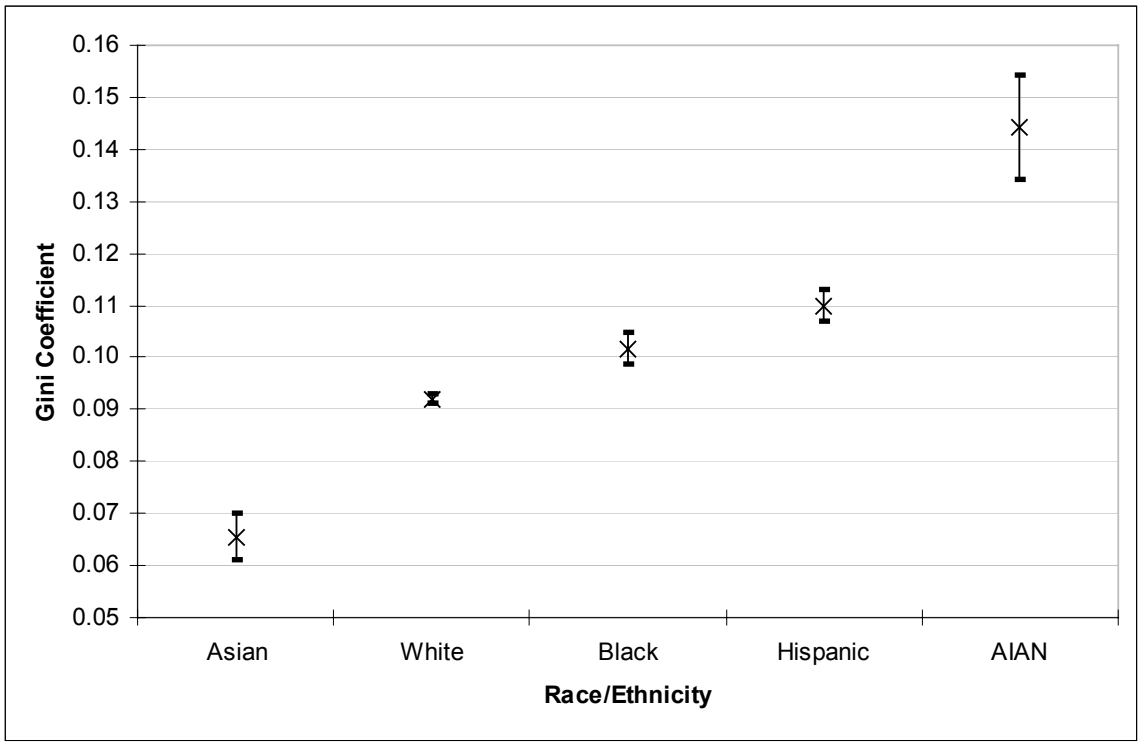


Figure 6. Health Lorenz Curve by Racial/Ethnic Groups

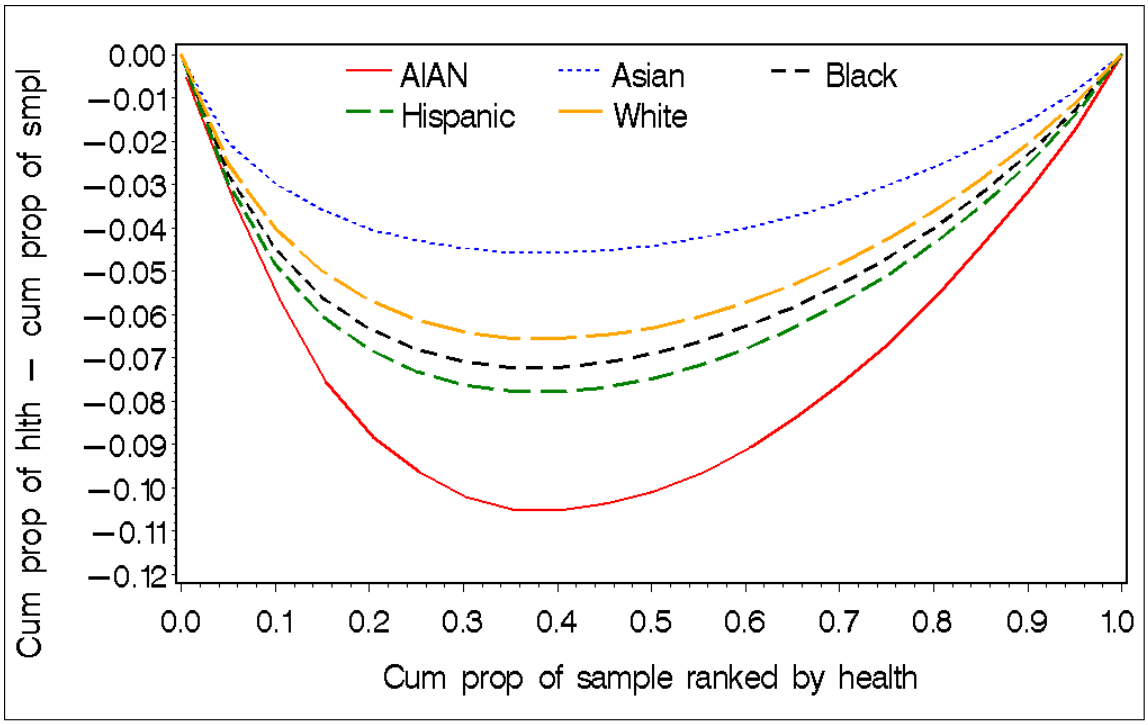


Figure 7. Health Inequality by Regions

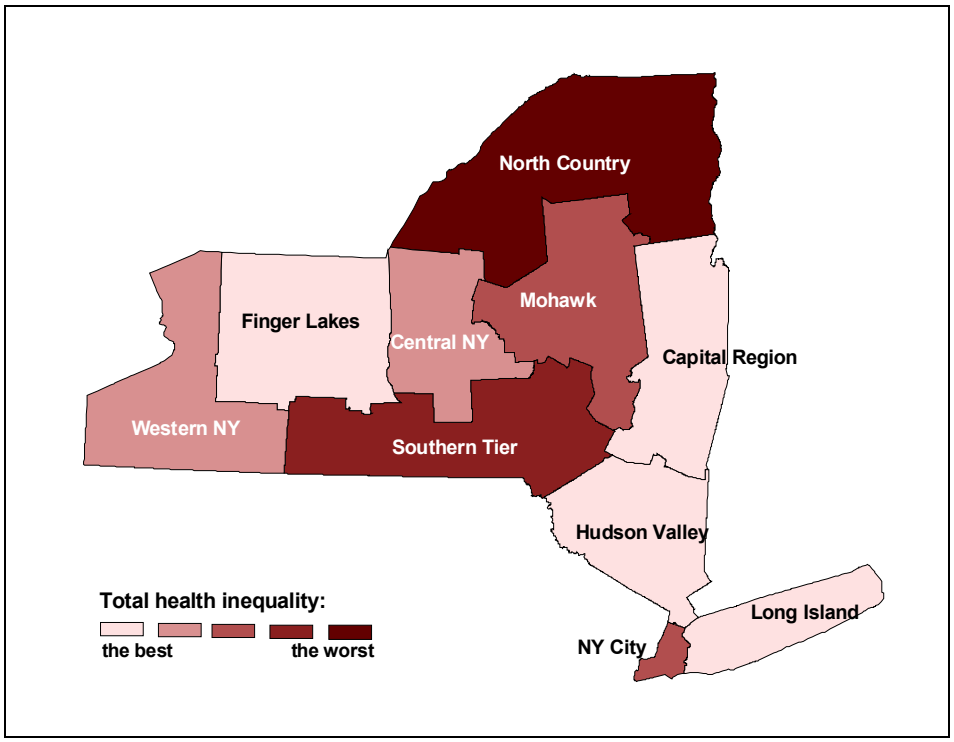
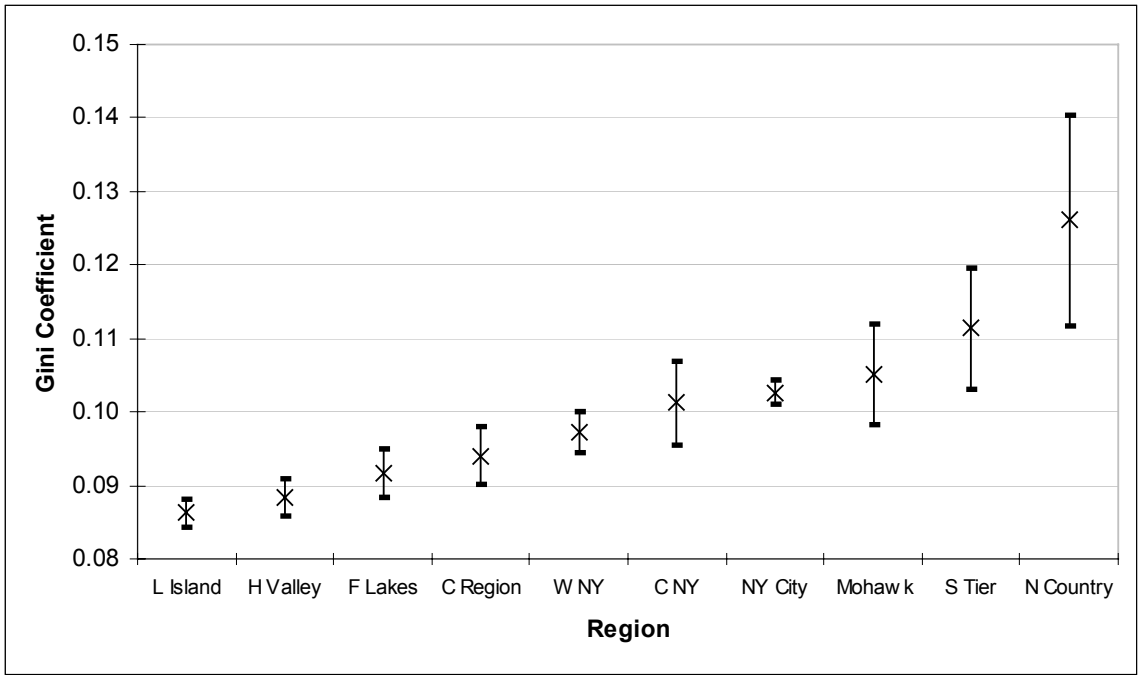


Figure 8. Gini Coefficient with Corresponding 95%-Confidence Interval by Regions



Appendix

Table A1. Counties in 10 Economic Development Regions of New York State

1 NY City	**4 Capital Region**	**6 North Country**	**9 Western New York**
Bronx	Albany	Clinton	Allegany
Kings	Columbia	Essex	Cattaraugus
New York	Greene	Franklin	Chautauqua
Queens	Rensselaer	Jefferson	Erie
Richmond	Saratoga	Lewis	Niagara
	Schenectady	St. Lawrence	
2 Long Island	Warren		**10 Finger Lakes**
Nassau	Washington	**7 Central New York**	Genesee
Suffolk		Cayuga	Livingston
	5 Mohawk	Cortland	Monroe
3 Hudson Valley	Fulton	Madison	Ontario
Dutchess	Hamilton	Onondaga	Orleans
Orange	Herkimer	Oswego	Seneca
Putnam	Montgomery		Wayne
Rockland	Oneida	**8 Southern Tier**	Wyoming
Sullivan	Schoharie	Broome	Yates
Ulster		Chemung	
Westchester		Chenango	
		Delaware	
		Otsego	
		Schuyler	
		Steuben	
		Tioga	
		Tompkins	

Table A2. Descriptive Statistics by Racial/Ethnic Groups

Variables	Race/Ethnicity						
	All	**White**	**Black**	**Hisp.**	**Asian**	**AIAN**	**Other**
Reported age in years	45.44	47.57	43.75	39.88	38.57	44.19	43.31
Gender (male=1)	0.488	0.489	0.435	0.476	0.590	0.545	0.595
Marital status	0.541	0.591	0.371	0.463	0.590	0.446	0.475
Education:							
Grade 8 or less	0.040	0.012	0.045	0.159	0.017	0.065	0.036
Grades 9 - 11 (Some high school)	0.072	0.047	0.100	0.165	0.019	0.184	0.047
Grade 12 or GED (High school graduate)	0.294	0.294	0.345	0.286	0.153	0.353	0.291
College 1 year to 3 years	0.259	0.266	0.289	0.217	0.197	0.270	0.263
College 4 years or more	0.336	0.380	0.221	0.173	0.614	0.127	0.363
Employment:							
Employed for wages	0.555	0.545	0.568	0.570	0.647	0.487	0.512
Self-employed	0.081	0.087	0.055	0.075	0.077	0.061	0.123
Out of work for more than 1 year	0.020	0.014	0.038	0.028	0.015	0.026	0.029
Out of work for less than 1 year	0.035	0.027	0.058	0.050	0.029	0.038	0.061
A homemaker	0.064	0.066	0.033	0.092	0.050	0.053	0.040
A student	0.045	0.039	0.049	0.049	0.119	0.034	0.057
Retired	0.159	0.194	0.130	0.070	0.048	0.160	0.103
Unable to work	0.041	0.028	0.070	0.067	0.016	0.140	0.076
Annual Household Income ($1,000)	55.71	62.21	42.74	36.74	62.47	34.39	48.05
Have health plan	0.860	0.914	0.836	0.685	0.787	0.755	0.765
Could not afford to see a doctor	0.110	0.078	0.141	0.209	0.114	0.182	0.200
Smoking	0.221	0.228	0.219	0.204	0.132	0.343	0.248
Heavy drinking	0.165	0.178	0.108	0.161	0.113	0.179	0.195
Self-assessed health status:							
Excellent	0.225	0.246	0.183	0.162	0.257	0.173	0.197
Very good	0.333	0.367	0.283	0.238	0.331	0.175	0.328
Good	0.298	0.273	0.356	0.346	0.322	0.381	0.287
Fair	0.110	0.084	0.138	0.205	0.064	0.158	0.127
Poor	0.035	0.030	0.039	0.048	0.026	0.113	0.061
Number of days physical health not good	3.343	3.264	3.632	3.765	1.650	5.362	3.478
Number of days mental health not good	3.227	3.039	3.483	3.783	2.468	5.131	4.297
Activities limited due to health problem	0.162	0.175	0.147	0.138	0.064	0.309	0.155
Body mass index (BMI)	26.48	26.23	28.05	27.00	23.97	27.61	25.81
Fruit and vegetable servings per day	4.032	4.031	3.914	3.938	4.287	4.280	4.758
Participate in any exercises	0.745	0.786	0.686	0.619	0.747	0.677	0.732
Ever had asthma	0.118	0.113	0.139	0.133	0.059	0.177	0.124
Ever told blood pressure high	0.243	0.245	0.302	0.221	0.122	0.312	0.195
Ever told had coronary heart disease	0.042	0.044	0.042	0.035	0.020	0.070	0.046
Ever told had myocardial infarction	0.035	0.038	0.026	0.032	0.000	0.076	0.022
Ever told had stroke	0.019	0.019	0.024	0.020	0.004	0.060	0.014
Ever told had diabetes	0.064	0.055	0.105	0.070	0.047	0.116	0.085
Ever told had arthritis	0.251	0.290	0.213	0.163	0.081	0.351	0.177
Had pain in or around a joint	0.383	0.417	0.310	0.316	0.237	0.525	0.466
Ever told blood cholesterol high	0.310	0.323	0.245	0.280	0.348	0.483	0.266

Source: Calculated from BRFSS 1999-2004

Table A3. Descriptive Statistics by Regions

Variable	Region*									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Reported age in years	43.38	47.04	46.25	47.33	49.49	44.19	47.50	45.03	46.18	45.64
Gender (male=1)	0.482	0.490	0.497	0.465	0.458	0.528	0.501	0.521	0.497	0.455
Marital status	0.432	0.620	0.621	0.576	0.556	0.573	0.591	0.552	0.565	0.569
Education:										
Grade 8 or less	0.072	0.025	0.026	0.022	0.019	0.000	0.017	0.012	0.015	0.010
Grades 9 - 11 (Some high school)	0.099	0.046	0.059	0.052	0.080	0.076	0.042	0.081	0.048	0.065
Grade 12 or GED (High school graduate)	0.269	0.281	0.247	0.270	0.320	0.291	0.304	0.297	0.346	0.267
College 1 year to 3 years	0.226	0.254	0.252	0.289	0.324	0.384	0.276	0.325	0.291	0.278
College 4 years or more	0.335	0.395	0.416	0.368	0.257	0.249	0.360	0.286	0.300	0.380
Employment:										
Employed for wages	0.536	0.550	0.576	0.566	0.530	0.647	0.561	0.567	0.580	0.587
Self-employed	0.090	0.093	0.096	0.063	0.071	0.031	0.044	0.047	0.056	0.062
Out of work for more than 1 year	0.028	0.011	0.014	0.014	0.017	0.005	0.021	0.010	0.017	0.021
Out of work for less than 1 year	0.046	0.023	0.023	0.037	0.040	0.014	0.031	0.041	0.041	0.034
A homemaker	0.063	0.074	0.070	0.068	0.035	0.014	0.059	0.057	0.063	0.072
A student	0.058	0.039	0.036	0.041	0.035	0.044	0.044	0.051	0.032	0.040
Retired	0.126	0.185	0.159	0.188	0.212	0.133	0.201	0.180	0.171	0.157
Unable to work	0.054	0.026	0.026	0.023	0.062	0.113	0.040	0.048	0.042	0.027
Annual Household Income ($1,000)	49.33	71.85	69.57	60.37	50.90	46.86	59.48	47.44	51.27	58.13
Have health plan	0.795	0.902	0.887	0.911	0.849	0.837	0.900	0.854	0.921	0.906
Could not afford to see doctor	0.143	0.095	0.106	0.070	0.096	0.114	0.084	0.135	0.089	0.092
Smoking	0.201	0.211	0.192	0.232	0.229	0.283	0.233	0.275	0.259	0.223
Heavy drinking	0.147	0.163	0.144	0.196	0.178	0.231	0.179	0.225	0.195	0.193
Self-assessed health status:										
Excellent	0.197	0.257	0.255	0.250	0.267	0.173	0.224	0.219	0.211	0.231
Very good	0.288	0.359	0.350	0.371	0.273	0.392	0.367	0.358	0.351	0.379
Good	0.326	0.261	0.275	0.258	0.288	0.307	0.290	0.280	0.316	0.292
Fair	0.144	0.100	0.086	0.091	0.144	0.071	0.083	0.097	0.098	0.076
Poor	0.044	0.023	0.034	0.031	0.028	0.057	0.036	0.046	0.024	0.021
Number of days physical health not good	3.428	3.081	2.895	3.299	4.264	3.244	3.414	3.692	3.617	3.104
Number of days mental health not good	3.566	2.900	3.228	2.874	3.391	1.303	2.822	4.019	3.259	3.011
Activities limited due to health problem	0.147	0.170	0.149	0.149	0.207	0.184	0.188	0.178	0.180	0.166
Body mass index (BMI)	26.37	26.29	26.35	26.45	26.46	27.90	26.84	26.54	26.83	26.54
Fruit and vegetable servings per day	4.004	4.058	4.110	4.037	3.822	4.204	4.159	3.765	4.075	4.058
Participate in any exercises	0.706	0.738	0.768	0.786	0.786	0.804	0.794	0.798	0.786	0.772
Ever had asthma	0.120	0.110	0.108	0.132	0.121	0.174	0.110	0.098	0.114	0.138
Ever told blood pressure high	0.222	0.242	0.253	0.283	0.335	0.320	0.261	0.275	0.254	0.244
Ever told had coronary heart disease	0.042	0.045	0.037	0.044	0.046	0.024	0.045	0.048	0.034	0.043
Ever told had myocardial infarction	0.031	0.033	0.032	0.024	0.045	0.009	0.053	0.038	0.033	0.028
Ever told had stroke	0.018	0.016	0.015	0.034	0.034	0.035	0.021	0.044	0.020	0.022
Ever told had diabetes	0.071	0.055	0.066	0.078	0.074	0.141	0.056	0.071	0.058	0.068
Ever told had arthritis	0.202	0.243	0.230	0.283	0.332	0.347	0.332	0.281	0.303	0.292
Had pain in or around a joint	0.342	0.371	0.365	0.366	0.382.		0.437	0.384	0.467	0.466
Ever told blood cholesterol high	0.284	0.341	0.299	0.310	0.274	0.334	0.335	0.255	0.329	0.289

Source: Calculated from BRFSS 1999-2004

Note:* 1 is New York City; 2 is Long Island; 3 is Hudson Valley; 4 is Capital Region; 5 is Mohawk; 6 is North Country; 7 is Central New York; 8 is Southern Tier; 9 is Western New York; and 10 is Finger Lakes.

Figure A1. Quality of Health
Based on Percentage of 'Very good' and 'Excellent' Health

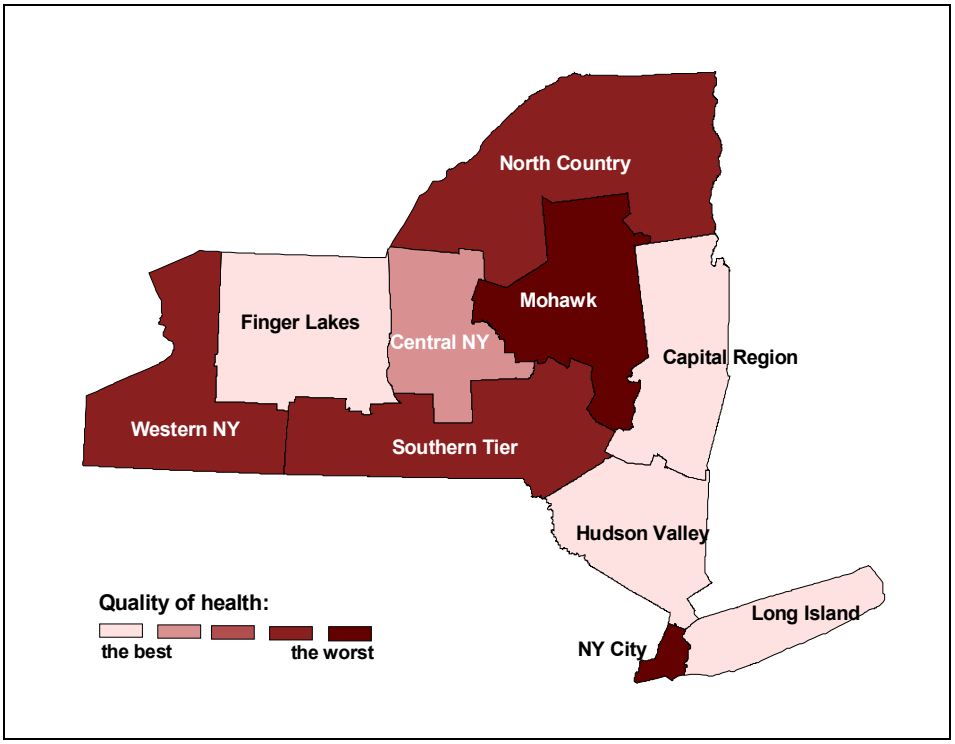
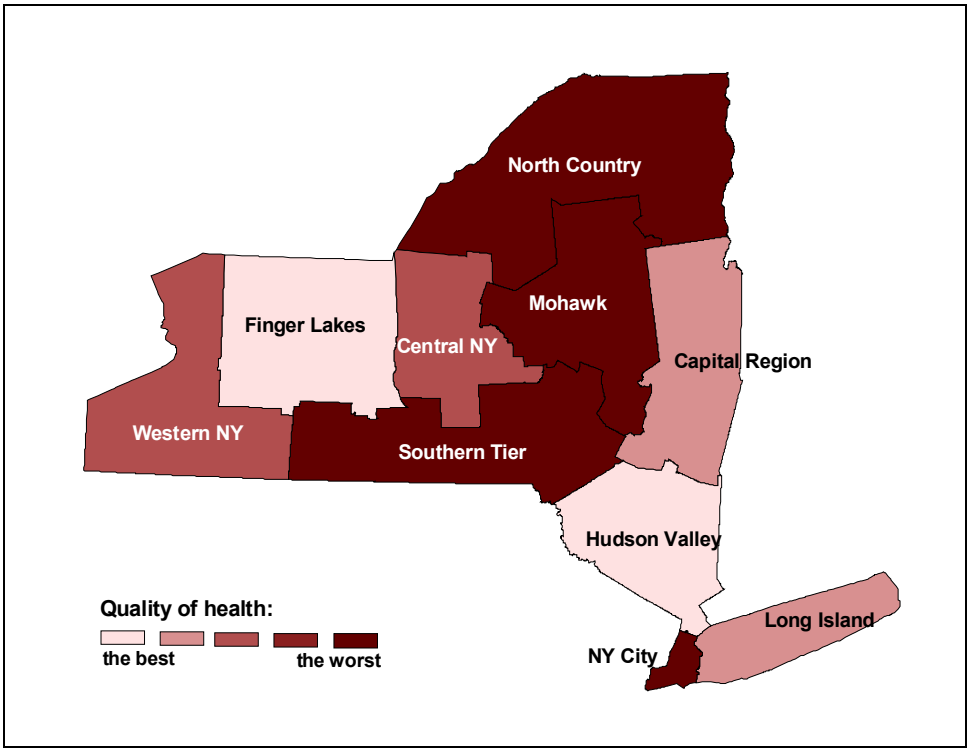


Figure A2. Quality of Health
Based on Percentage of 'Fair' and 'Poor' Health



References

Arias, E., 2004. United State life tables, 2002. National Vital Statistics Reports 53. National Center for Health Statistics, Hyattsville, Maryland.

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. NBER working paper 12674. Cambridge, MA.

Centers for Disease Control and Prevention (CDC). *Behavioral Risk Factor Surveillance System Survey Data*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 1999-2004.

Clarke, P.M., Ryan, C., 2006. Self-reported health: reliability and consequences for health inequality measurement. Health Economics 15, 645-352.

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.

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. J. 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.

Groot, W., 2000. Adaptation and scale of reference bias in self-assessments of quality of life. J. 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.

Hu, J., Lahiri, K., Vaughan, D.R., Wixon, B., 2001. A structural model of social security's disability determination process. The Review of Economics Statistics 83, 348-361.

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. J. of Health and Social Behavior 38, 21-37.

James, W., Nelson, M., 1997. The contribution of nutrition to inequalities in health, British Medical J., 314, 1545-1550.

Kakwani, N., Wagstaff, A., van Doorslaer, E., 1997. Socioeconomic inequalities in health: measurement, computation, and statistical inference. J. of Econometrics 77, 59-103.

Keppel, K.G., Pearcy, J.N., Klein, R.J., 2004. Measuring progress in healthy people 2010. Statistical Notes 25. National Center for Health Statistics, Hyattsville, Maryland.

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.

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. J. 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. J. 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. National Center for Health Statistics, Hyattsville, Maryland.

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. J. of American Medical Association 282, 1523-1529.

Rubin, D.B., 1987. Multiple imputation for nonresponse in surveys. John Wiley & Sons, New York.

Safaei, J., 2006. Income and health inequality across Canadian provinces. Health & Place, doi:10.1016/j.healthplace.2006.09.003.

Schafer, J.L., 1997. Analysis of incomplete multivariate data. Chapman and Hall, New York.

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. J. of American Medical Association 295, 784-792.

Sickles, R.C., Taubman, P., 1997. Mortality and morbidity among adults and elderly. In: Rozenzweig, M.R., Stark O. (Eds), Handbook of Population and Family Economics. North-Holland, Amsterdam.

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. Department of Health and Human Services, 2000. Healthy People 2010: Understanding and Improving Health. 2nd ed. Washington, DC: U.S. Government Printing Office, November 2000.

Wagstaff, A., VanDoorslaer, E., Watanabe N., 2003. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. J. of Econometrics 112, 207-223.

Wagstaff, A., VanDoorslaer, E., 2004. Overall versus socioeconomic health inequality: a measurement framework and two empirical illustrations. Health Economics 13, 297-301.

Wagstaff, A., Paci, P., van Doorslaer, E., 1991. On the measurement of inequalities in health. Social Science Medicine 33, 545-557.

Williams, D.R., Collins, C., 1995. US socioeconomic and racial differences in health: patterns and explanations. Annual Review of Sociology 21, 349-386.

Van Doorslaer, E., Jones, A.M., 2003. Inequalities in self-reported health: validation of a new approach to measurement. J. of Health Economics 22, 61-87.