Online Appendix for "Limited Life Expectancy, Human Capital and Health Investment"

Authors: Emily Oster, E. Ray Dorsey, Ira Shoulson

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 $\hat{\alpha}$ 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 $\hat{\alpha}$ are calculated as:

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

Column 1 of Table A1 reports this life year calculation by group. The figures in square parentheses in the symptom group labels indicate $\hat{\alpha}$ 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[1]. 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 $\hat{\alpha}$. Column 3 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

[1]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.

	Life Years (1)	*DALYs* (2)	*Return to College* (3)	*Return to Job Training* (4)
Groups for Gene Testing Analysis				
Without Gene [$\hat{\alpha}$ =25]	53.1	53.1	$412,199	
With Gene, No Symptoms at 35 [$\hat{\alpha}$ =25]	33.8	29.7	$311,485	
Groups for Education Symptom Analysis				
Symptoms 15-18 [$\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	$336,495	
Groups for Job Training Analysis				
Symptoms 20-29[$\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

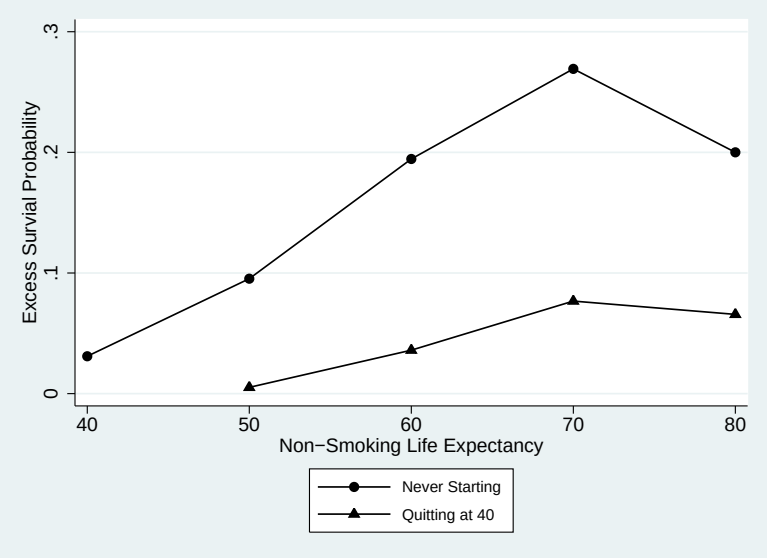
A.4. Smoking

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. 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

	Chance of Ever Developing	
	Breast Cancer	*Colon Cancer*
	(1)	(2)
Groups for Breast Cancer Analysis		
At risk at 40	9.4%	
Symptoms 30-40	2.3%	
Groups for Colon Cancer Analysis		
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**		
	15-18	*19-22*	*23-28*	*Over 30*	*20-29*	30-39	*Over 40*
	(n=25)	*(n=45)*	*(n=86)*	*(n=1991)*	*(n=26)*	*(n=60)*	*(n=255)*
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	$36.8^{a,b}$	$53.9^{a,b,c}$	37.9	46.2^{d}	$58.1^{d,e}$
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**	
	None	*Early*	*30-45*	*Over 45*
	(n=172)	*(n=359)*	*(n=140)*	*(n=950)*
Male (0/1)	0.32	0.46^{f}	0.34	0.39
White	0.89	0.97^{f}	0.91	0.91
Current Age	41.2	47.4^{f}	50.6	61.5^{g}
In US	0.96	0.93	0.95	0.86^{g}
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): ^{a} compared to onset 15-18;^{b}compared to onset 19-22; ^{c}compared to onset 23-28; ^{d}compared to onset 20-29; ^{e} compared to onset 30-39; ^{f}compared to no symptoms;^{g} compared to onset 30 to 45.

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

Panel A: Main Results and Robustness				
	Main Results		**Control Referral Method**	
	# Post-HS Years Education	Bachelor Degree	# Post-HS Years Education	Bachelor Degree
Sample	*Finish HS*	*Finish HS*	*Finish HS*	*Finish HS*
Tested, Negative	.677*	.251***	.605	.221**
	(.369)	(.096)	(.378)	(.103)
Standard Controls	YES	YES	YES	YES
# of Observations	126	126	126	126
Panel B: Falsification				
	Pre-Symptom Education		**Older Testers**	
	Complete High School		# Post-HS Years Education	Bachelor Degree
Sample			*Finish HS*	*Finish HS*
Tested, Negative	.005		.194	.073
	(.049)		(.370)	(.093)
Standard Controls	YES		YES	YES
# of Observations	135		135	135

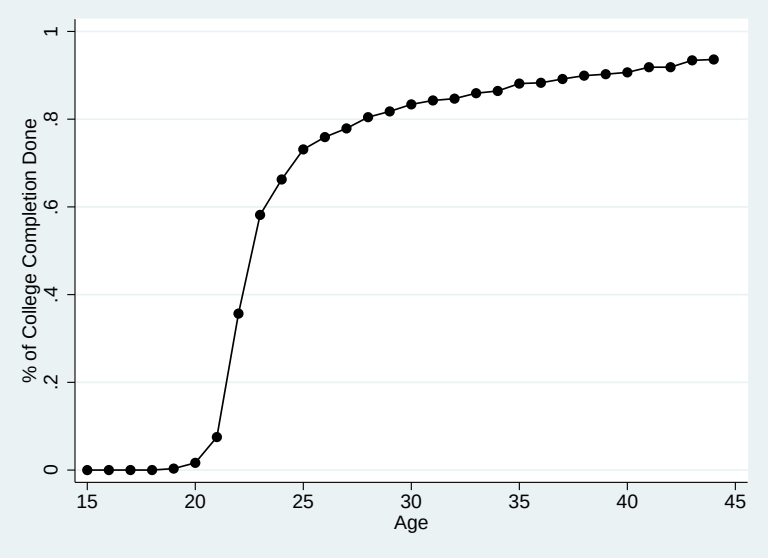
Notes: This table shows the impact of test results on educational attainment. The omitted category is individuals who have been tested and learn they *do* carry the HD mutation. The main results include individuals who are tested before or at the age of 28, and (if positive) do not develop symptoms before 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

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

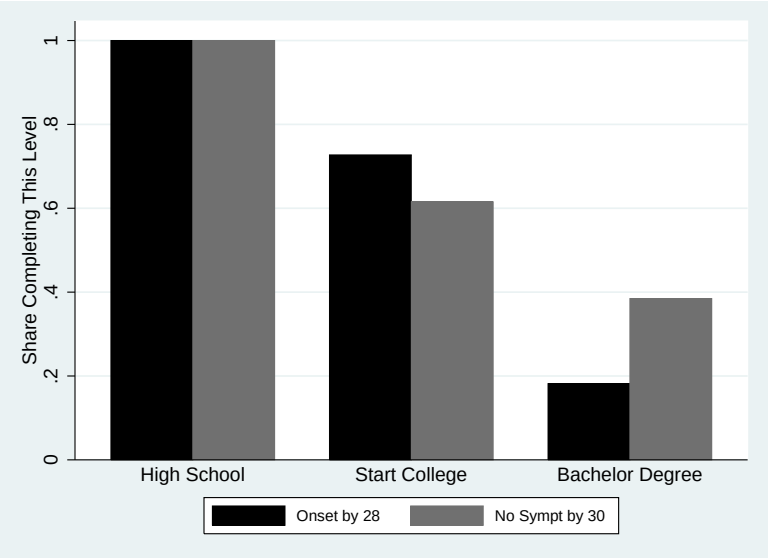
Notes: This table shows the impact of test results on smoking. The omitted category is individuals who have been tested and learn they *do* carry the HD mutation. Standard controls: gender, country dummies, race, age, dummies for education. Standard errors in parentheses.*significant at 10% **significant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