Health Kuznets Curve (HKC) Considering the Two Aspects of Health: Theoretical and Empirical Investigation

Draft on Nov 20th , 2017

Newest version can be reached here

Qingqing Yang
School of Economic Sciences
Washington State University
Pullman, WA 99163
qingqing.yang@wsu.edu

Robert Rosenman
School of Economic Sciences
Washington State University
Pullman, WA 99163
yamaka@wsu.edu

Abstract

The objective of this article is to show how the Kuznets Curve, which demonstrates the relationship between economic development and income inequality, can also be used to show the relationship between income and health (what we refer to as the "Health Kuznets Curve"). Previous studies have shown that as income increases health improves in both individual and aggregate. Our study introduces a more complex picture by showing that this positive correlation reverses when inequality increases. We first investigate the relationship between income and health theoretically using an optimal control dynamic model. Next, we use data from the National Health and Nutrition Examination Survey (NHANES) to test the model empirically. Based on the relationship of health and income at the individual level, we establish the relationship of income and health at the aggregate level, referring to these as "economic development" and "social health." We simulate economic development and social health with various sets of income levels and income distributions. The simulation result shows that a HKC exist in societies with all levels of equality. Health increases with economic development initially, and then decreases. However, during the economic development process, the health turning point comes earlier in an unequal society.

Keywords: Health Kuznets Curve, Health score, BMI, Economic development, Social health.

JEL code: I12, I14, I15

1. Introduction

It is generally accepted that health increases with income. Hence, we would expect the overall health of a population increases as the population's wealth goes up, and countries with higher national income generally have better health among its populous. However, a rising national income does not necessarily imply that all will enjoy better health. The relationship between income and health has personal aspects, so the distribution of income can have an important impact on how a rising national income affects a nation's health. In fact, little is known about how income inequality affects the relationship of health and income.

The objective of this research is to theoretically and empirically model a Health Kuznets Curve (HKC), the relationship between economic development and aggregate health in an economy, within the construct of income distribution. Kuznets first demonstrated that economic inequality will first increase and then decrease as an economy develops (Kuznets, 1955). His ideas have been extended to what is known as the Environment Kuznets Curve (EKC) (Lopez, 1994; Grossman and Krueger,1995; Dinda, 2005) which shows environmental degradation follows a path similar to income inequality, first increasing then decreasing as economic development proceeds. Unlike the EKC, empirical explorations of the HKC have had ambiguous results. To help explain the ambiguity, we first investigate the relationship between individual health and income level and then use those results to simulate the relationship of income distribution and health.

While there is ample empirical support that higher income improves health (Feinstein, 1973; Kitagawa & Hauser, 1973; Silver, 1972; Ettner, 1996), the role of income distribution (Kawachi & Kennedy, 1999) is less settled. Wilkinson and Pickett (2006) reviewed 168 analyses in 155 papers using regional or country level data about income distribution and health. and concluded that only about half of studies (87 out of 168 analyses) suggest that more egalitarian societies do have better health and longevity (Lynch, Smith, & Harper, 2004; Subramanian & Kawachi, 2004). In the rest of the studies, the results are either not significant or just partially supportive (Deaton, 2003; Deaton & Lubotsky, 2003; Beckfield, 2004). One possible explanation for the inconsistent empirical findings is that many studies measured inequality in areas too small to reflect the scale of income differences in a society (Wilkinson & Pickett, 2006). The conjecture in this paper is that both average income and income distribution affect health, but perhaps differently.

In this paper, we model the aggregate relationship between income and health starting with how health relates to individual income. We then formulate the aggregate relationship of health and income based on the individual relationship and simulate the relationship of income, income inequality and health with various sets of income levels and income distributions.

We model the HKC theoretically with an optimal control dynamic model. Our theoretical approach builds on models of the Environment Kuznets Curve (EKC) (Lopez, 1994; Grossman and Krueger,1995; Dinda, 2005). We extrapolate from those by assuming health is affected by two kinds of consumption, healthy and unhealthy consumption. The type of consumption at the individual level is affected and restricted by income level. Some types of consumption are detrimental to health in all levels, like smoking and drug use; other types of consumption, such as sugar or fat, can be bad for health when they are consumed in excess or not at all (Bolin & Lindgren, 2014). At very low income levels, people often suffer from bad health because of starvation, malnutrition, bad sanitation condition, and poor medical care. This situation gets better when the economy grows. However, higher levels of wealth may lead to an unhealthy lifestyle, resulting in various chronic diseases like obesity, cancers, diabetes and mental diseases, making the relationship between health and economy an inversed U-shape. This relationship derived from individual level health effects of income is verified by empirical analysis with the National Health and Nutrition Examination Survey (NHANES) data. Building the aggregate HKC from individuals' behavior allows us to include income distribution as a factor explaining the HKC. Increasing overall income in an economy does not necessary increase the overall health if the income distribution is heavily skewed.

Besides providing a theoretical foundation for how income distribution enters into the HKC, we offer two empirical extensions. Two common measures of health used in assessing the HKC are BMI and self-assessed health (SAH). BMI is frequently chosen due to its "simplicity and comparability," while SAH is used because it encompasses "total health." But how health is measured often gives conflicting results. For example, Costa-Font (2013) found that health inequality increased with income inequality when measuring health with SAH, but not with BMI. Each measure has its shortcomings. BMI is not considered a relevant indicator of health equality in the public health field, while SAH may be biased by relative income and other socioeconomic variables (Yang and Rosenman, 2016). To overcome these problems, we simulate the aggregate relationship based on the individual correlation between health and income for different income distributions. This method avoids the risk of inconsistent results due to different control variables and more clearly demarks the causal mechanism in the relationship between economic conditions and a society's aggregate health.

The next section introduces the theoretical model relating income and health at the individual level. The implications of the theory are explored in the third section, and parameters are estimated in the fourth section using NHANES data.In the fifth section we simulate the aggregate results based on the empirical model from section 4. Conclusion and discussion finish the paper.

2. Model

Following the basic structure of the Environment Kuznets Curve (EKC) (Lopez, 1994; Grossman and Krueger,1995; Dinda, 2005), where the environment is treated as a state variable, we treat health the same way. Individuals derive utility directly from both healthy and unhealthy consumption, and from being healthy. Our relationship between consumption and health is motivated by Grossman's (1972) seminal works on demand of health and subsequent work that focused on the relationship between health and health investment (Forster, 1989; Forster, 2001; Eisenring, 1999). We follow Forster (2001) by allowing two different types of consumption which have different and opposite impacts on health.

A representative individual maximizes his lifetime utility W by choosing the two types of consumption and health, hence

$$W = \int_0^\infty e^{-\rho t} \left[U(C_1(t), C_2(t), H(t)) \right] dt \tag{1}$$

$C_1(t)$ is consumption with generally healthy consequences, and $C_2(t)$ is consumption with generally unhealthy consequences. We term the first "healthy consumption" and the second "unhealthy consumption"[1]. $H(t)$ is health status. Consumption has direct utility, but also affects utility indirectly through health. The utility function is concave on arguments, so $U_{C_1} > 0$, $U_{C_2} > 0$, $U_{C_1C_1} < 0$, $U_{C_2C_2} < 0$, $U_{C_1C_2} > 0$, $U_H > 0$, $U_{HC_1} > 0$, $U_{HC_2} > 0$, $U_{HH} < 0$.

[1] Healthy consumption, or normal consumption, including appropriate food and nutrition, positively affects health, or correspondingly, decreases ill health. Unhealthy consumption, which might include things like smoking and overeating, decreases health by generating illnesses.

Let $S(t)$ be factors that denigrate health. Then the two types of consumption contribute to it by

$$S(t) = \frac{b}{C_1(t)} + aC_2(t) \tag{2}$$

and $S(t)$ affects the path of health by

$$\dot{H}(t) = \varphi(m(t)) - S(t) - \delta_H H(t) \tag{3}$$

where $m(t)$ is health investment[2], $\varphi(m(t))$ is how that investment translates into health, $S(t)$ captures the health consequences of the two types of consumption and δ_H is the natural depreciation of health.

Utility is maximized subject to an income constraint

$$Y = C_1(t) + pC_2(t) + p_m m(t) \tag{4}$$

where the price of C_1 is the numeraire, p and p_m are the relative price of $C_2(t)$ and $m(t)$ respectively. There is no saving or borrowing; at each point in time point income is allocated to normal consumption $C_1(t)$, unhealthy consumption $C_2(t)$, health investment $m(t)$.

[2] Our idea of health investment is quite general, including medical investment, equipment or facility to improve health, disease prevention and many other unlisted investments.

3. Income and Consumption

The current value Hamiltonian for the above model can be written as

$$\begin{aligned}\Psi &= U(C_1, C_2, H) + \lambda \dot{H} \\ &= U(C_1, C_2, H) + \lambda\left[\varphi(m(t)) - S(t) - \delta_H H(t)\right]\end{aligned} \tag{5}$$

Where $m(t) = \frac{1}{p_m}(Y - C_1(t) - pC_2(t))$ and $\lambda(t)$ is the co-state variable (the shadow price of $H(t)$) corresponding to state variables. Substituting for $m(t)$, the necessary conditions for an optimal solution are:

$$\frac{\partial \Psi}{\partial C_1} = U_{C_1} - \frac{1}{p_m}\lambda\varphi' + \lambda\frac{b}{C_1^2} = 0 \tag{6}$$

$$\frac{\partial \Psi}{\partial C_2} = U_{C_2} - \frac{p}{p_m}\lambda\varphi' - \lambda a = 0 \tag{7}$$

$$\dot{\lambda} = -\frac{\partial \Psi}{\partial H} + \rho\lambda = -U_H + \lambda(\delta_H + \rho) \tag{8}$$

From equation(6) and equation(7), it is easy to see

$$\lambda = \frac{U_{C_1}}{\frac{1}{p_m}\varphi' - \frac{b}{C_1^2}} = \frac{U_{C_2}}{\frac{p}{p_m}\varphi' + a} \tag{9}$$

Taking derivatives of (6) and (7) with respect to t, we have

$$U_{C_1C_1}\dot{C}_1 + U_{C_1C_2}\dot{C}_2 + U_{C_1H}\dot{H} - \frac{1}{p_m}\dot{\lambda}\varphi' - \frac{1}{p_m}\lambda\varphi''\cdot\left(-\frac{1}{p_m}\dot{C}_1 - \frac{p}{p_m}\dot{C}_2\right) + \dot{\lambda}\cdot\frac{b}{C_1^2} - \lambda\frac{2b}{C_1^3}\cdot\dot{C}_1 = 0 \tag{10}$$

$$U_{C_2C_1}\dot{C}_1 + U_{C_2C_2}\dot{C}_2 + U_{C_2H}\dot{H} - \frac{p}{p_m}\dot{\lambda}\varphi' - \frac{p}{p_m}\lambda\varphi''\cdot(-\frac{1}{p_m}\dot{C}_1 - \frac{p}{p_m}\dot{C}_2) + \dot{\lambda}a = 0 \qquad (11)$$

Combining these with equations (8) and (9) gives

$$\dot{H} = \frac{(U_{C_2C_2} - pU_{C_1C_2})\dot{C}_2 + (pU_{C_1C_1} - U_{C_2C_1} - \lambda p\frac{2b}{C_1^3})\dot{C}_1 + \left[-U_H + \lambda(\delta_H + \rho)\right](\frac{pb - aC_1^2}{C_1^2})}{U_{C_2H} - pU_{C_1H}} \qquad (12)$$

Now assume income $Y(t)$ changes every year. From equation (4), we have $\dot{Y} = \dot{C}_1 + p\dot{C}_2 + p_m\dot{m}$.

With simple reformulation,

$$\dot{C}_1 = \dot{Y} - p\dot{C}_2 - p_m\dot{m} \qquad (13)$$

Substituting (13) into (12) gives the relationship between $\dot{H}$ and $\dot{Y}$;

$$\dot{H} = \frac{(pU_{C_1C_1} - U_{C_2C_1} - \lambda p\frac{2b}{C_1^3})}{U_{C_2H} - pU_{C_1H}}\dot{Y} + \frac{(U_{C_2C_2} - p^2U_{C_1C_1} + \lambda p^2\frac{2b}{C_1^3})\dot{C}_2 - p_m(pU_{C_1C_1} - U_{C_2C_1} - \lambda p\frac{2b}{C_1^3})\dot{m} + \left[-U_H + \lambda(\delta_H + \rho)\right](\frac{pb - aC_1^2}{C_1^2})}{U_{C_2H} - pU_{C_1H}} \qquad (14)$$

Where λ is the shadow price of health and is defined in (9).

From (14), we can see the relationship of $\dot{H}$ and $\dot{Y}$ is decided by $pU_{C_1C_1} - U_{C_2C_1} - \lambda p\frac{2b}{C_1^3} < 0$ and $U_{C_2H} - pU_{C_1H}$, which has an ambiguous sign. The marginal utility of both normal consumption and unhealthy consumption increases when people become healthier, i.e., U_{C_1H} and U_{C_2H} are both positive. However, the sign of $U_{C_2H} - pU_{C_1H}$ is ambiguous.

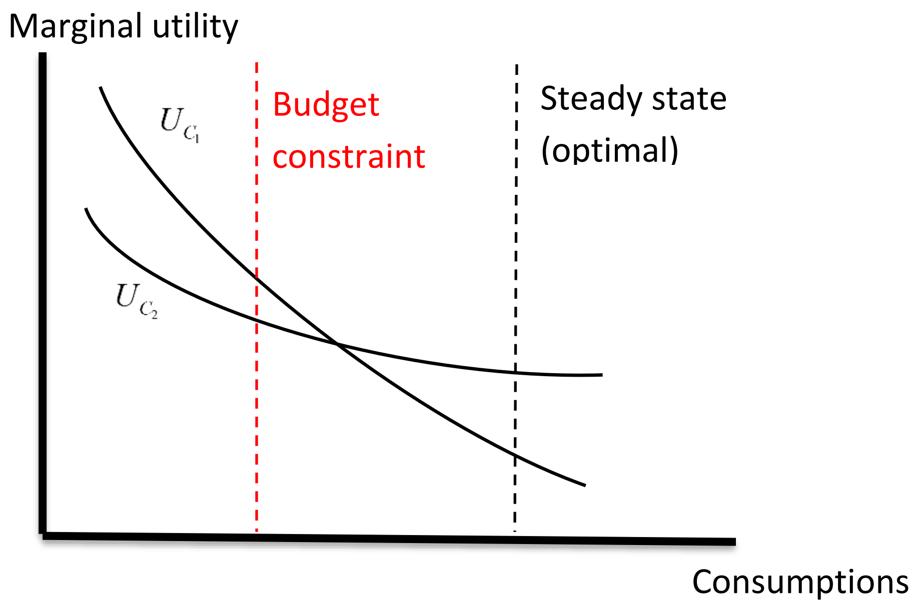

Figure 3-1. Marginal utility under Budget constraint and steady state

We note that equation (9) can be written as

$$\frac{U_{C_1}}{U_{C_2}} = \frac{\frac{1}{p_m}\varphi' - \frac{b}{C_1^2}}{\frac{p}{p_m}\varphi' + a} \tag{15}$$

The equation (15) shows that at the steady state $U_{C_1} < U_{C_2}$. Hence, to satisfy the optimal condition, from (15), consumption skewed towards unhealthy consumption. However, whether the steady state can be reached depends on the budget constraint, i.e. the income level. We add one additional intuitive assumption – at very low levels of consumption $U_{C_1} > U_{C_2}$. As shown in figure 3-1, at low incomes, $U_{C_1} > U_{C_2}$, consumption skewed towards healthy consumption. At higher levels of consumption, income is never a constrain, steady state is reached, $U_{C_1} < U_{C_2}$. Consumption shifts more towards unhealthy consumption, and we have the following possibilities:

1. When income is low, spending on both goods is low. We assume at this low level of consumption, people prefer to satisfy their basic need, healthy consumption generate more marginal utility of health, i.e., $\frac{U_{C_2H}}{p} < U_{C_1H}$ and we see a positive relationship between income and health.

2. When income level is high, both types of consumption increase, income is no longer an constraint, holding to (15) , eventually $\frac{U_{C_2H}}{p} > U_{C_1H}$; In this case, health status decreases while income increases.

In effect, we have an individual level Health Kuznets Curve (HKC) – as income grows, consumption skews towards unhealthy consumption. Higher levels of income results in less health. We explore this idea empirically in the next section.[3]

[3] A point of note is the role of *p*, the price of unhealthy consumption. It positions the turning point of consumption and health. If *p* is high, unhealthy consumption is relatively expensive. At any income level, less of it is done, and so a greater income is needed for the health impact of unhealthy consumption to dominate the health impact of healthy consumption. But if *p* is low, at any income level there is more unhealthy consumption so it's (unhealthy) health impacts dominate sooner. Increasing the relative price of unhealthy consumption is beneficial to the overall health of the society. This explains why unhealthy consumption d may dominate among the poor in developed countries, where unhealthy process foods are relatively cheap.

4. Empirical Analysis

4.1 Data

We used data from the National Health and Nutrition Examination Survey (NHANES). Although the NHANES program began in the early 1960s, a continuous program started from 1999, and has a changing focus on a variety of health and nutrition measurements. The continuous NHANES data set is released every two years. From 1999 to 2014, we have 8 waves of data (1999-2000, 2001-2002, 2003-2004, 2005-2006, 2007-2008, 2009-2010, 2011-2012, and 2013-2014). The survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in counties across the country. We choose the NHANES data to verify our hypothesis for two reasons. First, the wide range of income levels reflected in United States allows us to associate health status at different incomes. Second, the NHANES data includes very comprehensive information about human health, including both self-reported data and data collected from a physical examination.

4.2 Model

In section 3, we made some hypotheses about the non-linear relationship of health (H_i) and income (Y_i). We use the model with income and income square Y_i^2 (the income data is converted into 2014-dollar value with CPI information). If $\beta_1 > 0$, $\beta_2 < 0$, then we could say hypotheses 1) and 2) are verified.

$$H_i = \beta_0 + \beta_1 Y_i + \beta_2 Y_i^2 + \beta_3 Edu_i + \beta' X_i + \mu_i \tag{16}$$

4.3 Dependent variable

In this model, one of the key problems is to find an appropriate measure of health (H_i). Over all, we use two measurements of health, comprehensive health score and BMI. A comprehensive health score is built as the equation (17) based on a patented of health index scoring method (Patent CN104850758 A). This method we used to calculate the comprehensive health score is with a weighted average method that incorporates a set of health-related components, including blood pressure, Body Mass Index (BMI), diabetes, cancer and other diseases as basic items. For every basic health-related component, different algorithms are used to achieve a scores that are comparable and therefore able to be aggregated (that we call h_i , in 0-100 scale)[4]. See the appendix for algorithm details of the method used to build h_i . w_i is a set of weights. Each h_i is then weighted with w_i .

$$Health\_score = \frac{\sum_{i=1}^{n} w_i * h_i}{\sum_{i=1}^{n} w_i} \tag{17}$$

For a comparison, another measure of health (H_i) is the Body Mass Index (BMI). Since the BMI only reflects the body shape of the person, we anticipate these two sets of regression will have different results. BMI is an indicator of obesity level, and is merely one aspect of health. Although obesity cause many health issues, a lot of other severe health problem may not relate to BMI.

The equation (16) can be estimated by the ordinary least squares (OLS), but we must notice the potential endogeneity problem with education, and health. Education level affect the individual's health in the model. Health may also have influence on one's education levels. To solve this

[4] We slightly change some algorithms (e. g., algorithm of BMI score) to fit the data for the United States.

problem, we build a two-step least squares (2SLS) model to deal with these endogeneity issues. As per the rule of an instrument variable, we need to find the variable (or variables) that highly related to the person's education, but is not affected by his health. Considering the reasonability and data feasibility, the education levels of the household representative and his (or her) spouse are used as instrumental variables.

Table 1. Summary of the variables

Variable	Label	Mean	Std	Minimum	Maximum
Health_score	Health index	80.75	11.53	22.84	94.41
BMI	Body Mass Index (kg/m^2)	28.37	6.30	14.42	81.25
Work	whether have a job or business	0.6	0.49	0	1
act_vigor	Doing Vigorous work or recreation activity	0.39	0.49	0	1
act_moderate	Doing moderate work or recreation activity	0.57	0.49	0	1
bike_home	Walked, bicycled or Finish Tasks around home/yard	0.45	0.5	0	1
smoke100	Smoked at least 100 cigarettes in life	0.4	0.49	0	1
male	Gender	0.53	0.5	0	1
age	age	46.81	18.88	17	85
Hispanic	Race/Ethnicity	0.27	0.44	0	1
black	Race/Ethnicity	0.14	0.35	0	1
other_race	Race/Ethnicity	0.08	0.28	0	1
citizen	Citizenship Status	0.86	0.35	0	1
educyear	Education level in year	13.89	3.75	8	19
HR_educ	education level of Household representative	14.16	3.86	8	19
HS_educ	education level of Household representative's spouse	14.08	3.82	8	19
separated	divorce or separated	0.02	0.14	0	1
single	Never married	0.1	0.3	0	1
rhhinc	Real household income	83.89	64.68	1.08	203.76
rhhinc_ave	Average real household income per household member	27.52	24.65	0.15	101.88

All the other variables affecting individual health status are included in the vector X_{it}. Variables hypothesized to affect health outcomes are sex, age, race, ethnicity, marital status, behavior like smoking and activity, and education. Table 1 shows the summary of the variables in the model. “Health_score” is a continuous variable ranging from 0 to 100. Among the observations, the healthiest person is scored 94.41, while the unhealthiest person only got a score 22.84. BMI ranges from 14.42 to 81.25. And the average BMI is 28.37. In the US, those who has a BMI of more than 30 can be treated as obese. About 40% of the observations smoked at least 100 cigarettes in their life. The household income data in the survey is in a range form. In this model, we convert the range of the household income into a mean household income in a specific household income range based in the real income distribution in the US. The variable “rhhinc” is got by eliminating the inflation factor. By dividing the family size, we got the average real household income per household member in the model, denoted as “rhhinc_ave”. The mean of “rhhinc_ave” is about $27.52 thousand with a standard deviation of $24.65 thousand.

4.4 Results

Table 2 shows the results of the regressions. The education levels of the household representative and his (or her) spouse are used as instrumental variables considering the potential endogenous issue of health on education. In the first step relation of 2SLS model, the instruments are jointly significant and positive. It is reasonable that the family member’s education level is mutually affected.

Table 2. Regression results

	BMI		Health_score	
variable	OLS	2SLS	OLS	2SLS
Intercept	25.48135***	28.80790***	90.54303***	89.82216***
rhhinc_ave	-0.02951***	-0.02744***	0.03597***	0.05110***
rhhinc_ave2	0.00012***	0.00018***	-0.00018***	-0.00044***
Work	0.82448***	0.67448***	2.46301***	2.65015***
act_vigor	-1.19507***	-0.97351***	0.91436***	1.20222***
act_moderate	-0.04089	-0.11548	0.13569	0.04416
bike_home	-0.58990***	-0.60488***	1.17693***	1.26650***
smoke100	0.12784*	-0.03782	-0.95422***	-1.03257***
male	-0.66452***	-0.13247	0.97573***	0.53595***
age	0.03217***	0.01214***	-0.24670***	-0.25653***
Hispanic	1.05285***	0.48659***	1.14044***	1.60980***
black	1.79167***	1.55990***	0.22815*	0.15515
other_race	-1.84456***	-2.24592***	2.81482***	3.59580***
citizen	1.46599***	1.68807***	-2.71241***	-2.87621***
educyear	0.07219***	-0.10206***	0.02046	0.10385***
separated	-0.30106***	-0.77119**	-0.50982***	0.15723
single	-0.80568***	-2.19669***	-0.42700***	-1.16752***
Turning point		75.9814		57.7563

* $p < .1$, ** $p < .05$, *** $p < .01$

The first two columns in table 2 show the OLS and 2SLS regression result with the BMI as health measurement. First, assume all the variables are exogenous. Almost all the variables in OLS regression are significant. The coefficient of the education year (denoted as “educyear”) is surprisingly positive. The 2SLS shows a different significant coefficient of education year. The instrumental variable corrects the potential bias of education variable. Although the sign of the income variables does not change, the magnitude of the effect of income differs. The negative coefficient of income and positive coefficient of square of income imply a plausible U-shape relationship between income and BMI. But the turning point is around $75,981.40, which is higher than the 95% percentile of income range. Thus, in the normal range of income, BMI decreases when income increases.

The last two columns in table 2 are the regression results with health score. The coefficient of income is positive, while the coefficient of income square is negative. This means that the relationship between income and health is complex. The opposite signs of the income and income square terms indicate that the relation will have a turning point. Take first order derivative and set it equal to 0, we get the turning point will be at income level $57,756.30. When income is less than $57,756.30, health level increases when income increases, but the magnitude effect decreasing. When the income is higher than $57,756.30, an increase in income will not increase one's health. In the contrast, comparatively high income decreases people's health. According to the distribution of household income per person in the US, the turning point is about one standard deviation away from the mean (mean $29,818 and standard deviation $23,146).

Some of the variable have a consistent effect no matter which health measurement, BMI or health score, is used, while others don't. Doing Vigorous activity, and walking or bicycling regularly reduce BMI by approximately 1.2 and 0.6. These activities help people stay slim, and increase health score. The Hispanic on average has a little bit higher BMI than the white group, but they have higher health score. The other races (other than black and Hispanic) has lower BMI and is healthier than the base group (the white). And the citizen has on average a higher BMI and a lower health score. One more year of education decreases the BMI by 0.1, and increases his health score 0.1. although the marginal change is small, but if one had been in school more than 10 years, the changes becomes obvious and attractive. BMI only represent one aspect of health, other elements affected the individual's health status, e.g. blood pressure, blood sugar, and other diseases, may exceed the effects of obesity.

5. Simulate the income inequality and health in aggregate level

In the last section, we study the individual relationship of income and health. Using a comprehensive health score as a health measurement, we find an inverse U-shape relationship between income and health. Considering that health score is more comprehensive than BMI as a health measurement, the health score regression results will be the base of the simulation in this section.

In this section, we want to simulate the aggregate relationship of income inequality and health. Two issues need to be solved when simulate aggregate relationship from individual relationship. 1) Make an assumption of income distribution; 2) How to measure the inequality of income?

5.1. Income distribution

In the simulation, we assume income is a normal distribution. However, Normal distribution has a problem that it allows negative value for income. But income is non-negative. The following figure 5-1 shows the distribution of household income per person in the US. This is a shape of truncated normal distribution or Gamma distribution.

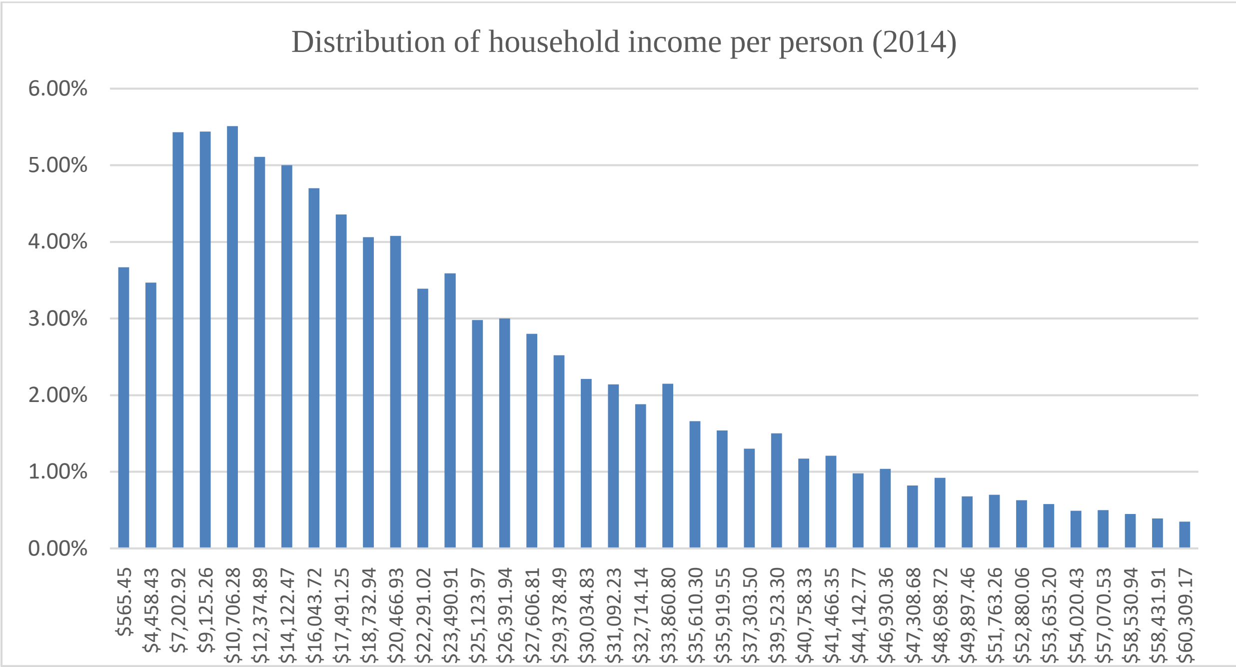

Figure 5-1. Income distribution in the US (2014 US Census data)

The real data shows that the household income has a mean of $75,738. The mean size of household is 2.54. On average, the household income per person is about $29,818.11, with the standard deviation of $23,146. The simulate income set should be close to the real situation. Thus, we pick the mean set center on $30,000, and the standard deviation set center on $23,000. To be specifically,

$$\mu_i \in (15000, 75000)$$
$$\sigma_j \in (13000, 35000)$$

5.2. Income inequality

The Gini coefficient is mostly commonly used measure of in equality. The Gini coefficient is defined mathematically based on the Lorenz curve. Lorenz curve plots the proportion of the total income of the population (y axis) that is cumulatively earned by the bottom x% of the population (Gini, 1912; 1921). A Gini coefficient of zero expresses perfect equality, where all values are the same (for example, where everyone has the same income). A Gini coefficient of 1 (or 100%)

expresses maximal inequality among values (e.g., for a large number of people, where only one person has all the income or consumption, and all others have none, the Gini coefficient will be very nearly one).

Many researches have provided alternative method to calculate the Gini Coefficient. In this paper, we use the R package “IC2” to calculate it. This package calculates the Gini coefficient follow the method in Cowell F.A. (2000) and Schechtman E., Yitzhaki S. (2008).

In the following part, we will do a simulation with an algorithm using a truncated normal distribution of income.

5.3. Algorithm for Monte Carlo simulation

The simulation is conducted as the following 6 steps:

1. Generate a sequence of mean μ_i , $i = 1, 2, ..., I$;
2. Generate a sequence of Standard Deviation σ_j , $j = 1, 2, ..., J$;
3. For each μ_i , generate J sets income with truncated normal distribution $TruncNorm(\mu_i, \sigma_j)$ or Gamma distribution $Gamma(k = \frac{\mu_i^2}{\sigma_j}, \theta = \frac{\sigma_j}{\mu_i})$;
4. Estimate the health level with the income generated in step 3;
5. Calculate the Gini coefficient for each scenario of income distribution;
6. Calculate the mean of the health for every scenario of income distribution.

With the above 6 steps, the output is $I*J$ sets of the income data, $I*J$ sets of the health score data, a set of Gini coefficient with $I*J$ elements, and a set of average health score with $I*J$ element.

5.4. Simulation results

The inequality level is controlled by the standard deviation of the income distribution. The following contour plot (figure 5-2) show the Gini Coefficient change as the mean and variance changes. Every line in the plot indicate the combination of mean and variance with same Gini coefficient level. Imagine drawing a horizontal line in the contour plot (Keep mean as a constant), the Gini coefficient increases as the standard deviation increases.

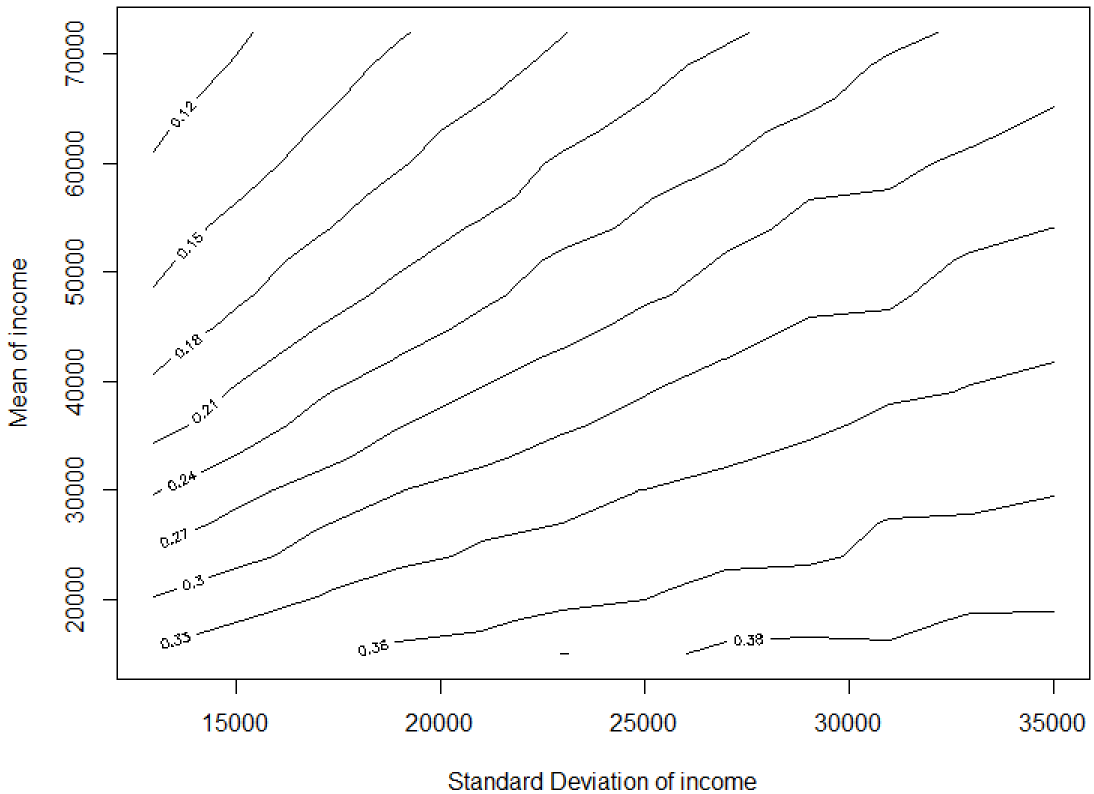

Figure 5-2. Contour plot of Gini Coefficient of income

In the figure 5-3, we can see that, when we assume a population with truncated normal distribution, the trend of average health is not monotonically changed with mean and variance. When income is in a relative low level, the average health increases when the standard deviation increases; when the income is in a relatively high level, the average health increases when the standard deviation decreases. The center of the contour plot, which representing the highest average health level, is located at the point with high income but low variance.

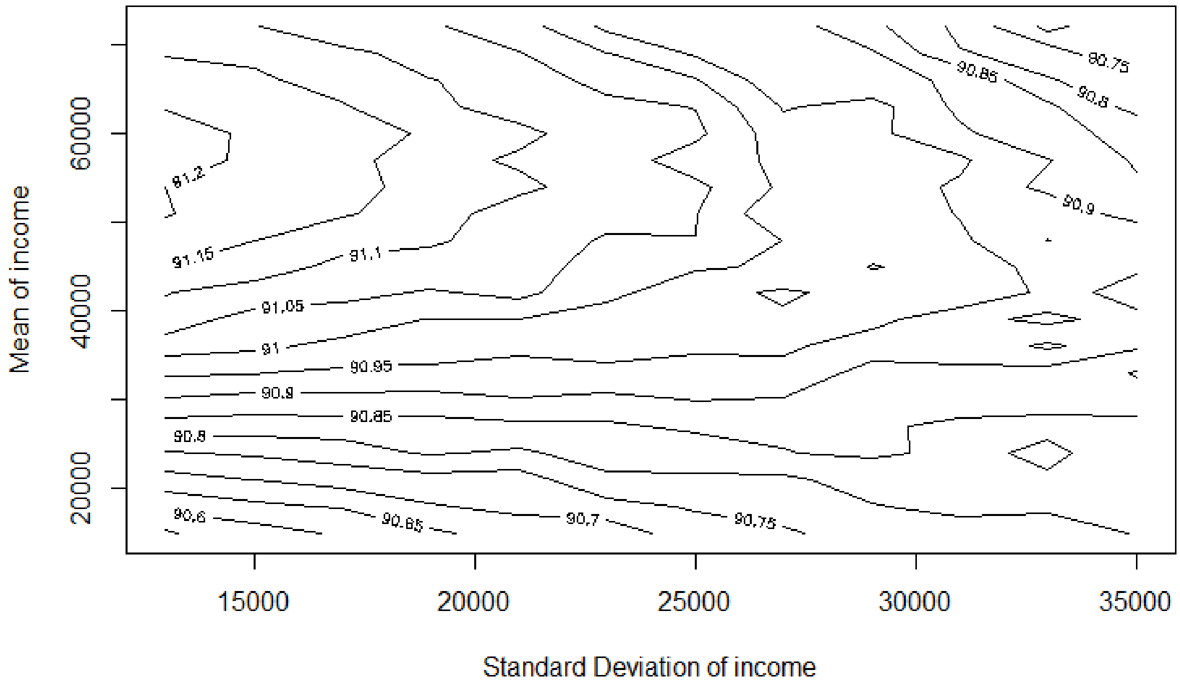

Figure 5-3. Contour plot of Average health (truncated normal population: 10,000)

We first investigate why there are inconsistent conclusions to the question which had been discussed in earlier empirical research, how does income inequality affect overall health? Income inequality can be represented by the Gini coefficient of income. In the figure 5-4, we compare the relationship of average health and Gini coefficient of income at various income level. In the first 4 graphs in figure 5-4, the mean of the income increases from $15,000 to $24,000. At each income level, the average health increases when the Gini Coefficient of income increases. When income level increases to$27,000 and $30,000, we can see that the trend is ambiguous. In the last several graphs, the average income level keeps increasing from $33,000 to $45,000. In contrast to the relationship in the first 4 graphs, the average level of health in the society decreases when the income Gini coefficient increases.

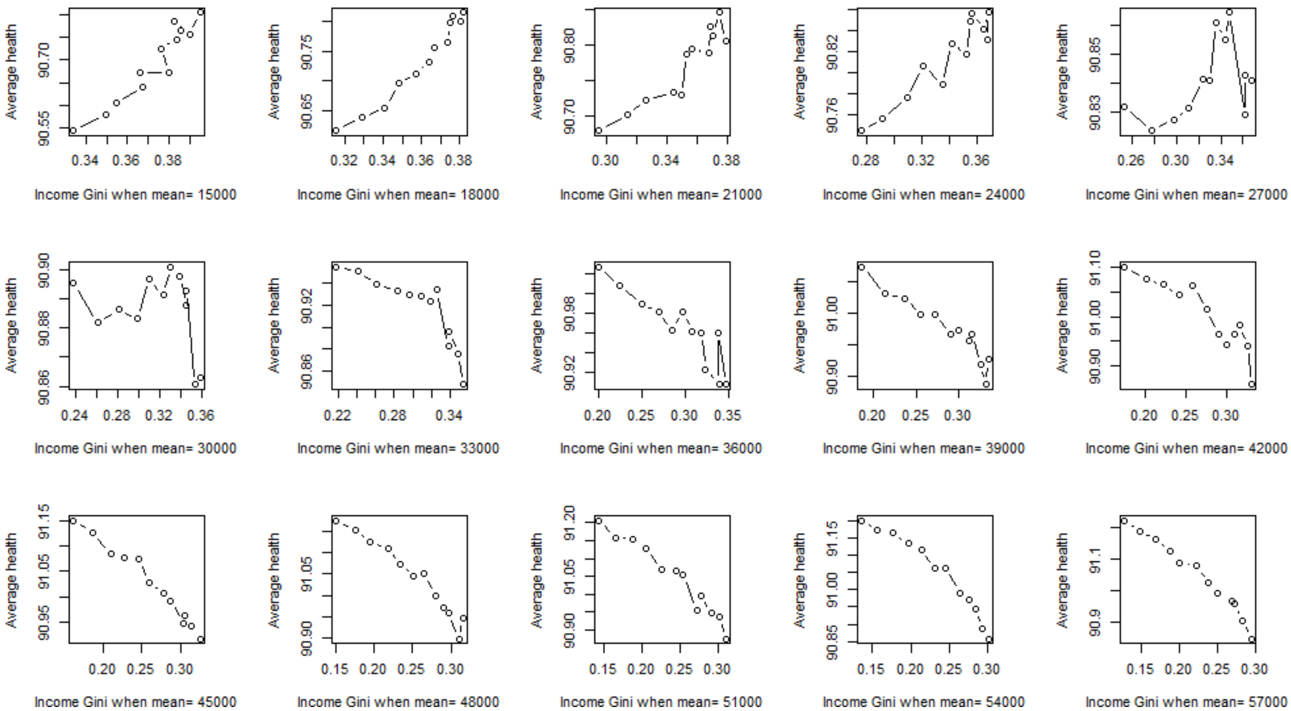

Figure 5-4. Average Health and income Gini coefficient (at different income level)

The relationship showed in figure 5-4 implies that in a poor country, it might be good to encourage a small portion of population to be wealthy first. But in a very rich country, enlarge the income gap between the poor and the rich groups hurt the health of the whole society.

On the other hand, the relationship of economic development of health is also affected by the income inequality level. The Figure 5-5 shows a relationship between income level and social average health level under different income inequality level. In figure 5-5, the variance increases from 13,000 in the first graph to 35,000 in the last graph. We can see a turning point in all levels of inequality. However, the turning point appear earlier in unequal societies. The first 4 or 5 graphs set income variance in a comparative low level. It is a comparative equal society. There is neither a high proportion of extremely rich people nor a high proportion of extremely poor people. In these communities, although we see that the overall average health increase when income level increases at first, then decreases at high average income level, the turning point shows up late at around $60,000 point. In the last several graphs, we set a very high variance of income. So overall it is an unequal society. In this society, wealth is allocated unequally with high proportion of people extremely poor and extremely rich. When economy develops, the rich group enjoy more benefit than the poor group. The turning point show up earlier in these graphs. In the second row, the turning point is around $50,000, while in the last 3 graphs, the curve of the average health level turning at the point of average income $40,000.

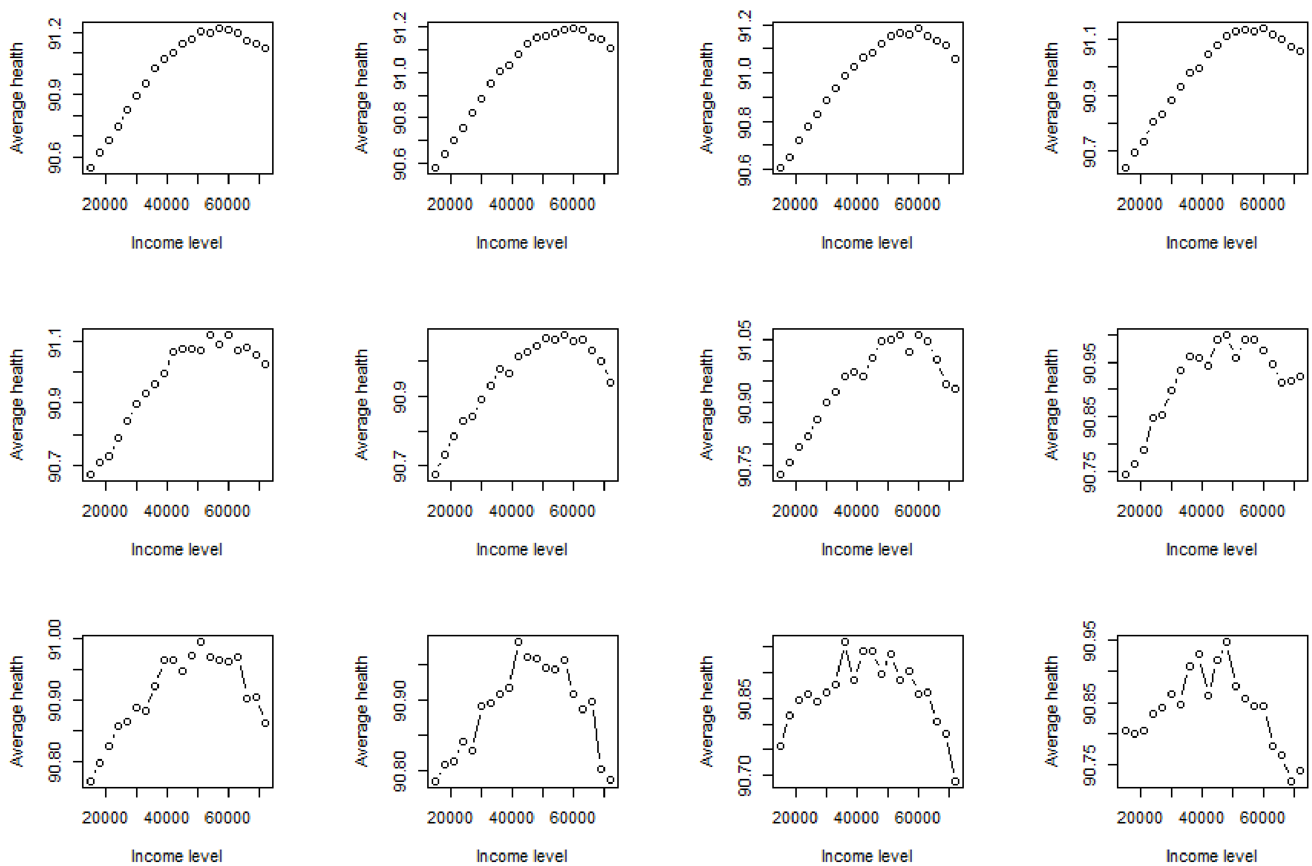

Figure 5-5. Relationship of income and average health (at different levels of income variance)

We call this reversed U-shape relationship between income and health a health Kuznets curve (HKC). The situation in the first 4 or 5 graphs in the figure 5-5 is the ideal countries or communities people prefer, while in realty, it is more likely to see the situation in the last several graphs in figure 5-5, e.g., South Africa, China, Haiti.

6. Conclusions

In this study, we aim to find the relationship of income and health in different inequality level with a more comprehensive measure of individual health. It is usually a macroeconomic problem and was estimated in an aggregated level. however, we believe that build a macroeconomic relationship based on the individual correlation is helpful to understand the inconsistent empirical conclusions in previous research. We first theoretically show the individual level relationship of income and health, which is an individual level HKC. The theoretical relationship is then verified by an empirical analysis with NHANES data, using BMI and Health score to measure the health level. We found that in the normal range of income, BMI decreases as income increases, but the overall health has a non-monotonic relationship with income. People with income level at the 70-90 percentile are the healthiest group. At last, we simulate the aggregate relationship of income inequality and average health, and the relationship of income and health at different income inequality level. The simulation result shows that a HKC exist in societies with all levels of equality. Health increases with economic development initially, and then decreases. However, during the economic development process, the health turning point comes earlier in an unequal society.

Although using this method give us a whole picture of the relationship between income, income inequality and health. It is inevitable that this research still has some limitation. First, the individual behavior may only explain partial of the health difference. Another important element affects one's health is public health investment, which is not discussed in this study. But we believe it won't affect the conclusion because an equal society usually has a better public health system. Second, only using the US data to estimate the individual relationship may be inadequate. If we have more dataset include the detail health information, then the result will be more convincing in the future.

Despite these limitation, this paper still provides a new idea to study on the aggregate macro health economic issue. It is not only useful to the evaluate the relationship of income, income inequality and health, but we can also apply this method to analyze the relationship of income, income inequality and health inequality, which is also a topic comes to an inconsistent conclusion because of various measure of health and various study regions (Costa-font et al, 2017, Sahn & Yang, 2009, Grecu & Rotthoff, 2015). Another application is to learn the correlation of education and health, which is also a very popular topic in health economics. And this method is also powerful to apply to a specific country or even regional health policy analysis. First, estimation what is the effect of the policy having on individual. Them estimate the total effect based on the population composition in the region or country.

Appendix.

Detail method to calculate a set of h_i

The health index we used to calculate the health score include BMI, blood pressure, blood sugar, cancer and other diseases. Among these indexes, BMI, blood pressure and blood sugar are assigned 30% weight, while cancer and other diseases are assigned 50% weight. Before taking the weight average of these indexes, they should be converted into same scale, which is a score from 0-100.

a. Convert BMI scoring:

BMI	(0,15)	(15,16)	(16,18.5)	(18.5,21.5)	(21.5,25)
BMI score	0~40	40~60	60~80	80~100	100~80
BMI	(25,30)	(30,35)	(35,40)	(40,45)	(45, max)
BMI score	80~60	60~40	40~20	20~0	0

b. Convert blood pressure scoring:

Diastolic	(0,50)	(50,60)	(60,70)	(70,90)	(90,100)	(100, max)
Diastolic score	0~40	40~95	95~100	100~95	95~40	40~0
Systolic	(0,80)	(80,90)	(90,105)	(105,120)	(120,180)	(180, max)
Systolic score	0~40	40~95	95~100	100~95	95~40	40~0

c. Convert blood sugar scoring: For simplicity, diabetes is used as a proxy of blood sugar, because diabetes is diagnosed based on the blood sugar level.

Blood sugar (diabetes)	No diabetes	borderline	Diabetes & take pill	Diabetes & take insulin
Diabetes score	90	60	40	10

d. Convert cancer and other disease scoring: Cancer is treated as a very severe disease. So, diagnosed with any kinds of cancer will be assign a very low score on this item. For other diseases, the score is based on how many kinds of diseases one has ever been diagnosed.

Cancer	No	Yes	-	-	-	-
Cancer score	100	10	-	-	-	-
Disease No.	0	1	2	3	4	>=5
Disease score	85	60	40	30	20	10

Reference

Beckfield, J. (2004). "Does Income Inequality Harm Health? New Cross-National Evidence." Journal of Health and Social Behavior 45(3): 231-248.

Bolin, K. and B. Lindgren (2014). Non-monotonic health behaviours - implications for individual health-related behaviour in a demand-for-health framework, Dept. of Economics, University of Gothenburg.

Cawley, J., et al. (2010). "The impact of income on the weight of elderly Americans." Health Economics 19(8): 979-993.

Costa-Font, J., et al., A 'Health Kuznets' Curve'? Cross-Country and Longitudinal Evidence (October, 2013). CESifo Working Paper Series No. 4446.

Costa-Font, J., et al. (2017). "A Health 'Kuznets' Curve'? Cross-Sectional and Longitudinal Evidence on Concentration Indices'." Social Indicators Research.

Cowell F.A. (2000) Measurement of Inequality. In Atkinson A.B., Bourguignon F. (Eds.) Handbook of Income Distribution. Amsterdam, Elsevier, Vol. 1, pp. 87-166.

Davis, B. and A.R. Karr, 1992, Bush to require regulators to weigh costs and impact on health, mortality, The Wall Street Journal, March 20.

Deaton, A. (2003). "Health, Inequality, and Economic Development." Journal of Economic Literature 41(1): 113-158.

Deaton, A. and D. Lubotsky (2003). "Mortality, inequality and race in American cities and states." Social Science & Medicine 56(6): 1139-1153.

Dinda, S. (2005). "A theoretical basis for the environmental Kuznets curve." Ecological Economics 53(3): 403-413.

Drabo, A. (2011). "Impact of income inequality on health: does environment quality matter?" Environment and Planning 43: 146- 165.

Ettner, S. L. (1996). "New evidence on the relationship between income and health." Journal of Health Economics 15(1): 67-85.

Fang, S. (2015). Health index scoring method, Google Patents.

Feinstein, J. S. (1993). "The Relationship between Socioeconomic Status and Health: A Review of the Literature." The Milbank Quarterly 71(2): 279-322.

Forster, B. A. (1989). "Optimal Health Investment Strategies." Bulletin of Economic Research 41(1): 45-57.

Forster, M. (2001). "The meaning of death: some simulations of a model of healthy and unhealthy consumption." Journal of Health Economics 20(4): 613-638.

Gangadharan, L. and M. R. Valenzuela (2001). "Interrelationships between income, health and the environment: extending the Environmental Kuznets Curve hypothesis." Ecological Economics 36(3): 513-531.

Gini, Corrado (1912). Variabilità e mutabilità. Reprinted in Pizetti, E.; Salvemini, T., eds. (1955). Memorie di metodologica statistica. Rome: Libreria Eredi Virgilio Veschi.

Gini, Corrado (1921). "Measurement of Inequality of Incomes". The Economic Journal. Blackwell Publishing. 31 (121): 124–126

Grecu, A. M. and K. W. Rotthoff (2015). "Economic growth and obesity: findings of an Obesity Kuznets curve." Applied Economics Letters 22(7): 539-543.

Grossman, G. M. and A. B. Krueger (1995). "Economic Growth and the Environment." The Quarterly Journal of Economics 110(2): 353-377.

Grossman, M. (1972). "On the Concept of Health Capital and the Demand for Health." Journal of Political Economy 80(2): 223-255.

Jolliffe, D. (2011). "Overweight and poor? On the relationship between income and the body mass index." Economics & Human Biology 9(4): 342-355.

Kawachi, I. and B. P. Kennedy (1999). "Income inequality and health: pathways and mechanisms." Health Services Research 34(1 Pt 2): 215-227.

Kitagawa, E.M., and P.M. Hauser. 1973. Differential Mortality in the United States:A Study in Socioeconomic Epidemiology. Cambridge: Harvard University Press.

Kuznets, S. (1955). "Economic Growth and Income Inequality." The American Economic Review 45(1): 1-28.

López, R. (1994). "The Environment as a Factor of Production: The Effects of Economic Growth and Trade Liberalization." Journal of Environmental Economics and Management 27(2): 163-184.

Marshall, J., 1992, Deadly consequences, San Francisco Chronicle, May 25.

Pampel, F. C., et al. (2012). "Obesity, SES, and economic development: A test of the reversal hypothesis." Social Science & Medicine 74(7): 1073-1081.

Sahn, D. E. (2009). "Measuring intra-household health inequality : explorations using the body mass index." Health Economics 18: 13-36.

Sen, Amartya (1977), On Economic Inequality (2nd ed.), Oxford: Oxford University Press

Schechtman E., Yitzhaki S. (2008) Calculating the Extended Gini Coefficient from Grouped Data: A Covariance Presentation. Bulletin of Statistics & Economics, 2(S08), pp. 64-69.

Silver, M. 1972. An Econometric Analysis of Spatial Variations in Mor- tality Rates by Age and Sex. In Essays in the Economics of Health and Medical Care, ed. V. Fuchs, 161-227. New York: Columbia University Press.

Wilkinson, R. G. and K. E. Pickett (2006). "Income inequality and population health: A review and explanation of the evidence." Social Science & Medicine 62(7): 1768-1784.