Title page

Title: Mortality among the healthy and the unhealthy, and the transition between those states: Which is more influential on the overall education-gradient of mortality?

Name:	Ruben Castro
Postal Address:	239 McNeil Building, 3718 Locust Walk, Philadelphia 19104, United States.
e-mail:	rcastro@sas.upenn.edu

Keywords: Mortality rate, Socioeconomic status, Health indicators.

"What this paper adds" box.

What is already known on this subject?

The process of mortality can be decomposed into the progression to different health states. Socioeconomic (SES) gradients in mortality are commonly found in both health state-specific mortality rates and in the transition rates between them. However, this methodology by itself does not identify which SES gradients are most influential on the overall SES gradient in mortality.

What does this study add?

The degree of influence of each SES gradient on mortality is a complex outcome of two factors: a) the influence of each health state and transition rate, *regardless* of their associated SES gradient, and b) the magnitude of the SES gradients. Slightly different measures of health states might dramatically change the degree of influence of each SES gradient.

Manuscript

Title: Mortality among the healthy and the unhealthy, and the transition between those states: Which is more influential on the overall-education gradient of mortality?

Abstract:

Objectives: To identify influential components of the socioeconomic status (SES) gradient in the mortality rate. A decomposition based in classifying individuals according to their health state is proposed and applied. The SES gradient in mortality is decomposed into gradients in: the transition rate from healthy to unhealthy, the mortality rate before the transition, and the mortality rate after the transition.

Methods: Follow-up data is used to compute those rates, and their SES gradients are estimated with Cox regressions. SES is measured by education, and unhealthiness is defined as being diagnosed with at least one chronic diseases.

Results: While the gradient in each of the three rates is similar, their contributions into the overall mortality gradient differ substantially; the gradient in the "mortality rate after the transition", in particular, explains the vast majority of the total gradient.

Conclusions: Upon defining health states, the methodology used here identifies the specific SES gradients that most influence the overall SES gradient. It is important to note that results show *both* the influence of each health state (regardless of SES gradients) and the magnitude of the associated SES gradients.

Introduction

Research has established a strong relationship between socioeconomic status (SES) and several outcomes, such as family formation, education, health, etc. Articles on the relationships between SES and mortality, and between SES and health in particular, present robust evidence showing higher death rates among lower SES individuals (see for example Goldman [1]). The gradient is observable during the whole life span of an individual[1], and it appears to have increased in recent years[2]. The gradient is present at all parts of the income distribution: The ratio of self-reported health among the upper to middle classes is comparable to the ratio of health among the middle to poor classes (Smith [7]).

A natural question is what is behind those gradients. The list of contributing factors is large. There is a huge body of research dealing with a wide range of underlying factors related to this phenomenon, such as risk/protective behaviors, access to medical care and early life nutrition. However useful and interesting establishing a measure of the relative influence of each of those gradients into the overall gradient in mortality may be, creating such a measure is complex. For example, estimations of SES gradients on each of those multiple factors, and of their interactions, are required.

Furthermore, a substantial fraction of the SES gradient in mortality and health cannot be attributed to specific factors[3]. De Vogli, Gimeno, Martini and Conforti [9] made the following

[1] During childhood, there is evidence of an effect of SES on health, as shown in Case, Lubotsky and Paxson [2]. In this case, the effect seems to be driven by household conditions and intrauterine nutrition. At the other extreme, at very old ages, a SES gradient is also found in numerous studies, such as in Haiyan Zhu and Yu Xie [3] and Huisman, Kunst and Andersen [4], who present evidence from various countries. It is worth noting that at old ages the gradient start to attenuate, as shown by Crimmins [5].

[2] Pappas, Queen, Hadden and Fisher [6] show that "among whites and blacks of both sexes, the differences in mortality according to educational level were greater for 1986 than for 1960*"* (results section.)

[3] This unexplained gap is actually one of the reasons cited for the prevalence of the *status syndrome* (Marmot [8]), which in brief states that psychosocial stress factors related to status in society influence health.

observation: "Although [major mechanisms proposed to explain the gradient in health] contribute to explain population health variations, large gradients, however, are found even in societies with favorable circumstances in terms of health determinants and health status" (page 143.) Hayward, Crimmins, Miles and Yang [10] studied the relationship of SES to the racial gap in chronic conditions and concluded that "the racial gap in health is spread across all domains of health, and socioeconomic conditions, not health risk behaviors, are the primary origins of the racial stratification of health" (p. 910.)

Decomposing a population-level process (for example, mortality) into individual health states, and transition rates among them, is commonly done to identify where SES gradients exist. Further research can then be conducted to analyze the factors behind those gradients. However, this methodology by itself does not identify which SES gradients are most influential on the overall SES gradient in mortality. In this paper I highlight the conceptual complexities of assessing the degree of influence of each SES gradient, which reflects a) how influential each health state and transition rate is, *regardless* of its associated SES gradient, and b) the magnitude of these SES gradients.

I first discuss conceptually a feasible methodology for addressing the above-mentioned aim. For simplicity, I use only two health states, which for ease of exposition I label as "good" and "bad." I then conduct an empirical analysis using this methodology, focusing on a particular measure of SES, education[4], and a particular measure of health state, reported chronic diseases[5]. This

[4] Education is selected as a salient measure of socioeconomic status because surveys usually include a reliable measure of years of education and/or highest completed degree. Besides, "the relation between SES, as measured by education, and health is one of the most powerful relationships in social science research, yet it is perhaps the most difficult to explain (Lynch [11], Pag. 309)".

[5] Measures of health states are less standard in the literature, and, as discussed in the conclusions of this article, each particular measure has double implications for the analysis: It sets the mortality rate of each health state and the transition rates between states. Thus, researchers with different interests may find different health state measures useful. I use just two health states, defined as being or not being diagnosed with at least one of the following conditions: hypertension, cancer, lung disease, heart problems and stroke.

definition, applied to the data used in this article, produced absorbing health states; this in turn considerably simplifies the exposition, yet at the same time preserves the major lessons learned from the exercise. Standard methodology is used to decompose the education gradient of mortality into three components: the education gradient in mortality *before* reaching a bad health state, *after* doing so, and the gradient of the transition rate between good to bad health[6]. The SES gradients on those rates are addressed using survival analysis techniques.

Finally, I repeat the empirical analysis, but use a slightly different measure of health states to highlight the complexities mentioned above and argue that health states should be chosen in light of the research question of interest.

Section 1 - Methodology

1.1 Data

The data used in this study comes from the Health and Retirement Survey (HRS) and the Study of Assets and Health Dynamics among the Oldest Old survey (AHEAD), which are provided jointly by the University of Michigan. All individuals included in the "Wave 2 release," that is, interviewed in 1994 (HRS) and 1993 (AHEAD), are included in this study[7]. Only the few individuals younger than age 50 in Wave 2 are dropped from the sample. Individuals included in Wave 2 are followed until Wave 6, in 2004.

[6] Certainly, the idea that "people in good health die" is difficult to accept, unless they die from accidental causes. The labeling of health states, in this sense, should be replaced by how health states are defined; in this article, health states should be labeled "with reported chronic conditions" and "without reported chronic conditions." I labeled the health states as "good" and "bad" only to facilitate the exposition of ideas.

[7] The data includes sampled individuals' spouses. The AHEAD cohort contains people born in 1923 or earlier. The initial HRS cohort, used here, contains people born between 1931 and 1941. The spouses of those cohorts were born between 1910 and 1969.

This panel data was selected because of the sample sizes (over 17,000 individuals) and the remarkable quality of the health measures. This survey, more importantly, has a long follow-up period of over 10 years. The sample also focuses on the older population, which, when conducting mortality analysis, can attain statistical power. Also, death occurrences can be double-checked by matching these reports with data from the National Death Index[8]. Extending the research to include younger people was rejected because it would require extremely large sample sizes, given that mortality is a rare event among the young.

1.2 Health and SES Measures

Measuring **health status** is complex, and there is no unique measure provided in the datasets. The selected measure depends on the issue that is being addressed and the kind of data available. Articles on healthy life expectancies, for example, use diverse measures, such as diseases, pain, cognition and mobility (Mathers, Sadana, Salomon, Murray and Lopez [12]), as well as self-assessed general health (Doblhammer and Kytir [13]), institutionalized population and disability (Crimmins and Saito [14]), and self-reported diseases (Banks, Marmot, Oldfield, and Smith [15]).

The methodology used in this article could, in principle, be applied to any of those measures; different research projects may find some of these measures more useful than others. The exercise to be conducted here is based on diseases, specifically, chronic diseases. This definition produces absorbing health states, which considerably simplifies the exposition. Diagnoses of chronic diseases are self reported, as are most health variables included in surveys. It is a fairly objective

[8] Repeatedly, HRS administrative staff checks the National Death Index for all individuals, even if they are not currently being interviewed. In this article, individuals are considered alive unless there is some evidence to the contrary, even if they are not being interviewed. This is not done in the case of health state, where individuals not interviewed are considered missing.

measure, as it is based on responses to the question, "Has a doctor ever told you that…"[9] An individual is classified as *unhealthy* if he or she reports any of the following: hypertension, cancer, lung disease, heart problems or stroke. Diabetes, psychiatric, and arthritis were not included because they seem to have a lower correlation with mortality[10].

As is the case with other health variables, a self-reported diagnosis is likely to contain SES-related non-random error. In particular, the diagnosis is related to access to and interactions with physicians, which introduces an *under diagnosis* problem if the poor are less likely to visit a doctor, and/or the doctor is less likely to diagnose them, and/or they are less likely to report the diagnosis in the survey. The definition of unhealthiness used here requires that just one of five illnesses is diagnosed, so that individuals with two types of illnesses, where only one is diagnosed, will still be detected. Thirty-five percent of unhealthy individuals in Wave 2 have two or more diagnosed sicknesses. The conclusions of the exercise conducted in this article, as discussed in the conclusion section, will be attenuated, rather than accentuated, if the poor are more likely to be diagnosed.

Regarding an **SES measure,** there is a general consensus that the educational level is a reliable indicator of an individual's SES (Lynch [11]). Education is set early on in life, which minimizes reverse causation issues from health states to this SES measure. Furthermore, Smith [7] analyzes different data sets, including HRS and AHEAD, and concludes that financial SES variables (household income, wealth and even exogenous changes in wealth) "are either not related or at best weakly related to the future onset of disease over the time span of eight years" (pag. 12)

[9] The exact question wording depends on whether this is a first interview, whether the person being interviewed is the same as in the prior interview, and whether the condition was reported at a prior interview. (HRS-AHEAD Codebook, Rand Center for the Study of Aging).

[10] The link from diabetes to mortality seems to be largely due to cardiovascular disease. In the words Chaturvedi, Jarrett, Shipley and Fuller [16], "we confirm the existence of an inverse socioeconomic mortality gradient in diabetic people and suggest that this is largely due to conventional cardiovascular risk factors".

Additional education, says Smith [7], "is strongly and significantly predictive of the new onset of major and minor disease" (pag. 12). In light of these observations, the SES gradient measure used in this article is obtained from "years of education," recoded in a scale from 1 to 5: 1 (0/8), 2 (9/11), 3 (12), 4 (13/15) and 5 (16+).

1.3 Concepts

From an individual-level perspective, the probability of dying and the probability of reaching a poor health state can be thought of as a joint random process. This process depends on a random component, and on several individual-level characteristics. The joint process can be decomposed into the "marginal" times the "conditional" distribution: the probability of dying given a particular health state (conditional distribution) times the probability of entering that particular health state (marginal distribution.).

Assuming only two health states, healthy and unhealthy, the probability of dying for any individual becomes:

P(dying)= P(dying while healthy)*P(health=healthy)+P(dying while unhealthy)
*P(health=unhealthy)

From a population-level perspective, the death rate is the outcome of three rates: F, the death rate among healthy individuals; B, the rate at which healthy people become unhealthy; and G, the

death rate among unhealthy individuals. *Unhealthy* is assumed here to be an absorbing state[11]. The death rate in a given period becomes:

$$\frac{M}{PY} = \frac{M_H + M_U}{PY_H + PY_U} = \frac{M_H}{PY_H} \cdot \left(\frac{PY_H}{PY_H + PY_U} \right) + \frac{M_U}{PY_U} \cdot \left(1 - \frac{PY_H}{PY_H + PY_U} \right) \qquad (1)$$

where M is the number of deaths observed in the period, and M_H (M_U) is the number of deaths among healthy (unhealthy) individuals. PY stands for the total person years, and PY_H (PY_U) are the person years lived by healthy (unhealthy) individuals. Thus F is equal to M_H/ PY_H and G is equal to M_U/ PY_U . The expression for B is not easily seen in equation (1). B influences only the term in parenthesis, in equation (1), which is also influenced by F and G.

This perspective will guide the rest of the article, where each rate and its gradient are estimated from sample data on the U.S.

1.4 Estimation of *F*, *B* and *G*

The methodology for estimating the *F*, *B* and *G* functions will be based on data from HRS and AHEAD (see the data section); each individual included in Wave 2, conducted in 1994, will be followed in each subsequent wave until the 2004 survey. This data will be used to estimate the risk in each of the three processes. In all cases, age-specific rates are calculated, and then an age standard is used to compute the overall death rate (see Section 1.5, decomposition procedure). Table 1 presents a brief description of the F, B and G rates.

[11] Based on the definition of unhealthiness used in this article, the data do not contain any cases of individuals reverting from an unhealthy to a healthy state.

[TABLE 1 HERE]

The sample data used in this article contains only people aged 50 or older. Because mortality prior to age 50 is a rare event, at least among the more recent cohorts in the U.S., this might not be crucial in the study of *F* and *G*; but, in the case of *B*, an earlier onset of a bad health state is certainly not uncommon[12]. In this study, the health state at age 50 is taken as given[13]. However, the observed proportion of people who are unhealthy at age 50 might be considered a good indicator of the incidence of *B* before age 50, given that mortality before age 50 is rare. The concluding section provides further discussion of this issue.

1.5 Decomposition procedure

The SES gradient of mortality is decomposed into three parts: the SES influence in *F*, in *B*, and in *G*. First, a *base case* is computed, in which there is no influence of SES. Then, *comparison cases* are computed, in which one at a time a SES gradient is included in *F*, *B* and *G*.

It is important to note that, in all cases, a "proportion of the population that is in poor health" is needed in each age interval. This prevalence is in fact the outcome of the incidence of *F*, *B* and *G*. The o*bserved* proportions of unhealthy individuals in the sample data should not be considered as valid parameters, due to internal consistency[14]. Instead, the population implied by *F*, *B* and *G* (or

[12] Of the people interviewed in the 1994 wave of HRS and AHEAD, about 38% of respondents aged 50 to 54 reported having been diagnosed with at least one of the illnesses considered in this article (see Section 1.4, health measures, and Section 1.1, data)

[13] Later in Section 2, simulations are used to assess the importance of the SES gradient in each function. But, in all cases, the proportion unhealthy in the first age interval, 50-54, is taken from the data instead of being simulated from the estimated values of F, B and G.

[14] The observed proportion unhealthy is the outcome of current and past values of F, B and G. Only if current and past values of those functions are equal, then the observed proportions unhealthy will be consistent with the proportion unhealthy implied by F, B and G. This article estimates only current values

their SES-influenced version, as in the comparison cases) provides consistent values for the proportion unhealthy. Methodology to derive this stable-population "proportion unhealthy" is done with standard demographic techniques (Preston et al. [17], Chapter 4).

Base case

Using sample data from HRS and AHEAD, the age-specific death rates are computed for *F, B* and *G*. Then, the proportion unhealthy, which is endogenous to F, B and G, is computed. Next, the total age-specific death rates are computed as in equation (1)[15]. Once the age-specific death rates are computed, they are collapsed into an overall death rate using the age composition of the U.S. in 1998 (Source: NCHS). Because the waves included in the analysis run from 1994 to 2004, 1998 seems to be a good central point to use as the age standard.

Comparison cases

The objective of each comparison case is to address, one at a time, the impact of using SES specific rates for *F*, *G* and *B*, instead of the ones used in the base case. But a big problem arises when computing age-specific, SES-specific rates: sample size. Estimates of rates might become very unstable, as is the case in this article. An option, used here, is to estimate a survival model whereby SES dummies are included as covariates; this will provide relative risks. Naturally, flexibility in this case is lower because the risks by SES *have to* follow a parametric rule between them; however, it is possible to obtain reasonable robust estimates.

---

of those functions (individuals are observed from 1993/1994 onwards). Figure 1 shows that the observed and implied values are substantially different.

[15] The values themselves, as well as the *observed* proportions in Wave 2, are shown in Figure 1.

Among the possible parametric and semi-parametric survival analysis approaches that could be used, a Cox regression is selected (Cox [17]). The Cox model has the advantage of not assuming any particular shape for the underlying hazard, except the relative risk by SES [16, 17]. Thus, this model also has the disadvantage that risks are assumed to be proportional across covariates (Kaplan-Meir's estimates of the hazard function do not, however, call this assumption into question.)

Again, the proportion unhealthy in the 50-54 age interval is taken as given, and interpreted as the results of SES-specific B rates.

Section 2 - Results

2.1 Descriptive statistics

The data contains 17,223 individuals, with women representing 57% of the total. The proportion unhealthy is 37% in Wave 2, and individuals range from 50 to 104 years old. The mean age is 66. During the period of observation, 28% of individuals died. The distribution by SES level (see description of this variable in Section 1.2) from SES 1 to SES 5 is 18%, 16%, 33%, 17% and 16%.

[16] Other than SES dummies, the only covariate used is age, which was included because the analysis in this article is age specific. The SES coefficients should reflect the relative risk and not the different age composition of each exposure group.

[17] A disadvantage is that Cox models are subject to bias when a significant proportion of individuals experience the risk event at the same time (Allison [18]). Though death dates are fairly precise, health states are only observed at the time of survey (i.e., they are discrete). In this case, a complementary log-log model (i.e., a discrete time model) was used, with substantially similar results.

The proportion unhealthy, shown in Figure 1, grows from 37% in the 50-54 age interval to a maximum of 70% in the 80-84 age interval. In the SES measure used here, years of education, the older population tends to have a much lower SES position.

[FIGURE 1 HERE]

2.2 Estimation of SES relative risks in *F*, *B* and *G*

Table 2 contains the outcome of the estimation[18]. At first glance, Table 2 seems to show that relative risks by SES follow a similar relationship in all functions. This is an important point: The SES gradient of the risk of dying in an unhealthy state appears to be somewhat similar to the gradient of the risk of dying in a healthy state, and also seems to be similar to the gradient of the risk of actually becoming unhealthy. The *B* function, however, has the smallest relative risks (closer to one). The *F* Function shows a relatively small gradient among high SES individuals, while the Function *G* shows a relatively big gradient among low SES levels.

[TABLE 2 HERE]

Some values shows low significance, but it has to be acknowledged that if one-side tests were used instead of two-side tests, most of the SES dummies would be significant. One-side tests, once the analyst is confident that the coefficient has to be positive (or negative), are appropriate. This is the case here, because SES 1 and 2 can be fairly assumed to have lower risk than SES 3, while the opposite is true for SES 4 and 5. In addition, the gradient implied by these parameters

[18] The proportion unhealthy at age 50-54 used in this article is 38%. When computing the influence of the gradient in *B*, the proportions unhealthy at age 50-54 are taken from the data, and are 50%, 48%, 37%, 36% and 31%.

does not look erratic. The conclusion here is that sample power is the reason for the wider confidence interval of some coefficients.

2.3 Decomposition results

Figure 2 shows the death rates by SES. The first group of columns shows the base case, in which by construction all SES groups have the same death rate. SES 3, the reference category, by construction remains the same in the base case and in all the comparison cases.

[FIGURE 2 HERE]

It can be seen that adding the SES relative risk of *F* introduces slightly higher death rates among individuals of SES Levels 1 and 2, and slightly smaller death rates among individuals from SES Levels 4 and 5. The same is true for Function *B*, though the former is more influential among higher SES groups, while the latter has a greater effect among the lower SES levels.

Function *G* is clearly the most influential of all the functions. The reasons behind these results are twofold: First, the gradient in the G rate is somewhat bigger than the one of F or B; and, second, the G death is, regardless of its associates SES gradient, very influential. Its influence comes from its high value[19] and the high prevalence of unhealthiness.

These results raise an interesting point. Even though the gradients in each function are fairly similar, their influences clearly are not. The reason lies in the overall prevalence of unhealthiness, regardless of the relative risks. This issue is discussed in greater detail in the concluding section.

[19] Crude deaths rates can be computed from Table 3: *F* has values of 633/58230=0.011 (deaths per year of exposure), whereas *G* has values of 4033/95504=0.042.

Section 3 - Discussion

Based on the definition of health states used in this article, we can see that SES gradients in state-specific mortality rates and transition-between-states rates are remarkably similar. While this finding is interesting in itself, this article focuses on the importance of each SES gradient. In terms of influence on the overall gradient in mortality, some of these SES gradients are more important than others.

Figure 2 shows simulations of the total death rate by SES when replacing *F* (mortality rate among people without reported chronic conditions) by an SES-specific *F* function. It also shows the same thing for *B* (mortality rate among people without reported chronic conditions) and *G* (transition rate between health states). Just by summing up the gradient produced by each function, it is possible to appreciate that most of the gradient comes from G[20]. Indeed, using individuals with 12 years of education as the reference point, we can see that 80% of the extra mortality in the lower two levels of SES comes from mortality among the unhealthy, while 67% of the lower mortality in the higher level of SES is also explained by this function.

Even though this analysis is done for people aged 50 or older, and so the proportion unhealthy at that age is taken as given, it is not unreasonable to assume that the role of *B* before age 50 is also taken into account. Because the influence of *F* and *G* are very small under this age, the proportion unhealthy at 50 might be largely an outcome of *B*.

[20] Instead of just summing up the gradients, the three SES-specific functions should be applied at the same time. Results are nearly the same.

The influence of *under diagnosis* among poorer individuals, if present, would introduce bias, overstating the gradient in the healthy,[21] and understating the gradient in the transition rate. But this paper concludes with the importance of the gradient among the unhealthy, so under diagnosis attenuates, rather than accentuates this conclusion. However, understating the gradient in the transition rate will also diminish the importance of B itself, which, unfortunately, accentuates the strength of the conclusions. Nevertheless, the definition of unhealthiness used here requires that any of five sicknesses be reported as diagnosed, so individuals with two sicknesses may be classified twice as unhealthy. Thirty-five percent of unhealthy individuals in Wave 2 have two or more diagnosed sicknesses.

Generally, this study has shown that it is possible to identify particular areas of the SES gradient in mortality in which SES gradients are most influential. Thus, future research or policy resources can be directed to those areas. If a researcher is interested in understanding the SES gradient in smoking-related mortality, for example, she could define health states as "fewer than 100 cigarettes," "smoke almost daily" and "smoke daily." After identifying the most influential SES gradients, she could devote her resources to understanding (sociologically, economically, psychologically, etc.) these particular gradients.

A key point is that each state-specific gradient's influence on the overall SES gradient in mortality reflect a) how influential each health state is, regardless of its associated SES gradient; and b) how big the SES gradients associated with each health state are. I avoid generalizing the conclusion of the particular application carried out in this article precisely because of this last point.

[21] Poor individuals in the healthy group will show bigger mortality rates than they should because some unhealthy individuals are mistakenly included in this group.

How influential each health state is depends on the combinations of all mortality and transition rates (F, B and G in this article). Both the "extra mortality" associated with a particular health state, as well as its prevalence, will determine its level of influence. For example, regardless of the SES gradient in *G* and *B*, both *G* and *B* will have very little influence if almost nobody enters into a bad health state. If, on the other hand, almost everyone becomes unhealthy, the gradient in *G* will be much more influential than the gradient in B or F.

[FIGURE 3 HERE]

As an exercise, the analysis in this article was repeated in its entirety using a more restrictive rule for classifying individuals as unhealthy. Instead of one sickness, at least two sicknesses must be diagnosed.

In this case (see Figure 3), fewer people fall into the unhealthy category. For example, let us assume that the 50-54 age interval proportion unhealthy is 8%, instead of the 38% cited in the original definition. At the same age interval, the SES-specific proportion unhealthy is equally small. But now, even though the relative risks are similar (not shown), the influence of *G* is much smaller, as is the influence of *B*. In this case, the SES gradient is shared in equal parts by *F*, *B* and *G*.

Acknowledgments

Special thanks to Samuel Preston and Jason Schnitkker, for invaluable help and support. I am also indebted to Hiram Beltran, for his useful comments and ideas.

Licence for Publication Statement

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in JECH and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence (http://jech.bmj.com/site/about/licence.pdf)."

References

[1.] Goldman, N. Social inequalities in health: Disentangling the underlying mechanisms. *Annals of the New York Academy of Sciences*. 2001; 954:118-139.

[2.] Case, A., Lubotsky, D. and Paxson, C. Economic Status and Health in Childhood: The Origins of the Gradient. *American Economic Review*. 2002; 92 (5): 1308–34.

[3.] Zhu, H. and Xie, X. Socioeconomic Differentials in Mortality Among the Oldest Old in China. Using data from the 1998, 2000, and 2002 waves of the Chinese Longitudinal Healthy Longevity Survey. *Research on Aging*. 2007; Vol. 29, No. 2, 125-143.

[4.] Huisman M, Kunst AE, Andersen O, Socioeconomic inequalities in mortality among elderly people in 11 European populations. *Journal of Epidemiology and Community Health*. 2004; 58:468–75.

[5.] Crimmins, E. Socioeconomic differentials in mortality and health at the older age. *Genus*. 2005; LXI(1), 163-178.

[6.] Pappas, G., Queen, S., Hadden, W., Fisher, G. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *The New England Journal of Medicine.* 1993; 329(2),103-109.

[7.] Smith, J. Unraveling the SES-Health Connection. *Aging, Health and Public Policy: Demographic and Economic Perspectives.* 2005*;* Vol. 30, pp. 108-132.

[8.] Marmot, M. *Status Syndrome: How Your Social Standing Directly Affects Your Health*. Bloomsbury Publishing PLC; 2004.

[9.] De Vogli, R., Gimeno, D., Martini, G., Conforti D. The pervasiveness of the socioeconomic gradient of health. *European Journal of Epidemiology*. 2007; 22:143-144.

[10.] Hayward, M.D., Crimmins, E.M., Miles, T.P, and Yang, Y. The Significance of Socioeconomic Status in Explaining the Race Gap in Chronic Health Conditions. *American Sociological Review*. 2000; 65:910-930.

[11.] Lynch, S.M. Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach. *Demography.* 2003; 40: 309-331.

[12.] Mathers CD, Sadana R, Salomon JA, Murray CJ, Lopez AD. Healthy life expectancy in 191 countries. *Lancet.* 2001; 357:1685-91.

[13.] Doblhammer G, Kytir J. Compression or expansion of morbidity? Trends in healthy-life expectancy in the elderly Austrian population between 1978 and 1998. *Soc Sci Med*. 2001; 52:385-391.

[14.] Crimmins, E., Saito, Y. Trends in healthy life expectancy in the United States, 1970–1990: gender, racial, and educational differences. *Soc Sci Med.* 2001; 52: 1629–41.

[15.] Banks, J., Marmot, M., Oldfield, Z., Smith, J., Disease and Disadvantage in the United States and in England. *Journal of the American Medical Association*. 2006; 295(17):2037-2045.

[16.] Chaturvedi, N., Jarrett, J., Shipley, M., Fuller, J. Socioeconomic gradient in morbidity and mortality in people with diabetes: cohort study findings from the Whitehall Study and the WHO Multinational Study of Vascular Disease in Diabetes. *British Medical Journal.*1998; 316:100-105.

[17.] Preston, S., Heuveline, P., Guillot, M. *Demography: measuring and modeling population processes*. Malden, Mass.: Blackwell Publishers; 2000.

[18.] Cox, D. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*. 1972; Series B34:187-220.

[19.] Allison, P.D. *Event History Analysis: Regression for Longitudinal Event Data*. Newbury Park, CA: Sage; 1984.

TABLES

Table 1: Exposure and occurrence in *F*, *B* and *G*

	F	***B****	***G***
Exposure starts in:	Person is healthy in Wave 2	Person is healthy in Wave 2	Person is unhealthy in any wave
Failure:	Dying	Person became unhealthy	Dying
Exposure ends with no failure if:	Person became unhealthy or reach Wave 7	Person die or reach Wave 7 in good health	Person reach Wave 7

**B* is assumed to be an absorbing state. In the data base and the health measure used in this article, no case of reversing health status from unhealthy to healthy was found.

Table 2: Cox model estimation of SES relative risks

Function *F*	HR	P-value	95% confidence interval		
SES 1	1.181	0.115	0.960	-	1.453
SES 2	1.130	0.301	0.896	-	1.424
SES 4	0.812	0.108	0.630	-	1.047
SES 5	0.761	0.037	0.589	-	0.984
Individuals	*7709*				
Failures	*633*				
Person years	*58230*				
Function *B*					
SES 1	1.178	0.001	1.070	-	1.296
SES 2	1.090	0.085	0.988	-	1.202
SES 4	0.986	0.772	0.899	-	1.082
SES 5	0.920	0.081	0.837	-	1.010
Individuals	*7711*				
Failures	*3873*				
Person years	*54720*				
Function *G*					
SES 1	1.237	0.000	1.141	-	1.341
SES 2	1.190	0.000	1.090	-	1.300
SES 4	0.957	0.373	0.869	-	1.054
SES 5	0.863	0.005	0.779	-	0.957
Individuals	*12612*				
Failures	*4033*				
Person years	*95504*				

"Function F" refers to mortality among healthy individuals, i.e, individuals without reported chronic conditions. "Function B" refers to the transition rate from healthy to unhealthy. "Function G" refers to mortality among unhealthy individuals, i.e., individuals with at least one reported chronic condition. The only other covariate included in the regressions is "age".

FIGURES

Figure 1: Proportion unhealthy: Stable population and observed values

Source**:** Author's calculations. The stable population is a computation explained in Section 1.5.

The observed values in Wave 2 are discussed in Section 2.1.

Figure 2: Death rates by SES

Source**:** Author's calculations.

Figure 3: Death rates by SES under a restrictive definition of unhealthiness

Source**:** Author's calculations.