

# Limited Life Expectancy, Human Capital and Health Investments

Emily Oster

Ira Shoulson

Ray Dorsey

University of Chicago and NBER

Georgetown University

Johns Hopkins University

September 10, 2012

## Abstract

Human capital theory predicts that life expectancy will impact human capital attainment. We estimate this relationship using variation in life expectancy driven by Huntington disease, an inherited neurological disorder. We compare investments for individuals who have *ex-ante* identical risks of HD but differ in disease realization. Individuals with the HD mutation complete less education and job training. The elasticity of demand for college attendance with respect to life expectancy is around 1.0. We relate this to cross-country and over-time differences in education. We use smoking and cancer screening data to test the corollary that health capital responds to life expectancy.

## 1 Introduction

Global life expectancy has increased by almost 20 years over the past five decades. These increases have contributed hugely to gains in well-being world wide (Becker, Philipson and Soares, 2005) and in the US (Murphy and Topel, 2006). They have also been linked theoretically<sup>1</sup> and empirically<sup>2</sup> to economic growth, though important recent work questions whether the connection is causal (Acemoglu and Johnson, 2007; Lorentzen, McMillan and Wacziarg, 2008).

One key channel through which life expectancy would impact growth is human capital investment. Because human capital theory predicts that a longer life expectancy strengthens the incentive to make investments in skill acquisition (Becker, 1964; Ben-Porath, 1967), and because human capital is a major input to growth (Mankiw, Romer and Weil, 1992), many economists have modeled a causal effect of life expectancy on growth via a human capital investment channel (i.e. Soares, 2005; Kalemli-Ozcan, 2002; Kalemli-Ozcan et al, 2005; de la Croix and Licandro, 1999).

---

<sup>1</sup>See, for example: Barro and Sala-i-Martin, 1995; Kalemli-Ozcan, 2002; Kalemli-Ozcan, Ryder and Weil, 2000; Young, 2005; Soares, 2005; Chakraborty, 2004; Hazan and Zoabi, 2006; Zhang, Zhang and Lee, 2003; de la Croix and Licandro, 1999; Echevarria, 2003; Weil, 2007; Meltzer, 1992

<sup>2</sup>Barro and Sala-i-Martin, 1995; Bloom, Canning and Sevilla, 2004; Meltzer, 1992

Such theories hinge critically on the strength of the microeconomic relationship between human capital investment and life expectancy, which has proved difficult to convincingly estimate. Cross sectional analyses show mixed results, and are subject to the concern that groups with lower mortality may have higher education for many reasons (Hazan, 2012; Hazan, 2009; Soares, 2006; Ram and Shultz, 1979; Meltzer, 1992). Fortson (2011) and Jayachandran and Lleras-Muney (2009) address these identification concerns using variation over time in mortality environment. But in both cases the changes in life expectancy are small and probabilistic; extrapolating from these small changes to larger differences across countries or over time may be difficult.

In this paper, we estimate the impact of life expectancy on human capital investment using data on individuals at risk for Huntington disease (HD).<sup>3</sup> HD is caused by an inherited genetic mutation; individuals with one parent with HD have a 50% chance of inheriting the affected copy of the gene and developing the disease. We can use variation across individuals *within* this population to estimate causal effects. Individuals with HD have a life expectancy of around 60. The scale of this variation is a close match to the macroeconomic changes of interest in the growth literature. Because the life expectancy differences are realized later in life, they also provide a good match to the recent gains in life expectancy, many of which have been at older ages (Eggleston and Fuchs, 2012).

This paper makes three contributions. First, we provide a strong and direct test of the qualitative assumption behind human capital theory; namely, that variation in life expectancy drives human capital investment. We find that it does. Second, we use our estimates to provide an elasticity of demand for education and job training with respect to life expectancy. We then use this elasticity to estimate the importance of life expectancy in explaining variation over time and across countries in educational attainment. Finally, we use this population to briefly consider whether limited life expectancy limits investments in long-term preventative health behaviors (Grossman, 1972; Dow, Philipson and Sala-i-Martin, 1999).

We begin with education and job training. Our sample includes individuals who have a 50% chance of inheriting the HD mutation. Our data is a cross section of individuals, but questions about

---

<sup>3</sup>Our analysis is similar in spirit to Stoler and Meltzer (2012) who use a population of individuals at risk for or with HD and compare educational attainment for individuals who learned about their risk before or after they turned 18. They find lower education for the early-learning group. Their results are intriguing, but their sample size of 56 individuals is limiting. Our study is able to go significantly further due to a much larger sample size and use of genetic testing and age of onset variation to identify precisely the pattern of information revelation. Further, we have a larger number of outcomes and are better able to adjust within sample for age and other controls. These advantages are crucial for making stronger causal statements and for calculating magnitudes, both of which are central contributions of the paper.

the history of their illness allow us to exploit variation in the timing of information revelation. We can estimate the impact of knowledge at the time decisions are made on those choices. We do this using two sources of variation in knowledge: genetic testing and early symptom development.

To analyze the impact of genetic testing, we limit to individuals who are tested at a young age prior to HD symptoms and therefore should perceive similar life expectancy before testing regardless of true status. Pre-testing characteristics are balanced across individuals with varying gene status. We compare subsequent educational attainment between those who learn they carry the HD mutation and those who learn they do not. We do not focus on the comparison between these groups and the non-testers, since the choice to test is unusual and such estimates would be confounded by selection into testing.

We find that individuals who learn they do carry the HD mutation complete much less education: they are 30 percentage points less likely to complete college. We undertake several robustness tests. There are no differences in high school completion, a decision which is made before testing for all individuals. Further, education does not differ across test results for individuals who engaged in predictive testing at older ages, after the time that education would be complete.

To analyze the impact of symptom onset we use the fact that individuals develop HD at different ages and, since onset is slow, we observe individuals who know they have HD but are not yet significantly disabled. We demonstrate that individuals with varying age of onset are similar on *ex ante* characteristics<sup>4</sup>, and that early symptoms themselves are unlikely to impact behavior. We ask whether behaviors differ for individuals who had symptom onset before they made education or job training decisions versus after.

Earlier symptom onset is associated with less education. Relative to individuals with no symptoms by 30, those with symptoms between 15 and 18 are less likely to begin college and much less likely to complete it. Individuals with symptoms between 19 and 22, and between 23 and 28, are no less likely to start college but less likely to complete. The effects scale with age of onset, indicating divergence of educational investment as HD status is revealed. The analysis of job training demonstrates similar results, even controlling for occupation.

As with genetic testing, we show that high school completion is unaffected by symptoms.

---

<sup>4</sup>The one exception to this is lack of balance on current age. Individuals who have earlier symptom onset are younger. This is mechanical: they can't be older, since they would be dead. We control for age, and argue that because this relationship is mechanical it doesn't suggest some larger issue.

Further, although the sample size is very small, we show some suggestive evidence of effects on education within sibling groups. We show that there is no impact of age of onset on routine job training – only training for advancement or promotion – suggesting those results are not driven by variation across job types.

These results provide strong qualitative evidence of one of human capital theory’s most basic predictions.. In Section 6, we discuss the quantitative implications. We use detailed information on survival and disability of HD individuals to estimate variation in life years and returns to human capital investment across groups. We combine this with our coefficients to produce elasticity estimates. We estimate the elasticity of demand for education with respect to life expectancy is around 1.0 and the elasticity of demand for job training is around 1.1.

We consider two applications. First, in Section 6.2 we apply the education elasticity to the model specified in Acemoglu and Johnson (2007) and show how they might be used to evaluate the appropriateness of the neoclassical growth model in their setting. Our magnitudes suggest it is unlikely that a neoclassical growth model explains their findings.

Second, in Section 6.3 we ask what share of the difference in college completion, either over time in the US or across countries, might be explained by differences in life expectancy. Over time within the US changes in life expectancy account for between 8 and 20% of the increase in college completion rates. Across countries life expectancy differences explain about 20% of the difference in tertiary enrollment rates. This suggests that changes in life expectancy would be expected to impact growth through the human capital channel.

In the final section of this paper we use our data to address investment in health capital (Grossman, 1972). Individuals in the sample who expect to die from HD have very little incentive to invest in costly behaviors which would prevent other fatal illnesses (Dow et al, 1999). The existing empirical evidence on health capital is even more limited than the evidence on education.<sup>5</sup> We use data on smoking and cancer screening. Consistent with theory, individuals who learn they carry the HD mutation through genetic testing or symptom onset are much less likely to quit smoking than comparable individuals without this information. Those with earlier symptom onset are less likely to have ever undergone cancer screening (conditional on age). From a policy standpoint, these results suggest positive spillovers from health improvements (Yarnoff, 2011).

---

<sup>5</sup>Dow et al (1999) provide suggestive evidence from a vaccination campaign. Oster (2012) estimates the impact of other diseases on HIV avoidance behavior and finds some evidence of these complementarities.

## 2 Background and Incentives<sup>6</sup>

HD is a degenerative neurological disorder that clinically affects an estimated 30,000 individuals in the United States. Symptoms include involuntary movement, impaired cognition and psychiatric disturbances. Typical onset is early middle age, but the range of onset is wide. This is illustrated in Figure 1. Individuals with HD will need increasing levels of supportive and institutional care for many years. Death follows approximately 20 years after onset. HD is a genetic disorder due to an excessive expansion in the *Huntingtin* gene on chromosome 4.

The disease is inherited in an autosomal dominant manner: individuals who have a parent with HD have a 50% chance of having inherited the genetic expansion and subsequently developing the disease.<sup>7</sup> There is no cure for HD or treatment that slows the progression, and symptomatic treatments are limited. The fact that HD has such clear and strong genetic predisposition means individuals are frequently aware of their family history and genetic risk.<sup>8</sup>

Importantly for our analysis, the progression of HD is slow. Early symptoms of HD are informative but not fully debilitating. Since 1993, a test for the HD genetic mutation has been available. Since everyone with the expansion will eventually develop HD, this test is predictive. However, testing rates are fairly low: 5-10% of the at-risk population report predictive testing (Shoulson and Young, 2011; Oster, Shoulson and Dorsey, *forthcoming*).

The analysis in this paper is based on the premise that individuals with HD or at risk for HD have limited incentives to invest in education or engage in preventative health behavior. In Appendix A we provide detailed calculations of life expectancy, disability-adjusted life expectancy and monetary and utility returns to human capital investment by HD group (Appendix Table A1 summarizes). This demonstrates clearly that individuals with the HD mutation have limited incentive to invest in human capital.

---

<sup>6</sup>In this section we provide only a brief overview of Huntington disease; for a fuller clinical discussion, please see Shoulson and Young (2011).

<sup>7</sup>HD development occurs with an expansion of 40 repeats or more. A small percentage of people (less than 3% of the HD population) have 36 to 39 repeats; these individuals may or may not live long enough to develop HD. In addition, men in this group may have children who “add” to the expansion so are in the 40+ category. These are both rare; from the standpoint of this paper it is reasonable to think of a simple inheritance pattern.

<sup>8</sup>It is, of course, possible that people may not know of their risk until they are older, since parents’ age of onset may be late or their parents may die of something other than HD before onset.

### 3 Data

#### Survey

Our data come from individuals enrolled in the Cooperative Huntington Observational Research Trial (COHORT), an ongoing observational study of individuals in the HD community conducted at 44 research sites. Data on two of our outcomes (education and smoking) are collected as part of the overall COHORT protocol and are available for all individuals. Data on job training and cancer screening were collected as part of a sub-study called the “Analysis of Life Decisions (ALD)” Survey, which was designed for the purposes of this paper. This data was collected from a subset of enrolled individuals.

COHORT enrollment is open to six types of people in the HD community: individuals who already have symptoms of HD (manifest HD), individuals who know they carry the HD gene but have not developed symptoms (pre-manifest HD), individuals who have a parent with HD but have not been tested and do not have symptoms (at-risk), individuals with a grandparent with HD (secondary risk), individuals who were at risk but have been tested and know they *do not* carry the HD mutation, and individuals who were never at risk for HD. The final category includes, for example, spouses of individuals affected by HD; we do not use this final group in our analysis.

The first column in Panel A of Table 1 shows counts of people in each group in the overall COHORT sample. The largest group is individuals with HD. There is significant variation within this group in whether they have symptoms, the degree of symptoms and the age of symptom onset; we will use this variation going forward. The group of individuals at secondary risk is so small that we will drop them. The second column in Panel A of Table 1 shows the number of people covered by the smaller ALD sub-study; this is the sample for which we will have data on cancer screening and job training. In this case the number of people who are tested and do not carry the mutation is too small to allow for comparison across tested individuals.

The selection of individuals into this sample is non-random. Most individuals were recruited at doctor’s visits or through enrolled family members. In some cases, recruitment was done at support group gatherings for individuals in the HD community, or through online chat boards. This brings the concern that any impacts we estimate may be specific to an especially engaged sub-sample of the HD population. This group could react more (or less) to their gene status. This is not a threat to

the *internal* validity of our estimates, but could affect our ability to make more general statements. There is also a concern that selection might vary across HD risk groups. We will address this in more detail below.

## **Human Capital Measures**

Our human capital outcomes are educational attainment and job training. In the primary COHORT survey individuals are asked about their educational attainment category, ranging from less than 9th grade through professional degree. We highlight decisions which are made after high school. To measure job training we asked individuals if they had ever enrolled in a job training program since starting to work and, if yes, what the reason was. We code individuals as engaging in job training as human capital investment if they report job training for “promotion or job advancement”. Summary statistics for the human capital variables appear in Panel B of Table 1.

## **Health Behavior Measures**

In the case of health behaviors (Section 7), we focus on three outcomes: smoking, mammogram (breast cancer screening) and colonoscopy (colon cancer screening). For smoking, we consider both whether the individual currently smokes and whether they currently smoke *conditional on* ever smoking. Smoking data is collected in the primary COHORT study.

Data on mammogram and colonoscopy come from the ALD survey. Individuals are asked whether they have ever undergone one of these screenings and, if yes, how recently. We define two outcomes for each screening: ever screened, and on-schedule screening (within the last year for mammograms, within the last 5 for colonoscopy). Because screening is recommended only for older individuals, we limit the mammogram analysis to individuals over 35, and the colonoscopy analysis to individuals over 40.<sup>9</sup> Summary statistics for the health measures appear in Panel B of Table 1.

## **Demographics and Disease Status**

Demographic variables (age, sex, etc.) come from the COHORT questionnaire. In addition to these demographics, we use information from investigator evaluations about current level of motor

---

<sup>9</sup>Many guidelines recommend starting cancer screening at 40, and colon cancer screening at 50. In our data, screening rates are similar for younger people. Given this, for sample size reasons we use the more inclusive age cutoffs, and include age controls to adjust for differences in screening rates by age.

symptoms and self-reported date of first symptoms. These variables are summarized in Panel C of Table 1. Of note in this sample, over 80% of individuals carry the HD mutation. The fact that this is greater than 50% is due to the construction of the sample, which includes people who already have symptoms.

## 4 Empirical Strategy

Broadly, the empirical strategy in this paper relies on variation in information revelation about gene status. Our population includes only individuals who begin life at 50% risk for HD. Information can be revealed in two ways. The first is through genetic testing among asymptomatic individuals. After testing, some individuals learn they carry the mutation and some learn they do not. Subsequent decisions should reflect this information difference. Second, information on HD status can be revealed by symptom onset. In this case the variation is in the *timing* of the information. An individual who learns they carry the HD mutation at 20 should be less likely to complete college than someone who doesn't learn until 30 since the latter is likely to have made their education decision prior to knowing that they would develop the disease.

The first of these analyses is closer to the ideal experiment since individuals differ only in the information they have, not in their symptoms. Making the second analysis convincing requires demonstrating that early HD symptoms themselves do not impact behavior; we will explore this in Section 4.2. However, both analyses fall slightly short of the ideal experiment because of the sample construction. The COHORT study is not a random sample of the HD population. If the type of individual who enrolls in the study varies by HD status (with the HD gene versus without, early versus late symptom onset) this could affect our results. We will provide a number of pieces of evidence, both in this section and with the results, that although this is a possible concern in theory, in practice it does not drive our results.

### 4.1 Primary Empirical Analysis

#### Identification based on Genetic Testing

Our simplest analysis uses data on genetic testing among asymptomatic individuals. This analysis will be possible only for education (not job training) since that information was collected as part of

the overall COHORT study, resulting in a larger sample size.

We identify individuals who undertook predictive testing (i.e. testing while asymptomatic) when they were young. We include individuals who were tested at or before the age of 28 and did not display symptoms before 35. Testing is sufficiently rare at young ages that we cannot separate within this group. We choose 28 as the cutoff because, based on census data, 80% of individuals who ever complete college will have done so by this age (only about 40% will have completed by age 22). Details of that calculation are in Appendix Figure B1.

Define an educational outcome for individual  $i$  as  $D_i$ . Define three groups: individuals who were tested and learned they do not carry the mutation, individuals who were tested and learned they do carry the mutation, and at-risk individuals who were not tested early. The last group is used as the omitted category, and we report results from a regression of the following form, where  $\mathbf{X}_i$  is a vector of controls:

$$D_i = \alpha + \beta_1(\textit{Tested Negative})_i + \beta_2(\textit{Tested Positive})_i + \mathbf{\Gamma}\mathbf{X}_i + \epsilon_i$$

The key test in the regression is whether the behavior of those who test negative differs from those who test positive (i.e. the test of the null  $\beta_1 = \beta_2$ ). Although we will discuss issues of balancing more below, it is useful to note here that the absolute magnitudes of  $\beta_1$  and  $\beta_2$  – that is, the difference between either tested group and those who are untested – are unlikely to be especially informative. The choice to test predicatively for HD is unusual; individuals who choose to be tested are likely different psychologically than those who do not. In other work (Oster et al, *forthcoming*; Oster et al, 2008) we have analyzed the genetic testing decision and shown that individuals who choose to be tested differ in their expectations and their reasons for initial testing avoidance. This makes the comparison between testers and non-testers difficult and likely not very meaningful.

Including the non-testers is nevertheless helpful because it allows us to control more flexibly for demographics (particularly age). In an appendix we will show results including only the tested individuals with a simpler age control; they are very similar.

Related to this, although we argue these estimates are internally valid, this is subject to the concern that because the testers are unusual they may also be more responsive. This is worth keeping in mind and we will find, in fact, that the magnitude of response is larger here than for the

symptom analysis.

## **Identification based on Symptoms**

Our second identification strategy relies on variation across individuals in the timing of symptom onset.

Consider the decision about college attendance: what is relevant is what the individual knows about her status at the time this decision is made. In our data we observe the timing of symptom onset, but do not observe the exact timing of schooling decisions. We attempt to construct an “event study” based on the age of symptom onset: groups with earlier onset should be more heavily affected since those individuals are most likely to have experienced symptoms before making the decisions.

Our analysis focuses on education after high school. We define four groups of individuals: those with symptom onset between 15 and 18, those with symptom onset between 19 and 22, those with symptoms between 23 and 28 and those with no symptoms by age 30. The first group should be mostly heavily affected; they should virtually all know their status before even starting college. The older the onset, the less affected individuals should be. The final group, with no symptoms by 30, should not have their education affected by HD status.

In the case of job training the timing is less well-defined, but we note that most training occurs when individuals are young. We define the most heavily affected group as those with onset between 20 and 30, the possibly affected group as those with onset between 30 and 40 and the unaffected group as those with no symptoms by 40.

## **4.2 Balancing and Impact of Symptoms**

### **Balancing and Sample Selection**

Our sample is not a random sample of the HD population. This leads to the concern that, for example, the kind of person who wants to be in the study despite testing negative may be different than the type who wants to be in having tested positive. Or, in the case of symptoms, the kind of person who wants to be in the sample when having symptoms young is different than those who want to be in the sample given older symptom onset. Table 2 begins to address this concern through balancing tests. We compare groups on several variables which are determined *before* any impact of HD (gender, age, location, race, which parent had HD).

Panel A of Table 2 shows balancing based for the genetic testing analysis. Columns 1 and 2 show that the gene-positive and gene-negative groups are very similar on pre-testing characteristics. In Column 3 we report means of these variables for the untested individuals who are included in these regressions. As we noted, the key test is between the two tested groups; this table illustrates one reason why comparing the tested to untested individuals is not helpful. They lack balance on gender, age and location. We see something similar when we look at ever-tested individuals in Columns 4, 5 and 6 of Panel A (these individuals will be used in the smoking analysis).

It may seem puzzling that anyone who tests negative would be willing to participate in a study of this sort – for most diseases, individuals who do not have the disease are not interested in taking time for a study of it – and this may lead to concerns that even if they look balanced these individuals may be very odd. In fact, in this case the continued participation is likely due to the fact that this disease runs in families. Individuals who test negative are likely to have family members (parents, siblings, cousins) who are affected by the disease. They are asked to participate in the study to help those individuals and the family connection likely increases the share of people willing to do that.

Panel B of Table 2 shows balancing on symptom onset age. To test for balance we run regressions of each variable (male, age, etc) on age-of-onset groups. This allows us to look for trends in the data: are individuals with earlier onset more white, or more male? The only consistent pattern we find is with age: the younger the symptom onset, the younger the individual at the survey time. This occurs since people only enroll when they are alive so those with earlier onset are on average younger. We address this by controlling for age. We note that this lack of balance is mechanical, so there is no reason to expect that it reflects some broader difference across groups.

In Appendix Table B1 we show summary statistics and t-tests for balancing across individual symptom groups, with identical conclusions.

### **Impact of Symptoms on Behavior**

A second concern which comes up in the results based on age of symptom onset is that the changes in behavior are due to symptoms themselves. Although we attempt to focus on decisions made when symptoms are fairly minor, this remains a concern.

One specific concern is with mental degeneration. Later stage HD is characterized by cognitive

declines. If this cognitive decline is significant earlier in the disease course, then it is possible that differences in (for example) college completion rates could be due to differences in ability rather than differences in choices. Figure 2 shows individual scores on two tests of cognitive ability (a “mini-mental state exam” and a broader battery of cognitive tests) and individual motor scores graphed against years from symptom onset.<sup>10</sup> Over the early years of symptoms we observe only limited declines in either measure of cognitive ability. This suggests mental abilities remain at least somewhat intact as information about status is revealed by physical symptoms.

Although this is encouraging, these tests provide only one measure of mental capacity. In addition, physical symptoms could limit education (for example, if individuals have trouble transporting themselves to school). We provide a partial solution to this by looking at working behavior among the early symptom onset groups. If these individuals do not attend college due to disability, we would not expect them to work instead, since working requires a similar level of functioning. If we do see employment history it suggests they choose to work rather than go to school.<sup>11</sup>

Using this methodology, we do find some evidence that this matters for the youngest symptom onset group: only 80% of them ever work, versus 97% of the group without early symptoms. The issue is much more limited for the older onset groups, where 95% of individuals ever work. This suggests the results for the latter two groups are probably not affected by this issue. To address this concern for the youngest symptom group, we will use these working figures to adjust our magnitude calculations when we discuss the symptom results. An alternative would be to simply drop individuals who never work from the analysis; the results when we do this are extremely similar (available from the author).

A final issue is the incidence of depression in this population. It stands to reason that knowing one faces limited life expectancy would impact mood, and it is possible mood impacts could drive changes in behavior. To the extent that our concern is about generating causal impacts, it is not as clear this is a confound: if lowering life expectancy in other ways also impacts mood, this effect *should* be included. Of course, the mechanism is different.

---

<sup>10</sup>The mini-mental state exam measures orientation to time and space, word recall, calculation skill (with subtraction by 7s from 100) and so on. Cognitive tests include a symbol digit modality test (pairing numbers with figures) and a verbal fluency test (say as many words as you can in 60 seconds).

<sup>11</sup>This may, in fact, overstate the extent of disability since people may choose not to work for the same reasons they choose not to go to school.

In practice, in our data, HD does not have as large an impact on depression as might be expected. Among individuals in the study with symptoms of HD, 50.3% of them report some depression. This is compared to 45.6% of individuals who are at risk but without symptoms, and 49.9% of the non-HD individuals in the study (spouses or partners of individuals in the study who are not themselves at risk). We see very similar figures for reported anxiety – 55.2% of those with symptoms, 54.7% of those at risk and 54.9% of the non-HD group. These figures suggest that the depression confound may be more limited than we would expect based on introspection.

## 5 Results: Impact of HD on Human Capital Investment

**Education** We begin with the analysis based on genetic testing. A simple comparison of means suggests a large impact of test results on educational attainment. Individuals who learn through genetic testing that they carry the HD mutation get, on average, 14.3 total years of schooling and 36% of them complete a bachelor degree. In contrast, those who learn they have a negative test result get an average of 15.03 years of schooling and 63% of them complete a bachelor degree. In simple t-tests the years-of-schooling differences are significant at the 5% level and the bachelor degree results are significant at the 1% level. These differences persist in regressions with demographic controls. Panel A of Table 3 shows our main results. Recall that the relevant test is the difference in coefficients for those who test and learn they do carry the mutation versus those who test and learn they do not. Columns 1 and 2 show large differences in both years of schooling and the chance of a bachelor degree, with those who learn they do not carry the HD mutation completing more schooling. Columns 3 and 4 control for method of recruitment into the study; the results are essentially unchanged. Appendix Table B2 replicates these results including only tested individuals with the same qualitative and quantitative conclusions.

In Panel B of Table 3 we turn to falsification. The first column here illustrates that, on an educational decision which is made prior to testing, test results have no impact. Both groups are similarly likely to finish high school. The second set of columns use data on individuals who underwent asymptomatic genetic testing when they were older. If we are worried about differences in selection of gene-positive versus gene-negative individuals into the sample, those concerns also apply to this population. However, these individuals were all untested at the time of education decisions.

Therefore, if the impacts we observe in Panel A reflect response to test results, we would not expect to see differences in this older group. This is what we find. Both tested groups are higher human capital than untested individuals (presumably due to selection) but there are no differences between them in educational attainment.

Figure 3 gives a visual sense of the symptom-based education results. We illustrate the chance of completing each level of schooling by age of symptom onset. All three groups are similarly likely to complete high school. The youngest age-of-onset group diverges at the start of college and continues to do so over time. The older age-of-onset groups diverge at the bachelor degree stage, to differing degrees. The figure starkly demonstrates the fanning out of educational attainment as HD status is revealed.

Table 4 shows corresponding regressions. The evidence is consistent with Figure 3. The youngest symptom onset group starts to diverge as early as starting college; this makes sense, since this group contains individuals with onset in the period before college enrollment. The two groups with onset after 18 do not diverge much on starting college, but do diverge when we look at bachelor degree completion. All groups complete fewer total years of education post-high school. The differences are quite large. As a falsification test Column 4 looks at high school completion. There are no differences across groups in this variable.

For some individuals in the COHORT study information was collected on their family members and, in a number of cases, members of the family are also in the study. This allows us to identify sibling groups and, in a few cases, we see groups in which siblings had onset at varying ages. The sample size here is very small, a total of about 25 individuals, which makes it very difficult to make any strong statements. However, Appendix Figure B2 shows educational choices within matching sibling pairs. The siblings with onset before 28 are no less likely to finish high school or start college, but are much less likely to complete a degree. Conditional on starting college, college completion rates are significantly lower (at the 5% level) for the early-onset group.

Tables 3 and 4 suggest qualitative support for human capital theory. Turning to magnitudes, we narrow in on college completion, which is the outcome we will analyze in the discussion of elasticities and applications. In order to interpret the magnitude of the coefficient it is important to take into account that the share of individuals *eligible* for the college completion decision will differ across groups.

Consider a population where we expect 60% of healthy individuals to complete college. Effectively none of them will have completed college by the age of 18. Therefore the effect for a group with symptom onset between 15 and 18 should be evaluated relative to a possible completion rate of 60%. When we look at an older onset group (say, those with onset between 23 and 28), however, many people will already have completed college by the time onset occurs. We should evaluate relative to a smaller eligible share.

We would like to report the share of people who are dissuaded from college completion by the HD information. The inputs to this magnitude calculation are the regression coefficients, an estimate of the completion rate of the healthiest (gene-negative or no-early-symptoms) group and information on the timing of college completion from Appendix Figure B1.

In the testing analysis, our coefficient implies an 82% reduction in completion among eligible individuals. Magnitude estimates are shown for the symptom onset groups at the bottom of Table 4. The first figures (outside of square brackets) show the basic calculation. The figures inside the square brackets adjust for disability. The need for this adjustment was noted in Section 4.2: to take into account the possibility that some people may be unable to complete college, we adjust the figures based on working behavior.

Although the figures vary a bit by group, something between 60 and 100% of eligible individuals seem to be dissuaded from college completion by HD. The disability adjustment makes a significant difference only for the youngest group and brings them in line with the 19 to 22 onset individuals.

This analysis also allows us to speak to the concern that the effects are “too large” for the oldest symptom onset group. In fact, they are not, although our estimate does suggest virtually everyone in this group who might have completed college is discouraged from doing so by the disease. The fact that this is the largest magnitude may not be surprising given this group is probably the one with the most limited attachment to higher education.

**Job Training** Our sample size for the analysis of job training is smaller, so we cannot rely on genetic testing for identification. The results based on age of symptom onset are shown in Table 5. Our outcome of interest is job training for advancement or promotion, and we find very large impacts. Relative to those with no symptoms by 40, individuals with symptom onset between 20 and

30 are about 27 percentage points less likely to report any job training for advancement or promotion. Individuals with symptom onset between 31 and 40 are also less likely to undertake training for advancement, although the coefficient is smaller, which is consistent with the fact that these individuals could have undergone job training in their twenties.

The second column of Table 5 shows similar results including a control for job category (managerial and professional, sales, service, etc). This leaves the coefficients virtually unchanged, suggesting differences across groups in job type do not drive our results.

The third column of the table estimates the impact of symptoms on other, non-advancement oriented job training (this includes, for example, training which is “required for your job”). If individuals with earlier onset are different in some other way – perhaps holding different jobs even within job category, or working less – then we should also see differences in this overall job training. In fact, if anything, those with earlier symptom onset are *more* likely to have engaged in other, non-promotion-associated, job training. This result also helps us partially rule out the concern that people with symptoms are not offered job training by their firm because of concerns about their departure from the firm.

## 6 Elasticity Estimates and Applications

The results above address the qualitative question of whether human capital is responsive to life expectancy. We argue that the answer is yes. This is of interest on its own. However, for many applications, knowing the qualitative relationship is not sufficient. Calibration of growth models, for example, requires taking a stand on the *quantitative* relationship between these variables.

On their own, the coefficients in the tables describe response to life expectancy changes of the magnitude produced by HD. In this section we combine these estimates with data on the impact of HD on life expectancy to derive more general estimates of magnitudes which can be used in other applications. Of course drawing quantitative conclusions from these data requires stronger assumptions (namely, that the response in this population is quantitatively similar to what we would see in the general population).

It is worth briefly discussing the appropriateness of this assumption and the possible direction of bias. Broadly, we see two issues. First, is the life cycle timing of the gains in life expectancy

reflective of changes over time or across countries? In this case we think our population is fairly well-matched. For example, when comparing individuals across gene test results, the life expectancy advantage for the individual with a negative test result accrues between ages 60 and 80. For the symptom analysis, the gains are earlier, but still well into adulthood. As Eggleston and Fuchs (2012) point out, many of the gains in life expectancy over the past decades have been in older adult survival. In this sense we think a strong case can be made that our estimates reflect the impact of the “right” change in life expectancy.

A second issue is that we measure the impact of getting a positive or negative signal about life expectancy. The fact that there is an explicit revelation of information may make this context different from the cross-country comparison where life expectancy simply varies across space. This issue comes up in all studies of this type – the change in information is necessary for making causal claims, but may make generalization hard. Our estimates may be more applicable in situations where the life expectancy changes are driven by discrete changes in health care – for example, vaccinations. Although it is difficult to say for sure, it seems to us that this is likely to bias our estimates upward, since individuals may be especially responsive when the information is presented to them.

A final note, moving away from external validity. Our estimates represent partial equilibrium effects of learning about limited life expectancy in early adulthood. From the standpoint of growth models, our estimates address the question of how higher education would be affected if individuals make that choice taking into account their life expectancy. We do not capture the possible impacts of increased or decreased life expectancy on early life investments in education by parents.<sup>12</sup> To the extent that these effects also matter our figures will be an under-estimate of the overall impacts.

## 6.1 Elasticity Estimates

In this section we calculate the elasticity of demand for college or job training with respect to life years, disability-adjusted life years (DALYs) and monetary returns to schooling.

A detailed discussion of how we calculate life years, DALYs and earnings is left to Appendix A. To summarize briefly, we use information from HD “life tables” (Newcombe, 1981) to calculate HD survival after symptom onset. We combine this with information from the overall US population on

---

<sup>12</sup>We also do not capture gains associated with the fact that individuals in better health may be more able to attend school (i.e. Miguel and Kremer, 2004; Bleakley, 2007)

death rates by age to calculate life expectancy given a particular age of onset. Disability adjustment is done using the COHORT data, which allows us to calculate the level of disability by time from symptom onset. Returns to a college education are calculated based on current wages for full-time, full-year workers with a college degree versus those with a high school degree. This is combined with information on working propensity by age in the general population, and data from COHORT on the reduction in employment as symptoms progress.

These calculations take into account the timing of these decisions. For example: when we consider the college-going decision for someone with onset between 23 and 28, we calculate their life expectancy and returns starting at the age we would expect them to complete college.

Appendix Table A1 reports post-decision life expectancy, DALYs and earnings by group. It is clear that groups with earlier symptom onset have lower life expectancy and lower returns to a college degree. These differences are large: the youngest symptom onset group has only about 7 years of life expected after college completion, whereas for the HD negative group this number is 53.

In Panel A of Table 6 we show elasticity estimates. These figures represent the percentage change in either schooling or job training for a percentage change in life years, DALYs or earnings. The first row shows education elasticities based on the symptom analysis. Because there are three symptom groups (and because we show results adjusting and not adjusting for disability), there is a range reported. On average, the data suggest an elasticity of around 1. The second row shows estimates for the elasticity based on the genetic testing analysis. The estimates here are larger. The elasticity figures for job training are in a similar range, around 1.1.

Our education elasticities are similar to those found by Jayachandran and Lleras-Muney (2009): their figures are between 0.6 and 1. This is interesting and encouraging, suggesting that varying populations and methodologies may nevertheless show similar results. This comparison is straightforward since Jayachandran and Lleras-Muney (2009) calculate an elasticity directly. In Appendix C we also compare to the magnitudes in Stoler and Meltzer (2012) by converting their figures to elasticities; again, we find the magnitudes are quite similar.

When we turn to using these education elasticity estimates in applications below, we focus on the estimates based on symptom onset. As we discussed above, because only a selected sample of individuals choose to test – and they may do so precisely because they are interested in using the information to guide their behavior – it is not surprising that this group is more elastic. Their

elasticity may be representative of the likely response of a group of individuals who seek out such information. The symptom onset results are likely to be more representative of the behavior of individuals who haven't sought out information of this type.

## 6.2 An Application: Acemoglu and Johnson (2007)

A key question for which these figures are relevant is the relationship between life expectancy and growth. Human capital investment is one channel through which this effect could operate, and finding that education is responsive to life expectancy provides another reason why we might expect a positive relationship between life expectancy and growth. In addition, our figures may be useful as an input to such models. Here, we consider an example.

Acemoglu and Johnson (2007) estimate the relationship between increases in life expectancy and economic growth, instrumenting for life expectancy with disease-specific advances which impact different areas differently. They present a simple growth model with fixed capital and land stock. In this model increases in life expectancy have two impacts. They increase income per capita through increased human capital investment and productivity growth. However, they decrease income per capita by increasing population. Their model produces the following relationship (from Section 6.C. of their paper):

$$\pi = \alpha(\gamma + \eta) - (1 - \alpha)\lambda$$

where  $\alpha$  is the labor share,  $\pi$  is the response of income per capita to life expectancy,  $\lambda$  is the response of population to life expectancy,  $\eta$  is the elasticity of human capital with respect to life expectancy and  $\gamma$  is the elasticity of TFP with respect to life expectancy.

Their paper estimates  $\pi$  and  $\lambda$ . They find that population increases with increases in life expectancy and GDP per capita actually decreases, at least over the horizons they estimate. They argue that, based on the estimates in the paper and with the fairly standard assumption that  $\alpha = \frac{1}{3}$ , it must be the case that  $(\gamma + \eta)$  is in the range between -0.5 and 0.1. Other estimates in the paper suggest even lower (i.e. more negative) figures.

Acemoglu and Johnson (2007) argue that this suggests their results are rationalized by the simple neoclassical growth model only if the impacts of health on TFP and education are small. They leave it at this, and without more information it is difficult to comment more strongly

positively or negatively. Our estimates (Panel A of Table 6), however, suggest a value for  $\eta$  of about 1. To harmonize their results with a neoclassical growth model the elasticity of TFP with respect to health must be in the range of -1.5 to -0.9. In other words, it must be negative and probably somewhat elastic.

Based on what we know about the relationship between health and productivity at an individual level, this seems unlikely. This suggests (consistent with what they say in their paper) that something other than a standard neoclassical growth story may be behind their results. Put differently, given the elasticity we estimate we *would* expect improvements in health to impact growth in a neoclassical model, and the fact that they find it does not suggests an offsetting factor is driving their findings.

### 6.3 Cross Sectional and Over Time Changes in Education

In addition to the contribution to the life expectancy-growth link, there is an interesting macroeconomic question of how important life expectancy changes over time or across countries are in explaining differences in education. One way to approach this is with cross country regression (for example, Barro and Sala-i-Martin, 1995; Hazan, 2012), but such analyses run the risk of confounding by omitted variables or reverse causality. An alternate approach, which we take here, is to combine an elasticity estimate which we think is well identified with information on changes in life expectancy over time and predict differences in educational attainment. We can then ask what share of the actual differences in education are explained by this life expectancy difference.

We look at two differences: changes over time in the US, and differences across countries. For the calculations, we use an elasticity of 1.04, which is the average of the symptom-based estimates (row 1 of Panel A of Table 6).<sup>13</sup> Were we to use the elasticity based on genetic testing, the shares explained would be about twice as large.

We begin looking at this over time in the US. Using decennial census data on college completion and life expectancy, we estimate what share of the “long differences” in college completion rates might be explained by changes in life expectancy. Our college completion rates are the share of individuals aged 30 to 35 at the time of the census who have completed college; we use this age range

---

<sup>13</sup>This is the average of the estimates without the disability adjustment, although the figure is very similar if we use the numbers with the adjustment. It would also be very similar if we used the elasticity with respect to DALYs. These numbers are all in the same range.

since these individuals are young enough that they likely made their decisions based on current life expectancy, but old enough to be largely done with college completion. The life expectancy we use is future life expectancy at age 25. This is obtained from Social Security Administration data (Bell and Miller, 2005).

These results are shown in Panel B of Table 6. We document the importance of life expectancy in explaining the 1960 versus 2010 difference, as well as the 1980 and 2000 versus 2010 differences. Changes in life expectancy account for a moderate share of the differences – between 6 and 20%, depending on the time frame. In the most recent decade it seems that increases in life expectancy play a relatively smaller role in the continued growth in college-going.

To look at the role of this relationship in explaining global differences we do two things. First, we look at aggregates in high income, middle income and low income countries. We use data from the World Development Indicators on mortality and the gross share enrollment in tertiary education to generate actual and predicted differences in college-going across regions. We base our life expectancy calculation on life table data so we can generate life expectancy at age 25 (rather than at birth). The results (in Panel C of Table 6) show that life expectancy explains between 20% and 30% of the difference across these broad groups. Second, we use the same World Development Indicators data at the country level to compare each pair of countries. The last row of Panel C of Table 6 reports the average share explained, which is about 18%.

These results suggest that life expectancy plays a similar role both across time within the US and across countries. Extrapolating out, this suggests that further increases in life expectancy would have some positive impacts on education and, by extension, on growth.

## **7 Health Capital: Smoking and Cancer Screening**

Thus far we have focused on human capital. As we note in the introduction, models of health capital also have implications for behavior in this population. Individuals who carry the HD mutation are less likely to benefit from cancer screening, since the lifetime risk of cancer is lower (because they are more likely to die from HD before they would have gotten cancer). Similarly for smoking, the benefits to quitting or never starting smoking are lower if your absent-smoking life expectancy is lower. In Appendix A we support this with explicit calculations of the lifetime risk of cancer and

mortality costs of smoking by HD status.

## 7.1 Empirical Strategy

The COHORT study overall collected data on smoking behavior, and the smaller ALD sub-study collected information on mammogram and colonoscopy. As above, we use two empirical approaches: identification based on genetic testing and identification based on symptom onset.

The genetic testing identification is possible only for smoking, due to the larger sample size from the overall COHORT study. We compare current smoking rates for asymptomatic individuals who have been tested and differ in test results.

When using symptom onset, we note that for several of the health decisions we consider (smoking, current cancer screening) the decision is made contemporaneously with the survey. In those cases, our independent variable of interest is defined by whether the individual is *currently* experiencing early symptoms.

The choice to begin cancer screening is not contemporaneous, and we therefore use information about symptoms at the time when screening should have begun. For mammography, initial screening should happen around 40; we define the affected group as women with symptoms between 30 and 40 and the unaffected group as those with symptom after. For colonoscopy, initial screening should happen around 45. We define the group with symptoms 35-45 as affected, and those with symptoms after 45 as unaffected.

### Balancing and Symptom Impacts

As above, one concern with this analysis is the non-random selection of individuals into the sample. Issues of balance are addressed in the discussion of Table 2. It is worth noting here that the one place we do see significant issues with balance is in the analysis of smoking based on early symptoms versus no symptoms. We will control for demographics in this analysis, of course, but the lack of balance suggests this analysis should perhaps be taken with more caution.

Also as in the human capital analysis, we face concerns in our analysis of symptoms that it may be the symptoms themselves which drive behavior. In this case, the largest concern is with the motor symptoms. Perhaps individuals are less likely to get a mammogram because the involuntary movements which characterize the disease make screening difficult (the individual would need to be

fully sedated). In the case of smoking this bias goes in the wrong direction: motor symptoms are likely to make it harder to smoke, *opposite* from the effect human capital theory predicts.

To test this concern directly, we limit our data to individuals who report being symptomatic, so they know for sure they are sick. Within this category, motor symptoms vary. We therefore observe individuals with the same information (they all know they have HD) but varying level of physical symptoms.

We graph smoking, mammography and colonoscopy probabilities against the motor scores for this group in Figure 4. Our concern would be that over the range of early symptoms (say, motor scores under 20) individuals with higher motor scores are more likely to smoke and less likely to engage in cancer screening. In practice we see, if anything, the opposite. Smoking rates are going down, not up, as symptoms get worse. Cancer screening is largely unaffected and, if anything, seems to be going up. It is true that at very high symptom levels cancer screening drops, but this is well outside the range we consider.

A second concern, for smoking in particular, is that individuals may be using cigarettes to “self-medicate”, due to anxiety about their HD status. One argument against this is the general figures on anxiety and depression we showed in Section 4: the differences in these psychological metrics are just not that big. In addition, at least in our data, the link between smoking and anxiety among those without symptoms is fairly weak. To show this we limit the data to the “control” individuals (who are spouses or partners and not at risk for HD) and at-risk individuals with no symptoms or testing. We regress smoking for this group on a dummy for whether they report any anxiety. The coefficient is 0.0002 and is not close to significant. Together, these facts suggest to us that anxiety is unlikely to drive results.

## 7.2 Results: Health Capital

**Smoking** We begin with the analysis based on genetic testing. Twelve percent of individuals who learn from testing that they carry the HD mutation currently smoke, versus 8% of those who learn they do not carry the mutation. Conditional on ever smoking only 32.5% of those who learn they do not carry the HD mutation are still smokers, versus 54.8% of those who learn they do carry the HD mutation. This latter difference is significant at the 5% level in a simple t-test.

Columns 1 and 2 of Panel A of Table 7 show the regression analog. As with the education

analysis, the important comparison is between the two tested groups (regressions without the at-risk individuals are in Appendix Table B3). We do not see a difference overall in current smoking in this population, but we do see a difference in current smoking for those individuals who ever smoked. Those who test positive are 17 percentage points more likely to still be smoking than those who test negative.

The second set of columns in Panel A of Table 7 show evidence based on symptom variation. The comparison is between those at risk (without symptoms) and those with early HD symptoms. Those with early symptoms are consistently more likely to smoke – both more likely to smoke at all and less likely to have quit if they ever smoked.<sup>14</sup>

**Cancer Screening** The effects on cancer screening are inherently more difficult to identify because of the age profile of symptom onset. Regular colonoscopy screening is supposed to start around age 45 or 50, at which point most individuals who will develop HD symptoms have already done so (see Figure 1). This will limit our sample size, but with this smaller sample of individuals with later symptom onset we can still ask the question of whether those with earlier symptom onset are less likely to ever screen. This analysis appears in Panel B of Table 7. Individuals with earlier symptom onset are significantly less likely to start cancer screening (this estimate is largely driven by differences in colonoscopy). The point estimates suggest that those with current symptoms are also less likely to be “on schedule” with their screening, but this is not significant.

## 8 Conclusion

We argue this paper makes three contributions. First, we provide a sharp test of human capital theory and find strong support for the qualitative predictions. Individuals with truncated life expectancy complete less education and are less likely to undertake job training. We argue that these effects are causal and that the simplicity of the setting allows us to be confident about the channels driving our results.

The estimates we derive can be combined with information on life expectancy with HD to make quantitative predictions about the impact of changes in life expectancy on educational

---

<sup>14</sup>This is consistent with the result in Stoler (2005) that, among 30 individuals with HD, the smoking rate is higher than the general population.

attainment. We find an elasticity of demand for college completion with respect to life expectancy of about 1.0. This figure, in turn, can be used to ask whether this channel is important in driving global patterns of human capital investment. We argue the data suggest differences in life expectancy explain about 20% of the cross-country variation in college completion, and a similar amount of the over-time variation within the US.

Finally, we use the same data to test a corollary of the human capital theory, namely the theory of health capital or competing mortality risks. We find some support. Individuals who carry the HD mutation are more likely to smoke and less likely to engage in cancer screening than those without the mutation. To the extent that this generalizes, it suggests that health improvements may be complementary. Improving people's health in one dimension may encourage them to invest in other ways (Yarnoff, 2011); on the flip side, of course, worsening health may have negative impacts on other health behaviors (as in Oster, 2012).

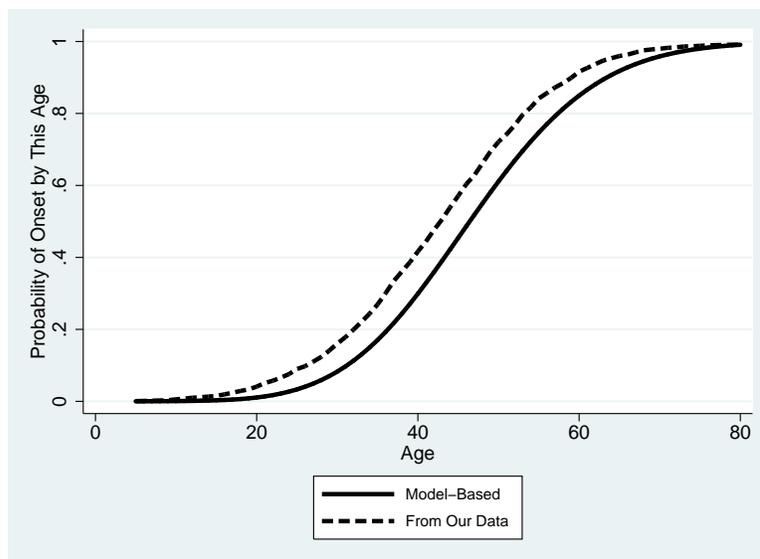
## References

- Acemoglu, Daron and Simon Johnson**, "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of Political Economy*, 2007, 115, 927–985.
- Barro, Robert and Xavier Sala i Martin**, *Economic Growth*, Cambridge, MA: The MIT Press, 1995.
- Becker, Gary**, *Human Capital*, Chicago, IL: University of Chicago Press, 1964.
- , **Thomas Philipson, and Rodrigo Soares**, "The Quantity and Quality of Life and the Evolution of World Inequality," *American Economic Review*, 2005, 95 (1), 277–291.
- Bell, Felicitie and Michael Miller**, "Life Tables for the US Social Security Area 1900-2100, ACTUARIAL STUDY NO. 120," SSA Report 2005.
- Ben-Porath, Yoram**, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, 75, 352–365.
- Bleakley, Hoyt**, "Disease and Development: Evidence from Hookworm Eradication in the American South," *Quarterly Journal of Economics*, 2007, 122, 73–117.
- Bloom, David, David Canning, and Jaypee Sevilla**, "The Effect of Health on Economic Growth: A Production Function Approach," *World Development*, 2004, 32 (1), 1–13.
- CDC**, "Worktable 23R. Death rates by 10-year age groups: United States and each state, 2007," National Vital Statistics System, Mortality Report 2010.
- Chakraborty, Shankha**, "Endogenous Lifetime and Economic Growth," *Journal of Economic Theory*, 2004, pp. 119–137.

- de la Croix, David and Omar Licandro**, “Life Expectancy and Endogenous Growth,” *Economics Letters*, 1999, *65*, 255–263.
- Dow, William, Tomas Philipson, and Xavier Sala i Martin**, “Health Investment Complementarities under Competing Risks,” *American Economic Review*, 1999, *89*, 1358–1371.
- Echevarria, Curz**, “Life Expectancy, Retirement and Endogenous Growth,” *Economic Modelling*, 2003, *21*, 147–174.
- Eggleston, Karen N. and Victor R. Fuchs**, “The New Demographic Transition: Most Gains in Life Expectancy Now Realized Late in Life,” *Journal of Economic Perspectives*, Summer 2012, *26* (3), 137–56.
- Fortson, Jane**, “Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa,” *Review of Economics and Statistics*, 2011, *93* (1), 1–15.
- Grossman, Michael**, “On the Concept of Health Capital and the Demand for Health,” *Journal of Political Economy*, 1972, *80*, 223–255.
- Hazan, Moshe**, “Longevity and Lifetime Labor Supply: Evidence and Implications,” *Econometrica*, November 2009, *77* (6), 1829–1863.
- , “Life Expectancy and Schooling: New Insights from Cross-Country Data,” *Journal of Population Economics*, 2012, *25*, 1237–1248.
- and **Hosny Zoabi**, “Does Longevity Cause Growth? A Theoretical Critique,” *mimeo*, Hebrew University, 2006.
- Heckman, James, Robert LaLonde, and Jeff Smith**, *The economics and econometrics of active labor market programs*, Elsevier, 1999.
- Jayachandran, Seema and Adriana Lleras-Muney**, “Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines,” *Quarterly Journal of Economics*, 2009, *124* (1), 349–397.
- Julian, Tiffany and Robert Kominski**, “Education and Synthetic Work-Life Earnings Estimates,” American Community Survey Reports 2011.
- Kalemli-Ozcan, Sebnem**, “Does the Mortality Decline Promote Economic Growth?,” *Journal of Economic Growth*, 2002, *7*, 411–439.
- , **Harl Ryder, and David Weil**, “Mortality Decline, Human Capital Investment, and Economic Growth,” *Journal of Development Economics*, 2000, *62*, 1–23.
- Langbehn, DR, RR Brinkman, D Falush, JS Paulsen, and MR Hayden**, “A New Model for Prediction of the Age of Onset and Penetrance for Huntington’s Disease Based on CAG Length,” *Clinical Genetics*, 2004, *65* (4), 267–277.
- Lorentzen, Peter, John McMillan, and Roman Wacziarg**, “Death and Development,” *Journal of Economic Growth*, 2008, *13*, 81–124.
- Mankiw, N Gregory, David Romer, and David N Weil**, “A Contribution to the Empirics of Economic Growth,” *The Quarterly Journal of Economics*, May 1992, *107* (2), 407–37.

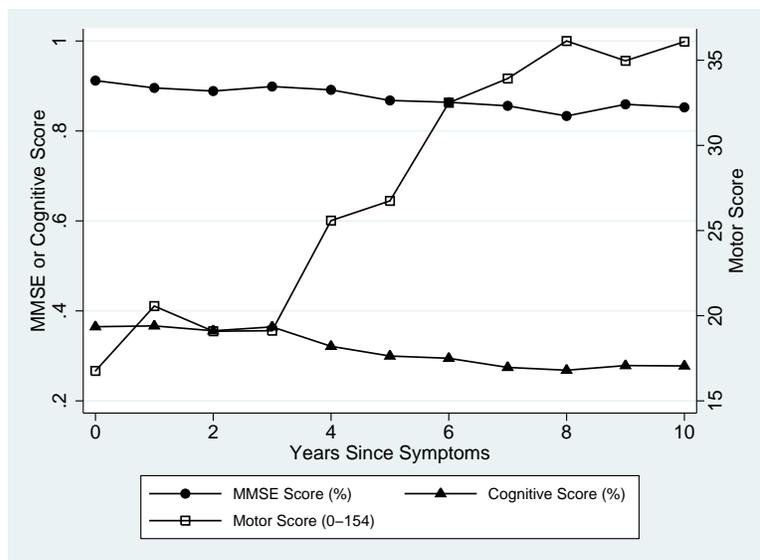
- Meltzer, David**, “Mortality Decline, the Demographic Transition and Economic Growth,” Dissertation, University of Chicago 1992.
- Miguel, Edward and Michael Kremer**, “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” *Econometrica*, 2004, 72, 159–217.
- Murphy, Kevin M. and Robert H. Topel**, “The Value of Health and Longevity,” *Journal of Political Economy*, October 2006, 114 (5), 871–904.
- Newcombe, RG**, “A life table for onset of Huntington’s chorea,” *Annals of Human Genetics*, 1981, 45, 375–385.
- Oster, Emily**, “HIV and Sexual Behavior Change: Why not Africa?,” *Journal of Health Economics*, 2012, 31, 35–49.
- , **E. Ray Dorsey, Jan Bausch, Aileen Shinaman, Elise Kayson, David Oakes, Ira Shoulson, and Kimberly Quaid**, “Fear of health insurance loss among individuals at risk for Huntington disease,” *American Journal of Medical Genetics, Part A*, 2008, 146 (16), 2070–2077.
- , **Ira Shoulson, and E. Ray Dorsey**, “Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease,” *American Economic Review*, forthcoming.
- Ram, Rati and Theodore W. Schultz**, “Life Span, Health, Savings, and Productivity,” *Economic Development and Cultural Change*, 1979, 27 (3), 399–421.
- Shoulson, Ira and Anne Young**, “Milestones in Huntington Disease,” *Movement Disorders*, 2011, 26 (6), 1127–1133.
- Soares, Rodrigo**, “Mortality Reductions, Educational Attainment and Fertility Choice,” *American Economic Review*, 2005, 95, 580–601.
- , “The effect of longevity on schooling and fertility: evidence from the Brazilian Demographic and Health Survey,” *Journal of Population Economics*, February 2006, 19 (1), 71–97.
- Stoler, Avraham**, “Mortality Risk and Economic Behavior,” Dissertation, University of Chicago 2005.
- **and David Meltzer**, “Mortality and Morbidity Risks and Economic Behavior,” *Health Economics*, 2012.
- US Census**, “Employment status of the civilian noninstitutional population by age, sex, and race,” CPS Report 2011.
- Weil, David**, “Accounting for the Effect of Health on Economic Growth,” *Quarterly Journal of Economics*, 2007, 122, 1265–1306.
- Yarnoff, Benjamin**, “Prevention versus Treatment with Competing Disease Risks,” *mimeo, RTI International*, 2011.
- Young, Alwyn**, “The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations,” *Quarterly Journal of Economics*, 2005, 120, 423–426.
- Zhang, Jie, Junsen Zhang, and Ronald Lee**, “Rising longevity, education, savings, and growth,” *Journal of Development Economics*, February 2003, 70 (1), 83–101.

**Figure 1: Age of HD Onset**



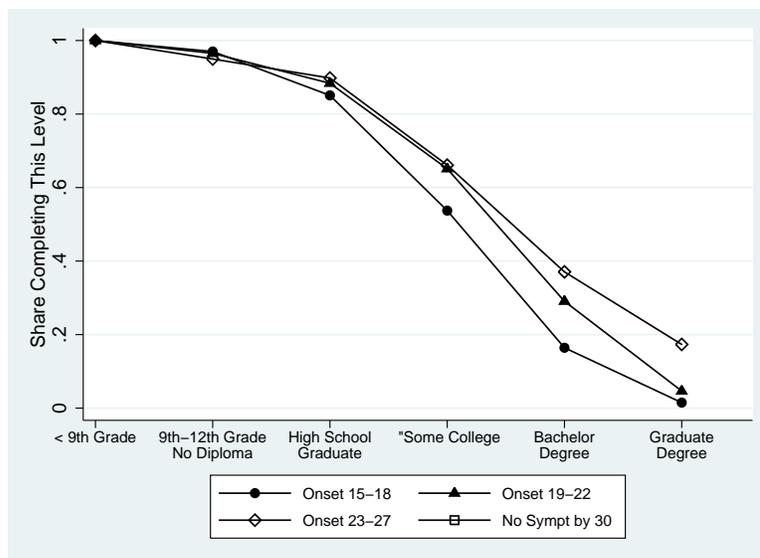
*Notes:* This graph shows information on the probability of HD onset by age. The solid line shows this probability based on the model developed in Langbehn et al (2004). This uses the CAG distribution from our data along with his equation for probability of onset by age given CAG repeats. The dotted line shows the distribution of age of onset from our sample of currently symptomatic individuals.

**Figure 2: Cognitive and Motor Symptoms by Time of Symptom Onset**



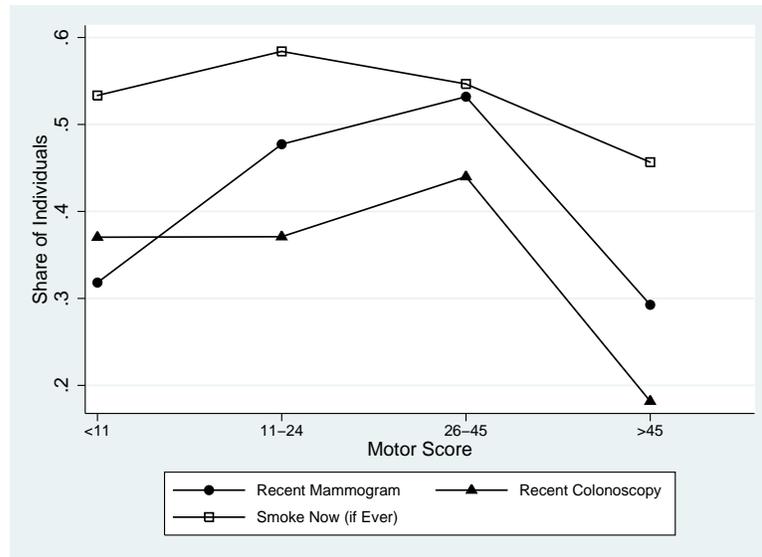
*Notes:* This graph shows motor scores and two cognitive test scores (on a mini mental state exam and a more general cognitive exam) by time from symptom onset. Time of symptom onset is identified by the participants. Motor scores are evaluated by the COHORT investigator.

**Figure 3: Educational Attainment by HD Symptom Timing**



*Notes:* This graph shows the share of individuals who have completed each level of education, by age of symptom onset. The oldest group is those who have no symptoms (and are not tested) by 30. Associate degree is coded as “Some College”.

**Figure 4: Impact of Symptoms on Health Behaviors**



*Notes:* This graph shows behavior among individuals who report that they are sure they have HD, by motor scores. Since everyone in the graph has the same information about HD status, differences in behavior can be attributed to differences in physical symptoms. In our analysis we only use individuals with limited symptoms (motor scores less than 20 or so). This graph therefore covers a much larger range of symptoms than our data.

Table 1: *Summary Statistics*

<b>Panel A: Count of COHORT Participants</b>			
	<i>All COHORT</i>	<i>ALD Sample</i>	
Manifest or Pre-Manifest HD	1,790	325	
50% Risk for HD	594	104	
25% Risk for HD	15	2	
Tested, Do not Carry Mutation	160	21	
<b>Panel B: Outcome Summary Statistics</b>			
	<i>Mean</i>	<i>Std. Dev.</i>	<i># of Obs</i>
Years of Education	13.74	2.33	2543
Bachelor Degree	0.359	0.479	2543
Job Training for Advancement (Age 20-60)	0.133	0.348	323
Smoke Now	0.154	0.361	2543
Smoke Now, if Smoke Ever=1	0.514	0.500	764
Ever Mammogram (Age $\geq$ 35)	0.781	0.414	238
Recent Mammogram (Age $\geq$ 35)	0.443	0.497	238
Ever Colonoscopy (Age $\geq$ 40)	0.452	0.498	304
Recent Colonoscopy (Age $\geq$ 40)	0.344	0.475	304
<b>Panel C: Demographic and HD Status Summary Statistics</b>			
	<i>Mean</i>	<i>Std. Dev.</i>	<i># of Obs</i>
Age	49.4	13.6	2542
Male (0/1)	0.418	0.493	2543
White	0.933	0.248	2543
Carry HD Mutation	0.818	0.385	2394
UHDRS Motor Score	22.8	22.3	2525
Any Symptoms (0/1)	0.625	0.484	2540
Age of First Symptoms (if any symptoms)	42.8	12.5	1575

Notes: This table shows simple summary statistics on sample sizes, outcomes and demographics. The ALD survey is the survey which contains questions on job training, mammogram and colonoscopy; education and smoking are covered in the primary COHORT surveys.

Table 2: *Balancing Tests*

		Panel A: Balancing with Testers		
		Categories for Smoking Analysis		
Categories for Education Analysis		Tested Ever	Not Tested	
Tested Young (<=28)	Untested at 35	(No Current Symptoms)	(No Current Symptoms)	(No Current Symptoms)
(No Symptoms by 35)	(No Symptoms by 35)	Positive	Negative	
		(n=101)	(n=34)	(n=1768)
Male (0/1)	0.306	0.294	0.424 <sup>a</sup>	0.325
White	0.960	0.941	0.929	0.977
Current Age	30.94	32.17	53.9 <sup>a,b</sup>	46.52 <sup>a</sup>
In US	0.871	0.794	0.931 <sup>a,b</sup>	0.79
Mother HD Parent	0.53	0.51	0.43	0.542
		(n=279)	(n=178)	(n=497)
		0.333	0.325	0.348
		0.959	0.977	0.900 <sup>a,b</sup>
		41.70	46.52 <sup>a</sup>	45.55 <sup>a</sup>
		0.82	0.79	0.969 <sup>a,b</sup>
		0.43	0.542	0.567
		Panel B: Balancing with Symptoms		
		Regression Coefficients: Outcome (Male, White, Etc.) on Symptom Group		
		Education	Job Training	Cancer Screening
		(Symptom Age)	(Symptom Age)	(Symptom Age)
		(n=2147)	(n=234)	(n=608)
Male (0/1)	-0.020	-0.0001	0.058	0.137*
White	-0.002	-0.020	-0.0004	0.066*
Current Age	9.97*	7.75*	10.79*	6.09*
In US	-0.018	-0.024	-0.091*	-0.032
Mother HD Parent	-0.014	0.018	.103	-0.074

Notes: This table shows balancing on early life variables. Panel A focuses on testers used in the education and smoking analysis and reports results from t-tests. In Panel A, “indicates significantly different from individuals with a positive test result and <sup>b</sup> indicates significant difference with people with a negative test result. Panel B analyzes balancing by age of onset using regressions. Each cell represents the coefficient on symptom group from a regression with the balancing variable (male, white, etc) on the left hand side. For education, job training and cancer screening the “symptom group” is created based on age of onset; for smoking, it is a dummy for having “early” symptoms versus none. For cancer we define two groups: onset age 30 to 45 and no onset by 45. The “early onset” group therefore captures the early onset group in both mammography and colonoscopy. In Panel B, \* indicates a significant coefficient at the 5% level. Appendix Table B1 shows the full balancing across all the individual groups.

Table 3: *Gene Testing and Educational Attainment*

<b>Panel A: Main Results and Robustness</b>				
	<b>Main Results</b>		<b>Control Referral Method</b>	
	# Post-HS Years Education	Bachelor Degree	# Post-HS Years Education	Bachelor Degree
<i>Sample</i>	<i>Finish HS</i>	<i>Finish HS</i>	<i>Finish HS</i>	<i>Finish HS</i>
Tested, Negative	.893** (.355)	.240*** (.091)	.876** (.355)	.235*** (.091)
Tested, Positive	.177 (.232)	-.018 (.059)	.220 (.233)	-.006 (.060)
<i>pos vs. neg p-value</i>	.06	.009	.09	.01
Standard Controls	YES	YES	YES	YES
# of Observations	1690	1690	1689	1690
<b>Panel B: Falsification</b>				
	<b>Pre-Symptom Education</b>		<b>Older Testers</b>	
	Complete High School		# Post-HS Years Education	Bachelor Degree
<i>Sample</i>			<i>Finish HS</i>	<i>Finish HS</i>
Tested, Negative	.046 (.056)		.494 (.307)	.158** (.078)
Tested, Positive	.045 (.036)		.450** (.215)	.126** (.055)
<i>pos vs. neg p-value</i>	.99		.90	.73
Standard Controls	YES		YES	YES
# of Observations	1903		1224	1224

Notes: This table shows the impact of test results on educational attainment. The omitted category in all cases is individuals who are in the HD risk group but are untested at the ages considered. The main results include individuals who are tested before the age of 30, and (if positive) do not develop symptoms before 35. Columns 3 and 4 of Panel B show results for individuals who were tested between 35 and 45, but did not have symptoms before 45. Standard controls: gender, country dummies, race, a cubic in age. Referral group control simply controls for whether the individual was recruited by a doctor (the primary recruitment method) or in another way (at a meeting, online, etc). Appendix Table B2 shows these regressions excluding the individuals who are at risk but untested. Standard errors in parentheses. \*significant at 10% \*\*significant at 5% \*\*\*significant at 1%

Table 4: *Symptom Onset and Educational Attainment*

	All			Falsification: High School Completion
	# Post-HS Years Education	Start College	Bachelor Degree	Complete High School
<i>Sample</i>	<i>Finish HS</i>	<i>Finish HS</i>	<i>Finish HS</i>	
Symptoms Age 15-18	-1.685*** (.462)	-.277*** (.105)	-.343*** (.118)	-.019 (.067)
Symptoms Age 19-22	-.992*** (.347)	-.061 (.079)	-.230*** (.088)	-.067 (.052)
Symptoms Age 23-28	-.598** (.242)	-.026 (.055)	-.118* (.061)	-.033 (.036)
<i>Impacts As Share of Eligible:</i>				
15-18			-82% [65%]	
19-22			-62% [59%]	
23-28			-101% [99%]	
Controls	YES	YES	YES	YES
# of Observations	1923	1923	1923	2146

Notes: This table reports estimates of the difference in education by age of symptom onset. The omitted category is individuals who are at risk but do not have HD symptoms by age 30. Controls: gender, country dummies, race, a cubic in age. The final rows report the impacts relative to what share of the population would be expected to still be eligible for this choice at the time of symptom onset. This re-scales the coefficients to reflect the fact that the older the onset the more likely that the education was completed before onset occurred. Details of this calculation are in the text. A value of -0.82 indicates that there we would expect an 82% reduction in college completion with symptom onset. Magnitude figures in square brackets represent magnitudes adjusted for disability, as measured by working behavior. Standard errors in parentheses.\*significant at 10% \*\*significant at 5% \*\*\*significant at 1%.

Table 5: *Symptom Onset and Job Training*

	Main Results: Job Training for Promotion or Advancement		Falsification: Other Job Training
<i>Sample</i>	<i>Age 20-60, Ever Worked</i>	<i>Age 20-60, Ever Worked</i>	<i>Age 20-60, Ever Worked</i>
Symptoms Age 20-30	-.269** (.109)	-.261** (.110)	.150 (.161)
Symptoms Age 31-40	-.127** (.062)	-.129** (.062)	.023 (.091)
Standard Controls	YES	YES	YES
Job Type Controls	NO	YES	YES
# of Observations	234	234	234

Standard errors in parentheses.\*significant at 10% \*\*significant at 5% \*\*\*significant at 1%

Notes: This table reports estimates of the difference in job training by age of symptom onset. The omitted category is individuals who are at risk but do not have HD symptoms by age 40. Controls: gender, country dummies, race, a cubic in age and dummies for category of education completed are included in all regressions. In Columns 2 and 3 we also include dummies for job category (managerial, service, etc).

Table 6: *Quantitative Implications*

<b>Panel A: Elasticities</b>			
<i>Elasticity With Respect to:</i>	<i>Life Years</i>	<i>Disability-Adjusted Life Years</i>	<i>Discounted Lifetime Return to Investment (in Dollars)</i>
College Completion (Symptoms)	0.78 to 1.31	0.69 to 1.18	0.66 to 1.12
College Completion (Gene Testing)	2.28	1.88	3.39
Job Training	1.24 to 1.42	1.07 to 1.26	1.07 to 1.26
<b>Panel B: Long Differences in College Completion in US</b>			
	<i>Actual Difference in College Share</i>	<i>Difference Predicted by Life Expectancy</i>	<i>Share of Difference Explained by LE</i>
2010 vs. 1960	23.6%	3.7%	16.2%
2010 vs. 1980	10.1%	2.0%	20.1%
2010 vs. 2000	8.3%	0.5%	6.2%
<b>Panel C: Global Differences in Tertiary Enrollment</b>			
	<i>Actual Difference in College Share</i>	<i>Difference Predicted by Life Expectancy</i>	<i>Share of Difference Explained by LE</i>
High income vs. Middle Income	45.5%	9.6%	21.2%
Middle Income vs. Low Income	17.4%	3.4%	19.5%
High Income vs. Low Income	63.0%	18.2%	28.9%
All Country Pairs (Median Share)			17.9%

Notes: Panel A shows the elasticity of demand for education and job training with respect to healthy life expectancy (life expectancy calculations described in Appendix A). Panel B shows the predicted impact of changes over time in life expectancy in the US. Panel C estimates the share of global differences in tertiary enrollment which could be explained by differences in life expectancy. The final row of this column shows the average amount explained after calculating the share explained for each pair of countries. Tertiary enrollment and life expectancy data is from the World Development Indicators. Calculations in Panels B and C use an elasticity value of 1.04.

Table 7: *Life Expectancy and Health Investments*

<b>Panel A: Smoking</b>				
<i>Sample</i>	<b>Genetic Testing</b>		<b>Symptom Onset</b>	
	<i>Smoke Now</i>	<i>Smoke Now</i>	<i>Smoke Now</i>	<i>Smoke Now</i>
	<i>All</i>	<i>Ever Smoke=1</i>	<i>All</i>	<i>Ever Smoke=1</i>
Tested, Negative	-.020 (.028)	-.143* (.085)		
Tested, Positive	.009 (.028)	.032 (.072)		
Have Early Symptoms			.116*** (.033)	.255*** (.082)
<i>positive vs. negative p-value</i>	.34	.06		
Standard Controls	YES	YES	YES	YES
# of Observations	948	233	608	189

<b>Panel B: Cancer Screening</b>			
	<i>Ever Screen</i>		<i>Recent Screen</i>
Symptom Onset Before Screening Age		-.141*** (.051)	
[Mam: 30-40, Colon: 35-45]			
Have Early Symptoms Now			-.134 (.095)
Colonoscopy		-.407*** (.046)	-.165** (.081)
Standard Controls		YES	YES
# of Observations		450	144

Standard errors in parentheses. \*significant at 10% \*\*significant at 5% \*\*\*significant at 1%

Notes: This table shows the impact of life expectancy on health behavior. In the genetic testing analysis in Panel A the omitted category is individuals who are in the HD risk group but are untested. Standard controls: gender, country dummies, race, a cubic in age, dummies for education. In the symptom analysis of smoking in Panel A the omitted category is individuals who are at risk but do not have HD symptoms. Panel B reports estimates of the differences in cancer screening across groups by timing of symptom onset. The omitted category in the “ever screen” analysis is individuals with symptoms after 40 (for mammography) or 45 (for colonoscopy) and, in the “recent screen” analysis, individuals with no symptoms. Mammography data is limited to women. Data used in Panel B may have multiple observations per individual if they reported on both colonoscopy and mammogram. Standard errors in Panel B are clustered by individual.

# Appendices: For Online Publication Only

## Appendix A: Incentive Calculations

In section 6 we calculate the elasticity of demand for education and job training with respect to three outcomes: life years, disability-adjusted life years and discounted monetary return to investment. In this appendix we outline the details of these calculations.

Appendix Table A1 provides the summary numbers for each group. For the symptom onset groups we simply do the described calculations for each particular age in the group and provide an average for the group. For the gene-positive and at-risk individuals we combine this type of calculation with information about age of onset. For example: for individuals who are gene-positive but without symptoms at 35 (the requirements for the education analysis) we calculate the probability of onset at each age after 35 given no symptoms at 35. We then calculate life expectancy for each age of onset, and average weighting based on the probability of onset by age. For at risk individuals we do a similar calculation but include as one of the possible outcomes the chance that they *do not* have the HD gene. The weight for this probability depends on the age at which they are still at risk without symptoms. Someone without symptoms at 40 is less likely to carry the gene than someone without symptoms at 30. This can be visualized based on Figure 1, and is taken into account in the analysis.

### A.1 Life Years

Life expectancy is calculated based on information on the life course of HD. We use life table information from Newcombe (1981) on survival by years of symptom onset. To give a sense, he calculates that 4% of individuals will have died in the first four years, half by the end of 15 years and 95% by 30 years from onset. In addition, we take into account the fact that people might die from something else prior to HD. We incorporate overall death rates from US life tables as our background death probabilities (CDC, 2010).

As we describe in Section 6, we calculate for each group the years of life after the possible completion of the human capital investment (either education or job training). This is important because, for example, the magnitude of the impact on education in the 23 to 27-year-old onset group is calculated relative to the share of individuals who are “at risk” for completion in that period. But those individuals will not accrue benefits to college starting at age 25; in fact, among individuals who have not completed college by this time, expected age of completion is 31. We calculate life years (or earnings returns) after this period.

Denote  $\hat{a}$  as the age at which the investment would be completed. Further, define the probability of dying at each age as  $d_a$ ; this probability is based on the time from symptoms and the background death rate. Life years after age  $\hat{a}$  are calculated as:

$$\left[ \sum_{a=\hat{a}}^T (d_a)(a) \right] - \hat{a}$$

Column 1 of Table A1 reports this life year calculation by group. The figures in square parentheses in the symptom group labels indicate  $\hat{a}$  used for that group.

## A.2 Disability-Adjusted Life Years

It is common in health economics to focus not only on life years but also on disability-adjusted life years. The idea is that individuals will not (and, hence, policy makers should not) value all years equally. If someone is in a vegetative state their quality of life is assumed to be low, and should not be counted as equal to a year of perfect health.

To calculate disability-adjusted life years in this case we need to take a stand on the quality of life for individuals with varying symptom levels. There is no standard way to do this, but our data provides a natural way to calculate these numbers. In particular, the COHORT data includes a measure of “Total Functional Capacity” which ranges on a scale from 0 to 13. A score of 13 is fully able, and a score of 0 indicates completely disabled. The questions making up the score include ones about ability to work, get around in every day life, do normal activities, etc. We divide this by 13 to get a measure of capacity on a range from 0 to 1 and model the relationship between total functional capacity and years from symptoms with a quadratic. We use this model to generate a measure of disability for all time from symptoms.

Disability adjusted life years are then simply calculated by adjusting the life year calculations so later years with greater symptom levels are worth less. Column 2 of Table A1 reports disability adjusted life years by group.

## A. 3 Earnings

We calculate the lifetime returns to a bachelor degree (for education) or to job training (for job training analyses). In addition to the information on time to death described in Subsection A.1 above, this return calculation uses data on earnings by group and working probability by year.

*Earnings* For education, we use Census Bureau calculations for full-time, full-year workers to calculate the earnings for individuals with a college degree and those without (Julian and Kominski, 2011). For job training, we rely on existing literature, and assume that a job training program has a benefit of 2.5% in wages in perpetuity<sup>15</sup>. We do the job training calculations assuming an individual has some college but no degree, which is the education level of the average individual in our sample. The expected earnings for someone with some college, no degree, are drawn again from the Census Bureau calculations.

*Probability of Working* The probability of working has two components. First, for individuals with HD symptoms we need to know the probability they will be working by time from symptom onset. In addition, we need to know the general chance of working by age. Since even non-HD individuals do not work with probability 1, if we assume that everyone without HD works all the time we will overstate the differences across groups.

For the general population we use data from the Census bureau on employment status by age (US Census, 2011). For the HD population we again use data from COHORT. We look at the probability of working by time since symptom onset. It is worth noting that this conflates leaving work because of being physically unable to continue with leaving work because of wanting more leisure time. From the standpoint of calculating monetary returns to investment, however, it isn't clear that it matters *why* individuals are not working, only that they are not.

We combine the earnings, probability of working and probability of death to calculate a return to getting a college degree or undertaking job training. We assume 3% discounting from  $\hat{c}$ . Column 3

---

<sup>15</sup>There are many varying estimates of the impact of job training (see a review in Heckman, Lalonde and Smith, 1999). The exact value is not crucial here; we take 2.5% as a conservative estimate of these impacts. Any positive impact of job training which has lasting impacts would give similar results.

of Table A1 reports the earnings return to college completion by group; Column 4 reports the earnings return to job training.

**Table A1: Incentives for Education and Job Training**

	<i>Life Years</i>	<i>DALYs</i>	<i>Return to College</i>	<i>Return to Job Training</i>
	(1)	(2)	(3)	(4)
<i>Groups for Gene Testing Analysis</i>				
Without Gene [ $\hat{\alpha} = 25$ ]	53.1	53.1	\$412,199	
With Gene, No Symptoms at 35 [ $\hat{\alpha} = 25$ ]	33.8	29.7	\$311,485	
<i>Groups for Education Symptom Analysis</i>				
Symptoms 15-18 [ $\hat{\alpha} = 25$ ]	7.44	4.26	\$23,161	
Symptoms 19-22 [ $\hat{\alpha} = 25$ ]	10.39	6.51	\$38,565	
Symptoms 23-28 [ $\hat{\alpha} = 31$ ]	8.46	5.01	\$28,517	
At risk at 30 [ $\hat{\alpha} = 25$ ]	41.67	39.66	\$336,495	
<i>Groups for Job Training Analysis</i>				
Symptoms 20-29 [ $\hat{\alpha} = 25$ ]	13.59	9.24		\$3534
Symptoms 30-39 [ $\hat{\alpha} = 35$ ]	13.27	9.04		\$3605
At risk at 40 [ $\hat{\alpha} = 25$ ]	40.62	38.92		\$15,138

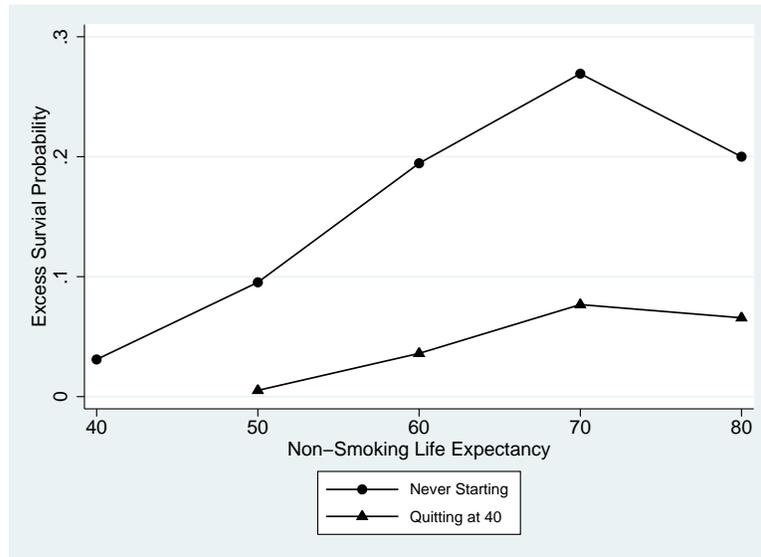
#### A.4. Smoking

Summarizing the incentives to quit smoking (or not start at all) is a bit more complicated. In general, cigarette smoking leads to an increased probability of death from a variety of causes, including lung cancer and emphysema. Much of this excess death occurs later in life, so a shortened lifespan will limit the incentive to avoid smoking. Calculating detailed returns by group as we did for education and job training is difficult, because there are many possible behaviors – never start smoking, start early and quit early, start early and don’t quit, start late, smoke more, smoke less, etc. What we can do, however, without needing such detailed information on exact smoking behavior, is give a sense of the benefits to either (a) never smoking or (b) quitting for people with varying life expectancy. The evidence in Table A1 demonstrates that life expectancy is limited for people with HD; if the benefits to not smoking or quitting are larger for those with greater life expectancy, we can conclude there is some added incentive to quit for those without HD.

We use data on total probability of death by age for smokers, non-smokers and former smokers from [www.smokefree.gov](http://www.smokefree.gov). As our benchmark, we consider someone who begins smoking at age 20, and smokes one pack a day. We calculate the excess probability of death for individuals with alternative life expectancy of 40, 50, 60, 70 and 80; that is, we calculate the excess chance of dying from smoking if you knew you would die at age 40 independent of the cigarettes. We look both at the benefits to never starting to smoke, and the benefits to quitting at age 40 (again, assuming an age 20 start).

The gains to not smoking or quitting are shown in Appendix Figure A1. Individuals with higher non-smoking life expectancy have higher benefits from either never smoking or quitting. Since this figure shows the difference in death in percentage *points*, it decreases at the oldest ages since, ultimately, everyone dies in both groups. The impact of smoking is large. For someone who would otherwise live only to 40, the benefit of not smoking is only about a 3 percentage point difference in the chance of death. For someone who would otherwise live to 70, it is 25 percentage points. The impacts of quitting are smaller, of course, but still very different based on life expectancy.

*Figure A1: Benefits to Not Smoking*



## A.5 Cancer Screening

The simplest way to document the varying benefits to cancer screening by group is to look at the chance of ever developing breast or colon cancer by group. Since screening is an effective way to detect cancer, the benefits to screening are higher if the incidence of cancer is higher. Put differently, since both breast and colon cancer are typically fatal without treatment and survival is fairly good with treatment, we can think of the benefits to screening as scaling with incidence.

For each type of cancer we compare the chance of developing that cancer by group, where the groups are as in the regression: individuals with symptoms by screening age versus those at risk but without symptoms by that age. For breast cancer, we compare the lifetime breast cancer probability for individuals with onset between 30 and 40 to those at risk without symptoms at 40. For colon cancer, the groups are those with symptoms between 35 and 45 versus no symptoms at 45. Cancer incidence by age is taken from the SEER cancer statistics (<http://seer.cancer.gov>). We compute the lifetime risk of cancer by combining these incidence numbers with the chance of still being alive at each age.

The results are shown in Appendix Table A2. Lifetime cancer risk is much higher for the at-risk individuals. Comparing the two cancer groups, we see the percentage increase is higher for colon cancer, although the absolute numbers are higher for breast cancer, which is more common.

**Table A2: Cancer Probabilities by Group**

	<i>Chance of Ever Developing</i>	
	<i>Breast Cancer</i>	<i>Colon Cancer</i>
	(1)	(2)
<i>Groups for Breast Cancer Analysis</i>		
At risk at 40	9.4%	
Symptoms 30-40	2.3%	
<i>Groups for Colon Cancer Analysis</i>		
At risk at 45		4.2%
Symptoms 35-45		0.7%

## Appendix B: Tables and Figures

*Table B1: Balancing Across Individual Groups for Symptom Analysis*

Panel A: Education and Job Training							
	Education				Job Training		
	Grouping: Onset Age				Grouping: Onset Age		
	<i>15-18</i>	<i>19-22</i>	<i>23-28</i>	<i>Over 30</i>	<i>20-29</i>	<i>30-39</i>	<i>Over 40</i>
	<i>(n=25)</i>	<i>(n=45)</i>	<i>(n=86)</i>	<i>(n=1991)</i>	<i>(n=26)</i>	<i>(n=60)</i>	<i>(n=255)</i>
Male (0/1)	0.48	0.40	0.52	0.42	0.35	0.45	0.41
White	0.92	0.97	0.92	0.93	0.96	0.93	0.92
Current Age	32.1	33.6	36.8 <sup>a,b</sup>	53.9 <sup>a,b,c</sup>	37.9	46.2 <sup>d</sup>	58.1 <sup>d,e</sup>
In US	1	0.91	0.96	0.91	0.92	0.95	0.86
Mom HD Parent	0.56	0.56	0.50	0.51	0.52	0.47	0.51

Panel B: Health				
	Smoking		Cancer Screening	
	Grouping: Current Symptoms		Grouping: Onset Age	
	<i>None</i>	<i>Early</i>	<i>30-45</i>	<i>Over 45</i>
	<i>(n=172)</i>	<i>(n=359)</i>	<i>(n=140)</i>	<i>(n=950)</i>
Male (0/1)	0.32	0.46 <sup>f</sup>	0.34	0.39
White	0.89	0.97 <sup>f</sup>	0.91	0.91
Current Age	41.2	47.4 <sup>f</sup>	50.6	61.5 <sup>g</sup>
In US	0.96	0.93	0.95	0.86 <sup>g</sup>
Mom HD Parent	0.58	0.51	0.45	0.55

Notes: This table shows balancing by the symptom onset groups. Significance marks (all at 5% level): <sup>a</sup> compared to onset 15-18; <sup>b</sup> compared to onset 19-22; <sup>c</sup> compared to onset 23-28; <sup>d</sup> compared to onset 20-29; <sup>e</sup> compared to onset 30-39; <sup>f</sup> compared to no symptoms; <sup>g</sup> compared to onset 30 to 45.

**Table B2: Gene Testing and Educational Attainment, Only Tested Individuals**

<b>Panel A: Main Results and Robustness</b>				
	<b>Main Results</b>		<b>Control Referral Method</b>	
	# Post-HS Years Education	Bachelor Degree	# Post-HS Years Education	Bachelor Degree
<i>Sample</i>	<i>Finish HS</i>	<i>Finish HS</i>	<i>Finish HS</i>	<i>Finish HS</i>
Tested, Negative	.677* (.369)	.251*** (.096)	.605 (.378)	.221** (.103)
Standard Controls	YES	YES	YES	YES
# of Observations	126	126	126	126

<b>Panel B: Falsification</b>				
	<b>Pre-Symptom Education</b>		<b>Older Testers</b>	
	Complete High School		# Post-HS Years Education	Bachelor Degree
<i>Sample</i>			<i>Finish HS</i>	<i>Finish HS</i>
Tested, Negative	.005 (.049)		.194 (.370)	.073 (.093)
Standard Controls	YES		YES	YES
# of Observations	135		135	135

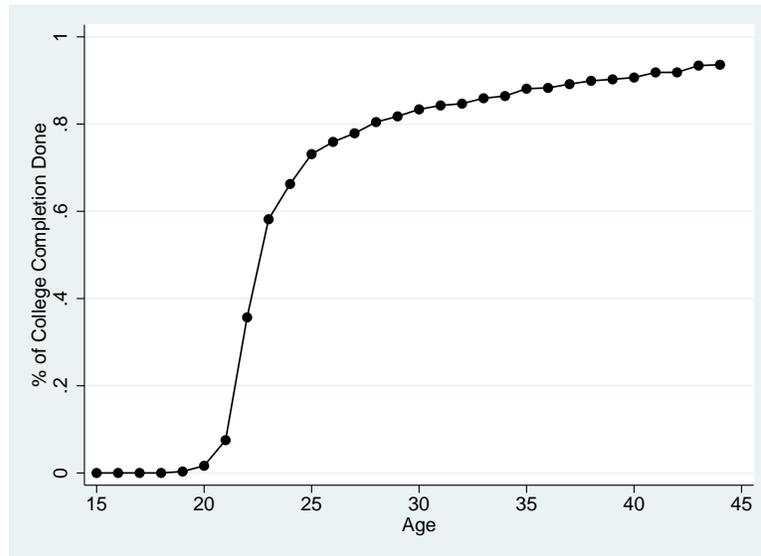
Notes: This table shows the impact of test results on educational attainment. The omitted category is individuals who have been tested and learn they *do* carry the HD mutation. The main results include individuals who are tested before or at the age of 28, and (if positive) do not develop symptoms before 35. Columns 3 and 4 of Panel B show results for individuals who were tested between 35 and 45, but did not have symptoms before 45. Standard controls: gender, country dummies, race and a control for age. Referral group control simply controls for whether the individual was recruited by a doctor (the primary recruitment method) or in another way (at a meeting, online, etc). Standard errors in parentheses. \*significant at 10% \*\*significant at 5% \*\*\*significant at 1%

**Table B3: Gene Testing and Smoking, Only Tested Individuals**

<i>Sample</i>	<i>Smoke Now All</i>	<i>Smoke Now Ever Smoke=1</i>
Tested, Negative	-.029 (.029)	-.186* (.101)
Standard Controls	YES	YES
# of Observations	451	105

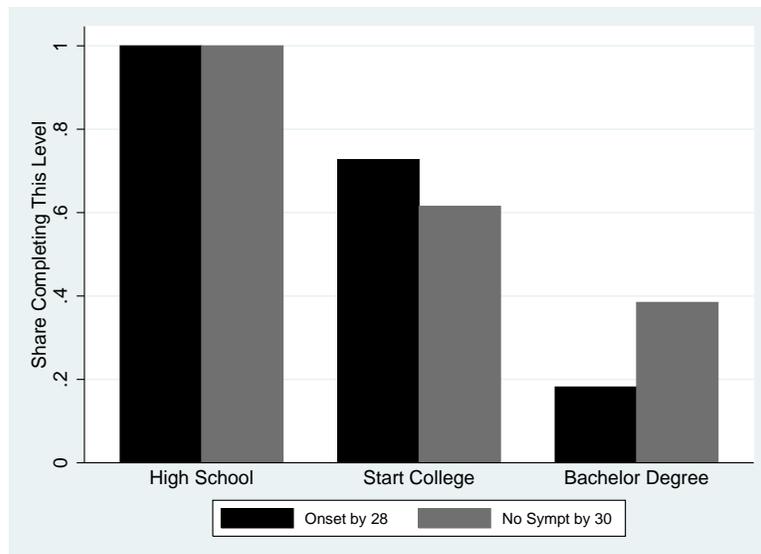
Notes: This table shows the impact of test results on smoking. The omitted category is individuals who have been tested and learn they *do* carry the HD mutation. Standard controls: gender, country dummies, race, age, dummies for education. Standard errors in parentheses. \*significant at 10% \*\*significant at 5% \*\*\*significant at 1%

*Figure B1: College Completion by Age*



Notes: This figure shows the timing of college completion by age. To construct this graph we use data from the 1980, 1990 and 2000 census and the 2010 American Community Survey. We begin with the 1980, 1990 and 2000 census. We take individuals who are 35 to 54 in the 2000 census and match them to the same birth cohort in the 1980 and 1990 census. We assume the 2000 data represents the “final” education for that cohort. We generate the share of college completion by age X as the level at age X divided by the final level in that group. We do a similar thing for the 1990, 2000 and 2010 combination. We average by age and graph.

*Figure B2: Within-Sibling Group Education Effects (N=24)*



Notes: This figure shows educational attainment for sibling groups in which we observe more than one sibling and one of the siblings has earlier HD onset. The sample size is very small.

## Appendix C: Magnitude Comparisons

This appendix briefly discusses the comparisons between our magnitudes and those in Stoler and Meltzer (2012). In the text of the paper we calculate an elasticity of demand for college completion with respect to life expectancy, disability-adjusted life years and earnings. As we discuss there, our magnitude is similar to the elasticity of demand for years of schooling with respect to life expectancy calculated by Jayachandran and Lleras-Muney (2009). Comparison with their paper is straightforward because they provide an elasticity. Stoler and Metlzer (2012) instead provide a calculation of the percent decrease in years of schooling for a year decrease in earnings capacity.

To compare, we use their data to calculate an elasticity.

We do this by adopting a version of our methodology applicable to their figures. To do so, we first use their results to calculate a percentage change in education. Their outcome is education de-meaned by the cohort average in the state. To calculate a percent change we need to reverse this calculation. The median education in the US for individuals aged 43 to 46 in 2000 (the closest year and broad age category for these data) is a high school degree: 12 years. Applying this to the data in the paper, we predict those who learn about their HD risk later complete 14.19 years of schooling, versus 11.74 (which is 14.19 minus their effect size of 2.45) for those who learn about their risk before age 18. This is a 17.2% decrease in years of schooling.

For disability-adjusted life expectancy, we calculate the disability-adjusted life years for someone who is HD negative versus someone who is at risk at age 18. We note this is slightly different than the calculation in their paper, which simply assumes everyone in the at-risk category will have HD onset at age 40, but the spirit is similar. We find that those who are HD negative should expect 53.0 disability-adjusted life years, versus 40.8 for individuals who are at risk at age 18. Combined with the education change, we find an elasticity of -0.74, very close to what we find in our paper. If we use earnings directly instead of disability-adjusted life years, we find an elasticity of -0.89, again very close to ours.