NBER WORKING PAPER SERIES

LIMITED LIFE EXPECTANCY, HUMAN CAPITAL AND HEALTH INVESTMENTS:
EVIDENCE FROM HUNTINGTON DISEASE

Emily Oster
Ira Shoulson
E. Ray Dorsey

Limited Life Expectancy, Human Capital and Health Investments: Evidence from Huntington Disease
Emily Oster, Ira Shoulson, and E. Ray Dorsey
NBER Working Paper No. 17931
March 2012
JEL No. I15,I25,J24

ABSTRACT

One of the most basic predictions of human capital theory is that life expectancy should impact human capital investment. Limited exogenous variation in life expectancy makes this difficult to test, especially in the contexts most relevant to the macroeconomic applications. We estimate the relationship between life expectancy and human capital investments using genetic variation in life expectancy driven by Huntington disease (HD), an inherited degenerative neurological disorder with large impacts on mortality. We compare investment levels for individuals who have ex ante identical risks of HD but learn (through early symptom development or genetic testing) that they do or do not carry the genetic mutation which causes the disease. We find strong qualitative support: individuals with more limited life expectancy complete less education and less job training. We estimate the elasticity of demand for college completion with respect to years of life expectancy of 0.40. This figure implies that differences in life expectancy explain about 10% of cross-country differences in college enrollment. Finally, we use smoking and cancer screening data to test the corollary that health capital is responsive to life expectancy.

Emily Oster
University of Chicago
Booth School of Business
5807 South Woodlawn Ave
Chicago, IL 60637
and NBER
eoster@uchicago.edu

Ira Shoulson
Georgetown University
2115 Wisconsin Ave NW
Suite 603
Washington DC 20007
ira@irashoulson.org

E. Ray Dorsey
Johns Hopkins University
Department of Neurology
Meyer Bldg, Room 6-181
600 N. Wolfe Street
Baltimore, MD 21287
Ray.Dorsey@jhmi.edu

Limited Life Expectancy, Human Capital and Health Investments: Evidence from Huntington Disease

Emily Oster

University of Chicago and NBER

Ira Shoulson

Georgetown University

Ray Dorsey

Johns Hopkins University

Draft: March 12, 2012

Abstract

One of the most basic predictions of human capital theory is that life expectancy should impact human capital investment. Limited exogenous variation in life expectancy makes this difficult to test, especially in the contexts most relevant to the macroeconomic applications. We estimate the relationship between life expectancy and human capital investments using genetic variation in life expectancy driven by Huntington disease (HD), an inherited degenerative neurological disorder with large impacts on mortality. We compare investment levels for individuals who have *ex ante* identical risks of HD but learn (through early symptom development or genetic testing) that they do or do not carry the genetic mutation which causes the disease. We find strong qualitative support: individuals with more limited life expectancy complete less education and less job training. We estimate the elasticity of demand for college completion with respect to years of life expectancy of 0.40. This figure implies that differences in life expectancy explain about 10% of cross-country differences in college enrollment. Finally, we use smoking and cancer screening data to test the corollary that health capital is responsive to life expectancy.

1 Introduction

Among the most basic predictions of human capital theory is that the life of the asset – individual life expectancy – will impact investment. Individuals with a limited life expectancy should be less likely to undertake any costly action which has only long-term payouts (Becker, 1964; Ben-Porath, 1967). For educational investments in particular this theory has important macroeconomic implications for understanding the path of the demographic transition and the impact of life expectancy on economic growth (Acemoglu and Johnson, 2007; Kalemli-Ozcan, 2002; Soares, 2005; Kalemli-Ozcan, Ryder and Weil, 2000; Lorentzen, McMillan and Wacziarg, 2008; Young, 2005; Weil, 2007).

Despite the importance of understanding the relationship between life expectancy and education, actually estimating it is made difficult by both reverse causality and omitted variable bias

concerns. This estimation challenge has prompted use of creative identification strategies. For example, Jayachandran and Lleras-Muney (2009) estimate impacts of reductions in maternal mortality on female literacy rates in Sri Lanka.[1] Although this and related papers provide a test of the general theory, they may fall short as inputs to the larger macroeconomic question. The search for compelling identification has meant that the life expectancy changes they analyze are small and uncertain. Further, the changes in investment considered (literacy, for example, or primary school) are somewhat removed from the "adult educational investments" that are of more central interest to the macro literature.

In this paper we estimate the impact of life expectancy on human capital investment using data on individuals at risk for Huntington disease (HD). HD is a degenerative neurological disorder with onset in early middle age. Individuals with this disease have a life expectancy of around 60, and a healthy life expectancy of 10 years fewer than that. 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. There is no cure, and only limited treatment, but there *is* a perfectly predictive genetic test.

We argue that we can use variation across *ex ante* similar individuals within this population to estimate causal effects.[2] Further, we argue that the estimates derived may be more useful than previous estimates in applications. This is true for several reasons. First, variations in life expectancy in this population are large – on the order of 20 years or more – and known to individuals. Second, the educational choices we consider – high school, college, graduate work – are exactly the type of investments that the growth literature highlights (e.g. Soares, 2005).

This paper makes three contributions. First, we provide a strong and direct test of the qualitative assumption behind human capital theory: does variation in life expectancy drive human capital investment? We find that it does. Second, we use our estimates to address the quantitative question: how *important* is life expectancy in driving large scale differences in human capital

---

[1] As another example, Fortson (2010) uses panel data on HIV rates in Africa to estimate the impact of HIV on education. In her case although investment motive is one possible channel, there are others and the experiment is not as direct as in Jayachandran and Lleras-Muney (2009). It is also worth noting a large literature on the relationship between health and education that focuses on children not getting schooling because they are too sick to attend (e.g. Bleakley, 2007; Miguel and Kremer, 2004). This is related in the sense that it is another way that health impacts education, but not in the sense of *why* this happens.

[2] Stoler (2005) addresses a subset of the outcomes here with a similar identification strategy but with only a very small number of individuals (30 to 50 people, depending on the outcome). His evidence is consistent with ours, but the limited sample size and sample selection make it difficult to draw compelling conclusions from his analysis.

attainment? We provide an explicit mapping between years of life expectancy and education or job training; we also estimate an elasticity. We ultimately argue that differences in life expectancy account for only a small share of variation over time or across countries in educational attainment. Finally, we use this population to briefly consider a related question: does limited life expectancy limit investments in long-term preventative health behaviors (Grossman, 1972; Dow et al, 1999)? We find support for this theory in the context of smoking and cancer screening.

We begin with our estimation of the impact of life expectancy on education and job training. Our sample includes individuals who have one parent with HD and, therefore, at birth have a 50% chance of inheriting the HD mutation. Our estimation strategy exploits variation across individuals in the timing of information revelation. We do this in two ways: with genetic testing and with early symptom development.

We begin by using the genetic testing data. We limit to individuals who are tested prior to HD symptoms and therefore should have similar life expectancy prior to testing regardless of true status. Consistent with this, we show that 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. This estimation is valid within-sample, although since testing is uncommon we may be concerned that the type of individuals who seeks genetic testing responds more than the average individual.[3]

Our second approach uses variation based on the timing of symptom onset. The larger sample size in this case allows us to add job training to our analysis. This approach relies on the observation 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 by it. We demonstrate that individuals with varying age of onset are similar on *ex ante* characteristics[4], and that early symptoms on their own are unlikely to impact these behaviors. Combining data on the individual age of symptom onset with an estimate of the age at which education and job training decisions are made we estimate whether behaviors differ for individuals who had symptom onset before they made a choice versus after.

Both approaches demonstrate impacts of life expectancy on human capital investment. The

---

[3]We explore the reasons for limited testing elsewhere (Oster et al, 2011). The limited sample of testers is the reason this analysis can only be run for the two outcomes for which we have the largest sample size.

[4]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.

analysis of genetic testing indicates that individuals who engaged in predictive testing before age 30 and learned they carried the HD mutation are about 30 percentage points less likely to complete college than those who were tested at a similar age and learned they do not carry the HD mutation. In exploring symptom onset, we find that those who have symptom onset between the ages of 17 and 24 are 14 percentage points less likely to begin college and 33 percentage points less likely to complete collect than those with no symptoms by 35. Individuals with symptom onset a bit later, between 25 and 30, are no less likely to begin college but are 14 percentage points less likely to complete it. This suggests divergence of educational investment as HD status is revealed. The analysis of job training demonstrates similar results. Earlier symptom onset dramatically limits investment in job training for advancement or promotion.

One concern with our analysis is that individuals must select into our sample; it is possible that this selection could vary across groups, which could threaten our causal interpretation. The balance across groups in *ex ante* characteristics limits this concern. In addition, we run several robustness an falsification tests. Our education results are robust to controlling for method of entry into the sample and for family fixed effects. Further, educational attainment dose not differ for individuals who engage in predictive testing at older ages, and there is no impact of onset age on *routine* job training.

These results provide strong qualitative evidence in favor of human capital theory. Following this, in Section 6, we discuss the quantitative implications of these estimates. As a first step, we use our coefficients alongside information on HD life expectancy to map out a relationship between years of life expectancy and behavior. Because our data cover individuals with a range of life expectancies, we can generate a somewhat non-parametric mapping of the life expectancy/human capital relationship. We can also provide a parametric summary in the form of an elasticity of demand for education or job training with respect to life expectancy. The education elasticity is small, in the range of 0.4 to 0.6. The job training elasticity is larger, around 1.4.

Either the parametric or non-parametric approach provides a concrete mapping between life expectancy and human capital which could be used in applications. As an example, we apply these estimates 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.

Perhaps the most ambitious use of our estimates is to apply them to large-scale variations in life expectancy and human capital. In Section 6.2 we take a first step towards this, asking what

share of the difference in college completion, either over time in the US or cross regions globally, might be explained by differences in life expectancy. Although this is inherently subject to somewhat heroic assumptions, it is precisely these large-scale differences that growth models seek to explain. Theory has connected changes in health to increases in human capital and subsequently to growth (Kalemli-Ozcan, 2002; Soares, 2005; Kalemli-Ozcan, Ryder and Weil, 2000; Young, 2005). What we can ask is, empirically, what share of variation in human capital might be accounted for by variation in life expectancy?

In general, we find the role of life expectancy is relatively small. Over time within the US changes in life expectancy account for only about 2 percent of the increase in college completion rates. Across regions or countries we find life expectancy differences are slightly more important but still small: they explain about 10% of difference in tertiary enrollment rates. This suggests that life-expectancy fueled desire to invest in human capital plays a fairly small role in driving increases in higher education (and, hence, growth). Of course, this does not rule out the possibility that improved health might drive growth through another channel.

In the final section of this paper we use our data to address the issue of 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. As pointed out in Dow et al (1999), when mortality risks "compete" with each other, individuals have less incentive to invest in one if the risk of dying from another increases. The existing empirical evidence on this is even more limited than in the human capital case.[5]

We use data on quitting smoking and undertaking cancer screening from our data; both actions fall in the category of being costly investments that have long term payoffs. Consistent with theory, individuals who learn they carry the HD mutation through either 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) and less likely to be "on-schedule" with their screening, although this latter result is only marginally significant. Overall, this suggests empirical support for the approach of modeling health investments as competing in this way. From a policy standpoint, this may suggest

---

[5]The existing literature on this is even more limited than in the human capital case. Dow et al (1999) provide suggestive evidence from a vaccination campaign, although causality is limited. Oster (2012) estimates the impact of other diseases on HIV avoidance behavior and finds some evidence of these complementarity.

positive spillovers from health improvements (Yarnoff, 2011).

The rest of the paper is organized as follows. Section 2 provides background on HD and the impact of the disease on investment incentives. Section 3 describes our data and Section 4 our empirical strategy. Section 5 shows our evidence on education and job training. Section 6 discusses magnitudes and cross-sectional predictions. Section 7 discusses the evidence on health investments, and Section 8 concludes.

2   Background and Incentives

This section discusses the background on Huntington disease (HD) in Subsection 2.1. In Section 2.2 we then provide some brief evidence on the differential incentive to invest in human capital and preventative health behaviors for individuals with and without HD.

2.1   Background on Huntington disease[6]

HD is a degenerative neurological disorder that clinically affects an estimated 30,000 individuals in the United States. Individuals with the disease typically begin to manifest symptoms in early middle age (30-50). Symptoms include involuntary movement, impaired cognition and psychiatric disturbances. Individuals will need increasing levels of supportive and institutional care for many years. Death follows approximately 20 years after onset. A test for the HD genetic expansion was developed in 1993. Since everyone with the expansion will eventually develop HD, this test is perfectly predictive.

HD is a genetic disorder due to an excessive expansion in the Huntingtin gene on chromosome 4; individuals with more than 40 repeats of a "C-A-G" (cytosine-adenine-guanine) sequence in this gene will inevitably develop HD unless they die from an unrelated cause prior to the expected onset of illness. 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. 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.[7]

---

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

[7]It is, of course, possible that people may not know of their risk until they are older, since parents' age of onset may

Importantly for our analysis, the progression of HD is slow. Early symptoms of HD may not be noticed by at-risk individuals, and in some cases what look like early symptoms are not. Put differently, a fairly large share of individuals who do not have HD nevertheless display some motor abnormalities. For this reason, early symptoms are not a perfect signal of HD, and individuals should update their probability of carrying the gene slowly. Since 1993, a test for the HD genetic mutation has been available. However, testing rates are fairly low: 5-10% of the at-risk population reports predictive testing (Shoulson and Young, 2011). The lack of any treatment or cure, and the fear of being unable to live with knowing that one carries the mutation are significant barriers (Oster et al, 2008; Oster, Dorsey and Shoulson, 2011).

2.2   Impact of HD on Incentives

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. The most basic way to demonstrate this is through the impact of HD on life expectancy and on healthy life expectancy. Long-term investments, be they in human or health capital, pay off more the longer an individual expects to live.

The identification of our effects relies on comparing asymptomatic individuals with and without the HD mutation and on comparing symptomatic individuals based on age of onset. Our conclusions rely on the assumption that life expectancy is lower for individuals with the HD mutation or with earlier symptom onset. Table 1 provides estimates of these parameters across groups.[8] This table makes it clear that HD limits both overall and healthy life expectancy (important because this give us a clue about impacts on time working). Consider two individuals at age 25 who do not have symptoms, one of whom knows they carry the HD mutation and the other who knows they do not. The life expectancy of someone who knows they carry the HD mutation is roughly 20 years less (60 versus 80) and their healthy life expectancy is about 35 years less.

The reductions in life expectancy are even larger for individuals who develop symptoms earlier. Someone who develops symptoms between 25 and 30 can expect an overall life expectancy of less than 50. The older someone is when they develop symptoms, the less severe are the limits on life

---

be late or their parents may die of something other than HD before onset. In our sample, everyone enrolled knows of their risk since this is a condition for enrollment.

[8]Appendix A provides details of these calculations.

expectancy. Overall, Table 1 demonstrates significant differences across groups life expectancy and, by extension, the incentive to invest in human capital. Appendix A provides more details.

3   Data

Survey

Our data comes 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 control 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, as they are likely not to be comparable to those at risk.

The first column in Panel A of Table 2 shows counts of people in each group in the overall COHORT sample. The largest group is individuals with manifest HD. It is worth noting that there is significant variation within this group in 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 2 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 up two concerns. The first is 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.

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 and their smoking behavior. Educational attainment is measured in categories, from less than 9th grade through professional degree. For education, our outcomes are number of years of education after high school, starting college and completion of a bachelor degree. We highlight the decisions which are made *after* high school since no individuals in our sample experience symptoms or engage in genetic testing before the end of high school.

To measure job training, which is done in the ALD sub-study, 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 2.

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. This latter outcome is simply whether or not the individual has quit smoking or not; this may be a more appropriate outcome. Since most individuals start smoking when they are young, and everyone in our sample has a similar level of risk at that point, they should be equally likely to begin smoking. Variation should arise only from the decision to quit. 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 recent (on-schedule) screening. In the case of mammogram, the recommendation is yearly screening so we define recent screening as screening in the last year. In the case of colonoscopy, the recommendation is every three to five years, so we define recent screening as screening in the last five years. Both of these types of screenings are recommended only for individuals who are above a certain age (or have other risks). We limit the mammogram analysis to individuals over 35, and the colonoscopy analysis to individuals over 40.[9] Summary statistics for the health measures appear in Panel B of Table 2.

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 symptoms and self-reported data on the date of first symptoms. The latter variable is useful when we look at decisions which happened in the past, as it allows us to define symptom status at younger ages. These variables are summarized in Panel C of Table 2. 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. In the at-risk group, the share with the mutation is 34%.[10]

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

---

[9]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.

[10]This is *less* than 50% because these individuals do not have symptoms, and as people age without symptoms the probability of carrying the mutation declines, as detailed in Section 2.3.

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 and, therefore, would 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. Some of these individuals test and learn they do carry the mutation (and therefore have a limited life expectancy); some test and learn they do not (and therefore have a normal life expectancy). We compare these groups. We are able to do only for education, and not for job training, since data on education was collected as part of the overall COHORT study so we have a larger sample size.

Define an educational investment decision made by individual i as D_i. We report results from regressions of the form below:

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

In this regression, $\mathbf{X_i}$ is a vector of controls (age, gender, etc). The omitted category of individuals is those who are at risk but untested. These individuals are included to pin down coefficients on controls; we will report p-values for the null hypothesis that $\beta_1 = \beta_2$. We should note that predictive testing is unusual; most people at risk choose not to be tested. By comparing within the tested

group we avoid selection concerns (although it is still possible that the impacts are larger in this group since these individuals sought information on their status). Including the at-risk individuals in the regression sample allows us to comment on the selection.

To complete this analysis we identify individuals who tested while asymptomatic and before they decided about college completion. Our sample therefore includes individuals who are tested before the age of 30 and were asymptomatic until at least 35.[11]

Identification based on Symptoms

Our second identification strategy relies on variation across individuals in the timing of symptom onset. Consider a decision D_{it} made by individual i at time t. Further, define HD_{it} as equal to 1 if individual i knows they have HD at time t and 0 otherwise. We run regressions of the following form:

$$D_{it} = \alpha + \beta_1(HD)_{it} + \mathbf{\Gamma X_{it}} + \epsilon_{it}$$

The coefficient β_1 tests for differences in behavior between individuals who know they have HD due to early symptom onset and those who are at risk for HD but with no information. The subscript t in this case makes clear that the variation is based on the relationship between symptom onset timing and decision timing. Consider the decision about college attendance: what is relevant is what the individual knows about their status *at the time this decision is made* (namely, around age 20 or so). Both education and job training decisions are made sometime in the past, although we do not observe the exact time that the decision is made. We therefore define groups of individuals based on 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.

To be more concrete, consider the case of education. Our analysis focuses on years of schooling after high school, college entrance and college completion. We define three groups of individuals: those with symptom onset between 17 and 24, those with symptom onset between 25 and 30 and those with no symptoms by age 30. The first group should be mostly heavily affected; some of them would know their HD status before starting college and virtually all before completing it. The second

---

[11]It might be better to limit to a younger sample but testing at very young ages is so unusual that the sample size becomes prohibitive. In addition, a fairly large share of people do take longer to complete college suggesting that extending the sample through 30 is appropriate.

group is unlikely to know their HD status before starting college and may or may not have known before completing it. The third group will have made all of these decisions without knowing if they had HD.

In analyzing job training we note that most job 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

Before moving to the results, we briefly consider two threats to identification.

Balancing

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. Panel A of Table 3 begin to address this concern through direct balancing tests. We compare groups (either based on test results or based on age of symptom onset) on several variables which are determined *before* any impact of HD (gender, age, location, race).

In the case of testing (Columns 1 and 2) the two groups are similar on all variables. In the case of age of symptom onset (Columns 3-8), we find these groups are similar on gender, race and location. The groups are unbalanced with respect to age, which 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 addition to this direct balancing test, in Section 5 we push the analysis further. We run robustness analysis with more comprehensive controls. This includes method of referral into the study and, for a subset of individuals with family histories, family fixed effects. In the latter analysis we are therefore comparing individuals within the same family who were both tested while

asymptomatic and learned different results. For education, we also run two additional falsification tests. Using the same individuals in our primary analysis, we estimate impacts on an educational decision made *before* symptom onset. These decisions should not be impacted by HD status. Second, we estimate impacts for individuals who tested when they were too old for the test to impact their educational decision. These individuals also should not differ in their educational attainment.

Impact of Symptoms on Behavior

When we consider the results based on symptom onset, in particular, we are also subject to the concern that changes in behavior are due to symptoms themselves, rather than the information which is conveyed by the symptoms. Although we attempt to focus on decisions made when symptoms are fairly minor, this remains a concern.

Our 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 1 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. Recall that our effects are identified off of individuals in the first few years of symptoms. Over these early years we observe only very limited declines in either measure of cognitive ability, even though there are large changes in physical symptoms. This suggests mental abilities remain at least largely intact as information about status is revealed by physical symptoms.

A related 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 than the human capital investment mechanism. In practice, HD does not have as large an impact on depression as might be expected. Comparing individuals with HD in our population to the controls individuals in the study, 49.7% of those with HD report some depression, versus 50% of controls. In terms of timing, the share of individuals reporting any depression spikes in the first year that symptoms are noticed, but declines

back to baseline by the second year. This suggests that the depression confound maybe more limited than we would expect *ex ante.*

5   Results: Impact of HD on Human Capital Investment

Education   We begin with the identification based on genetic testing. A simple comparison of means suggests some impact of test results on education. Individuals who learn through genetic testing that they do carry the HD mutation get, on average, 14.1 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 15.1 years of schooling and 66% of them complete a bachelor degree; these differences are both significant at the 1% level in simple t-tests.

These differences persist in regressions with demographic controls. Panel A of Table 4 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. Our primary results, in 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.[12] Columns 3-6 explore the robustness of our results to several additional controls. In Columns 3 and 4 we include family fixed effects (this limits the sample a bit since only a subset of individuals have family history data available). Despite the limited sample size, the results are similar. Columns 5 and 6 control for method of recruitment into the study, which could in principle be correlated with socioeconomic status; the results are essentially unchanged.

In Panel B of Table 4 we turn from robustness 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 issues should be similar in this population. However, these individuals were all untested at the time of education decisions; if the results for younger individuals are being driven by different choices in the

---

[12]We note that those who learn they do carry the HD mutation get a similar amount of education as the comparison group, who is untested. This likely reflects the selection of individuals into testing, reinforcing the view that we should focus on the difference between the two tested groups.

face of varying information, then we would not expect to see differences in the older group. This is, in fact, what we find: although both tested groups are higher human capital (due to selection) there are no significant differences between them in educational attainment. Overall, the results in Panel B reinforce the main results and suggest that differences in test results (and resulting life expectancy) do impact choice of education.

Figure 2 gives a visual sense of the education results based on age of symptom onset. We illustrate the chance of completing each level of schooling by age of symptom onset (young, medium and old). All three groups are similarly likely to complete high school. The youngest age-of-onset group diverges a bit at the start of college, and even more as time goes on. The group with onset age 25-30 has similar levels of initial college enrollment, but is less likely to complete college and much less likely to go on to graduate work. This divergence suggests investments diverge as individuals learn about their HD status.

Table 5 shows this impact statistically, testing whether educational choices begin to diverge with symptom onset. Consistent with Figure 2, we find that they do. 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 when people decide about college enrollment. Importantly, the individuals with onset between 25 and 30 *do not* diverge in the college enrollment period, but do diverge when we turn to bachelor degree completion. Overall, both groups complete fewer years of education post-high school. The differences are quite large.

When family fixed effects are included (Columns 4-6), the estimates are noisier, but we see the same general trend. Again, we can look at high school completion as a falsification test, since all of these groups have similar information at that time. The youngest onset group is no less likely to complete high school than those individuals with symptoms after 30; although the group with symptom onset between 25 and 30 is slightly less likely, there doesn't seem to be any trend.

Together with the results in Table 3, these results suggest that school enrollment is similar across individuals up until they learn that they do carry the HD mutation (in this case, learn through symptoms). At that point, differences begin to emerge, consistent with a theory in which people make different choices based on their impression of their life expectancy.

16

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 6. Our outcome of interest is job training for advancement or promotion, and there 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 (12.5%), which is consistent with the fact that these individuals could have undergone job training in their twenties.

The second column of this 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 – they have different kinds of jobs, or work less – then we should also see differences in this overall job training. In fact, if anything those with younger symptoms are *more* likely to have engaged in other, non-promotion-associated, job training.

6   Results: Estimating Magnitudes

As described, the results above address the qualitative question of whether human capital is responsive to life expectancy. We argue that the answer is yes. This qualitative question is of interest on its own; for example, knowing that this matters suggests that models of human capital formation should take life expectancy into account. However, for many applications knowing the qualitative relationship is not sufficient. Calibration of growth models which take this channel seriously, 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 conclusion from these data requires stronger assumptions (namely, that the response in this population is quantitatively similar, as opposed to just qualitatively similar, to what we would see in the general population. This caveat should be kept in mind when evaluating these results.

6.1 Mapping Life Expectancy to Behavior

The changes in life expectancy prompted by HD are large. Some individuals in our sample (namely, those who do not carry the HD mutation) have normal life expectancy. Others (namely, those with very early onset) have such limited life expectancy that there is virtually *no* investment incentive to complete higher education or engage in productive job training. Given this range, the simplest way to show a quantitative relationship is to graph predicted education and job training rates from the regressions against healthy life expectancy for each group. We focus in this section on the results based on age of symptom onset. Individuals who learn about their status from symptoms are likely to be more representative of the general population. Further, the symptom onset regressions have groups with the widest variation in life expectancy and, importantly ,in both education and job training the group with the earliest symptom onset has very limited life expectancy. This allows us to think about what these behaviors look like for someone with *no* investment incentive.

Figure 3 shows predicted college completion, years of education after high school and job training rates graphed against healthy life expectancy.[13] Details of the life expectancy calculations appear in Appendix A. We focus on *healthy* life expectancy since the long disability period associated with HD is not a general feature of mortality from other causes. In principle, from this graph, one could read off directly the expected college completion rate or job training rate for a given life expectancy. For example, our data indicate that someone with 10 years of healthy life expectancy has about a 5% chance of investing in job training and a 30% chance of completing college.

In the case of education, the slope of both series flatten out to the right; moving from 10 to 30 years of healthy life expectancy has a fairly limited impact on educational attainment. Further, although the earliest onset group has much lower college completion than others, it is not zero. Despite having virtually no incentive to invest, about 10% of this group completes a college degree. One interpretation of this is that there is some "consumption" component to education. Regardless of *why* it occurs, this does suggest an addition to models which relate life expectancy to educational investments. The canonical form of these models suggest that if there is no incentive to invest, individuals will not do so. These results suggest that is not the case, a fact which has implications for modeling the impact of reductions in life expectancy. Job training rates, in contrast, are zero for

---

[13]The education and job training values in this graph assume that all groups have the demographic characteristic of the average individual in the sample.

the group with lowest life expectancy. With little or no incentive to invest individuals do not appear to engage in job training behavior.

This figure provides one way to describe magnitudes. We can also use our coefficients, along with the life expectancy data, to generate an elasticity of demand for education nor job training with respect to life expectancy. Although this imposes more parametric assumptions, it provides a convenient summary of the magnitudes, and may be more useful as an input to theory. Panel A of Table 7 estimates the elasticities. Column 1 shows elasticities based on symptoms; Column 2 based on testing (not available for job training). We find the education response is fairly inelastic, and the response of job training is more elastic.

An Application

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 capital through increased human capital investment and productivity growth. However, they decrease income per capital by increasing population. Their model produces the following relationship (from Section 6.C. of their paper):

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

where α is the labor share, π is the response of income per capital to life expectancy, λ is the response of population to life expectancy, η is the elasticity of human capital with respect to life expectancy and γ is the elasticity of TFP with respect to life expectancy.

Their paper estimates π and λ. They find that population increases with increases in life expectancy and GDP per capital 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 (in Panel A of Table 7), however, suggest a particular value for η. If we focus on our symptom-based estimates, this value is around 0.3. We can therefore say, more concretely, that to harmonize their results with an neoclassical growth model the elasticity of TFP with respect to health must be in the range of -0.8 to -0.2. In other words, it must be negative, but small.

Based on what we know about the relationship between health and productivity at an individual level, this seems unlikely. This suggests (consistent with the suggestions in their paper) that something other than a standard neoclassical growth story may be behind their results. It is worth noting that this is not an inevitable conclusion; if we had estimated a much smaller elasticity, or even a zero elasticity, their results would have been explicable with a zero or even positive elasticity of TFP with respect to health. It is because our estimate of elasticity is positive and of a reasonable magnitude that we draw these conclusions.

6.2   Cross Sectional and Over Time Changes in Education

We consider now settings in which the magnitudes we describe above might help us better understand broader educational patterns. First, in this subsection, we relate our estimates to differences in the cross section and changes over time in educational attainment. We ask simply: based on the magnitudes in Section 6.1, what share of the changes in college completion over time within the US and or differences across regions globally might be explained by differences in life expectancy? We use an elasticity of 0.40, which is based on the symptom data (Column 1 of Panel A of Table 7); if we used the elasticity based on testing the share explained would be roughly twice as big.

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. These results are shown in Panel B of Table 7. We document the importance of life expectancy in explaining the 1960 versus 2010 difference, as well as the 1980 and 2000 versus 2010 differences. In all three cases, the share of the difference which can be accounted for by life expectancy is small. On average, only about 2 percent of the difference is explained by changes in life expectancy over time.

To look at the importance of this explanation globally 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 life expectancy and the gross share enrollment in tertiary education to generate actual and predicted differences in college-going across regions. The results (in Panel C of Table 7) show that life expectancy explains between 8% and 12% 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. For each pair we calculate the share of the difference in college-going which would be accounted for by the difference in life expectancy. The last row of Panel C of Table 7 reports the average share explained, which is 9.3%. These results suggest that life expectancy is more important in explaining cross-country or cross-region differences than differences over time within the US, but in either case the figures are small. Only about 10% of the cross-country differences could plausibly be explained by differences in life expectancy (note that this would be closer to 20% if we relied on the genetic-testing-based estimates).

From the standpoint of models human capital and economic growth, this result suggests that the channel of health leading to a greater desire for human capital investment is fairly minor. Much of the differences in human capital across countries or over time must be driven by other factors. It is worth noting, of course, that these factors *could* be health-related (by increasing capacity to learn, for example, as in Miguel and Kremer, 2004; Bleakley, 2007). What we suggest is only that the investment channel is relatively unimportant empirically.

7 Health Capital: Smoking and Cancer Screening

Thus far we have focused on human capital. As we note in the introduction, however, models of health capital also have implications for behavior in this population. Individuals who carry the HD mutation are very unlikely to benefit from cancer-prevention measures since it is very unlikely they will live long enough to actually develop cancer (see Appendix A for details). Given this limited incentive, individuals should be less willing to undertake costly actions to prevent these diseases. As we describe in Section 3, the COHORT study overall collected data on smoking behavior, and the smaller ALD sub-study collected information on mammogram and colonoscopy.

7.1 Empirical Strategy

As in the human capital case, we use two empirical approaches: identification based on genetic testing and identification based on symptom onset.

The genetic testing identification is possible for smoking (where we have more data). We compare current smoking rates for asymptomatic individuals who have been tested. When we turn to 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 variable of interest (HD_{it} in Equation (2)) is defined by whether the individual is *currently* experience early symptoms. For the third health outcome (ever undertaking cancer screening) the decision is made in the past, 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

Again, the non-random selection of individuals into our sample presents a possible challenge to this analysis. Panel B of Table 3 shows balancing across groups. The first two columns focus on the testing sample used for the analysis of smoking. The groups are largely balanced, with small age differences. The last set of columns compares individuals based on age of onset. As in the case of education, the only balancing issue here is current age, which is mechanical. Columns 3 and 4, which compare individuals with no symptoms to those with early symptoms (for our smoking and current cancer screening analysis) show the most significant balance issues. In particular, the early symptom group has many more men than the group with no symptoms and is significantly older. The latter result is, again, mechanical. We will control for age and gender. However, the lack of balance on these variables suggests these analyses should be taken with more caution than the others.

Symptom Impacts

In using symptoms as drivers of information here we again face the concern that symptoms themselves impact behavior. In this case it is the motor symptoms, themselves, which are the

concern. Perhaps individuals are less likely to get a colonoscopy because they are unable to be still for long enough for the procedure. 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; some individuals notice motor symptoms earlier than others. These individuals have the same information, but varying physical symptoms and we graph smoking, mammography and colonoscopy probabilities against their motor scores in Figure 4. Our concern would be that over the range of early symptoms (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 in Figure 4. Smoking rates are going down as symptoms get worse (actually pushing *against* the investment result). 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 of symptoms we consider.

7.2 Results: Health Capital

Smoking   We begin with the analysis based on genetic testing. Again, comparing means suggests some impact of test results. Twelve percent of individuals who learned they carry the HD mutation currently smoke, versus 8% of those who learn they do not carry the mutation. Perhaps more striking, conditional on ever smoking only 34.5% of those who learn they do not carry the HD mutation are still smokers, versus a full 62.5% of those who learn they do carry the HD mutation.

Columns 1 and 2 of Panel A of Table 8 shows regression analysis of smoking and genetic testing. In general we find that the results are stronger when we consider individuals who have ever smoked. This is not surprising: given that smoking is addictive and everyone in this sub-sample shared the same level of risk when young, we would expect differences to be muted. When we focus on all individuals who ever smoked, there is a 25 percentage point difference in the probability of quitting by test results.

The second set of columns in Panel A of Table 8 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. Overall, these results suggest that those individuals who know their lifespan is limited are more likely to smoke cigarettes. They are particularly more

likely to *continue* smoking, rather than trying to quit. This makes sense: individuals who have never smoked may not have any pull towards taking it up, but those who currently smoke (or did at one time) are likely to be more interested in either continuing or returning to the activity.

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. 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 8. Individuals with earlier symptom onset are significantly less likely to start cancer screening (this estimate is largely driven by differences in colonoscopy). Those with current symptoms are marginally significantly less likely to be on schedule with screening; the point estimate is similar if we separate mammograms and colonoscopies (available from the author).

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 our setting allows us to make causal statements about these relationships. Further, the simplicity of the setting and identification allows us to be confident about the channels driving our results. We argue that it is really the information about life expectancy that drives changes in behavior, rather than other circumstances.

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 attainment. These estimates, in turn, can be used to ask whether this channel is quantitatively important in driving global patterns of human capital investment. We argue the data suggest differences in life expectancy explain, perhaps, 10% of the cross-country variation in college completion, and substantially less than that of the over-time variation within this US. This implies, for example, that large increases in life expectancy in the worst off regions of the world might not

lead to rapid gains in higher education, at least not through this investment channel.

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. Again, we find 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 may have policy implications. It suggests that health improvements may be complementary with each other. 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.

Becker, Gary, *Human Capital*, Chicago, IL: University of Chicago Press, 1964.

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.

Dow, William, Tomas Philipson, and Xavier Sala i Martin, "Health Investment Complementarities under Competing Risks," *American Economic Review*, 1999, *89*, 1358–1371.

Fortson, Jane, "Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa," *Review of Economics and Statistics*, forthcoming.

Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 1972, *80*, 223–255.

Heckman, James, Robert LaLonde, and Jeff Smith, *The economics and econometrics of active labor market programs*, Elseveier, 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.

Kalemli-Ozcan, Sebnem, "Does the Mortality Decline Prmomote Economic Growth?," *Journal of Economic Growth*, 2002, *7*, 411–439.

——— **, Harl Ryder, and Davide Weil**, "Mortality Decline, Human Capital Investment and Economic Growth," *Journal of Development Economics*, 2000, *62*, 1–23.

Lorentzen, Peter, John McMillan, and Roman Wacziarg, "Death and Development," *Journal of Economic Growth*, 2008, *13*, 81–124.

Miguel, Edward and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 2004, *72*, 159–217.

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.

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.

Stoller, Avraham, "Mortality Risk and Economic Behavior." PhD dissertation, University of Chicago 2005.

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.

Figure 1:
Cognitive and Motor Scores by Time from Symptom Onset
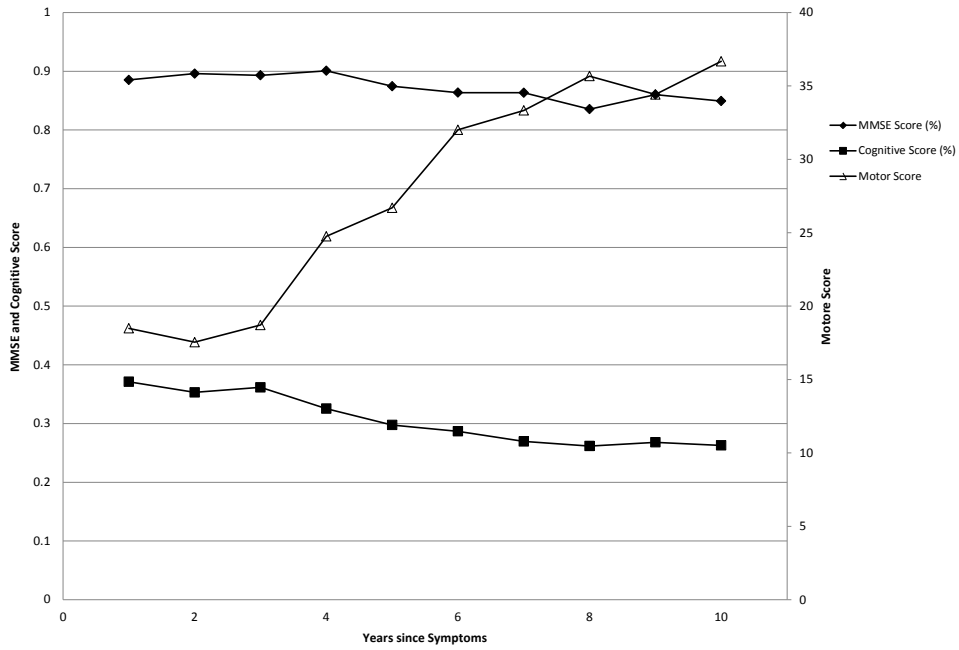
This graph shows motor scores and two cognitive test scores (on a mini mental state exam and a more general cognitive exam) by time from symptom onset. Time of symptom onset is identified by the participants.

Figure 2:
Educational Attainment by HD Symptom Timing
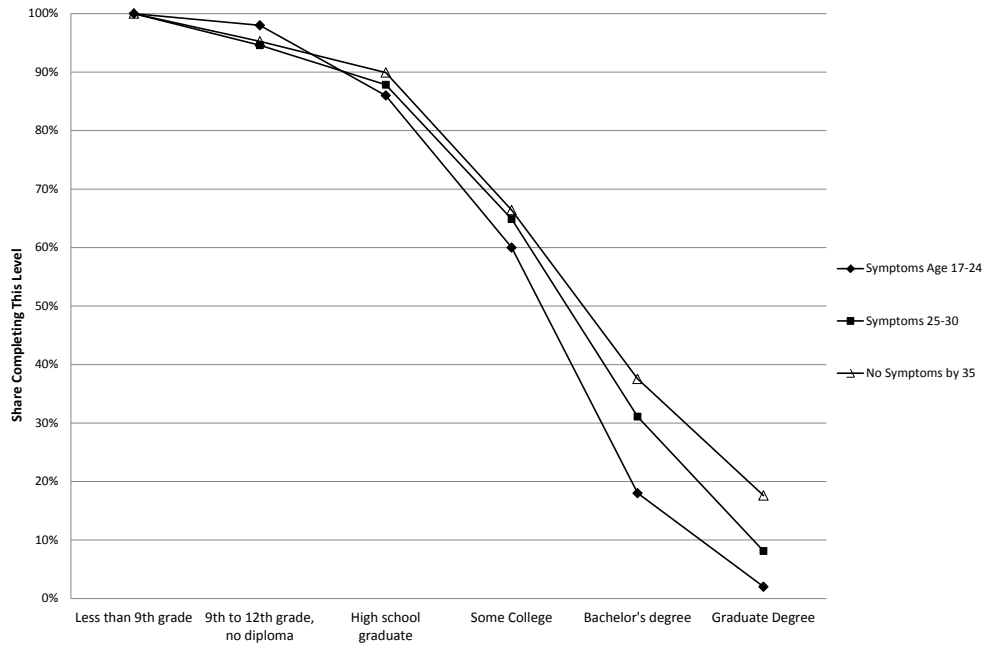
This graph shows the share of individuals who have completed each level of education, by age of symptom onset. The oldest group is those who have no symptoms by 35.

28

Figure 3
Life Expectancy, Predicted Education and Job Training

Notes: This graph shows predicted college completion, years after high school and job training by years of healthy life expectancy. This is based on our estimated coefficients and information about life expectancy with HD; details of the latter calculation are in Appendix A.

Figure 4:
Impact of Symptoms on Health Behaviors

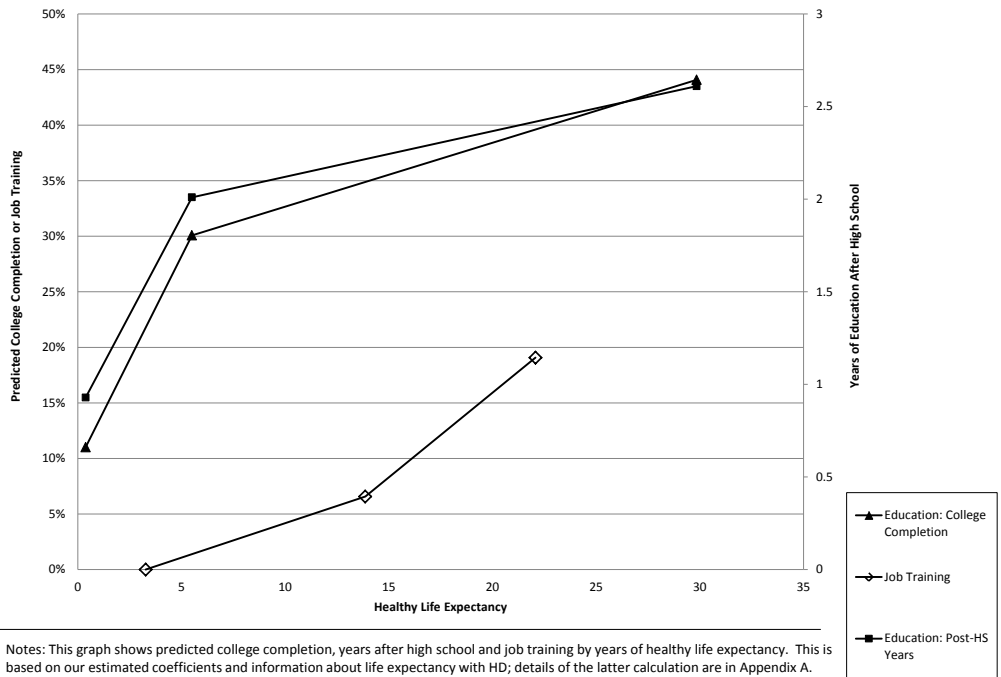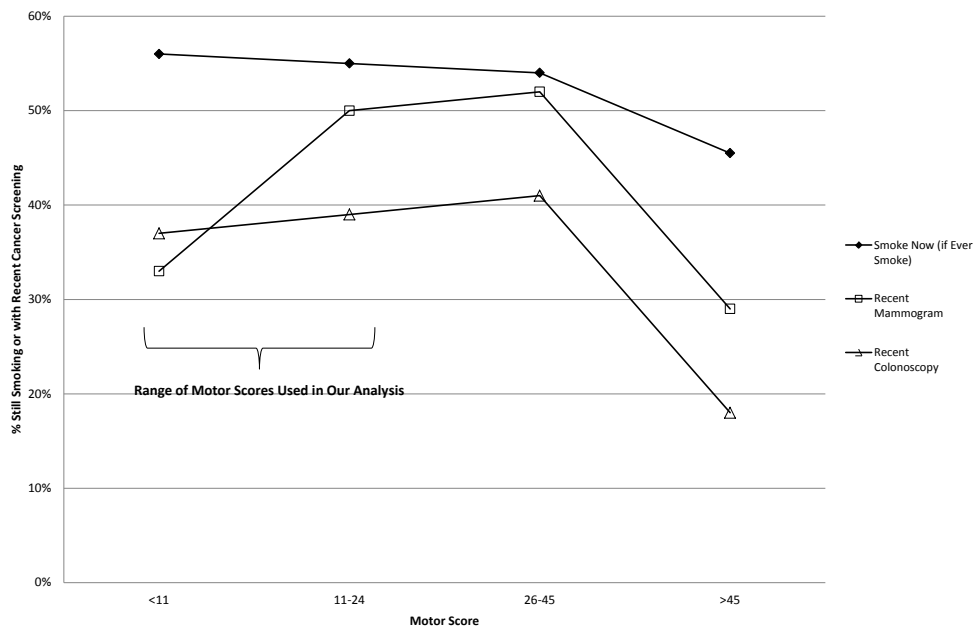This graph shows behavior among individuals who report that they are sure they have HD, by motor scores. Since everyone in the graph has the same information about HD status, differences in behavior can be attributed to differences in physical symptoms.

29

Table 1: *Life Expectancy and Working Life Expectancy by HD Status*

	Life Expectancy	Healthy Life Expectancy
No HD	79.1	79.1
HD, No Symptoms at 25	60.4	44.8
First Symptoms 17-24	41.1	24.5
First Symptoms 25-30	47.9	31.4
First Symptoms 31-35	53.1	36.8
First Symptoms 36-45	59.9	44.1
First Symptoms 46-55	68.0	53.5
First Symptoms 56-65	74.4	62.4

Notes: This table shows life expectancy and working life expectancy for individuals with and without HD. Calculations assume all individuals are alive at 17. The calculation for individuals with no symptoms at 25 assumes a distribution of possible age of onset. Working life expectancy is based on data (from our survey) on the chance of working by year from symptom onset. Details of these calculations appear in Appendix A.

Table 2: **Summary Statistics**

Panel A: Count of COHORT Participants		
	All COHORT	*ALD Sample*
Manifest HD	1,380	272
Pre-Manifest HD	241	52
50% Risk for HD	524	106
25% Risk for HD	11	0
Tested, Do not Carry Mutation	131	21

Panel B: Outcome Summary Statistics			
	Mean	*Std. Dev.*	*# of Obs*
Years of Education (Age>=18)	13.89	2.57	2846
Bachelor Degree (Age>=22)	0.364	0.481	2820
Job Training for Advancement (Age 35-60)	0.140	0.348	355
Smoke Now (Age>=20)	0.134	0.341	2841
Smoke Now, if Smoke Ever=1 (Age>=20)	0.473	0.499	807
Ever Mammogram (Age>=35)	0.823	0.382	294
Recent Mammogram (Age>=35)	0.491	0.550	295
Ever Colonsocopy (Age>=40)	0.510	0.500	413
Recent Colonsocopy (Age>=40)	0.402	0.491	412

Panel C: Demographic and Falsification Summary Statistics			
	Mean	*Std. Dev.*	*# of Obs*
Age	50.2	13.5	2276
Male (0/1)	0.415	0.492	2289
White	0.935	0.245	2289
Carry HD Mutation	0.812	0.390	2197
UHDRS Motor Score	23.4	22.4	2259
Any Symptoms (0/1)	0.635	0.481	2273
Age of First Symptoms (if any symptoms)	43.1	12.4	14.32

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, retirement savings, mammogram and colonsocopy; education and smoking are covered in the primary COHORT surveys. The summary of HD status in Panel C excludes control (i.e. never at risk) individuals.

Table 3: **Balancing Tests**

	Panel A: Human Capital Analysis							
	Tested before 30		*Analysis by Symptom Onset Age*					
	(Symptoms 35+)		Categories for Education			Categories for Job Training		
	Positive	*Negative*	*17-24*	*25-30*	*Over 30*	*20-29*	*30-39*	*Over 40*
	(n=114)	*(n=36)*	*(n=74)*	*(n=123)*	*(n=1670)*	*(n=154)*	*(n=346)*	*(n=1320)*
Male (0/1)	0.350	0.277	0.432	0.495	0.415	0.461	0.471	0.414
White	0.956	0.944	0.972	0.910	0.941	0.941	0.942	0.940
Current Age	33.36	33.72	33.94^a	39.58^a	55.3^a	36.33^a	45.64^a	58.49
In US	0.789	0.805	0.918	0.926	0.908	0.909	0.936	0.910
	Panel B: Health Analysis							
	Tested Ever		*Early Symptoms*		*Analysis by Symptom Onset Age*			
	(No Symptoms)							
	Positive	*Negative*	*No*	*Early*	*30-40*	*35-45*	*Over 45*	
	(n=167)	*(n=148)*	*Symptoms*	*Symptoms*	*(n=391)*	*(n=478)*	*(n=941)*	
Male (0/1)	0.347	0.297	0.306^a	0.472	0.465	0.453	0.415	
White	0.952	0.959	0.890	0.963	0.936	0.929	0.950	
Current Age	38.27^a	45.33	41.34^a	47.01	45.37^a	49.25^a	60.55	
In US	0.760	0.750	0.953	0.927	0.941	0.935	0.906	

Notes: This table shows balancing on early life variables. Panel A shows this for testers, Panel B by age of onset (for individuals who currently have symptoms) and Panel C by current level of symptoms. In Panel C, the No Symptom group is at risk but with a motor score of 0, early symptoms include people with a motor score less than 10 and later symptoms are those with a motor score between 11 and 20. [a] indicates different from the other test result group, the group one age group older, or the more advanced symptom group, at 1% level; [b] indicates different from one age group older, or other test result group, at 5% level.

Table 4: **Gene Testing and Educational Attainment**

	Panel A: Main Results and Robustness					
	Main Results		**Family Fixed Effects**		**Control Referral Method**	
	# Post-HS Years Education	Bachelor Degree	# Post-HS Years Education	Bachelor Degree	# Post-HS Years Education	Bachelor Degree
Sample	*Age>=18, Finish HS*	*Age>=22, Finish HS*	*Age>=18, Finish HS*	*Age>=22, Finish HS*	*Age>=18, Finish HS*	*Age>=22, Finish HS*
Tested, Negative	1.137***	.272***	2.294***	.510***	1.105**	.267***
	(.433)	(.090)	(.739)	(.159)	(.433)	(.091)
Tested, Positive	.0867	-.0294	1.230*	.170	.151	-.016
	(.269)	(.057)	(.681)	(.153)	(.269)	(.057)
pos vs. neg p-value	.02	.002	.18	.05	.04	.002
Standard Controls	YES	YES	YES	YES	YES	YES
# of Observations	1576	1565	872	866	1575	1564

	Panel B: Falsification			
	Pre-Symptom Education		**Older Testers**	
	Complete High School		# Post-HS Years Education	Bachelor Degree
Sample	*Age>=18,*		*Age>=18, Finish HS*	*Age>=22, Finish HS*
Tested, Negative	-.010		.736*	.208**
	(.052)		(.443)	(.091)
Tested, Positive	.048		.562**	.137**
	(.033)		(.287)	(.059)
pos vs. neg p-value	.29		.73	.49
Standard Controls	YES		YES	YES
# of Observations	1747		1107	1100

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

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 controls are dummies indicating the way the individual was recruited into the study (by site personal, online, at a meeting, etc).

Table 5: *Symptom Onset and Educational Attainment*

	All			Robustness: Family Fixed Effects			Falsification: High School Completion
Sample	# Post-HS Years Education	Start College	Bachelor Degree	# Years Education	Start College	Bachelor Degree	Complete High School
	Age>=18, Finished HS	*Age>=18, Finished HS*	*Age>=22, Finished HS*	*Age>=18, Finished HS*	*Age>=18, Finished HS*	*Age>=22, Finished HS*	*Age>=18*
Symptoms Age 17-24	-1.681***	-.140**	-.331***	-1.810**	-.017	-.435**	-.011
	(.335)	(.066)	(.070)	(.845)	(.195)	(.201)	(.044)
Symptoms Age 25-30	-.600**	.019	-.140***	-.297	.140	-.101	-.073**
	(.241)	(.047)	(.053)	(.748)	(.170)	(.177)	(.031)
P-VALUES:							
17-24 vs. 25-30	.003	.02	.02	.17	.53	.20	.14
Linear Trend	<.001	.23	<.001	.05	.68	.04	.12
Standard Controls	YES	YES	YES	YES	YES	YES	YES
Family FE	NO	NO	NO	YES	NO	YES	NO
# of Observations	1667	1667	1664	928	928	926	1860

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

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 (results are similar if we limit this latter group only to individuals who carry the mutation). Standard controls: gender, country dummies, race, a cubic in age. P-values for tests of equality between dummy variables is reported. We also report a p-value from a regression in which we estimate a coefficient on a linear control for group (3 groups).

Table 6: ***Symptom Onset and Job Training***

	Main Results:	Falsification:
	Job Training for Promotion or Advancement	**Other Job Training**
Sample	*Age 20-60,*	*Age 20-60,*
	Ever Worked	*Ever Worked*
Symptoms Age 20-30	-.267**	.135
	(.109)	(.160)
Symptoms Age 31-40	-.125**	.033
	(.044)	(.091)
P-VALUES:		
20-30 vs. 31-40	.19	.49
Standard Controls	YES	YES
# of Observations	235	235

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.

Table 7: ***Quantitative Implications***

Panel A: Elasticities			
	Elasticity WRT *Healthy Life Expectancy* *(Symptoms)*	*Elasticity WRT* *Healthy Life Expectancy* *(Testing)*	
College Completion	0.41	0.89	
Years of Education after High School	0.34	0.65	
Job Training for Advancement	1.41		
Panel B: Long Differences in College Completion in US			
	Actual Difference in College Share	Difference Predicted by Life Expectancy	*Share of Difference* *Explained by LE*
2010 vs. 1960	21.7%	0.49%	2.3%
2010 vs. 1980	10.4%	0.48%	4.6%
2010 vs. 2000	7.0%	0.13%	1.9%
Panel C: Global Differences in Tertiary Enrollment			
	Actual Difference in College Share	Difference Predicted by Life Expectancy	*Share of Difference* *Explained by LE*
High income vs. Middle Income	45.5%	3.7%	8.2%
Middle Income vs. Low Income	17.4%	1.5%	8.3%
High Income vs. Low Income	63.0%	7.4%	11.7%
All Country Pairs (Average Share)			9.3%

Notes: This table discusses magnitudes. 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 esimtates the share of global differnces in tertiary enrollment which could be explaind 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 eenrollment and life expetancy data is from the World Development Indicators. Calculations in Panels B and C use an elasticity value of 0.41.

Table 8: *Life Expectancy and Health Investments*

Panel A: Smoking				
	Genetic Testing		**Symptom Onset**	
	Smoke Now	*Smoke Now*	*Smoke Now*	*Smoke Now*
Sample	*All*	*Ever Smoke=1*	*All*	*Ever Smoke=1*
Tested, Negative	-.007	-.070		
	(.034)	(.122)		
Tested, Positive	.043	.255**		
	(.032)	(.121)		
Have Early Symptoms			.076*	.290***
			(.037)	(.110)
positive vs. negative p-value	.20	.03		
Standard Controls	YES	YES	YES	YES
# of Observations	601	124	422	115
Panel B: Cancer Screening				
		Ever Screen	*Recent Screen*	
Symptom Onset Before Screening Age		-.149***		
[Mam: 30-40, Colon: 35-45]		(.051)		
Have Early Symptoms Now			-.161*	
			(.09)	
Colonoscopy		-.398***	-.160*	
		(.047)	(.082)	
Standard Controls		YES	YES	
# of Observations		447	142	

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, education level (linear). 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 colonsocopy) 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. Controls in both panels: gender, country dummies, race, a cubic in age and educational attainment.

Appendix A: Incentive Calculations

A.1 Details of Life Expectancy and Healthy Life Expectancy.

Life expectancy is calculated based on information on the course of HD. We assume that after symptom onset individuals have 3 years of healthy, normal, life and 20 years of life total. In addition, we take into account the fact that it is possible they might die from something else prior to HD; we incorporate overall death rates from US life tables as our background death probabilities. The figures reported in Table 1 are simply the sum of these years of life expected. In mathematical terms, for someone with onset at age $\hat{a}$, life expectancy is:

$$\hat{a} + \sum_{t=\hat{a}}^{\hat{a}+20}(1 - p_{t-\hat{a}})(t - \hat{a})$$

where $p_{t-\hat{a}}$ is the chance of having died between ages $\hat{a}$ and t from a non-HD cause. Healthy life expectancy is similar, although with only three years of healthy life after symptom onset.

$$\hat{a} + \sum_{t=\hat{a}}^{\hat{a}+3}(1 - p_{t-\hat{a}})(t - \hat{a})$$

A.2. Detailed Incentive Calculations

In addition to calculating life expectancy and working life expectancy, we can more precisely calibrate investment incentives by calculating financial or years of life returns to these choices, by group.

Education and Job Training

Decisions about education and job training are made when young, with an expectation about future years of work. To give a sense of the variation in benefits, we define 3 groups: individuals without the HD gene, those with the gene but later symptom onset and those with the gene and earlier symptom onset.

We use data on earnings profiles by education group in the US Census and information on length of time working (as in Table 1). Columns 1-3 use earnings profiles from the Census to calculate present discounted value of earnings for someone currently age 25 with varying education levels. The longer lifespan of non-HD individuals gives them a much higher return to education, as illustrated by the larger differences in earnings across groups. To take one concrete comparison, the return to a college versus a high school degree is 25% higher for non-HD individuals versus those with later HD symptom onset, and 115% higher than those with early symptom onset.

Quantifying the benefits of job training requires more assumptions, particularly about the benefits of this training. We rely on existing literature, and assume that a job training program has a benefit of 2.5% in wages in perpetuity[14]. We do these calculations assuming an individual has some college but no degree, which is the education level of the average individual in our sample. Column 4 of Table A1 shows the discounted wage returns to job training at 30 by group. Clearly, the returns are higher (almost three times as high) for the control group versus the early symptom group.

---

[14]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.

Smoking

Cigarette smoking leads to an increased probability of death from a variety of causes, including lung cancer and emphysema. Since most of this excess death occurs later in life the incentives not to smoke will be limited by a shortened life expectancy. To get a sense of the magnitude of this impact, we use data on total probability of death by age for smokers, non-smokers and former smokers from www.smokefree.gov. We consider someone who begins smoking at age 20, and smokes one pack a day (results would be qualitatively similar for other assumptions). 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 regardless (in this case, from HD). In line with our later analyses, we focus on the decision to quit smoking and estimate the benefits to quitting at age 40, given that the individual started smoking at 20.

These probabilities are displayed in Appendix Figure 1. As expected, individuals with higher alternative life expectancy have higher benefits from quitting. The benefits diminish some when we consider an alternative life expectancy of 80, since death even for non-smokers increases a lot in this age range. The impacts are large. For someone who would otherwise live only to 50, the benefit of of quitting at 40 is less than 1 percentage point. In contrast, for someone who would otherwise live to 70, this figure is 7.5 percentage points.

Cancer Screening

There are two ways to think about quantifying the varying benefits to cancer screening. The first is to look at the chance of ever developing these types of cancers by group. Clearly, the benefit to doing any screening will vary with the chance of developing cancer, and this chance is significantly lower for individuals who expect to die earlier. The second approach is to quantify the years of survival gained by detection. People with an expectation of earlier mortality are not only less likely to ever develop the disease, but also they gain fewer years of life with early detection.

Columns 1 and 2 of Table A2 explore the benefits to screening for breast and colon cancer based on the chance of ever developing the disease. We distinguish between individuals who already have early symptoms at the first age for which screening is recommended, those who carry the gene but have symptoms only later, and non-HD individuals.[15] For both breast and colon cancer, we find the chance of developing the disease varies significantly across risk groups. In the case of breast cancer, individuals who already showing symptoms by 35 have a lifetime risk of 2%, versus over 13% for non-HD individuals; for colon cancer, these figures are 0.2% and 7.2%.

These numbers are even more extreme if we incorporate the lost life years in addition to the probabilities (in Columns 3 and 4 of Table A2).[16] Someone with early symptoms can expect to gain just 0.03 years of life from mammography, versus a full year of life for someone without HD. In the case of colon cancer, the corresponding numbers are 0.008 years and 0.36 years. Overall, the figures suggest there are significant difference in benefit of screening by HD status. Comparing the two cancer screenings, in absolute terms we see the difference in benefits is larger for mammography due to the higher incidence of breast cancer.

---

[15]For the symptomatic individuals, we assume death follows in 15 years. For the individuals who carry the gene but do not have symptoms, we use the data to inform us about the average age of symptom onset given no symptoms at the age of starting screening and assume death follows 15 years after.

[16]To do this calculation we assume that people would detect cancer at Stage 3 regardless of screening. We assume that if people survive five years they experience their normal lifespan, and otherwise life an average of 2.5 years. Obviously these assumptions are open to interpretation, but we argue they give us at least a broad sense of the magnitude of these effects.

Appendix Figure 1
Impact of Smoking on Death by Non-Smoking Life Expectancy
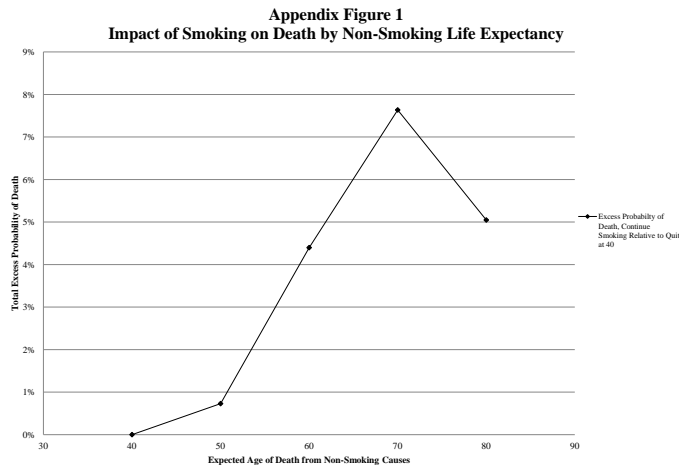
Notes: This figure shows the excess probability of death from smoking, grouped by non-smoking life expectancy. Someone in the "40" bin expects to die at 40 from non-smoking related causes for sure. For both lines we assume individuals start smoking at 20 and smoke 1 pack per day.

Table A1: Incentives for Education and Job Training

| | Present Discounted Value of Earnings | | | Returns to Job |
| | No HS Deg. | HS Deg. | Bachelor Deg. | Training at 30 |
	(2)	(3)	(4)	(5)
Non-HD	$431,039	$562,301	$892,464	$17,716
Gene Positive, Later Sympt.	$355,633	$465,852	$732,543	$13,894
Gene Positive, Early Sympt.	$225,137	$291,069	$444,077	$7,003

Notes: This table calculates the returns to education and job training for groups with varying HD status. The variation derives from the variation in expected retirement age, which we calculate based on our data. Earnings data comes from the US Census, assuming 3% discounting. For job training we assume a 2.5% wage return to job training in perpetuity, and use individuals with some college as the baseline, since this represents the average individual in our data.

Table A2: Incentives for Cancer Screening

| | Chance of Ever Developing | | Years of Life Saved by Screening | |
| | Breast Cancer | Colon Cancer | Breast Cancer | Colon Cancer |
	(1)	(2)	(3)	(4)
Non-HD, Life Exp=85	13.4%	7.2%	1.00	0.37
With Gene, No Sympt. at 35, LE=65	4.6%	1.6%	0.23	0.05
With Gene, Sympt. at 35, LE=50	2.0%	0.2%	0.03	0.01

Notes: This table calculates the benefit of cancer screening for groups with varying HD status. Data on cancer incidence comes from the SEER cancer statistics (http://seer.cancer.gov/). Life expectancy for HD individuals by symptom status is based on observations from our data. Years of life saved is based on the assumption that the disease is detected at stage 1 with screening and stage 3 without screening; these have varying survival probabilities.