Documentos de Trabajo 19

About the role of chronic conditions onto the US educational differences on mortality

Ruben Castro

UNIVERSIDAD DIEGO PORTALES
udp
facultad de economía y empresa

About the role of chronic conditions onto the US educational differences on mortality

Ruben Castro

Abstract. Educational differentials in the prevalence (ΔP) of any given condition can be simulated and decomposed as the outcome of differentials on incidence and mortality rates. The exact same methodology also decomposes the role of the chronic condition on the overall mortality (ΔOM), but this analysis is not conducted in the literature. Chronic conditions, first cause of death in the US, are studied here using US elderly follow up data about mortality and diagnosed diseases, including Hypertension, Hearth attack, Diabetes, Lung disease, Cancer and Stroke. On each case, all individuals are assumed to experience a baseline mortality, while the presence of a chronic conditions is associated with an "excess mortality" rate; along this lines, ΔP and ΔOM are decomposed as the outcome of educational differences on baseline mortality, incidence and excess mortality. I found that for most chronic conditions a) educational differences on the base line mortality explain most of ΔOM but not much of ΔP, b) differences on incidence are more important than differences on excess mortality, but only with regard to ΔP, not to ΔOM; furthermore, in the case of Hypertension and Hearth attack, excess mortality rates explain most of the ΔOM associated the chronic condition. In the case of Cancer, the lower incidence and higher excess mortality of individuals in the low education group predicts a relatively lower prevalence of Cancer but also a relatively higher overall mortality in this group. In terms of policy, this idea highlights that the relative importance of differentials on incidence and mortality rates depends on the target of interest, prevalence or overall mortality.

Introduction

Research on the relationship between education and mortality presents robust evidence showing higher death rates among individuals of low education[1]. This phenomenon is observable during the whole life span of an individual[2] and appears to have increased in recent years[3]; furthermore, the mentioned relationship is present at all levels of the income distribution (Smith 2005.) There is an extensive body of research dealing with a wide range of conditions that allegedly underlie the education-mortality relationship, yet, some fraction of the gradient in mortality and health has not being attributed to any specific factor[4].

---

[1] See for example Goldman (2001.)

[2] During childhood, there is evidence of an effect of SES on health, as shown in Case, Lubotsky and Paxson (2002.) In this case, the effect seems to be driven by household conditions and intrauterine nutrition. At the other extreme a SES gradient is also found at very old ages in numerous studies, such as in Haiyan Zhu and Yu Xie (2007) and Huisman, Kunst and Andersen (2004), who present evidence from various countries. It is worth noting that at old ages the gradient begins to attenuate, as shown by Crimmins (2005.)

[3] Pappas, Queen, Hadden and Fisher (1993) 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.)

[4] De Vogli, Gimeno, Martini and Conforti (2007) made the following 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 (2000) 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.)

A vast majority of the conditions that influence mortality show an association with education, in terms of its incidence rate and its associated mortality rates. With this in mind, different studies decompose the education differentials in the prevalence in terms of incidence and mortality differentials. In this study I highlight that is possible and straightforward to extend such decompositions to the overall mortality, with the aim of better understanding the role that any given condition plays on generating differences on the overall mortality across education groups, time, places or race. In other words, how much of poor people's higher mortality can be associated with a) mortality not associated with the factor being studied, b) the factor's incidence rate and c) the factor's associated excess mortality?.

Demographers and epidemiologists make use of multi-state tables[5] to simulate the dynamics of the prevalence. Along these lines, differences in the prevalence of any given condition across populations, time or places, can be simulated as the outcome of differences on the underlying incidence and mortality rates, such as Leveille et al. (2000) and Melzer et al. (2001) for sex and education differences in the prevalence of disability, respectively, and Lipscombe et al. (2007) for time-differences in the prevalence of Diabetes on Canada. And the idea behind this analysis has been extended even further, for example to analyze car accidents (where the "states" are defined by the action of being driving or not, in Dellinger et al. 2002), to analyze the incidence of disability given a chronic condition (Freedman et al. 2007) or to better portrait the dynamics of entry and exit from states of depression across life (Patten et al. 2005). Another extension is to simulate the dynamics of the overall mortality, which arises from the same model of that of the prevalence, as I discuss in section 1.1 and 1.2.

A large body of literature deals with education differentials on incidence and excess mortality of individuals with diagnosed chronic diseases, and such differentials are found significative in the vast majority of them; depending on the settings of each study, low versus high education is found associated with as much as 2 times as much risk of both incidence and mortality, in the case of the chronic conditions included in this study. Those associations between education and chronic diseases are the components behind the education differentials on the prevalence of chronic diseases. While several empirical analysis show that the prevalence of chronic conditions is higher among groups of lower education, this result is not obvious since their higher incidence is counteracted by their higher excess mortality. Besides, those components partially explain the role of any given chronic condition onto the educational differences on mortality. This analysis apparently has not being done before. Even though the exact same model and the same rates that decompose the prevalence do also decompose the overall mortality, the lessons from the former can not be plainly applied to the later: While higher excess mortality is unambiguously associated with lower prevalence, at least in models of reasonable simplicity, it is ambiguously associated with the overall mortality, because individuals with the condition rise the overall mortality but at the same time became less prevalent.

The decomposition I perform in this study aim to provide a measurement of the relative importance of base line mortality, incidence and excess mortality in terms of explaining the magnitudes of education differentials on the overall mortality and on the prevalence. The study is focused on chronic conditions. Although according to the US Center of Disease Control (CDC) "7 out of 10 deaths among Americans each year are from chronic diseases and Heart disease, Cancer and Stroke account for more than 50% of all deaths each year", apparently there are no studies that decompose the US's education differences on the prevalence of chronic conditions into its underlying incidence

[5] Multi-state life tables are also known as increment-decrement life tables; they are models for processes where individuals move between a finite number of states over time, including exit and reentry into the same state. See Preston et al (2000).

and excess mortality rates, let alone the role of this underlying forces into the observed education differentials on mortality. Furthermore, the prevalence of chronic conditions has been projected to growth from 118 million of Americans in 1995 to 171 millions in the years 2030 (Wu et al 2000).

This study use the concept of a base line mortality experienced by all individuals, and an excess mortality experienced on top of the base line by individuals with the chronic condition being analyzed. Excess mortality rates are used because of interpretation: probably highly educated individuals have lower mortality even in the absence of chronic conditions. Each chronic condition is analyzed separately. The empirical analysis focuses on education differentials on six chronic conditions among the elderly, as included in the Health and Retirement Study (HRS), namely, in the order of their prevalence, Hypertension, Hearth attack, Diabetes, Cancer, Stroke and Lung disease. Those conditions have virtually no reverse transition, which simplifies the discussion, and also are of widespread interest for researchers.

Multi-state table simulations, built upon follow up data on recent US elderly, show that incidence rates rather than excess mortality rates explain most of the education-prevalence association of the chronic conditions analyzed in this study. In the other hand, the education differentials in the overall mortality are found to be equally explained by incidence and excess mortality rates associated with chronic conditions. Furthermore, excess mortality rates explain most of the education-overall mortality association produced by hypertension and hearth attack.

Section 1 discusses the model, data and methodology; section 2 show results and last section discusses the results.

Section 1 – Model, data and methodology

1.1 The condition's prevalence versus the condition's influence on the overall mortality

For clarity of interpretation, mortality of individuals with a chronic condition is characterized here by "excess mortality", meaning the gap with respect to mortality among individuals without a chronic condition. The rationale has to do with ease of interpretation: probably highly educated individuals have lower mortality even in the absence of chronic conditions; therefore their lower mortality when having a chronic condition should not be associated entirely with chronic conditions. Note that the principal ground to choose among "excess mortality" or "mortality" is that of interpretation. Das Gupta (1993)'s decomposition formulas, used later in this study, can be equally applied to different measures and perspectives, but in this study, the interpretation is that of chronic conditions' *associated mortality* having something to do with a demographic process. For example, think of a chronic condition absolutely unrelated to mortality, called "X"; highly educated individuals' lower mortality would be partially attributed to X unless X's associated mortality is measured as "excess mortality".

Assume the existence of an irreversible chronic condition, from now on called unhealthiness for ease of exposition. Assume a cohort of individuals, some of them initially unhealthy, who goes under the following rates during one year: M, the base line annual mortality of all individuals; E is the annual excess mortality of unhealthy individuals; M+E, the annual mortality of unhealthy individuals; and T the transition rate into unhealthiness. It is easy to prove that the overall annual mortality (OM) can be expressed as in Equation 1, where P represents the proportion of total person years lived in an unhealthy state.

Therefore OM depends on two flux variables, M and E, and one stock variable, P, which in turns is a function of T, M and E. Thus E has two pathways into OM while T has one. This is the reason why the interplay of T and E behind P can be substantially different from the interplay of T and E behind the OM.

$$OM = M + PE \quad (1)$$
$$OM_T = P_T \cdot E \quad (2)$$
$$OM_E = P_E \cdot E + P \quad (3)$$

OM_T and OM_E, in Equation 2 and 3, are the derivatives of OM with respect to T and E respectively. Given that P_T and P_E are unambiguously positive and negative (see Appendix A.2), then: a) OM_E has two components, unambiguously of opposite sign, while OM_T is unambiguously positive. Therefore in this simple example, the effects of T and E on P are positive and negative, respectively, while the effect of T and E on OM are positive and ambiguous, respectively.

1.2 Interpreting P as the Prevalence of the chronic condition

The cohort just described can be followed not for one year but from many year years, and age-specific M, T and E rates can be introduced. The life-time overall mortality still can be expressed as in Equation 1, though the dynamics of the process make the algebra more complex. Still, E directly influences P and OM, while T influences only P.

The empirical analysis on this study is done using age-specific values of M, T and E to compute a multi-state life table; this synthetic cohort perspective is less data demanding than a short run decomposition, which requires a whole history of rates. The implications of the multi-state tables are interpreted next from the perspective of a stationary population, which provides an insightful theoretic perspective: the multi-state's healthy or unhealthy life expectancies can be re interpreted as the prevalence. Indeed, person years lived between ages x and $x+n$, as implied by a life table, is equivalent, on the stationary population, to the number of persons between those ages (Preston 2000). Therefore P, the ratio of person years, became the ratio of persons, i.e., the prevalence.

1.3 Decomposition analysis

This study computes the multi-state life table associated with the observed education differentials on incidence and excess mortality rates of different chronic conditions on the US. The equations of the multi-state model in this study are discussed in Appendix A.1. Basically, each individual have a probability of dying during one interval of time, which is higher if the individual has a chronic condition; individuals without chronic conditions have also the risk of transiting into "having a diagnosed chronic condition" during one interval if time. No reverse transition is assumed.

First, a reference group is defined. Based on age-specific estimations of M, T and E for this group, a multi-state life table is computed. The values of P (interpreted as the prevalence, in the stationary population implied by the multi-state life table) and overall mortality are computed. Second, a comparison group is defined. The differences between the reference and comparison groups are decomposed using Das Gupta (1993); thus, differences on P and overall mortality are decomposed as differences on M, T and E.

1.4 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[6]. 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.

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[7].

1.5 Measures

Different health conditions are provided in the datasets. Articles on healthy life expectancies, for example, use diseases, pain, cognition and mobility (Mathers, Sadana, Salomon, Murray and Lopez 2001), as well as self-assessed general health (Doblhammer and Kytir 2001), institutionalized population and disability (Crimmins and Saito 2001), and self-reported diseases (Banks, Marmot, Oldfield, and Smith 2006). The conceptuality discussed in this study can be applied to any of those measures.

The empirical exercise to be conducted here focuses on chronic diseases. Education differentials on six chronic conditions among the elderly, as included in the health and retirement study (HRS), namely Hypertension, Hearth attack, Diabetes, Lung disease, Cancer and Stroke, are analyzed in terms of prevalence and implications for the overall mortality[8]. All of them are documented to have high or moderate prevalence on recent years on the US[9], and to be associated with a higher mortality rate. However, the demographic dynamics behind their prevalence and specially their implications for the overall mortality are less known.

Diagnoses of chronic diseases are self reported, as are most health variables included in surveys. Nonetheless, it is a fairly objective measure, as it is based on responses to the question, "Has a doctor

---

[6] 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.

[7] Repeatedly, HRS administrative staff checks the National Death Index for all individuals, even if they are not currently being interviewed. In this study, 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.

[8] However, the link from diabetes to mortality seems to be largely due to cardiovascular disease. In the words Chaturvedi, Jarrett, Shipley and Fuller (1998), "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".

[9] The prevalence at age 50-54, according to the data used in this study, are 31%, 9%, 9%, 4%, 4% and 2% for Hypertension, Hearth attack, Diabetes, Lung disease, Cancer and Stroke. This statistics are comparable to the ones found in the literature. The prevalence of any of those chronic conditions at age 50-54 is 37%.

ever told you that…"[10] Psychiatric, and arthritis were not included because they seem to have a lower correlation with mortality.

As is the case with other health variables, a self-reported diagnosis is likely to contain education-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. This is a potential limitation of the application conducted in this study. Additional discussion about under-diagnosis can be found in the last section.

This article focuses on education-differentials, both because they are relevant for the study of chronic conditions and to enhance the implications of the results: education can be interpreted as a measure of socioeconomic status (Lynch 2003), and is set early on in life, which minimizes reverse causation issues. Furthermore, Smith (2005) analyzes different data sets, including HRS and AHEAD, and concludes that financial socioeconomic 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) Additional education, says Smith (2005), "is strongly and significantly predictive of the new onset of major and minor disease" (pag. 12). Education is measured in this study as "years of education," recoded in a scale from 1 to 3: 1 (0/11), 2 (12) and 3 (13+). Level 2 is the reference group.

1.6 Estimation of rates and differentials

Rates

The age-specific M, T and E rates of individuals in the reference group (individuals of 12 years of education) are estimated directly, as events over exposure, from follow up data in HRS and AHEAD. Table 1 presents a brief description of the M, T and E.

Table 1: Exposure and occurrence in M, T and E

	M	**T***	**M+E**
Exposure starts in:	Person has not a chronic condition in Wave 2	Person has not a chronic condition in Wave 2	Person has a chronic condition in any wave
Failure:	Dying	Person has a chronic condition	Dying
Exposure ends with no failure if:	Person has a chronic condition or reach Wave 7	Person die or reach Wave 7 without a chronic condition	Person reach Wave 7

*T is assumed to be an absorbing state.

The sample data used in this study 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 the long-term implications of M and E. But, in the case of T, an earlier onset of a chronic condition is certainly not uncommon: of the people interviewed in the 1994 wave of HRS and AHEAD, about 37% of respondents aged 50 to 54 reported having been diagnosed with at least

[10] 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).

one of the illnesses considered in this study. The prevalence of chronic conditions at age 50 is approximated in this study as the outcome of T alone.

Differentials

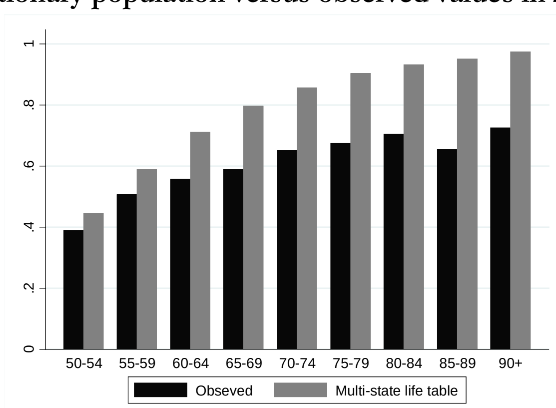
Because of sample sizes, education-specific age-specific estimates of rates become unstable. To gain stability, it is assumed that the ratio of rates (the hazard ratio) between the reference and the comparison group are constant across age intervals. Naturally, flexibility in this case is lower. To estimate the hazard ratios, one option is to pull together the data of individuals of all ages, and compute the ratios of OM or P by education, though this would also reflect the different age composition of those groups. A better option is a Cox model (Cox 1972), with age as the time variable, which also has the advantage of not assuming any particular shape for the underlying hazard; no covariates are included. Regarding the differences in the T rate before age 50, the differences in the prevalence at age 50 are included directly into the simulations and interpreted as the outcome of ΔT alone.

Section 2 - Results

2.1 Descriptive statistics

The data contains 17,223 individuals, with women representing 57% of the total. The mean age is 66. During the period of observation, 28% of individuals died. The distribution by education level is 34.2%, 33.4% and 32.4% for individuals of less than 12 years of education, 12 years (the reference group) and more than 12 years. The proportion with at least one chronic condition at the start of the follow period, shown in Figure 1, grows from 37% in the 50-54 age interval to a maximum of 70% in the 80-84 age interval.

Figure 1: Prevalence of individuals with at least one chronic condition: Stationary population versus observed values in 2002



Source: Author's calculations.

Nevertheless, it is important to emphasize that the long run implications of the currents rates are analyzed in this study by means of a multi-state life table. The observed proportion with a chronic condition, in the other hand, is the outcome of current and past values of M, T and E and only if current and past values of those rates are equal, then the observed proportions will be consistent with the proportions implied by the multi-state life table. Figure 1 shows the prevalence of individuals with at least one chronic condition, at the start of the follow up period, and the same prevalence but as implied by the stationary population implicit in the multi-state life table, which are clearly different.

2.2 Incidence and mortality rates by chronic condition

Regarding the magnitudes of the rates, Table 2 contains the number of total failures (either deaths or transitions) and total person years; the ratio of those values provides the "crude" values of M, T and E, and age-specific values of those rates can be found on Appendix A.3. The death rate of individuals without a chronic conditions is in all cases very close to 0.02 per person year (in the year 2000, the overall death rates of individuals of 50 or more years old was 0.025, according to data from the UN 2008 World Population Prospect database), while the death rate of individuals without *any* chronic condition is 0.01, probably because this last group is much more selected (indeed, person years without the condition is in this group half that of the other groups).

The excess mortality rate is more heterogeneous, from a 0.04 for hypertension, to 0.09 for stroke, therefore the mortality of individuals with a chronic condition (M+E) goes approximately from 0.06 to 0.11. Regarding incidence rates, the most likely transition is that of hypertension (0.04 per person year), and lung disease is the least one (0.01 per person year). The chronic conditions on Table 2 appear in the order of their prevalence; as expected, higher prevalence happens with higher incidence rates and lower excess mortality rates.

Table 2: Cox's Hazard ratios by education level[11]

	Hyper tension	Hearth attack	Diabetes	Lung disease	Cancer	Stroke	Any of them
Mortality without a chronic condition (M)[a]							
Failures	*1,949*	*2,580*	*3,596*	*3,851*	*3,654*	*3,647*	*550*
Person years	*82,571*	*119,832*	*131,987*	*141,163*	*136,938*	*142,134*	*54,344*
HR by education (reference group: medium education):							
Low	1.06	1.17*	1.16*	1.20*	1.24*	1.21*	1.10
High	0.80*	0.89*	0.87*	0.91*	0.87*	0.90*	0.76*
Incidence rate (T)							
Failures	*3,236*	*2,654*	*1,642*	*1,111*	*1,662*	*1,508*	*3,781*
Person years	*76,640*	*111,084*	*121,700*	*130,312*	*126,404*	*131,372*	*51,058*
HR by education (reference group: medium education):							
Low	1.10*	1.13*	1.48*	1.30*	0.88*	1.16*	1.10*
High	0.93	1.01	0.87*	0.79*	1.09	0.91	0.97
Excess mortality rate (E)[a]							
Failures	*2,754*	*2,078*	*1,111*	*851*	*991*	*1,021*	*4,132*
Person years	*70,319*	*33,105*	*20,879*	*11,765*	*15,919*	*10,819*	*98,655*
HR by education (reference group: medium education):							
Low	1.73*	1.20*	1.14*	1.09	1.23*	1.09	1.22*
High	1.33	0.90*	1.20	0.89	0.93	0.87	0.97*

* p-value<=5%. T refers to the transition rate from not having a chronic condition to having a chronic condition. E refers to the excess mortality among individuals with a chronic condition. The last category, "any of them" refers to individuals with one or more of the chronic conditions included in this study. [a]: For individuals with chronic conditions, the Cox's hazard ratios are estimated for M+E, and then the Cox's hazard ratios for M are used to derive the hazard ratios for E alone; basically, the ratio of high versus medium

[11] Table 2 displays values of hazard ratios estimated using data on US elderly, namely individuals of 50 or more years old. When computing the effect of T+ΔT, the prevalence of the chronic conditions at age 50-54 on each education group is taken directly from the data, which is assumed to reflect the action of T+ ΔT before age 50.

education's M+E can be expressed as the ratio on M, times R, plus the ratio on E1, times (1-R), where is R is M/(M+E). Therefore $HR_E=(HR_M+E - HR_M \cdot R)/(1-R)$.

Table 2 also contains the Cox model's estimation of hazard ratios (HR). On 16 out of 21 Cox regressions, the low to meddle education ratio is bigger than the rich to medium ratio. All the hazard ratios show a negative relationship between education and chronic conditions' incidence and mortality rates, except for three cases. Hypertension and Diabetes' excess mortality show a U-shape, and Cancer incidence rates' association with education is positive. Though apparently there are no comparable figures on the literature, De Gaudemaris et al (2002) found an inverted U-shape on prevalence, which corresponds with the U-shape on excess mortality found in this study (however, the levels of education used on De Gaudemaris et al (2002) are different than the ones in this study). In the case of Diabetes, a comparable figure is found in Dray-Spira (2010), where the mortality of adults with and without Diabetes by education level are shown on page 1202, and whose difference correspond to excess mortality; in the case of not cardio vascular diseases (the most common ones) there is a U shape, but not in the case of cardio vascular disease; this disagreement could be related to age standards (US population on Dray-Spira 2010, the implied stable population in this study) and the difference data sets (National Health Interview Survey from 1986 to 1996 on Dray-Spira 2010). The case of Cancer's incidence, finally, is also found previously in the literature; Smith (2005) points that "In all cases except cancer (which looks like an equal opportunity disease), the effects of schooling are preventative against disease onset"; lower incidence and higher excess should produce a lower prevalence in the long run; relatedly Johnson et al (2010) identified only two cancers, out of thirteen, that had link between higher prevalence and lower education level.

However, even in the case where the hazard ratios of T and E appear substantially similar, their influence on the long term prevalence and on the overall mortality can be substantially different.

2.3 Results

Table 3 show the results of using Das Gupta (1993) to decompose the education differentials on long-term prevalence and overall mortality produced by education differentials on M, T and E. Only the low versus medium education case is analyzed next. Appendix 4 contains the results for the high versus medium education case. This is both for simplicity and because the latter case raises issues regarding the significance of the hazard ratios showed on Table 2.

The total education differential corresponds exactly to the summation of each component of the decomposition. Each chronic condition is analyzed separately. Results on Table 3 should be interpreted from a long run perspective, as they pertain to hypothetic cohorts that go across life experiencing the age-specific rates estimated from the data.

Regarding the education differential on the long term prevalence of chronic conditions, the component associated with ΔM (the education differential on M) is very low, meaning that education differentials on the base line mortality, which could be interpreted as differentials not related to the particular chronic condition analyzed, are comparatively small. Also, the component associated with ΔT is substantially bigger than that of ΔE in all cases but Cancer; in other words, it is evident that incidence rather than excess mortality rates "explains" most of the education differentials on the long term prevalence of chronic conditions. Even in the case of the two most prevalent chronic conditions, Hypertension and Hearth attack, where the hazard ratios of E are bigger than that of T (see Table 2), the later are apparently the most influential.

Two reasons explain this result. First, incidence rates, unlike mortality rates, exert their influence since early stages of life; indeed, Hoffman et al. (1996) concluded that the majority of persons with chronic conditions are not elderly. Second, incidence reduces the "healthy" group and enlarges the "unhealthy" one, while excess mortality only reduces the "unhealthy" group.

It also worth noting that even though the order of the chronic conditions on Table 2 follows the order of the prevalence, the influences of T and E onto the prevalence does not show a clear pattern.

Table 3: Components of education differentials (Low versus medium education)

ΔP = Education differentials on the prevalence[a]				
	Total	ΔM	ΔT	ΔE
Hyperion	7.36	-0.35	9.06	-1.35
Hearth attack	3.67	-0.92	5.25	-0.66
Diabetes	8.63	-0.46	9.49	-0.4
Lung disease	3.14	-0.31	3.69	-0.24
Cancer	-1.39	-0.76	-0.14	-0.49
Stroke	0.77	-0.68	1.65	-0.2
Any of them	6.71	-0.29	8.04	-1.04
ΔOM = Education differentials on overall mortality[a]				
	Total	ΔM	ΔT	ΔE
Hyper tension	3.7	1.4	0.5	1.8
Hearth attack	4.9	3.7	0.7	0.5
Diabetes	5.4	3.7	1.4	0.3
Lung disease	5.6	4.6	0.8	0.2
Cancer	5.8	5.5	0.0	0.3
Stroke	5.6	5.0	0.5	0.1
Any of them	3.8	1.0	0.9	1.9

a: Prevalence is measured in percentual points, b: mortality is measured as number of deaths per thousand persona years. *The decomposition is based on Das Gupta (1993) . M refers to the basic mortality rate that people with and without the chronic conditions experience. T refers to the transition rate from not having a chronic condition to having a chronic condition. E refers to the excess mortality among individuals with a chronic condition. The last category, "any of them" refers to individuals with one or more chronic conditions.

In the case of the overall mortality, the situation is different. First, ΔM are the most influential components, meaning that no chronic condition by itself can explain the most of the educational differences in mortality. To some extent, this is due because the effect of incidence on the overall mortality is indirect, done throughout P, and the effect of excess mortality, in the other hand, is partially self counteracted, as discussed earlier, by its diminishing effect on the prevalence.

Second, the contribution of each chronic condition to the differential in the overall morality is much evenly distributed between ΔT and ΔE. Furthermore, in several cases excess mortality seems more responsible for education differentials on the over mortality. In the case of Hypertension and Hearth attack, excess mortality is more influential than incidence.

In Tables 2 and 3, each chronic condition is analyzed separately. But thirty-five percent of unhealthy individuals in Wave 2 have two or more diagnosed sicknesses, therefore the conclusions from each independent analysis might be distorted. An "any of them" category is added, meaning individuals with any one or more chronic conditions, is added at the bottom of Table 3; the results for this category go along the same lines than before: base line mortality explaining most of ΔOM but a not much of ΔP, and incidence being more important than excess mortality in the case of ΔP but not in the case of ΔOM.

Section 3 - Discussion

The incidence and excess mortality rates associated with chronic conditions, as well as the mortality rates not associated with the chronic conditions, differ by education groups. Understanding how those differentials articulate is fundamental to understand better the levels of prevalence by education, but also, by the same model is possible to understand better the levels of the overall mortality by education. However, this analysis is rarely done. In terms of policy, this idea highlights that the relative importance of differentials on incidence and excess mortality rates depends on the target of interest, prevalence or overall mortality.

The analysis in this study helps answering three questions. First, it is the incidence of a chronic condition or its associated extra mortality the most influential factor on the education differential on mortality?. Second, are the dynamics of the prevalence and the overall mortality substantially different?. And third, how associated, overall, is a given chronic condition with the education differential on mortality?.

The case of chronic conditions is especially interesting. Chronic conditions rank first on US causes of death, but the dynamics of the prevalence have not being analyzed, let alone the dynamic of its impact on the overall mortality, even though chronic condition have, for the most part, no reverse transition, which makes its multi-state analysis specially simple. I decompose the education-differentials on prevalence and overall mortality associated with six chronic conditions among the US elderly as included in the HRS survey, namely Hyper tension, Hearth attack, Diabetes, Cancer, Stroke and lung disease. Mortality among "unhealthy" individuals is conceptualized as the summation of mortality among "healthy" individuals (base line mortality) plus an extra mortality rate, in the spirit of interpreting only the extra mortality rate as the one associated with a chronic condition. Multi-state life tables provide long terms values of prevalence and mortality by education, that in turn are decomposed using Das Gupta (1993).

I found that a) educational differences on the base line mortality explain most of educational differences on the overall mortality (ΔOM) but a not much of educational differences on the prevalence (ΔP) and b) differences on incidence are more important than differences on excess mortality, but only with regard to ΔP, not to ΔOM; furthermore, in the case of Hypertension and Hearth attack, excess mortality rates explain most of the ΔOM associated the chronic condition.

In the long terms simulation of this study, Hypertension is the most prevalent chronic condition. Its low-to-medium education excess mortality ratio is the biggest of all chronic conditions. Its incidence ratio, in the other hand, is clearly lower than its excess mortality ratio. Yet, incidence is the main mechanism behind the educational differences on the prevalence of Hypertension, while excess mortality is the main influence of Hypertension into the educational differences on the overall mortality.

The case of Cancer appears instructive. The lower incidence and higher excess mortality of individuals in the low education group predicts a relatively *lower* prevalence but also a relatively *higher* overall mortality.

That incidence rates explain most of the educational differences in the prevalence of chronic conditions is not surprising. The conceptual section in this study shows that incidence modifies the size of both the "healthy" and the "unhealthy" group, while excess mortality modifies only the latter. In two of the few comparable studies, sex and education differences in the prevalence of disability were found to be mostly related to incidence rates Leveille et al. (2000) and Melzer et al. (2001).

The relative influence of incidence and excess mortality onto overall mortality is fundamentally different, because the former has only a direct effect on the overall mortality while the later has both a direct and an indirect effect on the overall morality.

Two important empirical limitations of this study are the dynamics before age 50 and the existence of under diagnosis. The data used in this study includes individuals of at least 50 years old at the start of the follow up period, and no mortality before that age is assumed. Although mortality rates before age 50 probably plays a minor role, this omission underscores the role of differentials on excess mortality rates, and overstates the role of differentials on incidence rates. Under-diagnosis among the poor, on the other hand, overstates the estimated mortality of healthy individuals and understates the rate of transition into unhealthiness. It likely also overstates mortality among the unhealthy. Thus, the effect of education differentials on incidence and excess mortality are under and over estimated respectively, which distorts the empirical analysis of this study.

Acknowledgments

Special thanks to Samuel Preston, Michel Guillot, Jere Behrman, Hans-Peter Kholer and Jason Schnittker, for invaluable help and support. I am also indebted to Hiram Beltran and Matthew Ruther.

References

[1.] Albano, JD., Ward, E., Jemal, A., Anderson, R., Cokkinides, VE., Murray, T., Henley, J., Liff, J., Thun, MJ. 2007. Cancer mortality in the United States by education level and race. *J Natl Cancer Inst.* (18):1384–94.

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

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

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

[5.] Chaturvedi, N., Jarrett, J., Shipley, M., and J. Fuller. 1998. 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.* 316:100-105.

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

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

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

[9.] Das Gupta P. Current Population Reports. U.S. Bureau of the Census; Washington, DC: 1993. "Standardization and Decomposition of Rates: A User's Manual" Series P–23, No 186.

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

[11.] Dellinger, A., Langlois, J., Li, G. 2002. Fatal Crashes among Older Drivers: Decomposition of Rates into Contributing Factors. *Am. J. Epidemiol.* 155(3): 234-241.

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

[13.] Dray-Spira, R., Gary-Webb, T., Brancati, F. 2010. Educational Disparities in Mortality Among Adults With Diabetes in the U.S. *Diabetes Care.* 33:1200-1205.

[14.] Freedman, V., Schoeni, R., Martin, L. 2007. Chronic Conditions and the Decline in Late-Life Disability. *Demography* 44.3 (2007): 459-477.

[15.] De Gaudemaris, R., Lang, T., Chatellier, G., et al. 2002. Socioeconomic inequalities in hypertension prevalence and care: the IHPAF study. *Hypertension* 2002; 39: 1119–25.

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

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

[18.] Hoffman, C., Rice, D., Sung, HY. 1996. Persons with chronic conditions: their prevalence and costs. *JAMA*. 276: 1473-1479.

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

[20.] Johnson, S., Corsten, MJ, McDonald, JT., Gupta, M. 2010. Cancer prevalence and education by cancer site: logistic regression analysis. *J Otolaryngol Head Neck Surg.* 39(5):555-60.

[21.] Leveille, S., Penninx, B., Melzer, D., Izmirlian, G., Guralnik, JM. 2000. Sex differences in the prevalence of mobility disability in old age: the dynamics of incidence, recovery, and mortality. *J Gerontol B Psychol Sci Soc Sci.* ;55(1):S41–S50.

[22.] Lipscombe, L., Hux, J. 2007. Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995–2005: a population-based study. *Lancet*. 369 (9563) 750–756.

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

[24.] Mathers CD, Sadana R, Salomon JA, Murray CJ, and A.D. Lopez. 2001. Healthy life expectancy in 191 countries. 357:1685-91.

[25.] Melzer, D., Izmirlian, G., Leveille, S., Guralnik JM. 2001. Educational differences in the prevalence of mobility disability in old age: the dynamics of incidence, mortality, and recovery. *J Gerontol B Psychol Sci Soc Sci*. ;56:S294–S301.

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

[27.] Patten, S., Lee, R. 2005. Describing the longitudinal course of major depression using Markov models: data integration across three national surveys. *Popul Health Metr* 2005, 3:11.

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

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

[30.] Wu, Shin-Yi and Anthony Green. Projection of Chronic Illness Prevalence and Cost Inflation. *RAND Corporation*. October 2000.

[31.] Zhu, H. and X. Xie. 2007. 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*. Vol. 29, No. 2, 125-143.

A1. Equation of the Multi-state model

The formulas used in the study are detailed here. Assume a generic chronic condition, and use "healthy" and "unhealthy" to notate individuals without and with the chronic condition, respectively. One year age intervals are used. At the beginning of each age interval, indexed by *a*, the number of healthy and unhealthy individuals are noted $l_{h,a}$ and $l_{u,a}$. During that interval, there is an age-specific probability that a healthy individual will die during the interval ($q_{h,a}$), an age-specific probability that an unhealthy individuals will die during the interval ($q_{u,a}$), and an age-specific probability that a healthy individual will became unhealthy during the interval ($q_{t,a}$). Only one event happens at the same interval, that is, healthy individuals do not transit into unhealthiness and then die during the same interval. No reverse transition is assumed.

As discussed for example in Preston, Hauveline and Guillot (2001), the transition probabilities are given by the age-specific rates of mortality among unhealthy individuals (M_a), excess mortality among unhealthy individuals (E_a) and the transition rate into unhealthiness (T_a):

$$q_{h,a} = M_a/(1+(M_a+T_a)/2) \quad (1.1)$$
$$q_{t,a} = T_a/(1+(M_a+T_a)/2) \quad (1.2)$$
$$q_{u,a} = (M_a+E_a)/(1+(M_a+E_a)/2) \quad (1.3)$$

The expressions for number of individuals surviving and person years lived depend on the assumption for the underlying *instantaneous* forces of mortality and transition. Given that one year intervals are used in this study, different assumptions probably will not substantially change the results. It is assumed here that events happen half the way through the interval to facilitate derivations. Therefore the person years and number of individuals alive at each age and health state are:

$l_{h,a} \quad =l_{h,a-1}(1-q_h-q_t)$ (1.4)
$l_{u,a} \quad =l_{u,a-1}(1-q_u)+ l_{h,a-1}(q_t)$ (1.5)
$PY_{h,a} \quad = l_{h,a-1}(1-q_h-q_t)+ l_{h,a-1}(q_h+q_t)/2$ (1.6)
$PY_{u,a} \quad = l_{u,a-1}(1-q_u)+ l_{u,a-1}(q_u)*1/2+l_{h,a-1}(q_t)/2$ (1.7)

The minimum age for the exercise in this study is 50 years old. The starting proportion unhealthy at that age is estimated from the proportions unhealthy in the 50-54 age interval as observed in the data.

A2. Derivatives of P

The goal of this Appendix is to prove that P's derivative is unambiguously positive with respect to T and unambiguously negative with respect to E.

Assume a cohort that goes under the same M, T and E rates every year, say, until year 100-th year, were the cohort is basically extinct. Assume that in year *x*-th, only, the T rate increases by ΔT. The derivative of $PY_{h,x}$ and $PY_{u,x}$ with respect to T are:

$d(PY_{h,x})/dT \quad = l_{h,x-1}(-dq_h/dT-dq_t/dT)/2 \quad <0$ (1.8)
$d(PY_{u,x})/dT \quad = l_{h,x-1}(dq_t/dT)/2 \quad >0$ (1.9)

It is fairly straightforward to prove that dq_t/dT is positive and $(-dq_h/dT-dq_t/dT)$ is negative. Therefore in the year *x*-th the effect of ΔT will increase the person years lived unhealthy and decrease the healthy ones. Thus the x-specific P will increase. Besides, $l_{h,x+1}$ decrease and $l_{u,x+1}$ increase, thus the (x+1)-specific P will also increase; indeed, it can be proved that each of the future l_h and l_u will experience the same kind of change, implying a higher n-th P. Thus, the overall P will be higher.

If T increases by Δ in all years, instead of only in one, then each effect adds to the other. In consequence dP/dT is positive. Along the same lines, this conclusion can be generalized to the case with *a*-specific M,T and E rates, which correspond to the empirical application in this exercise.

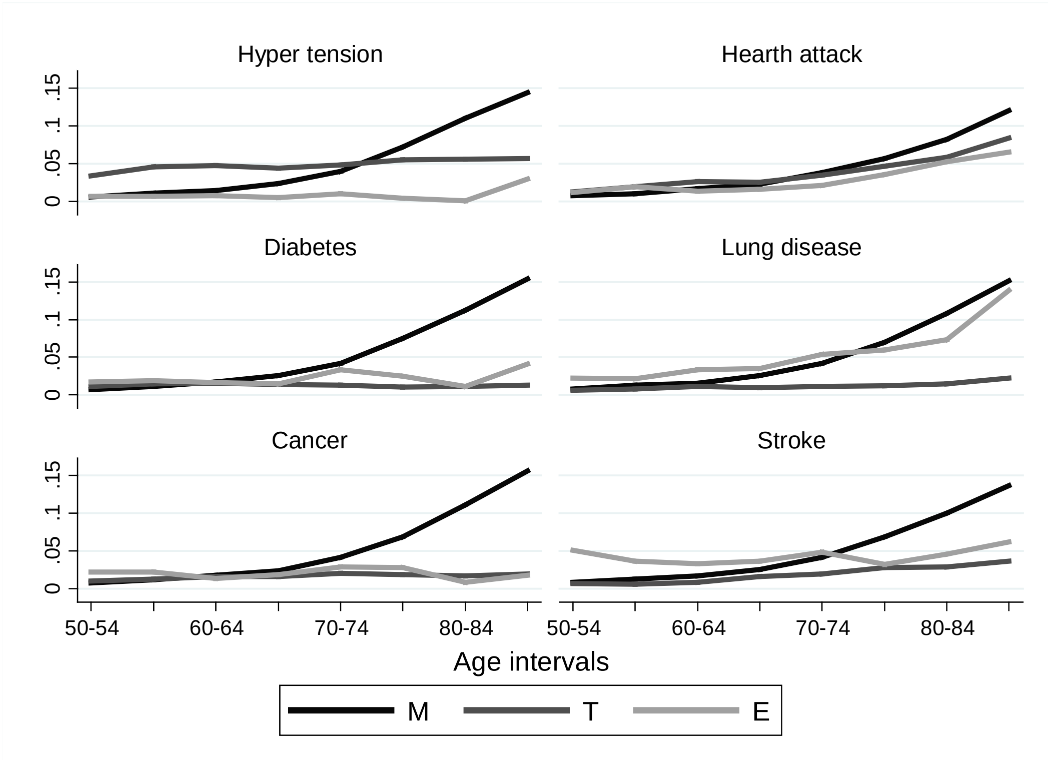
The effect of ΔE is simpler, because it influences only one the three probabilities. Assume that in year *x*-th, only, the E rate increases by ΔE. The derivative of $PY_{h,x}$ and $PY_{u,x}$ with respect to T are:

$d(PY_{h,x})/dE \quad = 0$ (1.10)
$d(PY_{u,x})/dE \quad = l_{u,a-1}(-dq_u/dE) \quad <0$ (1.11)

it is fairly straightforward to prove that $-dq_u/dT$ is negative. Given that only unhealthy person years are affected, *x*-th P will decrease. Besides, $l_{u,x+1}$ decrease. Therefore (*x+1*)-specific P decrease. Along these lines, it can be proved that each of the future l_u will experience the same kind of change, and therefore have a lower higher year-specific P. Thus, the overall P will be lower. If E increases by Δ in all years, then each effect adds to the other. In consequence dP/dE is negative. Again, this conclusion can be generalized to the case with *a*-specific M,T and E rates.

A3. Observed age-specific rates

Figure 2: Observed age-specific rates in the medium-education group.



*Source: author's calculations.

A4. Additional Tables

Table 4: Components of education differential (High versus medium education)

ΔP = Education differentials on the prevalence[a]				
	Total	ΔM	ΔT	ΔE
Hyperion	-2.4	1.4	-3.2	-0.7
Hearth attack	0.2	0.7	-0.9	0.4
Diabetes	-2.2	0.3	-2.0	-0.5
Lung disease	-1.0	0.1	-1.4	0.3
Cancer	1.9	0.6	1.1	0.2
Stroke	-0.2	0.4	-0.8	0.3
Any of them	-0.9	1.0	-2.1	0.2
ΔOM = Education differentials on overall mortality[a]				
	Total	ΔM	ΔT	ΔE
Hyper tension	-4.1	-4.6	-0.2	0.7
Hearth attack	-2.8	-2.5	-0.1	-0.2
Diabetes	-3.0	-3.0	-0.3	0.3
Lung disease	-2.6	-2.1	-0.3	-0.2
Cancer	-2.9	-3.0	0.2	-0.1
Stroke	-2.9	-2.5	-0.2	-0.2
Any of them	-3.1	-2.6	-0.2	-0.3

a: Prevalence is measured in percentual points, b: mortality is measured as number of deaths per thousand persona years. *The decomposition is based on Das Gupta (1993) . M refers to the basic mortality rate that people with and without the chronic conditions experience. T refers to the transition rate from not having a chronic condition to having a chronic condition. E refers to the excess mortality among individuals with a chronic condition. The last category, “any of them” refers to individuals with one or more chronic conditions.