ESSAYS ON HUMAN CAPITAL, EXPECTATIONS AND BEHAVIORS

A dissertation presented

by

Man Jia

to
The Department of Economics

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

in the field of

Economics

Northeastern University
Boston, Massachusetts
August, 2012
ESSAYS ON EXPECTATIONS, HUMAN CAPITAL AND LABOR MARKET OUTCOMES

by

Man Jia

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate School of Northeastern University
August, 2012
# TABLE OF CONTENTS

## Acknowledgements

## Abstract

### Chapter 1:
Can Differences in Mortality Expectations Explain Black-White Differences in Educational Attainment for U.S. Youth?  1-54

1. Introduction  
2. Literature Review  
3. Empirical Specification  
4. Data  
5. Results  
6. Discussion and Conclusion  
References  
Figures and Tables

### Chapter 2:
Do People Change Their Health-Related Behaviors When They Raise Their Survival Expectations?  ...55-91

1. Introduction  
2. Literature Review  
3. Empirical Approach  
4. Data  
5. Results  
6. Discussion  
7. Conclusion  
References  
Figures and Tables

### Chapter 3:
“The Bell Curve” Revisited: Does “Intelligence” Still Matter?  92-119

1. Introduction  
2. Literature Review  
3. Empirical Specification  
4. Data  
5. Results  
6. Discussion and Conclusion  
References  
Tables
Acknowledgements

I am deeply indebted to my advisor, Bill Dickens, for his constant inspiration, effective guidance, generous advice and steady support. The dissertation would not exist without his help. I am especially grateful to Professors Neil Alper and Andrew Sum for their constructive suggestions and their great patience. I would sincerely thank Professors Ni He, María José Luengo-Prado and Osborne Jackson for their constructive comments. I also benefit from discussions with my Northeastern fellow students. All remaining errors are mine own.
Abstract

While previous theoretical analysis suggests that racial differences in death rates might play an important role in explaining the black-white education gap in the U.S., there is little empirical research to test this implication. This paper estimates the extent to which differences in expected mortality risks prior to entering college can explain differences in adult educational attainment in the 2000s, using data from the 1997 National Longitudinal Survey of Youth (NLSY). This study finds that the impact of mortality is not as important as suggested by prior research. Specifically, of the total black-white education gap (roughly 1.12 schooling years), only about 0.05 years or less can be attributed to differences in mortality expectations. As this study confirms, the role of self-reported mortality expectations in explaining black-white education gap is small, and the impacts of death expectations from actual death rates on education are statistically insignificant for reference groups.

The second chapter examines whether individuals are likely to alter personal health-related behaviors once they increase their subjective longevity expectations. To determine if there is a relationship between health behaviors and longevity beliefs, I test one of implications of the Cutler-Glaeser (2009) smoking decision model, which suggests that nonsmokers whose expected survival probabilities have increased are unlikely to start smoking. This study uses data from the Health and Retirement Study (HRS), which is conducted every two years, from 1992 to 2010 (Waves 1-10). Specifically, the HRS data show that a certain share (2.13%) of nonsmokers at Wave t-1 whose subjective expected longevity beliefs increased across two waves did start smoking at Wave t. This small percentage is close to the fraction of new smokers who have steady or decreased survival beliefs (1.99% and 2.19%, respectively). This finding also holds true for other behaviors
including heavy drinking, obesity, and physical inactivity. Thus, the findings I present based on the HRS data contrasts with the Cutler-Glaeser model.

Using scores from the Armed Forces Qualification Test (AFQT), Herrnstein and Murray (1994) reported that intelligence can be a powerful predictor of a range of outcomes related to social behaviors (e.g., incarceration, marriage, out-of-wedlock birth, low birth weight and poverty). In contrast, a recent study found that measured intelligence using the same AFQT scores plays a considerably smaller role on an important socioeconomic indicator, namely, hourly wages as measured from 2000 to 2010. My third paper attempts to replicate the Herrnstein and Murray study using a different data set, the 1997 National Longitudinal Survey of Youth to look into several behaviorally-related social outcomes. The main finding is that, in general, the role of AFQT scores in predicting social behaviors has not substantially changed over the last 20 years. I provide a few possible explanations for this finding.
Chapter 1:
Can Differences in Mortality Expectations Explain Black-White Differences in Educational Attainment for U.S. Youth?

1. Introduction

One of the implications of human capital theory is that an investment is more valuable if the time span over which individuals may receive returns on that investment is longer, all else being equal (Becker, 1964; Ben-Porath, 1967). In particular, educational attainment, as an important human capital investment, is more valuable if one is able to reap educational returns over a longer time horizon. A related implication of human capital theory as it pertains to education is that individuals with higher expected mortality risks—and thus less time to benefit from educational investments—should spend less time in school. Considering the fact that the average mortality rate for black youth is higher than that for whites, differences in mortality would be likely to lead to lower educational attainment for blacks. Despite the growing body of literature documenting the strong association between mortality risks and educational attainment (e.g., Kalemli-Ozcan, 2002; Murphy and Topel, 2006; Lorentzen et al., 2008; Kuzmenko, 2011; Oster et al., forthcoming), the black-white race-based links remain largely unsubstantiated.

Education gaps between racial and ethnic groups have persisted for a long time, but have been gradually diminishing, especially since 2000. As illustrated in Figure 1A, in 2008 the proportion of 15-34 year-olds with at least high school diploma/GED (incarceration unadjusted), broken down by race and gender, is as follows: 76.42% for white men, 68.66% for black men, 78.54% for white women and 73.03% for black women (March Current Population Survey (CPS) data, 2008). Figure 1B shows that the college graduation rates for
these four racial/gender groups in 2008 are roughly 19.75%, 10.93%, 25.58% and 12.45%, respectively.

Racial and ethnic inequities in mortality risks also continue to persist in the U.S., even after controlling for socioeconomic and marital status. Although longevity for both blacks and whites has improved, especially since the 1960s – the black-white mortality gap remains. In fact, despite moderate narrowing of the racial mortality gap in the early 1990s, researchers continue to report on race-based differences in mortality (Satcher et al., 2005; Harper et al., 2007; Macinko and Elo, 2009; Geronimus et al., 2011; Hummer and Chinn, 2011; Jackson et al., 2011). Thus, two racially significant gaps are still evident in the U.S., despite some narrowing of both over the past decade: educational attainment levels (higher for whites), and mortality rates (higher for blacks).

We have, therefore, two longstanding (albeit narrowing) gaps between whites and blacks that have to date been minimally investigated with respect to their connectedness. While researchers have looked at whether mortality expectations play a role in educational attainment, the simple trends of education and mortality do not, as yet, indicate a race-based connection. Consequently, this paper examines whether differences in expected black-white mortality risks can explain differences in later educational attainment. This paper seeks to fill an important literature void by investigating the extent to which the racial inequalities in actual and anticipated mortality risks can explain the race-based educational differences among American youth. This paper is expected to contribute to the literature by providing empirical estimates of the prediction of expected mortality on educational attainment, as well as by directly using perceived probability of death as a measure of expected death risks—neither of which have been addressed in many previous studies.
As noted, a significant aspect of this study focused on racial differences in educational attainment. There are many factors that contribute to black-white education differentials, including family background, cognitive ability, peer effects, economic circumstances, education and labor market discrimination. This paper, however, investigates a potential factor that has received little attention in the literature—namely, the role of mortality expectations of adolescents in accounting for the black-white gap in educational attainment. In particular, I examine the extent to which youth expectations for future death predicts educational attainments when they are adults using the data from the National Longitudinal Survey of Youth (1997 cohort). One measure of the mortality expectations is self-reported probabilities of death available in NLSY97. The second data source that contributes to expectations on mortality is the observable mortality rates in race-sex-age-county reference groups from the national mortality data (detailed in Section 3).

The findings of this study suggest that in contrast to the important role of mortality differences suggested in previous literature (e.g., Kalemli-Ozcan et al., 2000; Kalemli-Ozcan, 2002; Lorentzen, et al., 2008), there is insufficient evidence to claim that mortality risk plays a role in the black-white education gap. Specifically, of the total black–white education gap (roughly 1.12 schooling years for respondents aged 25-28), expected mortality risk can account for, at most, 0.05 years. Moreover, a decrease of one percentage point in subjective mortality expectations is shown to lead to a highest-grade-completed decline of only about 0.004 years.¹ It should be noted that the coefficient associated with the actual death rate for a particular reference group is always statistically insignificant. In addition, I address the potential omitted variables bias by controlling for additional variables

¹ In this paper, subjective mortality expectations are measured using percentage points (ranging between 0 and 100).
compared with the basic estimation. While the impact of the subjective death rate on education in the benchmark estimation is significant, after adding a set of additional control variables, the coefficients become insignificant and their magnitudes decline slightly.

Identifying and understanding the role of expected mortality risks in educational attainment will improve our understanding of a potential source of social inequality. Educational stakeholders have long focused on closing the racial educational gap through a variety of means, principally involving programmatic changes. An inevitable outcome of such changes is discussion about which strategies or programs work, which ones do not. Thus, achieving a better understanding of how mortality differences affect educational attainment can provide insights into developing appropriate policies to help blacks improve their educational and employment outcomes in the future.

Two characteristics distinguish this paper from prior research. First, I estimate the impact of mortality expectations linked to a specific pre-college age cohort on educational attainment in adulthood using individual microdata, as well as aggregate data. Second, in addition to using the actual death rates of the same race-sex-age-county groups, this paper takes advantage of the self-reported death beliefs of 16 to 18 year olds as another measure of death expectations—a component not addressed in many previous studies. The subjective mortality probability is particularly important because it enables us to relax the usual assumption that individual death expectations are derived from actual death rates.

This paper makes two important contributions to the existing literature. First, I provide empirical estimates for the association between expected mortality risks and educational attainment—and, as noted, the role of mortality differentials is much less important than what previous literature has suggested.

Second, this study also adds to the literature concerning the role of subjective perceptions by investigating the predictive ability of the subjective beliefs of adolescents
with respect to the likelihood of future education. As a result, this study finds that the subjective mortality expectations of an individual can in some sense predict educational attainment. Moreover, this paper extends previous subjective mortality studies by looking at particular racial differences. Specifically, I find that among the many factors that account for black-white education differences, the role of subjective mortality perceptions is negligible.

The remainder of this paper is structured as follows. Section 2 provides a discussion of the relevant research associated with the mortality expectation-educational attainment theme. In Section 3, I introduce the underlying theoretical framework for this study and set out the empirical approach used to obtain the results reported herein. Section 4 presents data sets and sample description. In Section 5, I present the results and discuss robustness checks. Finally, Section 6 provides a discussion of several possible explanations for my findings, summarizes the conclusions, and discusses important limitations.

2. Literature Review

A significant body of literature has examined the effect of mortality risks on human capital investments. Many early studies employ actual mortality rates to measure expectations of mortality risks, and look at mortality risks from cross-national perspectives to examine the complex associations between mortality, education and long-term growth. Kalemli-Ozcan, et al. (2000) present a model implying that one percent reduction in mortality should lead to about one percent increase in schooling. Soares (2005) suggests that reductions in child mortality rates tend to play a significant role in the rise of education levels for both children and parents. His focus involves a theoretical analysis of long-term growth, using education as a mechanism. Lorentzen, et al. (2008) show that in their cross-national and subnational African data higher adult mortality is strongly
associated with lower investment in human capital by shortening time horizons. Additionally, a number of studies utilize microdata and causal analysis to examine specific developing countries. For example, Datar et al. (2007) find in rural India, and Jayachandran and Lleras–Muney (2009) claim in Sri Lanka that reductions in mortality played a significant role in educational increases. In contrast, Hazan (2011, 2012) use data from “post-transitioned countries” (countries that in 1960 had average life expectancies at birth above 50 years) to investigate the impact of increasing longevity on the acquisition of human capital between 1960 and 1990. Hazan concludes that a longer life expectancy is less important in terms of quantitative outcomes than previously thought. A recent study, Oster et al. (2012), also argue that in the U.S. the role of life expectancy in accounting for education is very small.

Along this vein, Gan and Gong (2004) is the most related research to this paper. The researchers use group-level mortality risk data from 1979-1981, which are obtained from the U.S. decennial life tables. Their data show that the average life expectancy—conditional on surviving to age 16—was 66.2 years for black males and 72.1 years for white males. Calibration results of Gan and Gong’s theoretical model show that the black-white differential in mortality risks would account for roughly two-thirds of the education gap between black and white males aged 26-36 in 1990. Specially, their model indicates that the predicted optimal schooling years for white and black males were 13.12 and 12.60 years, respectively. Compared with the observed 12.74 years of schooling for black men and 13.50 years for white men, they therefore conclude that the predicted gap in schooling represented 68.4% of the observed gap. Although noteworthy, this paper does not provide further empirical evidence using actual data.

Using objective mortality as mortality expectations requires the strong assumption that observable mortality rates are equivalent to the expected probabilities of death for
individuals. As a result, recently a growing number of studies have emerged that directly examine the impacts of self-reported expectations on future behaviors, instead of applying the traditional rational expectations.\(^2\)

Another literature strand of interest for this paper pertains to studies on death expectations, especially with respect to their predictive power. In general, findings in existing studies using NLSY97 expectations are puzzling. On the one hand, Braykov (2010) concludes that subjective mortality probabilities have “little relevance in predicting outcomes” (in terms of actual death—I add) based on significant forecast bias during estimation. Using the framework of rational expectations, his analysis focuses on devising mortality expectations by comparing death expectations with the actual death status of the respondents. Given the relatively low number of deaths among young respondents—coupled with the prevalence of high death expectations values—it comes as no surprise that relying on self-reported death expectations among the young is likely to be risky in terms of believable results.\(^3\) On the other hand, a sizable body of literature has found that perceived death or survival expectations contain useful private information. For instance, Sloan et al. (2010) find that youth tend to increase their mortality expectations after they become regular smokers. Similarly, Fischhoff et al. (2000, 2010) claim that the mortality beliefs that youth stated are, in general, “sensibly” related to other aspects of their life experiences (e.g., crime activities or a history of being bullied).

Furthermore, a growing body of literature finds that mortality expectations have strong predictive power for later outcomes, such as adult socioeconomic status (Nguyen et al., 2012), adult health behaviors (Duke et al., 2011; McDade et al., 2011) and risky

\(^2\) See for example, Manski (2004) and Klaauw (2012), for detailed reviews for earlier studies using expectations data.

\(^3\) Up to 2008, in the NLSY97 sample there were 90 prior deceased and 13 deceased, for a total of 103 deceased respondents (1.1% of the total observations in the survey).
behaviors (Harris et al., 2002; Valadez-Meltzer et al., 2005, Borowsky et al., 2009). Some of these studies also analyze the prediction of mortality expectations in relation to educational attainment. Duke, et al. (2011) use data from the National Longitudinal Study of Adolescent Health (hereafter, “Add Health”) and find that youth with higher early death perceptions tend to be less likely to achieve a high school diploma/GED by young adulthood. Recently, Nguyen et al. (2012) confirm the important role of early survival expectations in predicting adult educational attainment using the same data set. Specifically, the researchers examine three groups based on self-reported survival expectations to age 35: (a) “less than 50-50 chance,” (b) “a good chance,” and (c) “almost certain.” The corresponding odds ratios after logistic estimation between acquiring education for less than high school to college graduate were 1.73, 1.07, and 1.00, respectively. These results show that those who perceive lower chances of surviving would end up with lower levels of education.

3. Empirical Specification

This paper bridges aspects of two strands of existing literature—research on mortality risks (subjective and objective) and research having to do with educational attainment—to consider the possible impacts of mortality expectations on the decisions young adults make about acquiring education. The basic hypothesis that is tested in this paper is whether individual mortality expectations predict educational attainment among a certain age group of youth.

---

5 This paper’s analysis is completed before the publication of Nguyen, et al., (2012). The main disadvantage of Add Health data is that, in the surveys, respondents only are provided five categories for the perceived survival chances question. See Manski (2004) for a discussion of the disadvantage of categorical measures of expectations compared with percentages.

This study extends Duke, et al. (2011) and Nguyen et al. (2012) in at least two ways: a) I examine both objective and subjective mortality risks, and b), I employ the mechanism of time horizons, rather than depending on the notion of fatalism or pessimism.
This section justifies why mortality risks can be incorporated into estimations of educational attainment—a variable that is not typically seen in education production functions. The objective here is to derive an estimation specification using the human capital model.  

In the spirit of the human capital framework, which is subsequently applied in Lang (1993) and Acemoglu et al. (2001), I assume that an individual maximizes his lifetime utility at age \( a \) by choosing consumption \( c(t) \) and schooling \( s(t) \)

\[
\max_{c(t),s(t)} \int_a^\infty u(c(t))e^{-(\delta + \tilde{m})(t-a)} \, dt, 
\]

(1)

where \( u(t) \) is the instantaneous utility function and \( \delta \) stands for the discount rate (or time preference rate). The parameter of interest is \( \tilde{m} \), the constant flow mortality probability expected by an individual. We suppress the individual subscripts for now because they are not essential to the justification.

For simplification, I assume there are neither direct costs of education, nor consumption value associated with staying in school. The capital market is assumed to be perfect, with a constant interest rate \( \gamma \). Relaxing these assumptions will not change the conclusion.

The individual’s budget constraint is:

\[
A(a) + \int_a^\infty (w(t)[1-s(t)] - c(t))e^{-\gamma(t-a)} \, dt = 0.
\]

(2)

---

6 I acknowledge that the human capital model is not the only model that has been used in social science to explain the association of expectations and behaviors. For instance, in the social psychological field, the theory of reasoned action (Fishbein and Ajzen, 1975; Ajzen and Fishbein, 1980) is used to explain the association of survival expectations and behaviors, such as educational attainment and pursuing risky behaviors (e.g., Harris, et al., 2002). The theory of reasoned action argues that youth consider the benefits and costs associated with engaging in some behaviors, and have a tendency to adopt the behaviors in which the benefits outweigh the costs, while avoiding those in which the costs outweigh the benefits.
The initial value of assets (after tax and transfer) is indexed by $A(a)$. The wage rate for one unit of human capital is denoted by $w(t)$. The fraction of time spent on schooling denoted by $s(t)$ and $1 - s(t)$ represents the working time. Based on Fisher’s Separation Theorem (Fisher, 1930), the optimal schooling choice should maximize the lifetime budget which is represented as

$$\int_a^\infty w(t)[1 - s(t)]e^{-(\gamma + \tilde{m})(t - a)} dt.$$  

(3)

Thus, we can rewrite the individual maximization objective function as:

$$\max_S \int_S^\infty w(t)e^{-(\gamma + \tilde{m})t} dt.$$  

(4)

Suppose, for example, that over an individual's lifetime he or she obtains schooling until time $S$ and then chooses to work and does not return to school. At the end of time $S$ when the person reaches the highest level of schooling, his or her human capital level is $\eta(S)$. Assume that the schooling production function $\eta(.)$ is increasing, continuously differentiable and concave. We also assume wage growth rate $g_w$ as,

$$\dot{w}(t)/w(t) = g_w,$$  

(5)

with $w(0) > 0$. The net present discounted value of the earnings steam is assumed to be finite: $\gamma + \tilde{m} > g_w$. Plugging (5) into (4), the maximization function can be rewritten as

$$\max_S \frac{\eta(S)w(0)e^{-(\gamma + \tilde{m} - g_w)S}}{\gamma + \tilde{m} - g_w}. $$  

(6)

Accordingly, the optimal schooling $S^*$ is determined by:

$$\frac{\eta'(S^*)}{\eta(S^*)} = \gamma + \tilde{m} - g_w.$$  

(7)

Because we assume $\eta(.)$ is concave, equation (7) demonstrates that a higher expected mortality probability reduces one’s investment in schooling. Thus, the optimal total number
of schooling years for an individual with a higher probability of mortality will be smaller in comparison to that for a person with lower expected mortality probability, all else being equal.

Individual observable characteristics (e.g., race, gender, age, parental education, household income, etc.) may also influence expectations for educational attainment. Accordingly, one can approximate \( \frac{\eta'(S^*)}{\eta(S^*)} \) as \( XB - cS^* + u \), where \( X \) denotes characteristics, \( B \) for the impact of \( X \) on the left side of equation (7) and \( u \) for error term. We can rewrite (7) as

\[
S^* = \frac{1}{c} \left( XB - \bar{m} + (u - \gamma - g_w) \right).
\]

To help sort out paths of prediction, the analytic design I use is careful to take measures of mortality expectations into account prior to considering educational measures. Thus, the baseline specification estimate for predicting the expected probability of dying and adult educational years is as follows:

\[
S_{i2008}^{2008} = X^{1997}B^* + \mu m_i^{1997} + u_i^{1997}.
\]

(8)

where \( S_i \) represents the highest grade completed by individual \( i \) when he/she is 25–28 years old in the year 2008; \( m_i \) for expected mortality probability reported by respondents in 1997, which is the only year for which death probabilities were available before most respondents reached the typical age (17) for making college decisions in the U.S. It is noticeable that, in general, \( m_i \) is not necessarily equal to but is proportional to \( \hat{m}_i \) in the above theoretical discussion. In 1997, respondents in the sample were aged 16-18 years old. The main coefficient of interest, \( \mu \), indicates the impact of expected mortality risks on educational attainment. The difference between \( B \) and \( B^* \) picks up the shifting effects in \( \gamma \).
and $g_w$ between individuals. $u_i$ stands for error term. $X^{1997}$ stands for personal characteristics at age 17—again, the average age for an American youth to decide on college.\footnote{See, for example, Carneiro et al., 2010, Cooper and Luengo-Prado, 2011, Hryshko et al., 2011 for studies using age 17 as the typical age when one makes the college entry decisions in the U.S.}

To further explore the possible role of differences in mortality expectations in explaining education gaps between blacks and whites, we conduct the pooled Oaxaca-Blinder decomposition

$$
S_{white}^{2008} - S_{black}^{2008} = \mu (\bar{m}_{white}^{1997} - \bar{m}_{black}^{1997}) + [\bar{m}_{white}^{1997} (\mu_{white}^{1997} - \mu) + \bar{m}_{black}^{1997} (\mu - \mu_{black}^{1997})]
+ \Delta X^{1997} \beta + \bar{X}^{1997} \Delta \beta + v,
$$

where $\Delta X \beta$ and $\bar{X} \Delta \beta$ indicate the endowment effects and coefficient effects for other control variables, respectively. $\mu$ is the pooled coefficient of mortality on the difference in education for both blacks and whites together. $\bar{m}_{white}$ and $\bar{m}_{black}$ are the coefficients of mortality in the regressions for whites and blacks, respectively. The error term is $v$.

Before proceeding, it is important to note two important concerns about the estimation approach detailed above. First, it is worthwhile to remember that introducing subjective data into choice models has raised red flags among economists, and thus must be considered in making conclusions. Because this paper uses self-reported death expectations as mortality expectation measures, related concerns cannot be overlooked. As summarized by Bertrand and Mullainathan (2001), the key skepticism associated with subjective data is whether people give meaningful answers to subjective questions. Conversely, Manski (2004) claims that a “combination of choice data with other data should mitigate the credibility problem and improve our ability to predict behavior” (page 1330). Similarly, a growing body of research has confirmed the validity of subjective data and therefore including it in choice models (e.g., Bound et al., 2001; Krueger and Schkade, 2008; Delavande et al., 2011;
Zafar, 2011). In this study, supporting the use of subjective data is the fact that the regressions in Appendix Table 1 suggest that the elicited death expectations are by no means random. First, various factors typically associated with death expectations demonstrate some consistency with what we would expect. For example, those who have ever been alcohol abusers and have had multiple sexual partners are more likely to report higher death expectations; higher cognitive ability is associated with lower death probabilities for both 1997 and 2002. Furthermore, from 1997 to 2002 the death expectations are generally updated in an expected way. For instance, those who joined the army or other active service branch during the five-year period raise their death expectations. In summary, I would argue that the subjective questions I use do seem to reliably reflect mortality expectations and therefore are reasonable to be considered to have an effect on educational attainment.

Another concern of note is that the expectation measures might have picked up some unobservable effects, such as optimism/pessimism, inherent personality traits, and risk preferences that might have impacted education decisions independently. This paper addresses this problem as omitted variables bias. In subsection 5.2, I will examine this issue and demonstrate that, in general, the omitted variable bias would not threaten the basic conclusion of this study.

4. Data

The National Longitudinal Survey Year 1997 (NLSY97) was first administered to a representative sample of close to 9,000 youths who were 12 to 16 years old as of December 31, 1996. This paper used data about mortality expectations from the NLSY97 report (Round 1, 1997), as well as from the 2002 survey—the only other year during which self-reported mortality expectations are solicited. This study also incorporates data concerning
educational attainment obtained from the 2008 Survey (the final year of the NLSY when this study was completed), which is solicited from 7,490 adults aged 23 to 29. It should be noted that the primary data sample used in this paper (concerning mortality expectations) contains only responses from those who were asked the “death chances questions”—namely those 16 years of age or older. Thus 5,419 observations are dropped because they are younger than 16. Further, 129 observations are dropped because their answers are recorded as “invalid-skipped,” “don’t know,” or “refused.”

To focus on whites and blacks, this paper also exclude Hispanics (441 observations) and those who belong to other races (31 observations). To ensure that mortality expectations precede educational decisions, we exclude the survey responses of 103 whites and 77 blacks who do not attend school after 1997 because there is no way to determine whether their death expectations were reported before or after their decision to terminate. Also it is important to point out is that the sample size varies slightly across different specifications because some individuals did not answer other questions required for the construction of some of the regressors (e.g., parental education).

The actual mortality data used in this paper is obtained from the Compressed Mortality File (CMF), produced by the National Center for Health Statistics, a county-level national mortality and population database spanning the years 1968 to 2007. The numbers of deaths, crude death rates, and age-adjusted death rates are available by place of residence, age group, race, gender, year of death, and underlying cause of death. The mortality rate represents the total number of deaths divided by the total population for each particular group, which is then multiplied by 100. Although causes of death are reported in particular categories, we aggregated them into four broad causes of death: accidents, assaults, suicides and diseases. The main measurement used in this paper is
death from all causes, which thus includes all four categories.\textsuperscript{8} By matching age, race, sex and county with the NLSY97 sample, we construct expected mortality based on the grouped-aggregate of actual death rates for each reference group of individuals.

4.1 Self-Reported Probabilities of Death

Two death expectation questions were asked of a representative subsample of 3,436 respondents who are 16-18 years of age in 1997 (47.3% were born in 1980 and 52.7% were born in 1981).\textsuperscript{9} The options provided to the respondents are integers between 0% and 100%. In 2002, participants who took part in the 1997 survey were again asked about their chances of dying within the subsequent year and over the next five years. However, by 2002 most respondents had reached the age of 20-23, so there was a concern that they may have already concluded their education. We focus, therefore, on the two measures pertaining to death expectations that had been acquired in 1997.

Figures 3A and 3B illustrate the percentage distributions for the probabilities of dying. Three conclusions emerge from these distributions. First, the expectations exhibit substantial heterogeneity, using the full range of values for chances of death from 0% to 100%.\textsuperscript{10} As shown, in 1997, 1,037 people respond with “no chance” (0%), while the responses from 2,278 participants ranged between 1% and 50%, and 121 observations ranged from 51% to 100%. Second, there are two modes: 0% and 50%. Overall, approximately 30 percent of respondents provide 0% as answers to their death expectations; 20 percent respond with 50%. Third, the gap shown in the survey between the perceived probability of dying within the subsequent year and dying by age 20 is extremely small. For instance, let us suppose that the average respondent in the sample believes his probability of dying within

\textsuperscript{8} Other causes of death are also checked and the results of this paper pertain to those as well.

\textsuperscript{9} Specifically, in 1997 the survey asked “What is the percent chance that you will die from any cause — crime, illness, accident, and so on, in the next year?” , “…by the age 20?” . In 2002, “…in the next year?” and “…, in the next 5 years?”

\textsuperscript{10} Manski (2004) discusses the common assertion that instead of using the entire scale, respondents tend to focus on some particular points like 0, 50 and 100.
subsequent year 0.18. In order for the average probability of death over the ensuing five years to be close to 0.22, for years 2 to 5, that individual’s mortality probability would have to be as 0.01 for each of those years.\textsuperscript{11} That figure is incredibly low in comparison to the perceived probability of death. Instead, survey results indicated that the death probabilities of 86.53 percent of respondents by age 20 are higher than or equal to the expected probability of dying within the subsequent year. However, 62.63 percent of respondents gave the same answers to those two questions. This outcome is shown in Figures 3A and 3B, where the two frequency curves for dying in the subsequent year and by age 20 in both panels demonstrate similar distributions.\textsuperscript{12} This indicates that the individual perceptions for death probabilities are quite consistent across different target time points (by the next year, by age 20, and within the ensuing five years).

A comparison between these death expectations in 1997 and 2002 suggests that youth tend to revise their beliefs over the five year period. Although 25.92 percent of respondents did not change their expectations of dying between 1997 and 2002, 38.95 percent report declining expected probabilities, and 35.13 percent raise their expectations. The correlation coefficient between these two years is roughly 0.29 (significant at 0.01 level).

The subjective death probability that youth use to make lifetime schooling decisions is the probability of dying within the subsequent year. However, it is still worthwhile to consider the alternative possibility that youth believe they will die at an age closer to average death rates for their particular age group. For this purpose, I employ two approaches to examine the impact of actual death rates over longer time spans. First, I expand the age range for the reference groups from the age band of 15-34 (as the indicator

\textsuperscript{11} The survival rate in the next 5 years is calculated using $1 - 0.22 = (1 - 0.18)(1 - x)^4 \rightarrow x = 0.0063$, where $x$ is the death rate for each year in the following 4 years.

\textsuperscript{12} Similarly, studies have found that in the Health and Retirement Study many respondents fail to account for increases in yearly mortality rates with age (Elder, 2007).
for age 17 in the above analysis) to the age band of 15-67. In other words, the mortality rate that an individual uses to make schooling decisions is constructed by the number of deaths ages 15-67, divided by population size in each race-gender-age-county group, and multiplied by 100.

A second methodology used herein was to construct the expected life expectancy at age 17. The idea is that when teenagers make their schooling decisions, they consider how many additional years they may live before age 67, conditional on being alive at age 17. Since I sought to assess individual education decisions based on the expected mortality probability of youth in the ensuing years, I am principally interested in mortality rates during the “relevant ages.” Hence, I focus on life expectancy between age 17 and 64, conditional on surviving up to the age of 17. Following the method described in Jayachandran and Lleras-Muney (2009), I calculate the expected years of life between age 17 and 67, conditional on surviving until age 17 using the following formula:

$$e(17-67) = \left( \sum_{t=17,18,...}^{67} (t + \frac{1}{2}) \times \prod_{t=17}^{t-1} (1-m_t) \times m_t \right) + 67 \times \prod_{t=17}^{66} (1-m_t) \times (1-m_{67}) - 17,$$

(10)

where $m_t$ is the probability of dying at age $t$ (the mortality rate for age $t$). $\prod_{t=17}^{t-1} (1-m_t)$ is the probability of surviving from age 17 to age $t-1$. Death rate data are calculated for five-year (ages 15-19, 20-24) and ten-year (ages 25-34, ..., 55-64, 65-74) age bands. The calculation uses one year as the age increment between two consecutive years, and we assume the death rate is constant for each age in each age band. The summation accumulates the expected years of life for individuals who die at each age between 17 and 67. We then add

---

13 The upper band for age is determined to be 67 because currently typical retirement age is 67 in the U.S. The age bands in the mortality data are as follows: under 1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84 and 85+.  
14 Another way to present the expected life expectancy is using survival curves. In a graph, the area below the survival curve is equal to the life expectancy. For simplification, in this paper we only report the life expectancy as an alternative measure for actual death rates.
to each successive year, assuming death occurs in the middle of a year. The term following the summation is the expected years of life for those who have survived past age 67. Finally, 17 is subtracted because the construction of this equation is conditional on surviving until age 17. For this sample, therefore, the anticipated life expectancy for whites aged 17-67 is roughly 44.43 years of life, while for blacks it is 41.88 years.

4.2 Actual Mortality Rates Measures

Actual mortality rates are constructed using the same race-sex-age-county aggregate levels. Ages ranged from 15 to 34. From 1997 to 2000, the average mortality rate (per 100 population) for blacks males declined from 0.33 to 0.23, while for black females the rate decline from 0.12 to 0.08. In contrast, death rates for white men remained at 0.15, and for white women the rate also stayed constant at 0.05. Table 1 documents the comparison of actual death rates with the subjective death probabilities in the matched NLSY sample.

Despite typically held invulnerability beliefs among adolescents, Reyna and Farley (2006) found that youth tend to overestimate important risks. A number of studies on youth expectations also have confirmed over–pessimism in expectations of youth (Lochner, 2007; Fischhoff et al., 2010 Sloan, 2011). Similar to previous literature, this study finds that actual mortality rates for particular race-sex-age-county groups are noticeably lower than subjective mortality probabilities. The overestimation of subjective mortality probabilities is demonstrated in Table 1.

Two likely reasons explain why expectations may differ from actual rates. First, respondents are apt to incorporate private information that accounts for such differences. For example, a person with a personal health issue or a family history of serious disease is likely to report higher death expectations. Second, it is possible that there is some common inflation factor between actual mortality rates and elicited individual death beliefs among
certain ethic and/or gender groups. Table 1 (columns (5)-(6)) displays ratio patterns for subjective death probabilities compared to objective death probabilities according to racial/gender groups. By way of explanation, a high ratio of subjective death probability to objective death probability corresponds to increased pessimism toward longevity. In the groups we examined, this ratio was, on average, lowest for black men, followed by white men, black women, and white women. Thus, white women expected to live the longest.

4.3 Educational Attainment Measures

In this study, the primary measure for educational attainment is the highest grade completed for an individual aged 27-29 as of 2008. For those missing this 2008 variable, I use the highest grade completed as of 2007, and so forth. I also examine their chances for college attendance and high school graduation using probit estimations. To address the concern that state compulsory schooling laws and child labor laws may prevent students who want to drop out from doing so, additional schooling after the required age in school is also explored, with the results being quite similar.\textsuperscript{15} For simplification, I use the highest grade completed as the primary dependent variable.

4.4 Other Control Variables

In the following estimations we also control for other factors typically associated with educational attainment. First, we use the age-adjusted standardized the Armed Forces Qualification Test (AFQT) score as a proxy for cognitive ability. Second, we control for total real family incomes (in 1,000 dollars) and family size when respondents were 17 years of

\textsuperscript{15} This calculation follows Gan and Wang (2008). States vary with respect to compulsory years of schooling. Compulsory schooling years are calculated by using compulsory attendance laws and child labor laws. We choose the smaller value between two differentials: first is the differential of earliest age of leaving school and latest age for entering school; second differential is between age 16 and the latest age for entering school. The required ages for school entrance are 5 to 8 and school-leaving ages range from 16 to 18. The resultant years of compulsory schooling range from 9 to 11 years in our sample. The conclusion in this paper still holds true using additional school years as the dependent variable.
age. Additionally, parental educational background (measured by highest grade completed by each parent) is also included in the independent variables used in these estimations.

Finally, a set of variables indicating geographic controls at age 17 are included (resident region, central city, Metropolitan Statistical Area, and state and county of residence). If a respondent’s information at age 17 is missing, we use his/her information at age 16 or 18 instead. The county-level characteristics are taken from State and County Data Book 2000 (Haines, 2004). We use household median income, unemployment rates, and crime rates to capture county-level residential characteristics, since these are typical factors thought to affect the education choices of youth.

### 4.5 Statistical Description

Descriptive statistics for the primary variables according to racial group are reported in Table 2. The average age of the sample is approximately 16.32 in 1997 and 27.37 in 2008. Whites and blacks were of similar age (27 years old) and gender composition (approximately 50/50). In short, the demographic variables for blacks and whites were similar.

The main outcome variable is educational attainment, which is measured using the highest grade completed, ranging from 0 to 20. In 2008, the race-based education gaps were significant. On average, a white respondent acquires 1.09 additional years of schooling compared to his/her black counterpart, while the college attendance rates are 61.6% and 48.4% for whites and blacks, respectively.

The average self-reported death probabilities up to the next year are very close between racial groups: 22% (standard deviation=21.56) for whites and 24% (s.d.=23.68) for blacks. It should be noted that all the standard deviations were relatively large—indicating that the differences in subjective death expectations are significant within each racial group. With respect to actual “all-cause” death rates, the average for whites is 0.09 percent.
(s.d.=0.06) and 0.17 percent (s.d.=0.11) for blacks (group mean difference significant at 0.01 level).

As extensively documented, there are substantive gaps in widely-believed predictors of education across racial groups. Specifically, the average standardized age-adjusted AFQT scores are 0.40 (s.d. =0.94) for whites and -0.53 (s.d. =0.84) for blacks. Parental education and family income also contribute to significant racial gaps. The average number of educational years for white mothers is 13.39, compared to 12.56 for black mothers. Additionally, the average total household income for the families of respondents aged 17 is $53,280 for whites and $31,150 for blacks.

5. Results

This section details the estimation results from this study’s findings, beginning with benchmark specifications. It concludes with a discussion of robustness checks.

5.1 Benchmark Estimation Results

This following analysis pertains to the estimation results in connection with education equation (8), which was developed to predict the relationship between expected mortality and educational attainment. First, the ordinary least squares (OLS) regressions will be discussed, followed by an evaluation of the decomposition results based on the regressions. The goal of this analysis is to explore the role of black-white difference in mortality in education decisions for youth.

Before introducing mortality expectations, it is important to consider several aspects of the black-white education gap shown in Table 3, which pools the race-based results. In column (1) the coefficient for the categorical variable of black is -1.168, which represents the education gap we sought to explain in this study. When the mortality expectation measure is added to (1), the results displayed in columns (2)-(4) shows that the coefficients
for mortality expectations are negative. In looking at columns (5)-(7)—where AFQT and other controls (e.g., family background, demographic characteristics, state of residence and county characteristics at age 17) are added and mortality expectations are dropped—one can see that the coefficients for the black variable become positive, which principally results from controlling for AFQT scores. In other words, when AFQT scores are the same for blacks and whites, blacks acquire higher education than whites. This finding corresponds to results obtained by Lang and Manove (2010). In the last three columns (8)-(10), mortality expectations are added to the regressions shown in columns (5)-(7) in order to compare racial differences via the introduction of mortality expectations. The coefficient for the black variable became larger when we control for mortality expectations. This outcome implies that, all else being equal, when mortality expectations among blacks and whites are the same, blacks achieve educational attainment at higher levels than whites.

The basic regression results are presented in Table 4A. First, results for equation (8) are estimated for whites and blacks separately. Columns (1)-(6) display the results for subjective death rates, and columns (7)-(9) show actual death probabilities for whites and blacks, respectively. Since the objective mortality rates were measured at the county level, clustered standard errors associated with the respondents’ residence counties at age 17 are employed to allow for possible arbitrary correlations among individuals living in the same counties at age 17.

Columns (1), (4) and (7) of Table 4A show simple regressions of highest schooling years in 2008 (for those aged 27-29) on the categorical variable for blacks, as well as other factors associated with educational attainment. The coefficients for the black variable—0.842, 0.872 and 0.845—represent schooling gaps between blacks and whites, which we sought to explain.
Columns (2)-(3) show that one percentage point increase in the self-reported probability of dying up to next year results in decreased educational years by approximately 0.008 for whites and 0.007 years for blacks. It should be stressed that the impact on education is statistically significant at the 0.05 level. The coefficients for subjective mortality probability up to age 20 are reported in columns (5)-(6), which are -0.013 (s.e.=0.003) and -0.007 (s.e.=0.003) for whites and blacks, respectively.

In columns (7)-(9), the results for actual death probability are reported. As shown, the coefficients associated with actual death rates for the two reference groups are −0.14 (s.e.=1.07) for whites, and −0.65 (s.e.=1.013) for blacks. That is, the addition of one additional death per 100 population in the reference groups is correlated with −0.14 to −0.65 additional years of education for an average respondent (both coefficients are insignificant). Also important to note is that the coefficients are negative, which is consistent with what the human capital theory suggests. Comparing the regression results across both groups, we can see that the impact of mortality on educational attainment between whites and blacks is insignificantly different from zero. Additionally, the point estimates are very small, implying very little impact of large variations in mortality rates on outcomes.

Among other factors, the standardized AFQT variable is found to be highly significant in estimating years of schooling. For example, in the regressions using death probability during the ensuing year, an increase of 1.274 standard divisions of educational years resulted per standard deviation increase in the AFQT score for whites, in comparison to 1.491 for blacks. Other factors such as parental education and living arrangement at age 17 are also significant in most specifications. The important role of these factors is consistent with previous literature on educational attainment.
Table 4B summarizes decomposition results of estimates from Table 4A. In general, composition effects (i.e., the portion of differences explained by endowments) play an important role with respect to total education gaps. Nevertheless, self-reported mortality risks could, at most, account for only 0.05 years of educational attainment differences, all else being equal. Moreover, any unexplained educational gaps would only increase 0.019 years if the coefficients for subjective mortality expectations between whites and blacks would remain the same. As for actual death rates, if whites and blacks were to report identical actual mortality expectations based on actual death rates, the education gap would increase 0.021 years. This means that approximately 0.003 years of unexplained educational gaps can be linked to the coefficient associated with actual mortality rates. Moreover, the AFQT scores gap accounts for a significant portion of the black-white schooling gap of 1.21, which is tied to the “perceived death probability by the next year” measure. In summary, the AFQT variable is found to be the most important factor for explaining educational gaps, followed by parental education and family background—in contrast to geographic characteristics, which were found to be relatively insignificant.

Overall, the decomposition results shown in Table 4B confirm that gaps in reported subjective mortality or actual mortality risk data do little to explain race-based educational attainment disparities. Instead, the well-established predictors for educational attainment, such as parental education, cognitive ability, and family background continue to play important roles, irrespective of race.

5.2 Robustness Checks
This section confirms that the main findings from this analysis are satisfactorily robust when applied to a variety of subsamples, variable definitions, additional control variables, and alternative data sets. Therefore, I would argue that the results detailed
herein are not simply driven by the unique characteristics of this sample data, but rather can be more universally applied.

**Alternative samples.** Table 4A shows regression results based on data obtained from individuals who answered the survey’s subjective questions. In the ensuing discussion, however, the robustness of these results is explored using various samples, as shown in Tables 5A and 5B. First, by including results from those who were not asked about subjective death expectations, it was possible to construct a larger sample to examine the robustness of the estimated impacts of actual death probability on education. Second, in the benchmark estimation, the observations of respondents for whom AFQT scores were missing have been dropped. Third, due to the concern of any potential bias associated with systematic differences between those with valid AFQT data and those without, we include those respondents with missing AFQT scores into our analysis by introducing a dummy variable representing the missing of AFQT scores. Tables 5A and 5B show that these alternative samples result in qualitatively similar regression and decomposition results as the baseline samples.

**Alternative measures.** The probit estimation and the decomposition result of using the percentage of respondents with a high school diploma or higher are reported in Tables 6A and 6B, respectively. Using the same control variables used in Table 4A, the coefficient for subjective death probability up to age 20 was determined to be -0.003 (s.e.=0.002)—with significance at the 0.05 level. Moreover, according to our decomposition results, mortality expectation measures still only account for an insignificant segment of the gap associated with the rates of high school graduate or more.

---

16 The method proposed in Yun (2004) is used to conduct the non-linear decomposition estimates for binary dependent variables.
Furthermore, there is a concern that because the predictive ability of mortality expectations for education was non-linear, it would increase as mortality expectations increase. The nonlinearity is confirmed by the quantile regression results in our sample. We also investigate what could occur if the expected death probability measures are recoded using categorical variables to indicate the particular groups with higher mortality expectations: the self-reported death probabilities were higher than 50%, while the actual death rates for the reference groups were higher than 0.05. Panel A in Table 7A shows the regression results and Panel B shows the decomposition results. One can observe that in comparison to the control group (those who reported death expectations lower than 50% or lower than 0.05), despite the fact that some of the regression coefficients are significant, the mortality expectation variable still only accounts for a small portion of the race-based educational attainment gap.

Table 7B shows the estimation results for educational attainment using actual death rates based on reference groups with wider age ranges in comparison to those represented in the basic results. Thus, for reference groups ages 15-67 or 17-67, the coefficients associated with mortality rate remain insignificant, although the magnitude of the coefficients decline from those shown in Table 4A.

**Additional controls.** One important concern related to this study is that an individual’s self-reported perception of death probability might be correlated with more generally unobservable preferences or traits—which would result in biasing the estimated coefficients. In particular, one’s risk preferences (indicated by various risky behaviors) may lead a “risk lover” to have a higher probability of dying and a lower probability of investing in education.\(^\text{17}\) As an example, joining a branch of the armed services could certainly

---

\(^{17}\) See, for example, Cowany (2011), for a summary of economic literature exploring the link between risky behaviors of adolescents and educational attainment.
increase one’s death expectation, thereby hindering educational attainment before age 29 (the oldest age in the primary sample in 2008).

To deal with the problem of omitted variables bias, we create the following expanded model

\[ S_{i}^{2008} = X^{1997} B + \mu m_i^{1997} + a_i \rho + u_i^{1997}, \]

(11)

where \( a_i \) denotes a vector of variables thought to influence mortality risk and educational attainment simultaneously. This vector includes military history, substance abuse (marijuana use, alcohol abuse, and cocaine use), smoking, delinquency behavior (e.g., gang membership or a criminal record), having multiple sexual partners, or experiencing “hard times” early in life.\(^{18}\) All of these variables are also measured in 1997. The resulting estimates are documented in Table 7. It should be noted that we do not control for these additional variables from the outset due to the fact that they are not typically included in education estimation equations. Nonetheless, to account for potentially influential factors, we explored the possibility of omitted variables bias by examining an expanded model and the changes in the death probability coefficients before and after including these additional controls.

Another unusual variable that is added to the education estimation is the education level of respondents when the survey was first administered in 1997. This feature is included due to the concern that the prediction of subjective death on educational attainment might actually have resulted in education heterogeneity when they reported their expectations.

\(^{18}\) “Hard times” in the survey refers to experiences that some children go through such as living in a place without water or electricity, or in a homeless shelter.
One variable worthy of additional investigation is the measure for personality traits.\textsuperscript{19} When personality measures are factored into the regressions, the magnitude of coefficients associated with death expectations slightly decline. This outcome implies that personality can be considered to be associated with schooling decisions, as well as being correlated with death probabilities. However, it must be noted that personality measures are not available until 2002. Moreover, the variations in personality scores across the years were substantial, making it somewhat risky to use personality as an unassailable component of the relationship between mortality expectations and educational attainment. Hence, this paper avoided using personality measures in the analysis.

Measures of death expectations may reflect other unobservable information, such as whether respondents may anticipate any difficulties in pursuing additional schooling. Therefore, this study also controlled for respondents’ expectations for various adverse events during the upcoming year, including expectations of being arrested, being in jail, or becoming a victim of bullying. These expectations tend to be highly associated with actual experiences. That is, those who have experienced these adverse events in the past are likely to report a high probability for them in the future. Controlling for such “unobservable information” is expected to partially reduce the omitted variables bias.\textsuperscript{20}

In summary, after introducing the above additional controls, compared with the basic estimation results shown in Table 4A, the coefficients for death expectations slightly

\textsuperscript{19} The NLSY97 collected information on non-cognitive personality traits in five dimensions: 1) has trouble paying attention; 2) lies or cheats; 3) doesn’t get along well with others; 4) often unhappy; and 5) generally optimistic about the future. Thus, five categories for these five dimensions of personality traits were created based on related questions.

\textsuperscript{20} Alternatively, I also construct an index using factor analysis on the same battery of subjective probability variables to indicate the extent of optimism or pessimism. Following previous work (Lillard and Willis, 2001), all the “components” contributed positively to the principal component. This index is defined as the reciprocal of the principal component. It is important to note that this index is able to reflect other unobservable characteristics besides optimism. Nonetheless, it is not essential to this study whether the principal component was optimism or something else. When controlling for the optimism index in the regressions, the coefficient of death expectation was -0.006 and therefore statistically significant.
declined and the standard errors change little. The extent to which the coefficients decline suggests that the benchmark estimation of death expectations and educational attainment may reflect the influence of omitted variables bias. Nevertheless, the small extent of decline implies that one's death expectation remains an important component, irrespective of the influence of the additional controlling variables. Importantly, even when estimations are subject to omitted variables bias—and therefore the possibility that coefficients are overestimated—when one eliminates the omitted variables bias completely, the coefficients would reduce to even smaller values. Thus, the basic conclusion of this paper—that mortality expectations play a limited role in explaining the black-white educational attainment gap—is still valid even after controlling for some of the additional variables that have been examined.

**Add Health data set.** As noted in Section 2, previous researchers have used Add Health data to examine the prediction of survival expectations on future outcomes. I also examine categorical measures for self-reported survival chances from Wave 1 (1994-1995) to predict educational attainment (measured as the high school completion rate and the college completion rate) in Wave 4 (2008) available in Add Health data using logit estimation, also controlling for the same variables used by Nguyen et al. (2012). These results are documented in the left side of Tables 9A and 9B. Two categorical variables indicating low level of self-reported survival chances (chances less than 50%) and moderate level (“a good chance”) are of interest (the omitted category is high level “almost certain”). It is worthwhile to note that the categorical variable for low self-reported survival chances is

---

21 The Add Health data used in this paper is publically accessible data, which features data from 6,504 respondents who were randomly selected from 20,745 respondents -the restricted-use in-home file’s core sample and the high-education black sample (with at least one parent with a college degree). I drop the high-education black sample (520 respondents) from the analysis because of the high correlation of parental education and adolescent education. Therefore, two variables indicating residence county characteristics (poverty rate and adult violent crime arrests rates) in Nguyen, et al. (2012) are not available and therefore are not included in the estimation. Other control variables are very similar to Nguyen’s study.
shown to strongly predict educational attainment for whites and for the pooled sample (whites and blacks together). The portion of black-white educational disparities explained by this categorical variable is quite substantial.

Nevertheless, I notice that the covariates in Nguyen, et al. (2012) exclude an important variable: cognitive ability of adolescents, which we have seen plays the key role in predicting education in the NLSY97 data. Motivated by this concern, I incorporate the age-adjusted standardized Picture Vocabulary Test (AHPVT) score in the estimation (results presented in the right side of Tables 9A and 9B), which has been used in literature (e.g., Cheng and Udry, 2005) as a cognitive ability measure. Again, we see that cognitive ability is shown to be a strong predictor of education in adulthood. However, regarding the subjective survival expectations, introducing of AHPVT scores in the controls makes the association between having low survival expectations and educational attainment no longer very strong, especially when education is measured as the college graduation rates among youth.

Overall, with Add Health data we find that, if controlling for cognitive ability (which I think we should), the predictive power of self-reported chances of surviving to age 35 clearly declines, suggesting that Nguyen, et al.’s (2012) analysis might omit this important predictor for education.\(^{22}\) The role of self-reported chances of surviving in explaining black-white education gaps in terms of college graduation rate is not significant, although that is significant when education gap is measured using high school completion rate. In general, although the results we show here is by no means conclusive, we do not see an

\(^{22}\) Another thing I notice is that in Table 9A for blacks only, the self-reported survival expectations are insignificant in the specifications, regardless of controlling for cognitive ability or not. That raises the suspicion about the fatalism mechanism employed in related sociology studies (e.g. Nguyen, et al., 2012): Previous research has documented the presence of fatalism among black youth, as well as the black-white differences in terms of fatalism beliefs (e.g., Petterson, 1999). If that mechanism were true, we would expect to see the coefficients of low survival expectations for blacks would be larger than those for other ethnic groups, which is not shown in this data.
unambiguously important role of survival expectations plays in accounting for black-white education gap using Add Health data.

6. Discussion and Conclusion

This paper investigates the association between adolescents’ expectations of probable mortality and their educational attainment using yearly data from the National Longitudinal Survey (mortality expectations from 1997 and 2002; and educational attainment from 2008). In particular, I examine the predictive power of both discrepancies in actual black-white morality rates, as well as black-white mortality expectations, on differences in educational completion. In summary, this study finds that (1) actual mortality rates based on the reference group do not significantly affect education choices, and (2) even though the subjective mortality perceptions of youth have statistically significant effects on educational attainment in most specifications we have examined, the magnitude of the effects is quite small. Moreover, after controlling for a variety of additional variables, the impact of the mortality expectations of youth slightly declines. This finding indicates that after taking into account the possible omitted variables bias, the estimated coefficients of death expectations would decline further. Therefore, the results of this study strengthen the argument that the role of mortality differentials in explaining overall black-white disparities in educational attainment is not as substantial as Gan and Gong’s (2008) theoretical model suggests. Instead, the role of mortality risks discussed herein is really small, although the results are consistent with the Ben-Porath model (1967).

There are three likely reasons for the relatively low impact of expected mortality on educational attainment. The first possibility is that when individuals make schooling decisions, they are not likely to be thinking about morality risks. Young adults considering staying in high school and/or pursuing a college degree are paying attention to more
immediate concerns, such as the utility or disutility of school attendance, the financial implications of additional schooling, or how it will impact their earnings potential, among other factors.

Second, the role of mortality reduction depends on the initial mortality rates. For example, in Sri Lanka the maternal mortality ratio was dramatically reduced from 1.80% in 1946 to 0.53% in 1953 (Jayachandran and Lleras-Muney, 2009). Thus, while there have been some significant reductions in mortality in some developing countries, the substantive decline tends to be linked to a high starting level. In comparison, although the black-white differentials in mortality risks in the U.S. are obvious, the longevity of blacks does not differ dramatically from that of whites. In short, black-white morality differences are too small to contribute in any significant way to race-based education gaps.

Third, if we consider that returns on educational attainment have a strong monetary link in terms of lifespan earnings, then one crucial determinant of the length of payoff timespan is retirement age. Hazan (2009) argues that gains in life expectancy for American men born between 1840 and 1970 cannot be linked to significant increases in educational attainment during the same period because the rise in life expectancy did not lead to a rise in labor supply. Noting that during that period life expectancy increased and labor supply declined for men, he claims that noteworthy changes in labor supply must be a precondition for life expectancy to be positively correlated to educational attainment. Therefore, the lack of measures for expectations of retirement in this study might be a possible explanation the absence of a strong correlation between mortality expectations and education.

23 Maternal mortality ratio is the ratio of the number of maternal deaths per 100,000 live births.
24 If we take the non-pecuniary returns to education into consideration, individuals can still reap the benefits of education after retirement, such as better health and an enhanced social network. Thus, the time of retirement as a determinant for the length of the time span becomes less important.
An important question to ask is why do the results in this paper fail to support the Gan-Gong (2008) model? There are a few possible explanations. One potential reason is that human capital theory is not the only framework that can or should be used to explain the schooling decisions of adolescents. Therefore, although this paper documents the relationship between death expectations and education using the human capital model, I acknowledge that there could be alternative frameworks by which to investigate the association. For instance, the behavioral choice model may be useful in accounting for the limited attention adolescents pay to mortality expectations when making schooling decisions.

Gan and Gong (2008) propose that programs to reduce mortality in black neighborhoods would help increase educational attainment. While such policies may be desirable for a good many reasons, the results discussed herein suggest that such policies would have little or no effect on educational attainment.

One important limitation of this study is that the oldest respondents in the sample were only 29 years old. Thus, some of them are likely to return to school later in life to due altered economic circumstances, the natural progression of employment changes (some of which could necessitate additional schooling), or for other personal reasons that cannot be anticipated. Results in this paper suggest that they are highly unlikely to seek more educational opportunities due to a change in their mortality beliefs. However, a further exploration of this argument is at present not feasible due to data limitations. This association might be revisited if and when new data becomes available. In terms of future research, additional studies could look into the long-term impact of death expectations on actual death probabilities and labor market outcomes. Moreover, future research may be useful in expanding this study to examine predictions of death expectations for other human capital investments, such as on-the-job training and health.
References


Figures and Tables

Figures 1A and 1B:

Notes: CPS March data ages 15-34 in 1980-2010. The high school (highest grade completed is 12, with or without diploma or GED) graduation rates and college (with a bachelor’s degree) graduation rates are illustrated. Incarceration population not adjusted. Center for Economic and Policy Research. 2012. March CPS Uniform Extracts, Version 0.9.5.1. Washington, DC.
Figures 2A and 2B:

**Figure 2A: Trend of All-Cause Mortality Rates by gender and race (CMF data)**

![Graph showing trend of all-cause mortality rates by gender and race from 1980 to 2005.](image)

**Notes:** Compressed Mortality Data 1979-2007. Death rates include all death causes of people aged 15-34. The rates are constructed with total numbers of death for the age-race-sex groups in a particular year from 1979 to 2007 divided by the population size in the groups and then multiplied by 100.

**Figure 2B: Trend of Estimated Life Expectancy at Birth by gender and race (U.S. life tables data)**

![Graph showing trend of estimated life expectancy at birth from 1980 to 2005.](image)

Figure 3A and 3B:

**Figure 3A: Distribution of Perceived Death Probability 1997**
*Age 16-18 in 1997 (NLSY97)*

- Percent of Population (%)
- Perceived Death Probability 1997 (%)
- Next year (1997)
- By age 20 (1997)

**Figure 3B: Distribution of Perceived Death Probability 2002**
*Age 21-23 in 2002 (NLSY97)*

- Percent of Population (%)
- Perceived Death Probability 2002 (%)
- Next year (2002)
- In next 5 years (2002)

**Notes:** NLSY97 Data. Figures 3A and 3B illustrate the distributions of self-reported expected probabilities of death in 1997 and 2002 for blacks and whites, respectively. Sample 2002 is subsample of 1997 because only those who answered death expectation questions in 1997 were asked the same questions again in 2002.
Table 1: Subjective Probabilities of Death and Actual Death Rates of Reference Groups (NLSY97 and CMF data)

<table>
<thead>
<tr>
<th></th>
<th>(A) Self-Reported Death Probability(%)</th>
<th>(B) Actual Group Death rate (per 100 population)</th>
<th>(C) Ratio of (2)/(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) N</td>
<td>Mean (s.d.)</td>
<td>(3) N</td>
</tr>
<tr>
<td>White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,807</td>
<td>16.54 (20.41)</td>
<td>1,807</td>
</tr>
<tr>
<td>Female</td>
<td>916</td>
<td>15.10 (19.69)</td>
<td>916</td>
</tr>
<tr>
<td>Black</td>
<td>891</td>
<td>18.02 (21.03)</td>
<td>891</td>
</tr>
<tr>
<td>Male</td>
<td>892</td>
<td>22.92 (25.31)</td>
<td>888</td>
</tr>
<tr>
<td>Female</td>
<td>431</td>
<td>22.70 (25.57)</td>
<td>431</td>
</tr>
<tr>
<td>Total</td>
<td>2,699</td>
<td>17.01 (21.65)</td>
<td>2,695</td>
</tr>
<tr>
<td>Male</td>
<td>1,347</td>
<td>16.11 (20.74)</td>
<td>1,347</td>
</tr>
<tr>
<td>Female</td>
<td>1,352</td>
<td>19.24 (22.32)</td>
<td>1,348</td>
</tr>
</tbody>
</table>

Notes: NLSY97 and CMF data. Means are reported, with standard deviations in parentheses. NLSY97 sample weights are used. The subjective death probabilities are the death chances percentage in next year in 1997. Actual death rates include all causes, which is the actual death rates of reference groups for individuals in the NLSY97 sample. Column (6) is the average of the reported mortality risks divided by the actual death rates for each individual. The numbers of observations in column (5) are slightly smaller than (1) or (3) because some observations have actual death probabilities with zero and are dropped from the ratio calculation.
Table 2: Details and Summary Statistics of the Primary Variables (NLSY97)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>White</th>
<th>Black</th>
<th>Combined</th>
<th>Comparison (1) &amp; (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EDUCATION:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest Grade Completed by 2008</td>
<td>14.04</td>
<td>12.97</td>
<td>13.92</td>
<td>***</td>
</tr>
<tr>
<td>College attendance by 2008</td>
<td>0.616</td>
<td>0.484</td>
<td>0.604</td>
<td>***</td>
</tr>
<tr>
<td>High school graduates by 2008</td>
<td>0.859</td>
<td>0.727</td>
<td>0.832</td>
<td></td>
</tr>
<tr>
<td><strong>DEMOGRAPHIC VARIABLES:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at interview date 2008</td>
<td>27.37</td>
<td>27.38</td>
<td>27.37</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.497</td>
<td>0.524</td>
<td>0.502</td>
<td></td>
</tr>
<tr>
<td>Health limitation</td>
<td>0.135</td>
<td>0.111</td>
<td>0.131</td>
<td></td>
</tr>
<tr>
<td><strong>EDUCATION DETERMINANTS:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age-adjusted standard AFQT scores</td>
<td>0.399</td>
<td>-0.530</td>
<td>0.274</td>
<td>***</td>
</tr>
<tr>
<td>Mother highest grade completed</td>
<td>13.39</td>
<td>12.56</td>
<td>13.28</td>
<td>***</td>
</tr>
<tr>
<td>Father highest grade completed</td>
<td>13.56</td>
<td>12.35</td>
<td>13.43</td>
<td>***</td>
</tr>
<tr>
<td>Live with both mother and father, age 17</td>
<td>0.583</td>
<td>0.261</td>
<td>0.531</td>
<td>***</td>
</tr>
<tr>
<td>Family size, age 17</td>
<td>4.132</td>
<td>4.368</td>
<td>4.149</td>
<td>**</td>
</tr>
<tr>
<td>Gross household real income age, 17</td>
<td>53.28</td>
<td>31.15</td>
<td>52.83</td>
<td>***</td>
</tr>
<tr>
<td><strong>GEOGRAPHIC CONTROLS:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residence in Northeast age, 17</td>
<td>0.186</td>
<td>0.113</td>
<td>0.168</td>
<td>***</td>
</tr>
<tr>
<td>Residence in orth central age, 17</td>
<td>0.322</td>
<td>0.164</td>
<td>0.299</td>
<td>***</td>
</tr>
<tr>
<td>Residence in West age, 17</td>
<td>0.177</td>
<td>0.0635</td>
<td>0.173</td>
<td>***</td>
</tr>
<tr>
<td>Residence in Central city age, 17</td>
<td>0.190</td>
<td>0.415</td>
<td>0.237</td>
<td>***</td>
</tr>
<tr>
<td>Residence in MSA, age 17</td>
<td>0.736</td>
<td>0.771</td>
<td>0.761</td>
<td></td>
</tr>
<tr>
<td>Local unemployment rate %, age 17</td>
<td>4.322</td>
<td>4.320</td>
<td>4.242</td>
<td></td>
</tr>
<tr>
<td>Local crime rate (per million population), age 17</td>
<td>4050.5</td>
<td>5389.9</td>
<td>4222.1</td>
<td>***</td>
</tr>
<tr>
<td>Local median household income /1,000, age 17</td>
<td>47.49</td>
<td>43.04</td>
<td>46.49</td>
<td>***</td>
</tr>
<tr>
<td>Local Bachelor’s Degree or higher percentage, age 17</td>
<td>19.38</td>
<td>18.83</td>
<td>19.17</td>
<td>*</td>
</tr>
</tbody>
</table>

Observations | 1,049 | 507 | 1,556 |
Percentage (%) | (67.4%) | (32.6%) | (100%) |

Notes: NLSY97 data. All measures are computed using NLSY97 sample weights. Means are reported, with standard deviations in parentheses. Parental highest grade and AFQT have fewer numbers of observations because of missing values. The two-group mean t tests or Chi-square tests are conducted between whites and blacks. *** significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.
Table 3: OLS Estimation of Highest Grade Completed (whites and blacks pooled together)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported probability of</td>
<td>no</td>
<td>-0.019***</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>-0.008***</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>death by next year</td>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported probability of</td>
<td>no</td>
<td>no</td>
<td>-0.023***</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>-0.012***</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>death by next year</td>
<td></td>
<td></td>
<td>(0.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death expectations based on</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>-2.506***</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>-0.420</td>
<td></td>
</tr>
<tr>
<td>actual death probability</td>
<td></td>
<td></td>
<td></td>
<td>(0.791)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.805)</td>
</tr>
<tr>
<td>Black</td>
<td>-1.168***</td>
<td>-1.117***</td>
<td>-1.026***</td>
<td>-1.089***</td>
<td>0.810***</td>
<td>0.845***</td>
<td>0.804***</td>
<td>0.842***</td>
<td>0.872***</td>
<td>0.845***</td>
</tr>
<tr>
<td></td>
<td>(0.190)</td>
<td>(0.186)</td>
<td>(0.188)</td>
<td>(0.166)</td>
<td>(0.190)</td>
<td>(0.197)</td>
<td>(0.165)</td>
<td>(0.188)</td>
<td>(0.194)</td>
<td>(0.190)</td>
</tr>
<tr>
<td>AFQT scores</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.079)</td>
<td>(0.080)</td>
<td>(0.078)</td>
<td>(0.080)</td>
<td>(0.080)</td>
<td>(0.083)</td>
</tr>
<tr>
<td>Family background and</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>demographic characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State dummies at age 17</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>County characteristics age 17</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>(0.108)</td>
<td>(0.123)</td>
<td>(0.135)</td>
<td>(0.116)</td>
<td>(2.404)</td>
<td>(2.335)</td>
<td>(2.795)</td>
<td>(2.390)</td>
<td>(2.335)</td>
<td>(2.951)</td>
</tr>
<tr>
<td>N</td>
<td>1,556</td>
<td>1,556</td>
<td>1,555</td>
<td>1,550</td>
<td>1,556</td>
<td>1,555</td>
<td>1,550</td>
<td>1,556</td>
<td>1,555</td>
<td>1,455</td>
</tr>
<tr>
<td>R²</td>
<td>0.023</td>
<td>0.043</td>
<td>0.052</td>
<td>0.023</td>
<td>0.465</td>
<td>0.469</td>
<td>0.439</td>
<td>0.468</td>
<td>0.469</td>
<td>0.464</td>
</tr>
</tbody>
</table>

Notes: NLSY97 data. The dependent variable is the highest grade completed in 2008. Other control variables not reported are age, female, bad health and family size at age 17. For columns (4), (7) and (10), clustered standard errors by race-sex-age-county groups at age 17 are reported. Observations with missing AFQT scores are dropped. Missing parental highest grades are addressed by the dummy variable method. Geographic controls include three indicators for residence region, indicators for central city residence and MSA residence at age 17. County’s characteristics include the unemployment rate, crime rate, and real median household income in the residence county at age 17. Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level,* significant at the 10 percent level.
Table 4A: Benchmark OLS Estimation of Highest Grade Completed by Racial Groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Both</th>
<th>(2) White</th>
<th>(3) Black</th>
<th>(4) Both</th>
<th>(5) White</th>
<th>(6) Black</th>
<th>(7) Both</th>
<th>(8) White</th>
<th>(9) Black</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported probability of death by next year</td>
<td>-0.008***</td>
<td>-0.008**</td>
<td>-0.007**</td>
<td>-0.012***</td>
<td>-0.013***</td>
<td>-0.007**</td>
<td>-0.420</td>
<td>-0.140</td>
<td>-0.650</td>
</tr>
<tr>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.805)</td>
<td>(1.070)</td>
<td>(1.013)</td>
</tr>
<tr>
<td>Self-reported probability of death by age 20</td>
<td>0.842***</td>
<td>0.597***</td>
<td>0.658***</td>
<td>0.606***</td>
<td>0.607***</td>
<td>0.694***</td>
<td>0.845**</td>
<td>0.576***</td>
<td>0.602***</td>
</tr>
<tr>
<td>(0.188)</td>
<td>(0.135)</td>
<td>(0.207)</td>
<td>(0.113)</td>
<td>(0.133)</td>
<td>(0.205)</td>
<td>(0.137)</td>
<td>(0.158)</td>
<td>(0.251)</td>
<td></td>
</tr>
<tr>
<td>AFQT scores</td>
<td>1.318***</td>
<td>1.274***</td>
<td>1.491***</td>
<td>1.320***</td>
<td>1.274***</td>
<td>1.495***</td>
<td>1.326***</td>
<td>1.275***</td>
<td>1.492***</td>
</tr>
<tr>
<td>(0.080)</td>
<td>(0.082)</td>
<td>(0.150)</td>
<td>(0.080)</td>
<td>(0.082)</td>
<td>(0.152)</td>
<td>(0.083)</td>
<td>(0.095)</td>
<td>(0.161)</td>
<td></td>
</tr>
<tr>
<td>Mother highest grade completed</td>
<td>0.104***</td>
<td>0.113***</td>
<td>0.194***</td>
<td>0.100***</td>
<td>0.112***</td>
<td>0.191***</td>
<td>0.114***</td>
<td>0.107***</td>
<td>0.179***</td>
</tr>
<tr>
<td>(0.033)</td>
<td>(0.037)</td>
<td>(0.055)</td>
<td>(0.033)</td>
<td>(0.037)</td>
<td>(0.055)</td>
<td>(0.036)</td>
<td>(0.041)</td>
<td>(0.065)</td>
<td></td>
</tr>
<tr>
<td>Father highest grade completed</td>
<td>0.134***</td>
<td>0.134***</td>
<td>0.056</td>
<td>0.131***</td>
<td>0.125***</td>
<td>0.053</td>
<td>0.132***</td>
<td>0.145***</td>
<td>0.079</td>
</tr>
<tr>
<td>(0.033)</td>
<td>(0.036)</td>
<td>(0.057)</td>
<td>(0.033)</td>
<td>(0.035)</td>
<td>(0.057)</td>
<td>(0.035)</td>
<td>(0.039)</td>
<td>(0.064)</td>
<td></td>
</tr>
<tr>
<td>Household income ($1000) age 17</td>
<td>0.001</td>
<td>0.002</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>-0.001</td>
<td>0.002*</td>
<td>0.002*</td>
<td>0.001</td>
</tr>
<tr>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Live with both parents age 17</td>
<td>0.825***</td>
<td>0.801***</td>
<td>0.751***</td>
<td>0.821***</td>
<td>0.789***</td>
<td>0.748***</td>
<td>0.808***</td>
<td>0.848***</td>
<td>0.567*</td>
</tr>
<tr>
<td>(0.147)</td>
<td>(0.157)</td>
<td>(0.258)</td>
<td>(0.147)</td>
<td>(0.158)</td>
<td>(0.259)</td>
<td>(0.151)</td>
<td>(0.167)</td>
<td>(0.322)</td>
<td></td>
</tr>
<tr>
<td>State dummies age 17, County characteristics age 17</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Notes: NLSY97 data. The dependent variable is the highest grade completed in 2008. Other control variables not reported are age, female, bad health and family size at age 17. For columns (7)-(9), clustered standard errors by race-sex-age-county groups at age 17 are reported. Observations with missing AFQT scores are dropped. Missing parental highest grades are addressed by the dummy variable method. Geographic controls include three indicators for residence county, indicators for central city residence and MSA residence at age 17. County’s characteristics include the unemployment rate, crime rate, and median household income in 2006 dollars in the residence county at age 17.

Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level,* significant at the 10 percent level. ***significant at the 1 percent level, **significant at the 5 percent level,* significant at the 10 percent level.
Table 4B: Oaxaca Decomposition of the Black–White Educational Years Gaps from Table 4A

<table>
<thead>
<tr>
<th></th>
<th>(1) Self-reported probability of death next year</th>
<th>(2) Self-reported probability of death by age 20</th>
<th>(3) Death expectations based on actual death probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.111)</td>
<td>(0.111)</td>
<td>(0.100)</td>
</tr>
<tr>
<td>Predicted schooling years: black</td>
<td>12.932***</td>
<td>12.977***</td>
<td>12.958***</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.116)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>Predicted total difference</td>
<td>1.173***</td>
<td>1.124***</td>
<td>1.139***</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
<td>(0.167)</td>
<td>(0.146)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Endowment effects</th>
<th>Coefficient effects</th>
<th>Endowment effects</th>
<th>Coefficient effects</th>
<th>Endowment effects</th>
<th>Coefficient effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total endowments or coefficients effects</td>
<td>1.942***</td>
<td>-0.768***</td>
<td>1.912***</td>
<td>-0.788***</td>
<td>1.900***</td>
<td>-0.761***</td>
</tr>
<tr>
<td></td>
<td>(0.169)</td>
<td>(0.153)</td>
<td>(0.168)</td>
<td>(0.156)</td>
<td>(0.150)</td>
<td>(0.148)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Endowment effects</th>
<th>Coefficient effects</th>
<th>Endowment effects</th>
<th>Coefficient effects</th>
<th>Endowment effects</th>
<th>Coefficient effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed decompositions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality risks</td>
<td>0.047***</td>
<td>-0.019</td>
<td>0.040**</td>
<td>-0.126</td>
<td>-0.021</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.086)</td>
<td>(0.016)</td>
<td>(0.088)</td>
<td>(0.026)</td>
<td>(0.148)</td>
</tr>
<tr>
<td>AFQT scores</td>
<td>1.247***</td>
<td>0.074**</td>
<td>1.237***</td>
<td>0.074**</td>
<td>1.266***</td>
<td>0.071**</td>
</tr>
<tr>
<td></td>
<td>(0.107)</td>
<td>(0.034)</td>
<td>(0.106)</td>
<td>(0.034)</td>
<td>(0.100)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>Female, age, health</td>
<td>-0.021</td>
<td>-0.132</td>
<td>-0.025</td>
<td>-0.379</td>
<td>-0.025</td>
<td>-1.254</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(3.583)</td>
<td>(0.020)</td>
<td>(3.561)</td>
<td>(0.026)</td>
<td>(4.089)</td>
</tr>
<tr>
<td>Parental education</td>
<td>0.290***</td>
<td>0.086</td>
<td>0.277***</td>
<td>0.028</td>
<td>0.282***</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.735)</td>
<td>(0.059)</td>
<td>(0.733)</td>
<td>(0.055)</td>
<td>(0.698)</td>
</tr>
<tr>
<td>Family income, family size</td>
<td>0.312***</td>
<td>0.458</td>
<td>0.308***</td>
<td>0.542*</td>
<td>0.302***</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.330)</td>
<td>(0.053)</td>
<td>(0.325)</td>
<td>(0.053)</td>
<td>(0.304)</td>
</tr>
<tr>
<td>Geographic variables</td>
<td>0.067</td>
<td>0.640</td>
<td>0.076</td>
<td>0.675</td>
<td>0.083</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.517)</td>
<td>(0.087)</td>
<td>(0.499)</td>
<td>(0.080)</td>
<td>(0.579)</td>
</tr>
<tr>
<td>intercept</td>
<td>-1.875</td>
<td>-1.601</td>
<td>-1.601</td>
<td>-1.119</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.670)</td>
<td>(3.632)</td>
<td>(3.632)</td>
<td></td>
<td></td>
<td>(4.161)</td>
</tr>
</tbody>
</table>

N: 1,558 1,556 1,548

Notes: NLSY97 data. Decomposition uses the coefficients from a pooled model over both groups as the reference coefficients. The dependent variable is black-white differential in highest grade completed in 2008. Other control variables not reported are age, female, bad health and family size at age 17. For actual death rates, clustered standard errors by race-sex-age-county groups at age 17 are reported. Observations with missing AFQT scores are dropped. Missing parental highest grades are imputed with constant 0 and dummy variables indicating missing are included as control variables. Geographic controls include three indicators for residence regions, indicators for central city residence and MSA residence at age 17. Characteristics of counties include the unemployment rate, crime rate, and real median household income at the residence county at age 17.

Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Table 5A: Robustness Check: Alternative Samples of Mortality and Education

<table>
<thead>
<tr>
<th></th>
<th>(1) Sample dropping those who have completed schooling in 1997</th>
<th>(2) Sample including those who do not have subjective death expectations</th>
<th>(3) Sample including those missing AFQT scores are dummy variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Black</td>
<td>White</td>
</tr>
<tr>
<td><strong>Self-reported probability of death next year</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.006*</td>
<td>0.006*</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td><strong>Self-reported probability of death by age 20</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.011***</td>
<td>0.004</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td><strong>Actual Mortality rates from reference group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.291</td>
<td>-0.407</td>
<td>1.195</td>
</tr>
<tr>
<td></td>
<td>(1.544)</td>
<td>(1.432)</td>
<td>(0.913)</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>1,037</td>
<td>470</td>
<td>2,907</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.431</td>
<td>0.515</td>
<td>0.454</td>
</tr>
</tbody>
</table>

Notes: NLSY97 data. Control variables are same as Table 4A. The dependent variable is in highest grade completed in 2008. Other control variables not reported are age, female, bad health and family size at age 17. For actual death rates, clustered standard errors by race-sex-age-county groups at age 17 are reported. Missing parental highest grades are imputed with constant 0 and dummy variables indicating missing are included as control variables. Geographic controls include three indicators for residence regions, indicators for central city residence and MSA residence at age 17. County’s characteristics include the unemployment rate, crime rate, and median household income in 2006 dollars at the residence county at age 17.

Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Table 5B: Decomposition Results for Robustness Check from Table 5A

<table>
<thead>
<tr>
<th>Sample including those who have two-wave subjective death expectations</th>
<th>Sample including those who do not have subjective death expectations</th>
<th>Sample including those missing AFQT scores are dummy variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Endowment gap</td>
<td>Endowment effect by each factor</td>
<td>Endowment gap</td>
</tr>
<tr>
<td>Total Endowment gap</td>
<td>Effect by each factor</td>
<td>Effect by each factor</td>
</tr>
<tr>
<td>Total</td>
<td>1.786***</td>
<td>1.816***</td>
</tr>
<tr>
<td>Reported probability of death next year</td>
<td>(0.158)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>Reported probability of death by age 20</td>
<td>0.035**</td>
<td>0.003*</td>
</tr>
<tr>
<td>Death expectations based on actual death probability</td>
<td>(0.017)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Death expectations based on actual death probability</td>
<td>-0.139</td>
<td>0.028*</td>
</tr>
<tr>
<td>Death expectations based on actual death probability</td>
<td>(0.093)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Death expectations based on actual death probability</td>
<td>0.134</td>
<td>0.028*</td>
</tr>
<tr>
<td>Death expectations based on actual death probability</td>
<td>(0.155)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>N</td>
<td>1,597</td>
<td>4,493</td>
</tr>
</tbody>
</table>

Notes: NLSY97 data. Decomposition uses the coefficients from a pooled model over both groups as the reference coefficients in Table 5A. The dependent variable is black-white differential in years of schooling in 2008. Other control variables not reported are age, female, bad health and family size at age 17. For actual death rates, clustered standard errors by race-sex-county groups at age 17 are reported. Observations with missing AFQT scores are dropped. Missing parental highest grades are imputed with constant 0 and dummy variables indicating missing are included as control variables. Geographic controls include three indicators for residence regions, indicators for central city residence and MSA residence at age 17. County’s characteristics include the unemployment rate, crime rate, and median household income in 2006 dollars at the residence county at age 17. Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Table 6A: Probit Estimation Results for High School Graduates –(Marginal Effects Reported)

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) White</th>
<th>(2) Black</th>
<th>(3) White</th>
<th>(4) Black</th>
<th>(5) White</th>
<th>(6) Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported probability of death next year</td>
<td>-0.003</td>
<td>-0.005*</td>
<td>-0.004*</td>
<td>-0.006**</td>
<td>0.399</td>
<td>1.314</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(1.100)</td>
<td>(0.941)</td>
</tr>
<tr>
<td>Self-reported probability of death up to age 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death expectations based on actual death probability</td>
<td>0.004</td>
<td>0.006**</td>
<td>0.006</td>
<td>0.008**</td>
<td>0.006**</td>
<td>0.008**</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>AFQT scores</td>
<td>0.584***</td>
<td>0.999***</td>
<td>0.582***</td>
<td>0.990***</td>
<td>0.554***</td>
<td>0.980***</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.196)</td>
<td>(0.085)</td>
<td>(0.193)</td>
<td>(0.082)</td>
<td>(0.183)</td>
</tr>
<tr>
<td>Mother highest grade completed</td>
<td>0.039</td>
<td>0.128**</td>
<td>0.042*</td>
<td>0.135***</td>
<td>0.034</td>
<td>0.108**</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.050)</td>
<td>(0.024)</td>
<td>(0.049)</td>
<td>(0.029)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>Father highest grade completed</td>
<td>0.066**</td>
<td>0.064</td>
<td>0.055**</td>
<td>0.066</td>
<td>0.068**</td>
<td>0.114**</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.050)</td>
<td>(0.026)</td>
<td>(0.051)</td>
<td>(0.029)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>Household income ($1000) age 17</td>
<td>-0.001</td>
<td>-0.005***</td>
<td>-0.001</td>
<td>-0.005***</td>
<td>-0.000</td>
<td>-0.004**</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Live with both parents age 17</td>
<td>0.616***</td>
<td>0.385*</td>
<td>0.617***</td>
<td>0.390*</td>
<td>0.568***</td>
<td>0.483**</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
<td>(0.212)</td>
<td>(0.128)</td>
<td>(0.209)</td>
<td>(0.128)</td>
<td>(0.218)</td>
</tr>
<tr>
<td>State dummies age 17</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>County characteristics age 17</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>N</td>
<td>1,110</td>
<td>510</td>
<td>1,110</td>
<td>510</td>
<td>1,054</td>
<td>458</td>
</tr>
</tbody>
</table>

Notes: NLSY97 data. Marginal average effects after probit model are reported. The dependent variable is total years of schooling in 2008. Other control variables not reported are age, female, bad health and family size at age 17. For columns (7)-(9), clustered standard error by race-sex-age-county groups at age 17 are reported. Observations with missing AFQT scores are dropped. Missing parental highest grades are imputed with constant 0 and dummy variables indicating missing are included as control variables. Geographic controls include three indicators for residence regions, indicators for central city residence and MSA residence at age 17. County characteristics include the unemployment rate, crime rate, and median household income in 2006 dollars at the residence county at age 17. Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level,* significant at the 10 percent level.
Table 6B: Decomposition of Black–White Gaps in Percentage of High School Graduates from Table 5A

<table>
<thead>
<tr>
<th></th>
<th>NON-LINEAR DECOMPOSITION FOR PROBIT REGRESSION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Self-reported probability of death next year</td>
</tr>
<tr>
<td></td>
<td>Predicted schooling years white</td>
</tr>
<tr>
<td></td>
<td>0.871*** (0.010)</td>
</tr>
<tr>
<td></td>
<td>Predicted schooling years black</td>
</tr>
<tr>
<td></td>
<td>0.872*** (0.010)</td>
</tr>
<tr>
<td></td>
<td>Predicted total difference</td>
</tr>
<tr>
<td></td>
<td>0.871*** (0.010)</td>
</tr>
</tbody>
</table>

|                          | Endowments effects                           | Coefficients effects                        | Endowments effects                           | Coefficients effects                        | Endowments effects                           | Coefficients effects                        |
| Total endowments or coefficients effects | 0.244*** (0.030)                              | -0.088*** (0.033)                           | 0.235*** (0.030)                              | -0.085*** (0.033)                           | 0.238*** (0.029)                              | -0.085*** (0.034)                           |

**Detailed decompositions**

**Mortality risks**

- Predicted schooling years white: 0.004 (0.003) 0.008 (0.015) 0.004* (0.002) 0.008 (0.016) -0.002 (0.004) -0.014 (0.022)
- Predicted schooling years black: 0.130*** (0.018) 0.035*** (0.013) 0.127*** (0.018) 0.034*** (0.013) 0.131*** (0.017) 0.034*** (0.012)
- Predicted total difference: -0.001 (0.003) 1.295 (0.913) -0.002 (0.002) 1.251 (0.902) -0.002 (0.003) 1.119 (0.819)

**Age-adjusted standard AFQT**

- Predicted schooling years white: 0.039*** (0.009) -0.197 (0.147) 0.036*** (0.009) -0.231 (0.146) 0.033*** (0.009) -0.197 (0.156)
- Predicted schooling years black: 0.120* (0.010) 0.065 (0.065) 0.121* (0.010) 0.065 (0.010) 0.034*** (0.010) 0.110** (0.056)
- Predicted total difference: 0.036*** (0.017) 0.056 (0.063) 0.035** (0.016) 0.066 (0.062) 0.038** (0.015) 0.070 (0.079)

**Female, age, health**

- Predicted schooling years white: -1.406 (0.944) -1.335 (0.927) -1.209 (0.851)
- Predicted schooling years black: -1.345 (0.928) -1.210 (0.853)
- Predicted total difference: -1.219 (0.856)

**Parental education**

- Predicted schooling years white: 0.039*** (0.009) -0.197 (0.147) 0.036*** (0.009) -0.231 (0.146) 0.033*** (0.009) -0.197 (0.156)
- Predicted schooling years black: 0.120* (0.010) 0.065 (0.065) 0.121* (0.010) 0.065 (0.010) 0.034*** (0.010) 0.110** (0.056)
- Predicted total difference: 0.036*** (0.017) 0.056 (0.063) 0.035** (0.016) 0.066 (0.062) 0.038** (0.015) 0.070 (0.079)

**Family income, family size**

- Predicted schooling years white: -1.406 (0.944) -1.335 (0.927) -1.209 (0.851)
- Predicted schooling years black: -1.345 (0.928) -1.210 (0.853)
- Predicted total difference: -1.219 (0.856)

**Geographic variables**

- Predicted schooling years white: 0.039*** (0.009) -0.197 (0.147) 0.036*** (0.009) -0.231 (0.146) 0.033*** (0.009) -0.197 (0.156)
- Predicted schooling years black: 0.120* (0.010) 0.065 (0.065) 0.121* (0.010) 0.065 (0.010) 0.034*** (0.010) 0.110** (0.056)
- Predicted total difference: 0.036*** (0.017) 0.056 (0.063) 0.035** (0.016) 0.066 (0.062) 0.038** (0.015) 0.070 (0.079)

**Intercept**

- Predicted schooling years white: -1.406 (0.944) -1.335 (0.927) -1.209 (0.851)
- Predicted schooling years black: -1.345 (0.928) -1.210 (0.853)
- Predicted total difference: -1.219 (0.856)

**Notes:** NLSY97 data. The non-linear decomposition for binary dependent variables uses the method in Yun (2004). The coefficients used in the decomposition are from a pooled probit model over both groups as the reference coefficients. The dependent variable is black-white difference in the share of high school graduates in 2008. Other control variables not reported are age, female, bad health and family size at age 17. Observations with missing AFQT scores are dropped. Missing parental highest grades are imputed with constant 0 and dummy variables indicating missing are included as control variables. Geographic controls include three indicators for residence regions, indicators for central city residence and MSA residence at age 17. Characteristics of counties include the unemployment rate, crime rate, and median household income in 2006 dollars at the residence county at age 17. Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Table 7A: Results for Categorical Variable of Mortality Expectations (NLSY97)

<table>
<thead>
<tr>
<th>Panel A: Regression Results</th>
<th>Both</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficients (s.e.)</td>
<td>N</td>
<td>R2</td>
</tr>
<tr>
<td>Dummy variable:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported death probability by next year &gt;=50%</td>
<td>-0.341**</td>
<td>1,556</td>
<td>0.467</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td></td>
<td>(0.170)</td>
</tr>
<tr>
<td>Dummy variable:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported death probability by next year &gt;=50%</td>
<td>-0.460***</td>
<td>1,555</td>
<td>0.466</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td></td>
<td>(0.166)</td>
</tr>
<tr>
<td>Death expectations based on actual death probability &gt;=0.05</td>
<td>-0.094</td>
<td>1541</td>
<td>0.462</td>
</tr>
<tr>
<td></td>
<td>(0.188)</td>
<td></td>
<td>(0.156)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Decomposition Results</th>
<th>Education gap (total effects)</th>
<th>Endowment effects</th>
<th>Coefficient effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.170***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.173)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy variable:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported death probability by next year &gt;=50%</td>
<td></td>
<td>0.033**</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.016)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>AFQT scores</td>
<td></td>
<td>1.226***</td>
<td>0.083***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.114)</td>
<td>(0.032)</td>
</tr>
<tr>
<td></td>
<td>1.120***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy variable:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported death probability by next year &gt;=50%</td>
<td></td>
<td>0.037**</td>
<td>-0.087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.016)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>AFQT scores</td>
<td></td>
<td>1.205***</td>
<td>0.092***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.112)</td>
<td>(0.032)</td>
</tr>
<tr>
<td></td>
<td>1.158***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy variable:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death expectations based on actual death probability &gt;=0.05</td>
<td></td>
<td>0.006</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.009)</td>
<td>(0.133)</td>
</tr>
<tr>
<td>AFQT scores</td>
<td></td>
<td>1.267***</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.106)</td>
<td>(0.032)</td>
</tr>
</tbody>
</table>

Notes: The mortality expectations are coded as a dummy variable for those whose subjective expectations are above 50% and actual mortality rates are higher than 0.05 per hundred population. These estimations are for whites and blacks together. All the other control variables that are not reported in this table are the same as in Table 4A and Table 4B. Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Table 7B: OLS Regression Using Alternative Measures for Actual Death Rates

<table>
<thead>
<tr>
<th></th>
<th>(1) White</th>
<th>(2) Black</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual mortality rate aged 15-67</td>
<td>-0.140</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>(0.864)</td>
<td>(0.599)</td>
</tr>
<tr>
<td>N</td>
<td>1,140</td>
<td>547</td>
</tr>
<tr>
<td>R²</td>
<td>0.459</td>
<td>0.500</td>
</tr>
<tr>
<td><strong>Panel B:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life expectancy aged 17-67</td>
<td>.056</td>
<td>.124</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.090)</td>
</tr>
<tr>
<td>N</td>
<td>1,087</td>
<td>502</td>
</tr>
<tr>
<td>R²</td>
<td>0.468</td>
<td>0.510</td>
</tr>
</tbody>
</table>

Notes: NLSY97 data. The dependent variable is highest grade completed in 2008 as defined in Table 1. Clustered standard errors are in parentheses. Other control variables not reported are age, female, bad health and family size at age 17, which are the same as baseline specification. Dummies for residence states and county characteristics at age 17 are also controlled.

Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Table 8: OLS Regression for Schooling Years Adding Additional Controls

<table>
<thead>
<tr>
<th></th>
<th>White (1)</th>
<th>Black (1)</th>
<th>White (2)</th>
<th>Black (2)</th>
<th>White (3)</th>
<th>Black (3)</th>
<th>White (4)</th>
<th>Black (4)</th>
<th>White (5)</th>
<th>Black (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported probability of death in next year</td>
<td>-0.006**</td>
<td>-0.004</td>
<td>-0.006**</td>
<td>-0.005</td>
<td>-0.006**</td>
<td>-0.004*</td>
<td>-0.007**</td>
<td>-0.005*</td>
<td>-0.004</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Reported probability of death by age 20</td>
<td>-0.010***</td>
<td>-0.004</td>
<td>-0.007*</td>
<td>-0.004</td>
<td>-0.007*</td>
<td>-0.004</td>
<td>-0.008***</td>
<td>-0.004</td>
<td>-0.007**</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Death expectations based on actual death probability</td>
<td>1.863</td>
<td>0.027</td>
<td>1.261</td>
<td>-0.521</td>
<td>1.761</td>
<td>1.432</td>
<td>0.441</td>
<td>1.945</td>
<td>0.199</td>
<td>1.038</td>
</tr>
<tr>
<td></td>
<td>(1.461)</td>
<td>(1.165)</td>
<td>(1.257)</td>
<td>(1.926)</td>
<td>(1.373)</td>
<td>(1.264)</td>
<td>(1.123)</td>
<td>(1.274)</td>
<td>(1.023)</td>
<td>(1.004)</td>
</tr>
<tr>
<td>(a) Education at age 17</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>(b) Early-year experience; Substance abuse; Delinquency</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>(c) Optimism or expected difficulty in doing well in school</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>(d) Personality</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Notes: NLSY97 data. Decomposition uses the coefficients from a pooled model over both groups as the reference coefficients. The dependent variable is highest grade completed in 2008. Other control variables not reported are age, female, bad health and family size at age 17. For actual death probability, clustered standard error by local same race-sex-age groups at age 17 are reported. Observations with missing AFQT scores are dropped. Missing parental highest grades are imputed with constant 0 and dummy variables indicating missing are included as control variables. Geographic controls include three indicators for residence regions, indicators for central city residence and MSA residence at age 17. County’s characteristics include the unemployment rate, crime rate, and median household income in 2006 dollars at the residence county at age 17. Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Table 9A: Add Health Data: Logistic Regression for Educational Attainment

<table>
<thead>
<tr>
<th>Marginal effect</th>
<th>Without controlling for cognitive ability</th>
<th>Controlling for cognitive ability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) % High school graduate+</td>
<td>(2) % college graduate+</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>White</td>
</tr>
<tr>
<td>Reported survival chance: less than 50%</td>
<td>-0.053*** (0.013)</td>
<td>0.055*** (0.014)</td>
</tr>
<tr>
<td>a good chance</td>
<td>-0.001 (0.012)</td>
<td>0.003 (0.013)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.030** (0.012)</td>
<td>0.058*** (0.020)</td>
</tr>
<tr>
<td>Cognitive test scores</td>
<td>0.060*** (0.018)</td>
<td>0.055*** (0.019)</td>
</tr>
<tr>
<td>Parent high school</td>
<td>0.063** (0.027)</td>
<td>0.045 (0.027)</td>
</tr>
<tr>
<td>Parent some college</td>
<td>0.077*** (0.020)</td>
<td>0.066*** (0.020)</td>
</tr>
<tr>
<td>Parent college</td>
<td>-0.067*** (0.013)</td>
<td>0.065*** (0.015)</td>
</tr>
<tr>
<td>Mother received welfare</td>
<td>0.029*** (0.011)</td>
<td>0.023** (0.011)</td>
</tr>
<tr>
<td>Live with both parents</td>
<td>0.004 (0.005)</td>
<td>0.010* (0.006)</td>
</tr>
<tr>
<td>Attachment to parents</td>
<td>4.235 (0.006)</td>
<td>3.396 (0.009)</td>
</tr>
</tbody>
</table>

Notes: Add Health public-use data. Marginal effects at the mean after logistic estimation are reported. Standard errors in parentheses. The dependent variables are the percentages of high school graduate or more and college graduate or more in Wave 4 respectively. Other control variables not reported are age, female, bad health and family size at Wave 1. Cognitive ability test scores are measured using the Picture Vocabulary Test age-adjusted standardized scores. Control variables that are incorporated in the estimation but are not reported are: risky behaviors include violence behavior scale, the number of days smoke in past 30 days, whether have ever used any illicit drugs and heavy drinking in past 12 months. Other factors that are included in the decomposition but not reported are female, age, born as U.S. citizen, safe in neighborhood, self-reported health status indicators and depression measure, and whether enrolled in school in Wave 4 (all the other impenent variables are in Wave 1). ***significant at the 1 percent level, **significant at the 5 percent level,* significant at the 10 percent level.
Table 9B: Add Health Data: Oaxaca Decomposition for Educational Attainment (after logit estimation)

<table>
<thead>
<tr>
<th>Coefficients decomposition</th>
<th>Without controlling for cognitive ability</th>
<th>Controlling for cognitive ability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) %High school graduate+ (2) %college graduate+</td>
<td>(3) %High school graduate+ (4) %college graduate+</td>
</tr>
<tr>
<td>Predicted education: whites</td>
<td>0.923*** (0.005) 0.332*** (0.009)</td>
<td>0.921*** (0.006) 0.333*** (0.009)</td>
</tr>
<tr>
<td>Predicted education: blacks</td>
<td>0.854*** (0.015) 0.206*** (0.015)</td>
<td>0.850*** (0.015) 0.198*** (0.015)</td>
</tr>
<tr>
<td>Predicted total difference</td>
<td>0.069*** (0.015) 0.126*** (0.017)</td>
<td>0.071*** (0.016) 0.135*** (0.018)</td>
</tr>
<tr>
<td>Explained difference</td>
<td>0.038*** (0.008) 0.066*** (0.011)</td>
<td>0.076*** (0.010) 0.160*** (0.014)</td>
</tr>
<tr>
<td>Unexplained difference</td>
<td>0.032** (0.016) 0.060*** (0.018)</td>
<td>-0.004 (0.017) -0.025 (0.019)</td>
</tr>
<tr>
<td></td>
<td><strong>Total endowments or coefficients effects</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endowments effects Coefficients effects</td>
<td>Endowments effects Coefficients effects</td>
</tr>
<tr>
<td>Reported chance of living to age 35: less than 50%</td>
<td>0.012*** (0.003) -0.001 (0.010)</td>
<td>0.011*** (0.003) -0.003 (0.010)</td>
</tr>
<tr>
<td></td>
<td>0.000 (0.000) 0.001 (0.010)</td>
<td>0.002 (0.012) 0.000 (0.000)</td>
</tr>
<tr>
<td>A good chance</td>
<td></td>
<td>0.010*** (0.003) -0.006 (0.010)</td>
</tr>
<tr>
<td></td>
<td>0.004 (0.001) 0.005 (0.012)</td>
<td>0.004 (0.003) 0.005 (0.011)</td>
</tr>
<tr>
<td>Cognitive test scores</td>
<td></td>
<td>0.005 (0.006) 0.120*** (0.008)</td>
</tr>
<tr>
<td></td>
<td>0.004* (0.002) 0.000 (0.010)</td>
<td>0.003 (0.002) 0.005** (0.003)</td>
</tr>
<tr>
<td></td>
<td>-0.003** (0.014) -0.002** (0.013)</td>
<td>-0.003** (0.010) -0.212 (0.010)</td>
</tr>
<tr>
<td></td>
<td>-0.001 (0.006) 0.075*** (0.007)</td>
<td>0.100 (0.006) 0.025*** (0.007)</td>
</tr>
<tr>
<td></td>
<td>-0.005* (0.003) -0.017*** (0.020)</td>
<td>0.040** (0.018) 0.003 (0.020)</td>
</tr>
<tr>
<td></td>
<td>0.028 (0.211) -0.522** (0.241)</td>
<td>0.326 (0.212) -0.429* (0.233)</td>
</tr>
</tbody>
</table>

Notes: Add Health public-use data. The non-linear decomposition for binary dependent variables uses the method in Yun (2004) and the coefficients in the decomposition are from a pooled logistic model over both groups as the reference coefficients. The dependent variables are the percentages of high school graduate or more and college graduate or more in Wave 4 respectively. Other control variables not reported are age, female, bad health and family size at Wave 1. In the decomposition, family background includes parental education indicators, parental receiving welfare indicator, attachment to parent scale, and living with both parents indicator. Risky behaviors include violence behavior scale, the number of days smoke in past 30 days, whether have ever used any illicit drugs and heavy drinking in past 12 months. Other factors that are included in the decomposition but not reported are female, Born as U.S. citizen, self-reported health status indicators, depression measure, and whether enrolled in school Wave 4 (all the other impendent variables are in Wave 1). ***significant at the 1 percent level, **significant at the 5 percent level,* significant at the 10 percent level.
## Appendix

### Appendix Table 1: OLS regression Results on Death Expectations

<table>
<thead>
<tr>
<th></th>
<th>(1) 1997 OLS</th>
<th>(2) 1997 OLS</th>
<th>(3) 1997-2002 OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self-reported probability of death next year</td>
<td>Self-reported probability of death by age 20</td>
<td>Self-reported probability of death next year</td>
</tr>
<tr>
<td>AFQT scores</td>
<td>-2.071***</td>
<td>-1.641***</td>
<td>-1.855***</td>
</tr>
<tr>
<td></td>
<td>(0.576)</td>
<td>(0.605)</td>
<td>(0.552)</td>
</tr>
<tr>
<td>Experience hard times</td>
<td>9.443***</td>
<td>8.166**</td>
<td>6.607***</td>
</tr>
<tr>
<td></td>
<td>(3.342)</td>
<td>(3.211)</td>
<td>(2.088)</td>
</tr>
<tr>
<td>Multiple sexual partners</td>
<td>5.702***</td>
<td>4.673*</td>
<td>2.969***</td>
</tr>
<tr>
<td></td>
<td>(1.790)</td>
<td>(2.443)</td>
<td>(1.131)</td>
</tr>
<tr>
<td>Ever alcohol abuse</td>
<td>1.800</td>
<td>2.123*</td>
<td>2.278**</td>
</tr>
<tr>
<td></td>
<td>(1.109)</td>
<td>(1.237)</td>
<td>(1.068)</td>
</tr>
<tr>
<td>Ever active in armed forces</td>
<td>n.a.</td>
<td>n.a.</td>
<td>7.451*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.753)</td>
</tr>
<tr>
<td>N</td>
<td>1,556</td>
<td>1,556</td>
<td>1,556</td>
</tr>
</tbody>
</table>

**Notes:** NLSY97 data. The dependent variable for columns (1) is death probabilities by next year in 1997 and (2) is death chance by age 20. Column (3) is the change of death chances up to next year from 1997 to 2002. Other control variables not reported are age, female, bad health and family size at age 17, parental education, family income, county characteristics and residence state dummies. Coefficients for variable "ever active in armed force" are not available for column (1) and (2) because in 1997 only a few respondents are active in armed force.

Because the county-level characteristics variables are measured at the county level, the standard errors in parentheses allow for arbitrary correlation among individuals living in the same county at age 17 (i.e. clustering). ***significant at the 1 percent level, **significant at the 5 percent level, * significant at the 10 percent level.
Chapter 2:
Do People Change Their Health-Related Behaviors When They Raise Their Survival Expectations?

1. Introduction

Do people change their health-related behaviors when they change their survival beliefs? In particular, when an individual’s subjective longevity increases, does he or she then adopt an unhealthy lifestyle as a conscious or unconscious “invincibility response?” Although health investment theory predicts that a longer prospective lifespan tends to reinforce healthy lifestyle choices (Grossman, 1972), the link between subjective longevity beliefs and health behaviors has received little attention in the economic literature. Previous studies on the determinants of health behaviors typically have focused on factors such as education, socioeconomic status and health status. This paper seeks to add to the literature by testing whether changes (in particular, increases) in survival beliefs are associated with changes in health behaviors in terms of smoking, heavy drinking, obesity and physical inactivity.

It is imperative to clarify the factors associated with health behaviors due to their documented association with individual morbidity and mortality outcomes (e.g. Balia and Jones, 2008). Mokdad et al. (2004) document that smoking, poor diet/physical inactivity, and heavy drinking are the three primary sources of all modifiable mortality risks in the U.S. in 2000. Exploring the links between subjective longevity and health behaviors may clarify a potential rationale for the heterogeneity in health behaviors across individuals. Moreover, smoking, heavy drinking, obesity and inactivity are all important proximate
inputs into later health status. The focus on health behaviors underscores potential pathways through which survival beliefs may eventually affect later health outcomes.

At least two reasons are tied to the belief that subjective longevity might affect health behaviors. First, the health investment model suggests that, all else being equal, individuals with higher longevity expectations will invest more in health (for instance, by adopting a healthy lifestyle), because of the longer time-span to reap the benefits.¹

The second reason for expecting survival beliefs to have effects on health activities is that changes in subjective longevity probably also reflect aspects of heterogeneity within individuals, such as individualized health choices, degrees of risk aversion, rates of time preference, and varying abilities to acquire and process relevant information. One interesting linkage mentioned in Balia (2011) is the perception that an anticipated long lifespan enables people to have more time to compensate for negative effects of unhealthy behaviors by diversifying their investments in health.

Although a few recent studies have investigated how changes in health behaviors are associated with changes in subjective expectations of survival, they have largely focused on the formation of subjective longevity expectations that result from health-related changes. By contrast, this paper examines the link between survival beliefs and health behaviors in a reversed direction—how changes in subjective survival probabilities lead to changes in health behaviors. This correlation will be determined by testing one of the predictions of the smoking determination model in Cutler and Glaeser (2009) (hereafter, the Cutler-Glaeser model), which predicts that nonsmokers who have increased expectations of subjective longevity are unlikely to start smoking; similarly, in a population,

¹ The quote by Eubie Blake, who was a life-long smoker but lived a long life, in some sense reflects this investment view, “If I’d known I was going to live this long, I would have taken better care of myself” (see link http://en.wikipedia.org/wiki/Eubie_Blake).
the share of new smokers with increased survival expectations will be lower in comparison to individuals with unchanged or decreased survival expectations. I test this prediction by using a sample from the Health and Retirement Study 1992-2010, which include the responses of people who are 50 years or older in 1992. By examining an older sample it is possible to discount some of the typical confounding effects seen in younger smokers, such as experimentation, habit formation, and reinforcement/persistence (Orphanides and Zervos, 1995). In contrast to the implications of Cutler-Glaeser model, this study supports the finding that between two time periods, a percentage of previous nonsmokers whose expected survival probabilities increase start smoking between two consecutive waves (i.e., every two years)—and that their smoking-onset choices are not significantly different from those who expressed steady or declining survival beliefs. In other words, this study supports the fact that the decision to start smoking is unrelated to any changes in longevity expectations among nonsmokers. This finding holds also true for other health behaviors, such as heavy drinking, overeating and physical inactivity.

This paper contributes to the literature in two important ways. First, by using qualitative and quantitative analysis to test whether health behaviors change in response to alterations in survival beliefs, this study suggests a potential linkage that has been neglected in previous literature—namely, how differences in subjective longevity beliefs can explain the heterogeneity in health behaviors across individuals. Second, this paper adds to the literature by discussing the validity of self-reported subjective survival data in choice models.

The remainder of the paper is organized as follows. In Section 2, I briefly review the related literature. In Section 3, I discuss the empirical approach I selected for this study, along with estimate specifications. Section 4 presents a description of the data set, key
variables and sample definitions. Section 5 reports the estimation results, after which possible interpretations for the findings are reviewed in Section 6. Section 7 suggests some limitations to the study, and summarizes the findings.

2. Literature Review

The link between longevity and health behaviors has long been of interest in health and medical fields; in fact, research investigating the relationship between the two is expanding into heretofore unexplored areas. In particular, health researchers are beginning to look into whether subjective life expectancy beliefs can be considered a significant predictor of health behaviors for both men and women (e.g., Scott-Sheldon et al., 2010).

Within the economic literature, this study bridges two strands: studies on subjective expectations data and studies on health behaviors. With respect to the first body of evidence, subjective expectations have long been considered to be important in the decision-making process. Specifically, any intertemporal decision—or more generally decision-making under uncertainty—would require the decision-maker to have some beliefs about the probability of pertinent future events, regardless of whether those beliefs are consistent with actual probabilities. In part related to the aging of the American population, a growing number of studies are starting to examine whether the subjective mortality or longevity beliefs of older people can be tied to their behavioral choices. So far the results have been mixed: survival probabilities have been linked to retirement decisions (Bloom et al., 2007), savings and consumption patterns (Wong, 2009) and bequest choices (Gan et al., 2004); in contrast, no significant predictive relationship has been identified for the value of Social Security or the optimal claim age (Sun and Webb, 2011).
The second branch of related literature has to do with studies on the determinants of health behaviors. Among them, many have analyzed easily-quantifiable factors such as education and income levels, while only a few have explored the impact of subjective survival expectations on personal health behaviors. A study by Hurd and McGarry (2002), using data from the 1998-2010 waves of the HRS, claim that a person’s subjective mortality expectations increased with the death of a parent—since it had the effect of reminding the child of the inevitability of dying. The researchers also reported that that age of death (i.e., dying younger) also negatively impacted subjective survival beliefs. Khwaja et al. (2007) find that, in the aggregate, individuals are able to accurately predict longevity regardless of smoking status. In contract, at the individual level, current smokers are relatively optimistic about longevity, while people who have never smoked are relatively pessimistic in their subjective mortality beliefs.

Studies more closely related to this paper have used the Bayesian updating framework to investigate the impacts of starting and quitting health-related behaviors on subjective longevity beliefs. For instance, using a comparison of real survival probabilities with estimated survival probabilities, Wang (2008) finds that smokers overestimate their survival probabilities. In a later study, Novak (2010) investigates the influence of personal health problems related to excess body weight on individual risk perceptions. Although these papers look at different health behaviors or factors, they share the goal of trying to account for the formation of subjective survival expectations. That is, they have examined how individuals adjust their longevity expectations in response to changes in personal health behaviors or external stimuli (e.g., the death of a parent). This study, in contrast, looks at how individuals adjust their health behaviors relative to their subjective longevity beliefs—with a particular emphasis on smoking as the behavioral health choice in question.
3. Empirical Approach

In this section, I first discuss the qualitative implications of the Cutler and Glaeser (2009) model. Subsequently, I present the basic empirical specifications for estimating subjective survival beliefs and health behaviors utilized in this paper, followed by discussing potential estimation problems.

3.1 Implication of the Cutler-Glaeser Model

Economic models of smoking behavior can be divided into three groups: imperfectly rational addiction, myopic addiction, and rational addiction (Chaloupka and Warner, 2000). The one that is most related to this study is the third group, rationale addiction, which assumes that prospective smokers decide to smoke only if they believe the benefits outweigh the costs of smoking (Becker and Murphy, 1988). Cutler and Glaeser (2009) propose a smoking-decision model in the spirit of Becker’s rational addiction theory. Following their work, I use smoking as a behavioral choice to derive the implication of their model. It should be stressed, however, that this model also is applicable to other risky behaviors.

Consider a two-period binary smoking choice between $t\cdot 1$ and $t$. Let $U(Y)$ denote the flow utility function for nonsmokers for each time period, where $Y$ stands for income. Smoking, as a behavior, is defined as a binary behavior; any additional utility it brings to the individual is denoted as $R$. The normalized out-of-pocket financial cost of smoking is denoted as $P_c$. Thus, the flow utility function for smokers is represented by $U(Y - P_c) + R$. The individual time discount rate is denoted by $\delta$. The probability of survival from $t\cdot 1$ to $t$ is denoted by $s$. Hence the total present value of utility for a nonsmoker beginning at a certain age $a$ until the prospective age $a+n$ is as following
\[ V_{\text{nonsmoke}} = U(Y) + \sum_{k=1}^{n} (\delta s)^k U(Y) = \frac{U(Y)}{1-\delta}, \]

where \( k=1,2,...,n, \ n \to +\infty. \)\(^2\) The term following the summation is the cumulative present utility (after discounting) since age \( a. \) Adding the utility at age \( a, \ U(Y), \) the sum is the total present utility since age \( a. \)

Now let’s turn to the case of smoking. One expects a decrease in the magnitude of \( \tau \) in longevity if a person smokes. Similarly, total present value for smoking is denoted by

\[ V_{\text{smoke}} = [U(Y-P_c) + R] + \sum_{k=1}^{n} [\delta(s-\tau)]^k [U(Y-P_c) + R] = \frac{U(Y-P_c) + R}{1-\delta(s-\tau)}. \]

### 3.2 Qualitative Implications

Of particular interest to this study is the effect of personal longevity beliefs, denoted by \( s, \) on an individual’s smoking decision. We now turn to the comparative statics of \( s. \) Any changes in survival expectations, denoted by \( \Delta s, \) lead to utility changes for both nonsmokers (3) and smokers (4), as depicted in the following equations:

\[ \frac{\Delta V_{\text{nonsmoke}}}{\Delta s} = \frac{\delta U(Y)}{(1-\delta)^2}; \]

\[ \frac{\Delta V_{\text{smoke}}}{\Delta s} = \frac{\delta[U(Y-P_c) + R]}{(1-\delta(s-\tau))^2}. \]

We first examine the comparative statics for nonsmokers. The condition

\[ V_{\text{nonsmoke}} \geq V_{\text{smoke}} \]

holds true for a nonsmoker, i.e.

\[ \frac{U(Y)}{1-\delta} \geq \frac{U(Y-P_c) + R}{1-\delta(s-\tau)}. \]

---

\(^2\) For simplification, we discuss an infinite and discrete case. The implication of this model of interest still holds in the finite or in continuous cases.
A simple math transformation derived by multiplying (5) by \( \frac{\delta}{1-\delta s} \) on both sides will give us the first inequality in (6):

\[
\frac{\delta U(Y)}{(1-\delta s)^2} \geq \frac{\delta[U(Y-P_s^c)+R]}{(1-\delta(s-\tau))(1-\delta s)} \geq \frac{\delta[U(Y-P_s^c)+R]}{(1-\delta(s-\tau))^2}.
\]

The second inequality follows because \( 1-\delta s < 1-\delta(s-\tau) \). That is, a same-unit change in longevity beliefs would lead to a larger utility change among nonsmokers compared to smokers (i.e., \( \frac{\Delta V_{\text{nonsmoke}}}{\Delta s} > \frac{\Delta V_{\text{smoke}}}{\Delta s} \)).

If there is an increase in subjective survival probability (\( \Delta s > 0 \)), this model confirms that the utility of nonsmoking would be greater than that of smoking. That is, nonsmokers would not start smoking. However, if subjective longevity beliefs are to decrease (\( \Delta s < 0 \)), the results are uncertain. Nonsmokers might either start smoking or decide to remain nonsmokers.

We now turn to smokers, whose baseline condition is \( V_{\text{nonsmoke}} < V_{\text{smoke}} \). The comparison of \( \Delta V_{\text{nonsmoke}} \) and \( \Delta V_{\text{smoke}} \) is ambiguous, because multiplying the baseline condition by \( \frac{\delta}{1-\delta s} \) gives us

\[
\frac{\delta U(Y)}{(1-\delta s)^2} < \frac{\delta[U(Y-P_s^c)+R]}{(1-\delta(s-\tau))(1-\delta s)}.
\]

Based on \( 1-\delta(s-\tau) \geq 1-\delta s \), we have

\[
\frac{\delta[U(Y-P_s^c)+R]}{(1-\delta(s-\tau))^2} < \frac{\delta[U(Y-P_s^c)+R]}{(1-\delta(s-\tau))(1-\delta s)}.
\]

The dilemma is that we do not know for initial smokers (i.e., respondents who start off with \( V_{\text{smoke}} > V_{\text{nonsmoke}} \)) which is larger: the change in utility of smoking associated with the
change in the per unit measure of expected survival expectations \( \frac{\Delta V_{\text{non-smoke}}}{\Delta s} = \frac{\delta U(Y)}{(1 - \delta s)^2} \), or

the change in the utility of nonsmoking \( \frac{\Delta V_{\text{smoke}}}{\Delta s} = \frac{\delta[U(Y - P_s) + R]}{[1 - \delta(s - \tau)]^2} \).

Whether smokers would become nonsmokers depends on two factors. The first is the magnitude of the gap in present values between smoking and nonsmoking associated with the initial time period. The second is the magnitude of the changes for the present values of utility in response to a unit change in survival probabilities for both smoking and nonsmoking. Because both factors are unknown, the result for smokers is undetermined.

As a reminder, the Cutler-Glaeser model suggests that nonsmokers whose expected survival probabilities have increased are less likely to start smoking in comparison to nonsmokers who indicate static or reduced personal longevity beliefs. This linkage, however, is very much along the lines of the classic “chicken-or-egg” scenario—in short, which came first? Because changes in health behaviors and changes in beliefs are difficult to separate along a change continuum, there might be the possibility of reverse causality: that is, health activities can also influence survival beliefs. When taking reverse causality into account, nonsmokers who start smoking would therefore be expected to report a reduction in expected longevity, based on findings in previous studies that people increase their mortality expectations after they become regular smokers (e.g., Sloan et al., 2010). Thus, between two time periods, nonsmokers who report increased longevity expectations should, in theory, be much less likely to start smoking in comparison to those who report static or especially decreased survival expectations. In contrast, the possible changes in health behaviors for smokers who are prompted by changes in survival beliefs are not
predictable—that is, smokers could continue to smoke or quit smoking in the second time period.

The remainder of this paper will focus on nonsmokers in Wave t-1 to test these predictions because that is the particular group for which the Cutler-Glaeser model has unambiguous implications.

3.3 Estimation Specification

To ascertain likely changes in health behaviors in response to fluctuations in subjective survival expectations, the following basic estimation equation is utilized. This predictive relationship is represented by a two-period change model where we controlled for the measures at their initial levels and changes over two consecutive waves, as follows:

\[
(9) \ H_t - H_{t-1} = (X_t - X_{t-1})B + X_{t,t-1}\tilde{B} + \alpha_{UP,t}UP_t^* + \alpha_{DOWN,t}DOWN_t^* + \bar{\mu}s_{t,t-1} + u_t,
\]

where \( H_t \) is a type of health behavior for an individual \( i \) in year \( t \). So the dependent variable is the change in the health-related behavior. The basic regression controls for a vector of variables for observable individual characteristics \( X_t \), including typical demographic characteristics, work status and region of residence. \( s_{t(t-1)} \) and \( s_t \) stand for self-reported survival probabilities in period \( t-1 \) and \( t \). \( X_{t,t-1} \) and \( s_{t,t-1} \) represents the initial characteristics and survival beliefs that we controlled for. \( u_t \) indicates the error term for the estimation at time \( t \) of the change from \( t-1 \) to \( t \).

In order to test the implication of the Cutler-Glaeser model more easily, I divided the sample of nonsmokers in \( t-1 \) into three groups based on changes in survival beliefs over two consecutive waves: (a) those who raised their beliefs \( (UP_t^* , a binary variable equal to 1 if expectations increased) \); (b) those who reduced beliefs \( (DOWN_t^*) \); and (c) those whose
beliefs did not change. In order to reduce the effect of measurement error, we defined those whose changes in expected survival probabilities are less than 10 percentage points as the group with unchanged survival beliefs. In other words, if the differential in survival beliefs reported by an individual across two waves is less than 10%, then this respondent is defined as belonging to the “unchanged beliefs” group. The other two groups consisted of people whose beliefs increase by 10 percentage points, and those whose beliefs decrease by 10 percentage points, respectively.

Then, assuming the Cutler-Glaeser model to be valid, if we are to take reverse causality into account, in estimating equation (14) we would see that for initial nonsmokers \((H_{i,t-1}=0)\), \(\alpha_{UP}\) should be much smaller than \(\alpha_{DOWN}\). That is, the model predicts that \(\alpha_{UP} < \alpha_{DOWN}\). This is the hypothesis that the regressions are going to test.

We start with two consecutive waves rather than checking changes across multiple waves, mainly because the latter would reduce the sample size because our main estimation method is pooled cross-sectional estimation. Because this study is not designed to provide an accurate estimation of the impact of survival expectations on health behaviors, the possible unobserved heterogeneity problem in estimating equation (14) should not be considered a major threat to the results described herein.

---

3 The results are robust if using alternative cutoff values such as 5, 15 or 20 percentage points.
4 We also checked longer time intervals than two consecutive waves. See details in Section 5.
5 A typical method used in previous literature to address unobservable heterogeneity problem in this context is the instrumental variable (IV) method, which utilizes the death of a family member or relative as the instrumental variable for survival beliefs. For instance, Biró (2011) use the recent death of a sibling as the instrumental variable for the European data; similarly, Liu and Rettenmaier (2007) use the death of a same-sex parent. In contrast, in the HRS data the survival expectations are barely correlated with the recent death of a sibling, although recent parental deaths are shown to directly affect health behaviors. If we want to address unobservable heterogeneity, the lack of good instruments in the data means that we have to rely on panel estimations to cancel out time invariant of unobservable effects. By focusing on changes over two waves, we can cancel out time invariant effects. In addition, there are a few proxies available in the data for time preference and risk preference (see details in Subsection 4.3). We control for them in some specifications.
Another general concern that must be addressed is that economists have long been dubious about the validity of expectations data. This issue will be discussed in Subsection 4.1.

4. Data

The data used in this study are from 10 waves (1992-2010) of the Health and Retirement Study (HRS) data. The survey is repeated every two years. Initial interviews, conducted in 1992 and 1993, provide detailed information on the health status and socioeconomic status of a nationally representative sample of non-institutionalized persons born between 1931 and 1941 and their spouses. A total of 12,652 individuals were included in the HRS sample in 1992. Variables of particular importance for this paper include expectations of survival to age 75 and age 85, and various health behaviors, as well as indicators of the characteristics of the respondents. The analysis is robust to alternative methods of defining the measures that will be presented in this section. A detailed sensitivity analysis is in subsequent Subsection 4.2.

4.1 Variables for Survival Beliefs

An essential variable for this study is the measure of survival expectations. From Wave 1 to Wave 10, respondents are asked if they expect to survive to the age of 75 (“P75”). Additionally, there are two other measures: survival probability to the age of 85 (denoted as “P85,” which is asked from Wave 1 to Wave 4), as well as the probability of surviving at

---

6 An alternative data set is the 1997 National Longitudinal Survey of Youth (NLSY). This paper utilizes HRS data rather than NLSY97 data because youth smoking is correlated with various other factors, such as the legal age of smoking. In spite of this shortcoming, I assessed the NLSY97 data set as a supplement to this study. In 1997 and 2002, respondents aged 15-18 (in 1997) are asked the question “What is the percent chance that you will die from any cause — crime, illness, accident, and so on, in the next year?” By considering subjective death probability to be the value of subtracting the survival probability from 1, I find results qualitatively similar to what is reported in this paper. Specifically, of those with increased survival probabilities, 28.88% start smoking. The share of new smokers with steady or increased death probabilities is 32.65% and 35.39%, respectively.
least 10 more years (denoted as “P10”, which is asked from Wave 2 to Wave 10 but unavailable for Wave 4). For this study, probability to age 75 is used as the primary measure because it is available for each wave. However, the other two measures are examined as well.

Figures 1A and 1B illustrate the distributions of the reported expectations associated with the first wave. Respondents are requested to report an integer probability from 0 to 100. The modes are 0, 50 and 100 percentage points. In the first wave, the average survival probability to age 75 is 64.4% (s.d.= 29.33) and for age 85 the corresponding percentage is 42.7% (s.d.= 31.72). Although economists have expressed concerns about introducing subjective data into a survival probability choice model due to problems such as lack of sufficient understanding of the interview questions, selecting answers due to social desirability, and the existence of rounding problems, extensive literature supports the validity of subjective expectations data (e.g. Zafar, 2011). Nevertheless, the validity of the elicited subjective longevity data might still be in question for the following two reasons: (1) adjusting for different target ages; and (2) consistency of beliefs between different target ages.

Footnotes:

7 The question about survival beliefs in the HRS survey is the following: “I would like to ask you about the chance that various events will happen in the future. Using any number from 0 to 10, where zero equals absolutely no chance and 10 equals absolutely certain, what do you think are the chances that you will live to be 75 or more?” The second measure is: “And how about the chances that you will live to be 85 or more?” To accommodate older respondents, starting from Wave 4 this question has an alternative version, which was, depending on the age of the respondents: “The probability of living up to an age from 80 to 100.” The question asked in 1992 is unique to that baseline year: “Using any number from 0 to 10, where 0 equals absolutely no chance and 10 equals absolutely certain, what do you think are the chances that you will live to be 75 (85) or more?” The first two waves are given scales from 0 to 10; HRS recoded them into 0-100% to be comparable with other waves. A sizable literature such as Smith et al. (2001) have used this measure. From Wave 3 to Wave 10 the probability have the range from 0-100%.

8 There are 2.57% respondents (302 out of 11,758 observations) whose reported probabilities of surviving to age 75 are smaller than expectations of survival to age 85. Some studies indicate that these respondents lack a sufficient understanding of the questions and therefore drop them from their analysis. I keep them in the sample instead, but the analysis of this study remains robust even when dropping them.
One way to adjust for different targeted ages is described by Gan et al. (2003) and Salm (2010)—namely, by calculating a constructed survival hazard rate. The subjective survival probability of individual $i$ can be calculated using the following method (equations 15 and 16). I use self-reported survival probability data until age 75 as an example, but the calculation method can also be applied to age 85, or to the probability of surviving up to the next ten years:

\[
(s_{it}^{t+2})^{\frac{1}{2}}(75-a) = s_{it}^{75};
\]

(10)

\[
\frac{\ln s_{it}^{t+2}}{\ln s_{it}} = \exp(\ln s_{it}^{t+2} - \ln s_{it}^{t+2}).
\]

(11)

We used $t+2$ because the data are available every two years. Let $s_{it}^{75}$ be the probability that respondents believe they will be alive up to age 75, 85, or up to the next ten years. For simplicity, we assumed that the calculated survival probability is constant through all the years between the interview date and the target age or years. Let $s_{it}^{t+2}$ denote survival probabilities based on U.S. longevity data for the same age-gender reference groups.\(^{9}\)

\[
\frac{\ln s_{it}^{t+2}}{\ln s_{it}}
\]

denotes the extent of pessimism an individual has about his/her survival chances. $s_{it}^{t+2}$ stands for the probability of surviving until the next wave. Instead of using the self-reported probabilities $s_{it}^{75}$, using the survival probabilities up to the next wave, $s_{it}^{t+2}$, gives qualitatively equivalent results.\(^{10}\)

---

\(^{9}\) In the HRS users’ guide, the interpretation of the survival probabilities based on the life tables is as follows: “The life tables used are the annual life tables for the year of the wave…. As the 2008 life table is not yet available, the 2006 table is used for wave 9 and wave 10. For Waves 3A and 3H, an alternate set of life tables, similar to those from the NCHS web page but providing estimates beyond age 85, are taken from the Berkeley Mortality Database. The probability of survival is calculated as the number surviving at age 75 divided by the number surviving to R’s age, for males or females, as appropriate.”

\(^{10}\) Hence, for simplification, in this paper I just report survival beliefs.
Consistency concerns regarding the validity of the survival expectations data are associated with the fact that for some respondents, survival expectations until age 75 are smaller or equal to those until age 85. Specifically, in the first four waves when pooled cross-sectional data (i.e., for both P75 and P85) are available, there are 2.3% of respondents (that is, 839 out of 37,027 person-year observations) for whom the value of P75 is lower than P85; additionally, 27.1% of respondents answered the same probability for P75 and P85. Since respondents had to be 50 years or older when the HRS was first given in 1992, most individuals still had 20-30 years ahead of them before reaching the ages of 75 or 85. Thus, the equivalence of elicited expectations to age 75 and age 85 may simply have reflected that respondents are unclear about future events; they misunderstood the question; or they are unable to form internally consistent probabilities of survival. Including them in the sample, however, does not change the basic results, which is why their responses are incorporated in the analysis.

4.2 Health behavior variables

Some important behavioral choices that are known contributors to severe and/or chronic health problems include smoking, heavy drinking, overeating/obesity and physical inactivity. For simplification, they are coded as binary variables.

**Smoking.** According to the Centers for Disease Control and Prevention, “Tobacco use is the single most preventable cause of disease, disability, and death in the United States.” In each wave, the HRS asked respondents if they had ever smoked or currently smoked. “Smoking,” according to the survey, referred to the consumption of more than 100

---

cigarettes in a respondent's lifetime, excluding pipes and cigars. This study primarily focused on the dichotomous indicator for current smoking.

**Heavy drinking.** The HRS asked respondents about the number of drinks they had consumed on the days they drank within the prior three-month period. Heavy drinking is defined as three or more drinks per day for both men and women, which is the same definition used by Cutler and Glaeser (2005).

**Obesity.** Height and weight are included as self-reported data for each of the ten waves, which are then converted to a body mass index (BMI) number. BMI is calculated as weight in kilograms divided by height in squared meters. Obesity is defined by the National Institutes of Health as having a BMI of 30 kg/m² or greater.

**Inactivity.** Physical activity has been shown to be an important factor in keeping morbidity and mortality at bay, and most Americans do not engage in sufficient amounts of physical activity (Pratt et al., 1999; Danaei et al., 2009). From Wave 1 to Wave 6, respondents were asked if they engage in any vigorous activities three or more times per week. Beginning in Wave 7, the HRS redefined the question and asked respondents to categorize their physical activity according to vigorous, moderate or light physical activity. For this study, we focus on vigorous activities, including sports, heavy housework, or a job that involved physical labor. We use two variables as measures. For data obtained from the first six waves, inactivity is coded as 1 if an individual’s frequency of vigorous activity is less than three times per week. Starting with Wave 7, inactivity is coded as 1 if the respondents do not engage in any vigorous activities at all.

### 4.3 Other Explanatory Variables

In this study, we control for the following demographic variables: age, gender, race, and education, census region of residence, marital status, number of children, total real
household income, household financial wealth, household size, and work status. Moreover, because death of family members might possibly impact a person’s expectations of survival, we also control for the death of a sibling(s). We also included an indicator for the death of the same-gender parent in three age categories: (a) younger than 75, (b) 75-84, and (c) older than 85.

Current health status is important for both a person’s health behaviors, as well as for his or her longevity beliefs. In fact, Smith, et al. (2001) argue that a change in personal health status represents the key factor for a change in an individual’s survival beliefs. In order to measure change of health status for this research, I utilize the total number of physician-diagnosed diseases to assess this variable, including high blood pressure, diabetes, cancer, chronic lung disease (i.e., chronic bronchitis or emphysema, but not asthma), heart problems, stroke, psychiatric problems and arthritis. Clark and Etilé, 2002, 2006) report that people smoke less or quit smoking when they know more clearly about the consequences of smoking by observing the deterioration in their own health caused by smoking, or by observing the deterioration of their spouse or other household members who smoke. Similarly, Cutler et al. (2010) find that someone who is diagnosed with cancer may quit smoking.

Furthermore, I control for measures of self-reported overall health. Self-reported health has been well documented as having important implications for actual physical and mental health outcomes (Benítez-Silva and Dwyer, 2003; Hill et al., 2011). The HRS provided the following possible answers for that question: (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor. I recode this self-assessed health status variable as a binary variable of 0 or 1: 0 for excellent, very good, and good, or 1 for fair or poor.

12 Income and financial wealth are adjusted for personal consumption expenditures (chain-type price index) and represent real 1992 prices.
I also examine two measures of mental health. The first is associated with the Center for Epidemiologic Studies Depression Scale scores (CESD). HRS respondents are asked whether they agreed or disagreed with eight statements about their mental status during the week prior to completing the survey, such as whether they felt depressed much of the time. CESD scores range from 0 (good mental health) to 8 (bad mental health). I recode the CESD scores as three categorical variables: (a) good mental health with a CESD score equalizing 0 (reference group); (b) having one mental problem (1); and (c) having two to eight mental problems. The second measure of mental health is represented by a binary variable indicating whether there is any incidence of a doctor-diagnosed mental health condition between interview waves.

In some specifications, I control for two sets of variables by considering them as proxies for unobservable characteristics of time preference for individuals. First, a measure for risk preference is also included in the HRS survey, which is based on a set of “income gamble” questions. Again, I use a binary variable of either 1 or 0: 1 to indicate individuals who had the tendency to avoid high risks; and 0 for everyone else. Second, it should be noted that for Waves 4 to 8, a measure indicating a “length of time horizon” for formulating a financial investment plan is available. In general, however, three categories are defined:

---

13 The Center for Epidemiologic Studies Depression Scale (CESD), developed by Lenore Radloff, is one of the most common methods for allowing an individual to determine his or her depression quotient.
14 Using different ways to code the CESD measure afford qualitatively similar results.
15 The HRS risk preference question is as follows: “Suppose that you are the only income earner in the family, and you have a good job guaranteed to give you your current (family) income every year for life. You are given the opportunity to take a new and equally good job, with a 50-50 chance it will double your (family) income and a 50-50 chance that it will cut your (family) income by a third. Would you take the new job? If yes, then: Suppose the chances were 50-50 that it would double your (family) income, and 50-50 that it would cut it in half. Would you still take the new job? If no, then: Suppose the chances were 50-50 that it would double your (family) income and 50-50 that it would cut it by 20 percent. Would you then take the new job?” In Waves 2 and 3, these questions are not asked.
16 In the HRS survey, respondents are asked which option best described their financial planning horizon: (1) Next few months, (2) Next year, (3) Next few years, (4) Next 5-10 years, or (5) Longer than 10 years.
a short financial horizon if the answer is under 5 years (omitted group), the medium financial horizon (5-10 years); and long horizon (longer than 10 years).

Another important variable is knowledge or beliefs about the consequences of health behaviors. A significant body of literature has documented the link between level of education/general health knowledge and health behaviors (Cutler & Lleras-Muney, 2008; Altindag et al., 2011). Unfortunately, the HRS did not obtain such information and thus it could not be directly incorporated into this study. However, an approximation is that health knowledge is highly correlated with educational attainment; since we controlled for education, we partially controlled for this factor.

Finally, in addition to family income, this study also controlled for a number of other factors indicating personal or family characteristics, such as work status, health insurance coverage, region of residence, family size and marital status.17

4.4 Description of the Estimation Sample

Important to note is that we exclude any observations with missing values for any of the variables mentioned above. To exploit the panel nature of the data, we restrict the sample to individuals for whom at least two waves of data are available.18 The description of the sample is reported in Table 1. As shown, approximately 90.3% are white, 62.3% are female, 26.4% had bachelor’s degrees, and on average the respondents have completed 13.3 years of schooling.19 The average age is 58.3 years.

Table 2 describes the frequency distributions of health behaviors and the changes in reported survival expectations over two consecutive waves. As shown in columns from left

---

17 In the Cutler-Glaeser model, cigarette prices are included; however, no information is available on residence states or counties. Since cigarette prices vary by location, it is not possible to accurately assess changes in cigarette prices.

18 In the data, only a small number of respondents died between waves (around 2% of the sample).

19 The demographic characteristics of our sample are not similar to the national representative figures for the total population because HRS data contains a senior population (ages 51 and 61 in 1992 and their spouses).
to right, the frequency distributions are divided according to the following survival expectations groups: a decrease in probability of more than 10 percentage points (left), a change (either direction, up or down) in probability of less than 10 percentage points (middle), and an increase in probability of more than 10 percentage points (right). The rows indicate the status of health behaviors.

It is immediately apparent that these frequencies are inconsistent with the implications of the Cutler-Glaeser model—namely, that in a given population the share of initial nonsmokers who express higher survival probabilities will be lower in comparison to individuals with unchanged or decreased survival expectations. This table also confirms that 2.13% of those with elevated subjective longevity beliefs started smoking, while a slightly higher percentage (2.19%) whose longevity expectations decreased also started smoking. Additionally, the percentage of former nonsmokers with steady longevity expectations who started to smoke is also similar – about 1.99%. In summary, there is no appreciable difference among former non-smokers who started to smoke based on changes (or lack of changes) in their survival expectations—each group is approximately 2%. This pattern seems to be true for other health behaviors as well. Nevertheless, I acknowledge that using just summary statistics alone does not fully explain the link between changes in survival beliefs and changes in health behaviors. A regression controlling for observable individual characteristics (detailed in Section 5), is also essential for ruling out the role that other factors might play.

Table 3 illustrates the mean and standard deviations for the self-reported probabilities, as well as their changes according to health behaviors. Panel (A) represents those who did not engage in unhealthy behaviors in either wave; Panel (B) signifies participants who began unhealthy behaviors during the second wave; and Panel (C)
represents the total. We can see that, in general, for those who start unhealthy behaviors during the second wave, the percentage of those who indicate increased survival beliefs is higher than individuals who report decreased probabilities. For instance, among people who report increased beliefs of survival, those who start smoking is 34.5%; this fraction contrasts markedly with the 23.6% of people who indicate reduced survival expectations and start smoking (0.236 in Table 3). This pattern differs from what the Cutler-Glaeser model implies—that the share of new smokers with increased survival beliefs will be lower than new smokers with reduced survival expectations.

5. Results

This section documents and discusses the baseline estimation results, followed by a review of the results from the robustness checks.

5.1 Regression Results

Table 4 displays the results of the pooled probit regressions on health behaviors. In each regression, the dependent variable represents a binary variable for an unhealthy behavior, which equals to 1 if the unhealthy behavior was initiated during time t. Those who engaged in the unhealthy behavior during the previous wave were dropped. The survival probabilities are indexed by three binary variables: (a) an increase of more than 10 percentage points; (b) a change of smaller than 10 (omitted group); and (c) a decrease of more than 10. For easier interpretation, the marginal effects after probit estimation are included (where the coefficients indicate any changes in the probability from a per-unit change in the relevant variable).

20 Recall that the Cutler-Glaeser model has unambiguous implications for initial nonsmokers only. This is why this study focused on them and dropped the initial smokers.
Table 4 presents the results when we controlled for a comprehensive set of variables, including age, sex, race, parental education, household income, household size, marital status, children, work status, region of residence, mental health, difficulty in activities of daily living, health insurance, recent death of family members over the two waves, and wave dummies. Because we were targeting the changes over two consecutive survey waves, we used the lagged measures in Wave t-1 for the control variables, and then documented any changes in the variables over the two waves. Additionally, as shown in column (2), the following additional factors were also included: self-reported health status, doctor-diagnosed health problems, and a mental health measure. In assessing other possible factors influencing the relationship between survival probabilities and health behaviors, we also considered the crucial importance of “health shocks,” which for this study are defined as serious illnesses or accidents. Hence, controlling for health shocks might result in an underestimation of the effects of survival probabilities on health behaviors. An important reason for controlling for the influence of this factor is associated with fact that health shocks may affect health behaviors in other ways (besides through changes in lifespan expectations). While this possibility is acknowledged, we would argue that the likely existence of other mechanisms is small. Column (3) shows the results when we controlled for diagnosed diseases of a spouse, based on the idea that a respondent may have acquired additional health awareness if a spouse was diagnosed with certain diseases. Finally, as shown in column (4), length of financial planning horizons and time preference indices were also added to partially control for unobservable heterogeneity. The standard errors are clustered by individuals.

As shown in Table 4, the coefficients do not change to any significant degree after adding these extra controls. In all specifications, we used the Wald test for comparing the
coefficient associated with the binary variable for having increased survival probability (P75) with the variable for “decreases in P75;” results demonstrated that they were not statistically significantly different (p-value greater than 0.3). Therefore, the regression results suggest that the likelihood of switching to smoking for those with increased survival probabilities was not significantly different from those with steady or reduced probabilities.

Among other control variables, recent marriage since last wave, higher family income and college education were associated with a lower tendency to begin unhealthy behaviors. Conversely, veteran status (being a veteran), poor self-reported health status, mental health problems and having certain diagnosed diseases were linked with a higher probability of starting risky behaviors. Work status, financial horizon and risk preference generally did not have any significant effects on the health behaviors of respondents.

To summarize, the results shown in Table 4 confirm that among initial nonsmokers, the likelihood of starting to smoke among people with increased survival beliefs is not significantly different from people with unchanged or decreased prospective survival beliefs. Compared with people whose survival probabilities change little over two waves, this finding contradicts what the Cutler-Glaeser model suggests.

5.2 Robustness Checks

The following discussion provides an analysis of whether our regression results were sensitive to the sample definition, as well as alternative measures and estimation methods.

A). Different cutoff values. As noted in the above analysis, the cutoff value for dividing the sample based on changes in survival probability is 10 percentage points. We also experiment with five alternative cutoff values: 0, 5, 10, 20 or 30 percentage points. Because the results are not significantly different for any of these values, for simplification
the results associated with using 20 percentage points are shown in Table 5.1. The basic pattern is similar to the above analysis.

B). **Alternative measures.** Table 5.2 reports the calculated survival probability until the next wave (every two years), rather than using the reported probability of living until age 75. As shown, there was no significant difference between the coefficients for the group with increased calculated survival probabilities compared to the group with decreased probabilities.

C) **Panel probit estimation.** The results of the random effects probit estimation are documented in Table 5.3. We expect that when using panel data estimations, the unobservable heterogeneity outcomes would be further reduced. In contrast, however, the resulting coefficients displayed larger absolute values compared with those in Table 4. In other words, people who reported increased survival expectations were less likely to smoke compared with those who indicated steady longevity expectations. Nonetheless, the qualitative results are equivalent to what we observed from the pooled estimation data.

D) **Different time span.** In the above analysis, we used two consecutive waves. We also consider whether using longer time intervals—say, across 4, 6, 8 or 10 years—would change the outcome if people have more time to change their health behaviors. One motivation of experimenting with longer time span is to reduce the role of measurement error in estimation. Interestingly, the patterns are consistent across the other time intervals as well. For convenience, the results for across 8 years (i.e., 4 waves) are detailed in Table 5.4 but our conclusion also holds for other intervals between the two time periods examined by this paper.

Overall, we find that an increase or decrease (or no change) in reported survival probability is essentially irrelevant to whether a former nonsmoker began to smoke or not.
As shown in the above discussion, this conclusion is robust to different cutoff values, alternative measures for assessing subjective survival probability, and different estimation methods.

6. Discussion

The question that begs an answer, therefore, is why—in opposition with the theoretical outcomes associated with the Cutler-Glaeser model—a small percentage of former nonsmokers (~2%) in the HRS survey start to smoke when their self-reported survival probabilities increased? One obvious possibility is that there may be something inherently wrong with the Cutler-Glaeser model, which is beyond the scope of this study. But there are also at least three other options: A) the expectation measures may be problematic; B) people’s tastes for unhealthy behaviors (indicated by the utility function U(.), or the extra utility afforded by smoking $R$, in the Cutler-Glaeser model) might change over time for reasons that cannot be easily assessed; and C), beliefs about the consequences of unhealthy behaviors (represented by perceived loss of longevity by smoking $\tau$ ) may have changed across two waves. These three possibilities are discussed below.

A) Expectations measures. One possible interpretation is that the expectations measures may not accurately represent the beliefs that respondents actually used when they made their smoking decisions during each time period. However, our examination of the validity of this subjective measure (Section 3) does not really raise any red flags with respect to its validity. If this were the case, then our findings would suggest that we needed to consider these subjective measures as actual objective beliefs of future events that individuals use when they make economic decisions. Unfortunately, we lack reliable methods for further testing this possibility.
B) Changes in beliefs about health effects. Another hypothesis is that the new smokers who indicate increased survival beliefs represent people whose beliefs about the consequences of unhealthy behaviors have changed over time for a variety of unknown reasons not assessed herein. Although we control for potentially influential variables such as diagnosed diseases of their own or their spouses (both initial status and changes), it is still possible that these people change their beliefs about behavioral consequences through different mechanisms, such as watching health-related TV programs or reading health-based articles online or in newspapers.

Unlike some other data sets that solicit targeted information about the consequences of smoking, such as data from the Survey on Smoking (SOS), the HRS data did not generate analogous information concerning self-reported beliefs about the consequences of smoking or other unhealthy behaviors. Therefore, changes in subjective health beliefs over time could not be included as a variable.

C) Changes in taste. Another possibility is that those new smokers who reported increased survival beliefs simply changed their tastes toward smoking across two waves for unknown reasons. Consider, for example, that the utility function changes over time. However, such a proposition is contrary of the standard economic assumption of constant tastes.

7. Conclusion

This paper examines the link between survival beliefs and health behaviors to test the following implication of the Cutler-Glaeser model: at the individual level, a nonsmoker who reports a higher survival probability is unlikely to start smoking; at the aggregate level in a population, the share of new smokers with increased survival expectations will be lower in comparison to individuals with unchanged or decreased survival expectations. The
main finding describes herein is that a certain share of former nonsmokers (2.13%) who raised their survival expectations began to smoking—and that this percentage was similar to new smokers with unchanged (1.99%) or decreased (2.19%) survival beliefs. This conclusion contradicts what the Cutler-Glaeser model predicts. We also find similar patterns for other health-related behaviors, including heavy drinking, overeating/obesity and physical inactivity.

However, certain limitations must be noted. First is the possible influence of other theoretical models. In addition to the health investment framework, there may be other theories (e.g., behavioral economics theory) that can be used to account for the link between survival expectations and health behaviors. Second, the health behaviors that were investigated herein are based on self-reported data, rather than on third-party reported data (with the exception of the BMI measurement). Thus, to the extent that bias in self-reporting varies across behaviors, the use of multiple health behaviors was expected to mitigate this bias. In conclusion, although this paper is believed to represent a viable exploration of the association between self-reported longevity expectations and health behaviors, future research should include the influence of other mechanisms or theoretical models not assessed herein.
References


Figures and Tables

Figure 1: Distribution of survival expectations in Wave 1

Notes: HRS 1992-2010 data. Panel A includes all racial groups, whites, blacks, Hispanics and other races.
Table 1: Sample Statistical Description (Pooled Cross-Sectional HRS data)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>s.d.</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current smoker</td>
<td>0.018</td>
<td>0.134</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Heavy drinker(&gt;3 drinks)</td>
<td>0.060</td>
<td>0.238</td>
<td>0</td>
<td>1</td>
<td>15390</td>
</tr>
<tr>
<td>Obesity</td>
<td>0.286</td>
<td>0.452</td>
<td>0</td>
<td>1</td>
<td>15177</td>
</tr>
<tr>
<td>Physical inactivity</td>
<td>0.548</td>
<td>0.498</td>
<td>0</td>
<td>1</td>
<td>15002</td>
</tr>
<tr>
<td>Age</td>
<td>58.322</td>
<td>3.308</td>
<td>30</td>
<td>66</td>
<td>15421</td>
</tr>
<tr>
<td>Black</td>
<td>0.063</td>
<td>0.243</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.011</td>
<td>0.104</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Other races</td>
<td>0.023</td>
<td>0.150</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Not born in U.S.</td>
<td>0.071</td>
<td>0.258</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Female</td>
<td>0.623</td>
<td>0.485</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Veteran</td>
<td>0.215</td>
<td>0.411</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>High school graduate or GED</td>
<td>0.390</td>
<td>0.488</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Some college</td>
<td>0.241</td>
<td>0.428</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>College and above</td>
<td>0.264</td>
<td>0.441</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Mother’s education (years)</td>
<td>10.287</td>
<td>3.178</td>
<td>0</td>
<td>17</td>
<td>15421</td>
</tr>
<tr>
<td>Father’s education (years)</td>
<td>10.025</td>
<td>3.653</td>
<td>0</td>
<td>17</td>
<td>15421</td>
</tr>
<tr>
<td>No children (respondent and spouse)</td>
<td>0.065</td>
<td>0.247</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Lagged + northeast</td>
<td>0.198</td>
<td>0.399</td>
<td>0</td>
<td>1</td>
<td>15418</td>
</tr>
<tr>
<td>Lagged midwest</td>
<td>0.263</td>
<td>0.440</td>
<td>0</td>
<td>1</td>
<td>15418</td>
</tr>
<tr>
<td>Lagged south</td>
<td>0.328</td>
<td>0.469</td>
<td>0</td>
<td>1</td>
<td>15418</td>
</tr>
<tr>
<td>Lagged real earnings (1992$)</td>
<td>15594</td>
<td>36114</td>
<td>0</td>
<td>3.95e+6</td>
<td>15421</td>
</tr>
<tr>
<td>Lagged real total non-house financial assets(1992$)</td>
<td>118474</td>
<td>716390</td>
<td>-1981053</td>
<td>9.66e+7</td>
<td>14097</td>
</tr>
<tr>
<td>Lagged married</td>
<td>0.789</td>
<td>0.408</td>
<td>0</td>
<td>1</td>
<td>15402</td>
</tr>
<tr>
<td>Lagged work for pay</td>
<td>0.634</td>
<td>0.482</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Lagged difficulty of activity</td>
<td>0.071</td>
<td>0.292</td>
<td>0</td>
<td>5</td>
<td>15421</td>
</tr>
<tr>
<td>Lagged CESD 1 problem</td>
<td>0.211</td>
<td>0.408</td>
<td>0</td>
<td>1</td>
<td>14200</td>
</tr>
<tr>
<td>Lagged CESD 2-8 problems</td>
<td>0.230</td>
<td>0.421</td>
<td>0</td>
<td>1</td>
<td>14200</td>
</tr>
<tr>
<td>New $^k$ northeast</td>
<td>0.001</td>
<td>0.034</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>New midwest</td>
<td>0.002</td>
<td>0.045</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>New south</td>
<td>0.008</td>
<td>0.092</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Real earnings change (1992$)</td>
<td>15032</td>
<td>36057</td>
<td>0</td>
<td>3.95e+6</td>
<td>15421</td>
</tr>
<tr>
<td>Real total non-house financial assets(1992$) change</td>
<td>123523</td>
<td>708083</td>
<td>-1981053</td>
<td>9.66e+7</td>
<td>14097</td>
</tr>
<tr>
<td>New married</td>
<td>0.009</td>
<td>0.097</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>New work for pay</td>
<td>0.042</td>
<td>0.200</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>New difficulty of activity</td>
<td>0.039</td>
<td>0.194</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>New mental health CESD 1 problem</td>
<td>0.152</td>
<td>0.359</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>New mental health CESD 2-8 problems</td>
<td>0.122</td>
<td>0.328</td>
<td>0</td>
<td>1</td>
<td>15421</td>
</tr>
<tr>
<td>Lagged self-reported bad health</td>
<td>0.139</td>
<td>0.346</td>
<td>0</td>
<td>1</td>
<td>15418</td>
</tr>
<tr>
<td>Self-reported health better</td>
<td>0.118</td>
<td>0.323</td>
<td>0</td>
<td>1</td>
<td>15409</td>
</tr>
<tr>
<td>Self-reported health worse</td>
<td>0.221</td>
<td>0.415</td>
<td>0</td>
<td>1</td>
<td>15409</td>
</tr>
<tr>
<td>Lagged high blood pressure ever</td>
<td>0.363</td>
<td>0.481</td>
<td>0</td>
<td>1</td>
<td>12152</td>
</tr>
<tr>
<td>Lagged diabetes ever</td>
<td>0.099</td>
<td>0.298</td>
<td>0</td>
<td>1</td>
<td>12150</td>
</tr>
<tr>
<td>Lagged cancer ever</td>
<td>0.077</td>
<td>0.267</td>
<td>0</td>
<td>1</td>
<td>12149</td>
</tr>
<tr>
<td>Lagged lung problems ever</td>
<td>0.051</td>
<td>0.220</td>
<td>0</td>
<td>1</td>
<td>12154</td>
</tr>
<tr>
<td>Lagged heart problems ever</td>
<td>0.143</td>
<td>0.350</td>
<td>0</td>
<td>1</td>
<td>12158</td>
</tr>
<tr>
<td>Lagged stroke ever</td>
<td>0.030</td>
<td>0.171</td>
<td>0</td>
<td>1</td>
<td>12157</td>
</tr>
<tr>
<td>Lagged psychiatric problems ever</td>
<td>0.070</td>
<td>0.256</td>
<td>0</td>
<td>1</td>
<td>12153</td>
</tr>
<tr>
<td>Lagged arthritis ever</td>
<td>0.406</td>
<td>0.491</td>
<td>0</td>
<td>1</td>
<td>12143</td>
</tr>
<tr>
<td>New high blood pressure</td>
<td>0.041</td>
<td>0.199</td>
<td>0</td>
<td>1</td>
<td>15416</td>
</tr>
<tr>
<td>New diabetes since last interview</td>
<td>0.020</td>
<td>0.140</td>
<td>0</td>
<td>1</td>
<td>15414</td>
</tr>
<tr>
<td>New cancer since last interview</td>
<td>0.016</td>
<td>0.125</td>
<td>0</td>
<td>1</td>
<td>15406</td>
</tr>
<tr>
<td>New lung problems since last interview</td>
<td>0.006</td>
<td>0.078</td>
<td>0</td>
<td>1</td>
<td>15413</td>
</tr>
<tr>
<td>New heart problems since last interview</td>
<td>0.021</td>
<td>0.142</td>
<td>0</td>
<td>1</td>
<td>15415</td>
</tr>
<tr>
<td>New stroke since last interview</td>
<td>0.005</td>
<td>0.072</td>
<td>0</td>
<td>1</td>
<td>15420</td>
</tr>
<tr>
<td>New psychiatric problems since last interview</td>
<td>0.012</td>
<td>0.109</td>
<td>0</td>
<td>1</td>
<td>15416</td>
</tr>
<tr>
<td>New arthritis since last interview</td>
<td>0.049</td>
<td>0.216</td>
<td>0</td>
<td>1</td>
<td>15391</td>
</tr>
</tbody>
</table>

Notes: HRS 1992-2010 data. HRS personal weights are used. For each particular health-related behavior, only those who did not conduct that behavior in Wave t-1 are included. Income figures are in real 1992 dollars.

$^k$ “Lagged” measures indicate the measures in previous wave.

$^*$ “New” measures denote the changes over two consecutive waves.
Table 2: Frequency Distributions for Changes in Survival Expectations and Changes in Health Behaviors between Two Waves (HRS data)

<table>
<thead>
<tr>
<th>Change of Health behaviors</th>
<th>Changes in survival expectations between two waves</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Decrease&gt;=10 percentage</td>
<td>(2) Change&lt;10 percentage</td>
<td>(3) Increase&gt;=10 percentage</td>
<td>(4) Total</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>N</td>
<td>Column %</td>
<td>N</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not smoke t-1; not smoke t</td>
<td>97.81</td>
<td>9001</td>
<td>98.01</td>
<td>11338</td>
</tr>
<tr>
<td></td>
<td>(32.53)</td>
<td></td>
<td>(42.09)</td>
<td></td>
</tr>
<tr>
<td>Not smoke t-1; smoke t</td>
<td>2.19</td>
<td>200</td>
<td>1.99</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td>(34.16)</td>
<td></td>
<td>(39.99)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>9201</td>
<td>100</td>
<td>11584</td>
</tr>
<tr>
<td></td>
<td>(32.57)</td>
<td></td>
<td>(42.05)</td>
<td></td>
</tr>
<tr>
<td>Heavy drinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not drink t-1; not drink t</td>
<td>96.34</td>
<td>10937</td>
<td>96.48</td>
<td>13687</td>
</tr>
<tr>
<td></td>
<td>(32.56)</td>
<td></td>
<td>(41.46)</td>
<td></td>
</tr>
<tr>
<td>Not drink t-1; drink t</td>
<td>3.66</td>
<td>407</td>
<td>3.52</td>
<td>492</td>
</tr>
<tr>
<td></td>
<td>(32.81)</td>
<td></td>
<td>(40.14)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>11344</td>
<td>100</td>
<td>14179</td>
</tr>
<tr>
<td></td>
<td>(32.57)</td>
<td></td>
<td>(41.41)</td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not obese t-1; not obese t</td>
<td>93.76</td>
<td>8076</td>
<td>93.9</td>
<td>10295</td>
</tr>
<tr>
<td></td>
<td>(32.42)</td>
<td></td>
<td>(42.22)</td>
<td></td>
</tr>
<tr>
<td>Not obese t-1; obese t</td>
<td>6.24</td>
<td>563</td>
<td>6.1</td>
<td>720</td>
</tr>
<tr>
<td></td>
<td>(32.15)</td>
<td></td>
<td>(40.83)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>8639</td>
<td>100</td>
<td>11015</td>
</tr>
<tr>
<td></td>
<td>(32.40)</td>
<td></td>
<td>(42.13)</td>
<td></td>
</tr>
<tr>
<td>Inactivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not inactive t-1; not inactive t</td>
<td>62.11</td>
<td>238</td>
<td>66.02</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td>(28.18)</td>
<td></td>
<td>(48.37)</td>
<td></td>
</tr>
<tr>
<td>Not inactive t-1; inactive t</td>
<td>37.89</td>
<td>127</td>
<td>33.98</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>(32.17)</td>
<td></td>
<td>(46.59)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>365</td>
<td>100</td>
<td>540</td>
</tr>
<tr>
<td></td>
<td>(29.57)</td>
<td></td>
<td>(47.75)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: HRS 1992-2010 pooled cross-sectional data. N indicates the number of individuals multiplied by waves. The variation of N across health behaviors is due to the difference in available years for the health behaviors. Percentages in parentheses are row percentages.
Table 3: Statistical Description of Survival Expectations and Health Behaviors

<table>
<thead>
<tr>
<th>Smoking</th>
<th>(A) Not engaged in the unhealthy behavior at both t-1 and t</th>
<th>(B) Engaged in the unhealthy behavior at t, but not engage in the unhealthy behavior at t-1</th>
<th>(C) Total not engaged in the unhealthy behavior at t-1 (A)+(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>N</td>
</tr>
<tr>
<td>Reported probability to live 75+ (denoted by P75) %</td>
<td>70.237</td>
<td>25.598</td>
<td>15134</td>
</tr>
<tr>
<td>P75 change over two waves %</td>
<td>1.073</td>
<td>24.548</td>
<td>15134</td>
</tr>
<tr>
<td>Calculated survival probability until next wave (based on P75) %</td>
<td>1.844</td>
<td>0.282</td>
<td>14747</td>
</tr>
<tr>
<td>Change of calculated survival probability %</td>
<td>0.154</td>
<td>0.148</td>
<td>13349</td>
</tr>
<tr>
<td>P75 increase 10%+ (dummy)</td>
<td>0.324</td>
<td>0.468</td>
<td>15134</td>
</tr>
<tr>
<td>P75 change 10% (dummy)</td>
<td>0.439</td>
<td>0.496</td>
<td>15134</td>
</tr>
<tr>
<td>P75 decrease 10% (dummy)</td>
<td>0.237</td>
<td>0.425</td>
<td>15134</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Heavy drinking</th>
<th>(A) Not engaged in the unhealthy behavior at both t-1 and t</th>
<th>(B) Engaged in the unhealthy behavior at t, but not engage in the unhealthy behavior at t-1</th>
<th>(C) Total not engaged in the unhealthy behavior at t-1 (A)+(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>N</td>
</tr>
<tr>
<td>Reported probability to live 75+ (denoted by P75) %</td>
<td>70.191</td>
<td>25.782</td>
<td>14149</td>
</tr>
<tr>
<td>P75 change over two waves %</td>
<td>1.101</td>
<td>24.738</td>
<td>14149</td>
</tr>
<tr>
<td>Calculated survival probability until next wave (based on P75) %</td>
<td>1.841</td>
<td>0.282</td>
<td>13770</td>
</tr>
<tr>
<td>Change of calculated survival probability %</td>
<td>0.153</td>
<td>0.148</td>
<td>12406</td>
</tr>
<tr>
<td>P75 increase 10% (dummy)</td>
<td>0.323</td>
<td>0.468</td>
<td>14149</td>
</tr>
<tr>
<td>P75 change 10% (dummy)</td>
<td>0.440</td>
<td>0.496</td>
<td>14149</td>
</tr>
<tr>
<td>P75 decrease 10% (dummy)</td>
<td>0.237</td>
<td>0.425</td>
<td>14149</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obesity</th>
<th>(A) Not engaged in the unhealthy behavior at both t-1 and t</th>
<th>(B) Engaged in the unhealthy behavior at t, but not engage in the unhealthy behavior at t-1</th>
<th>(C) Total not engaged in the unhealthy behavior at t-1 (A)+(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>N</td>
</tr>
<tr>
<td>Reported probability to live 75+ (denoted by P75) %</td>
<td>71.342</td>
<td>24.912</td>
<td>10098</td>
</tr>
<tr>
<td>P75 change over two waves %</td>
<td>1.107</td>
<td>23.804</td>
<td>10098</td>
</tr>
<tr>
<td>Calculated survival probability until next wave (based on P75) %</td>
<td>1.847</td>
<td>0.280</td>
<td>9879</td>
</tr>
<tr>
<td>Change of calculated survival probability %</td>
<td>0.154</td>
<td>0.144</td>
<td>8912</td>
</tr>
<tr>
<td>P75 increase 10% (dummy)</td>
<td>0.321</td>
<td>0.467</td>
<td>10098</td>
</tr>
<tr>
<td>P75 change 10% (dummy)</td>
<td>0.451</td>
<td>0.498</td>
<td>10098</td>
</tr>
<tr>
<td>P75 decrease 10% (dummy)</td>
<td>0.229</td>
<td>0.420</td>
<td>10098</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inactivity</th>
<th>(A) Not engaged in the unhealthy behavior at both t-1 and t</th>
<th>(B) Engaged in the unhealthy behavior at t, but not engage in the unhealthy behavior at t-1</th>
<th>(C) Total not engaged in the unhealthy behavior at t-1 (A)+(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>N</td>
</tr>
<tr>
<td>Reported probability to live 75+ (denoted by P75) %</td>
<td>74.293</td>
<td>23.742</td>
<td>3673</td>
</tr>
<tr>
<td>P75 change over two waves %</td>
<td>1.010</td>
<td>23.235</td>
<td>3673</td>
</tr>
<tr>
<td>Calculated survival probability until next wave (based on P75) %</td>
<td>1.850</td>
<td>0.274</td>
<td>3621</td>
</tr>
<tr>
<td>Change of calculated survival probability %</td>
<td>0.149</td>
<td>0.134</td>
<td>3466</td>
</tr>
<tr>
<td>P75 increase 10% (dummy)</td>
<td>0.314</td>
<td>0.464</td>
<td>3673</td>
</tr>
<tr>
<td>P75 change 10% (dummy)</td>
<td>0.454</td>
<td>0.498</td>
<td>3673</td>
</tr>
<tr>
<td>P75 decrease 10% (dummy)</td>
<td>0.232</td>
<td>0.422</td>
<td>3673</td>
</tr>
</tbody>
</table>

Notes: HRS 1992-2010 data. Sample only include those who do not engage the unhealthy behavior in Wave t-1. Income figures are in real 1992 dollars.
### Table 4: Pooled Probit Estimation for Health Behaviors

<table>
<thead>
<tr>
<th>Health Behavior</th>
<th>Pooled cross-sectional estimation</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dependent variable: risky health behavior in t</td>
<td>(1) Control for basic variables</td>
<td>(2) Own health status added</td>
<td>(3) Health status of spouse added</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference group: self-reported probability (P75) change&lt;10%</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.007</td>
</tr>
<tr>
<td>P75 increase 10%+</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td>P75 decrease 10%+</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>N</td>
<td>14,591</td>
<td>14,519</td>
<td>10,985</td>
<td>2,500</td>
</tr>
<tr>
<td><strong>Heavy drinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P75 change&lt;10%</td>
<td>-0.002</td>
<td>-0.002</td>
<td>0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td>P75 increase 10%+</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>-0.004</td>
</tr>
<tr>
<td>P75 decrease 10%+</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>N</td>
<td>17,418</td>
<td>17,335</td>
<td>12,718</td>
<td>3,000</td>
</tr>
<tr>
<td><strong>Obesity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P75 change&lt;10%</td>
<td>0.003</td>
<td>0.002</td>
<td>0.003</td>
<td>-0.002</td>
</tr>
<tr>
<td>P75 increase 10%+</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.011)</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.019</td>
</tr>
<tr>
<td>P75 decrease 10%+</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>N</td>
<td>13,593</td>
<td>13,536</td>
<td>9,953</td>
<td>2,358</td>
</tr>
<tr>
<td><strong>Inactivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P75 change&lt;10%</td>
<td>0.017</td>
<td>0.035</td>
<td>0.018</td>
<td>-0.009</td>
</tr>
<tr>
<td>P75 increase 10%+</td>
<td>(0.047)</td>
<td>(0.048)</td>
<td>(0.054)</td>
<td>(0.070)</td>
</tr>
<tr>
<td></td>
<td>-0.036</td>
<td>-0.042</td>
<td>0.001</td>
<td>-0.166**</td>
</tr>
<tr>
<td>P75 decrease 10%+</td>
<td>(0.051)</td>
<td>(0.053)</td>
<td>(0.060)</td>
<td>(0.065)</td>
</tr>
<tr>
<td>N</td>
<td>634</td>
<td>624</td>
<td>503</td>
<td>327</td>
</tr>
</tbody>
</table>

**Notes:** HRS 1992-2010 data. Sample only includes those who do not engage the unhealthy behavior in Wave t-1. For each particular health-related behavior, only those who do not engage that behavior in Wave t-1 are included. The pooled cross-sectional probit model is estimated. Marginal effects are reported. Robust standard errors clustered by the individuals are in parentheses. HRS individual weights are used. Basic controls include age, gender, marital status, having children or not, residence region, life insurance, work status, family income, parental education, difficulty in activities, mental health measure, the death of a family member and wave dummies. The level measures in previous wave and the changes over two consecutive waves for all the control variables are included. In all specifications, the Wald tests for comparisons of the coefficients of “survival probability P75 increase” with those of “P75 decrease” indicate that there are no statistically significant differences between those two coefficients (p-value is greater than 0.3). **significant at the 5 percent level.
### Table 5.1 Robustness check: Pooled Probit Estimation on Health Behaviors

| Smoking | | | | |
| --- | --- | --- | --- | |
| | Control for basic variables | Own health status added | Health status of spouse added | Financial horizon and time preference added |
| | Dependent variable: risky health behavior in $t$ | | | |
| | (1) | (2) | (3) | (4) |
| P75 increase 20%+ | 0.001 | 0.001 | -0.000 | -0.001 |
| (0.003) | (0.003) | (0.002) | (0.006) |
| P75 decrease 20%+ | -0.002 | -0.002 | -0.003 | 0.000 |
| (0.003) | (0.003) | (0.003) | (0.006) |
| N | 15,687 | 15,610 | 11,826 | 2,728 |

| Heavy drinking | | | | |
| Reference group: self-reported probability (P75) change<20% | | | | |
| P75 increase 20%+ | -0.002 | -0.002 | -0.000 | 0.003 |
| (0.003) | (0.003) | (0.003) | (0.006) |
| P75 decrease 20%+ | 0.007* | 0.007* | 0.004 | 0.002 |
| (0.004) | (0.004) | (0.004) | (0.008) |
| N | 18,665 | 18,576 | 13,643 | 3,257 |

| Obesity | | | | |
| Reference group: P75 change<20% | | | | |
| P75 increase 20%+ | -0.004 | -0.004 | -0.007 | -0.017 |
| (0.005) | (0.005) | (0.006) | (0.010) |
| P75 decrease 20%+ | 0.001 | -0.000 | -0.003 | 0.005 |
| (0.005) | (0.005) | (0.006) | (0.012) |
| N | 14,630 | 14,569 | 10,725 | 2,571 |

| Inactivity | | | | |
| Reference group: P75 change<20% | | | | |
| P75 increase 20%+ | 0.069 | 0.072 | 0.078 | 0.058 |
| (0.054) | (0.056) | (0.065) | (0.081) |
| P75 decrease 20%+ | -0.084 | -0.101* | -0.056 | -0.188*** |
| (0.053) | (0.053) | (0.060) | (0.059) |
| N | 704 | 700 | 555 | 373 |

**Notes:** HRS 1992-2010 data. Sample only includes those who do not engage the unhealthy behavior in Wave t-1. The pooled cross-sectional probit model is estimated. Marginal effects are reported. Robust standard errors clustered by the individuals are in parentheses. HRS individual weights are used. Basic controls include age, gender, marital status, having children or not, residence regions, life insurance, work status, family income, parental education, difficulty in activities, mental health measure, the death of a family member and wave dummies. The level measures in previous wave and the changes over two consecutive waves for all the control variables are included. In all specifications, the Wald tests for comparisons of the coefficients of “survival probability P75 increase” with those of “P75 decrease” indicate that there are no statistically significant differences between those two coefficients (p-value is greater than 0.3).
***significant at the 1 percent level.

### Table 5.2 Robustness check: Pooled Probit Estimation on Health Behaviors

(Using self-Calculated survival probability based on P75 (denoted as Pt), rather than reported P75)

<table>
<thead>
<tr>
<th>Reference group: Calculated probability until next wave</th>
<th>Pooled cross-sectional estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt change&lt;20%</td>
<td>(1)</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
</tr>
<tr>
<td>Survival probability until next wave Pt increase 20%+</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>Survival probability until next wave Pt decrease 20%+</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>N</td>
<td>13,856</td>
</tr>
<tr>
<td>Heavy drinking</td>
<td></td>
</tr>
<tr>
<td>Survival probability until next wave Pt increase 20%+</td>
<td>-0.015*</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
</tr>
<tr>
<td>Survival probability until next wave Pt decrease 20%+</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>N</td>
<td>16,112</td>
</tr>
<tr>
<td>Obesity</td>
<td></td>
</tr>
<tr>
<td>Survival probability until next wave Pt increase 20%+</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
</tr>
<tr>
<td>Survival probability until next wave Pt decrease 20%+</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
</tr>
<tr>
<td>N</td>
<td>13,343</td>
</tr>
<tr>
<td>Inactivity</td>
<td></td>
</tr>
<tr>
<td>Survival probability until next wave Pt increase 20%+</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
</tr>
<tr>
<td>Survival probability till next wave Pt decrease 20%+</td>
<td>-0.065</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
</tr>
<tr>
<td>N</td>
<td>693</td>
</tr>
</tbody>
</table>

**Notes:** HRS 1992-2010 data. Sample only includes those who do not engage the unhealthy behavior in Wave t-1. The pooled cross-sectional probit model is estimated. Marginal effects are reported. Robust standard errors clustered by the individuals are in parentheses. HRS individual weights are used.

Basic controls include age, gender, marital status, having children or not, residence region, life insurance, work status, family income, parental education, difficulty in activities, mental health measure, the death of a family member and wave dummies. The level measures in previous wave and the changes over two consecutive waves for all the control variables are included.

In all specifications, the Wald tests for comparisons of the coefficients of “survival probability P75 increase” with those of “P75 decrease” indicate that there are no statistically significant differences between those two coefficients (p-value is greater than 0.3).

**significant at the 5 percent level,** significant at the 10 percent level.
### Table 5.3: Random Effect Probit Estimation on Health Behaviors

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control for basic variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own health status added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health status of spouse added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial horizon and time preference added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference group: self-reported probabilities (P75) change&lt;10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P75 increase 10%+</td>
<td>-0.026</td>
<td>-0.070</td>
<td>-0.089</td>
<td>-0.079</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.121)</td>
<td>(0.107)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>P75 decrease 10%+</td>
<td>0.042</td>
<td>-0.138</td>
<td>-0.136</td>
<td>-0.134</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.135)</td>
<td>(0.119)</td>
<td>(0.120)</td>
</tr>
<tr>
<td>N</td>
<td>14,591</td>
<td>14,519</td>
<td>10,985</td>
<td>2,500</td>
</tr>
<tr>
<td>id</td>
<td>5,731</td>
<td>5,711</td>
<td>3,613</td>
<td>843</td>
</tr>
<tr>
<td>rho(fraction of variance due to u_i)</td>
<td>0.813</td>
<td>0.873</td>
<td>0.816</td>
<td>0.810</td>
</tr>
</tbody>
</table>

**Notes:** HRS 1992-2010 data. Sample only include those who do not engage the unhealthy behavior in Wave t-1. The random effect probit model is estimated. Marginal effects are reported. Robust standard errors clustered by the individuals are in parentheses. HRS individual weights are used.

### Table 5.4: Pooled Probit Estimation on Health Behaviors between Longer Time Span

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control for basic variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own health status added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health status of spouse added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial horizon and time preference added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference group: self-reported probabilities (P75) change&lt;10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P75 increase 10%+</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>P75 decrease 10%+</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>N</td>
<td>2416</td>
<td>2324</td>
<td>2292</td>
<td>1556</td>
</tr>
</tbody>
</table>

**Notes:** HRS 1992-2010 data. Sample only include those who do not engage the unhealthy behavior in Wave t-1. The random effect probit model is estimated. Marginal effects are reported. Robust standard errors clustered by the individuals are in parentheses. HRS individual weights are used.
Chapter 3:

“The Bell Curve” Revisited: Does “Intelligence” Still Matter?

1. Introduction

In their book, *The Bell Curve*, Herrnstein and Murray (1994) (hereafter, H&M) used data from the 1979 National Longitudinal Survey of Youth (NLSY) to demonstrate that intelligence, as captured by achievement test scores from the Armed Forces Qualification Test (AFQT), is one of the most important factors related to economic, social, and overall success. Specifically, the researchers argued that intelligence can be a powerful predictor of a variety of performance dimensions (e.g., income, job performance, chance of unwanted pregnancy in comparison to other traditional predictors, such as schooling and family background. In making their case, the researchers presented evidence on the correlation between levels of cognitive ability and widely different dimensions of social behaviors (e.g., marriage, poverty, and crime). In their analyses, the effect of AFQT scores is more than twice as large as the effect of parents’ socioeconomic status.

By contrast, some recent studies have found that cognitive ability measured with the same AFQT scores had dramatically declined impact on one typical measure of success, namely hourly wage as calculated in the 2000s (e.g., Castex and Dechter, 2011). These contradictory reports regarding intelligence/cognitive ability and outcomes of social behavior merit further investigation. In particular, is the effect of intelligence on social behaviors as reported by Herrnstein and Murray still valid nearly 20 years later, or has it diminished as Castex and Dechter would seem to indicate? In order to answer this

---

1 More details of NLSY data are provided in Section 4.
question, this study assessed the same measures for “social behavior” that H&M evaluated in 1994, rates of incarceration, marriage, out-of-wedlock birth, low birth weight and poverty—but used data from the 1997 National Longitudinal Survey of Youth (NLSY) for comparative purposes.

By contrast, some recent studies, such as Castex and Dechter (2011), find that cognitive abilities measured with the same scores have a substantially smaller impact on hourly wage in the 2000s. The diminishing role of cognitive ability motivates us to wonder: does the effect of intelligence on social behaviors, which is also examined by H&M, remain as important in the 2000s as 20 years ago, or does it also diminish?

We also looked at possible changes in the role of formal education, not only because of the well-established fact that intelligence and education tend to be highly correlated, but also because Castex and Dechter (2011) reported that during the decade from 2000 to 2010, education had a large impact on higher hourly wages, which is contrasted with the smaller effect that AFQT scores seemed to play on hourly wages. Overall, our primary finding is that, in general, the role of AFQT scores in predicting social behaviors has not dramatically changed over the last 20 years. This outcome is discussed, along with some possible explanations for this study’s findings.

The role of intelligence in explaining social behaviors is important for a number of reasons. First, most of the social behaviors this paper examines (e.g., poverty and high incarceration rates) have been widely recognized as severely limiting individual achievement, as well as the collective progress of American society. Therefore, understanding the role of intelligence as one of the crucial predictors for these potentially-crippling social problems is vital for public policymakers if they are to address them.

---

2 Following H&M, the term “social behaviors” in this paper refers to incarceration, marriage, out-of-wedlock birth, low birth weight and poverty.
effectively. Second, this paper will shed light on changes in the impact of human intelligence over time. As Flynn (1994, 1999) has documented, there is convincing evidence that substantial and long-sustained increases in intelligence test scores have occurred over time. This means that if Herrnstein and Murray (1994) are to be believed, we would then expect to see the social problems that increased intelligence is purported to diminish also decreasing—all else being equal. In contrast, however, most of the social problems the researchers investigated almost 20 years ago have not uniformly decreased; in fact, many have increased. This paradoxical outcome results in a dilemma that merits further investigation: If, indeed, human intelligence is increasing, is its impact increasing, remaining the same, or decreasing? This is a broad question whose many ramifications cannot be addressed within the scope of this paper. Nonetheless, an aspect of the relationship between intelligence and social behaviors based on two important studies above will be addressed herein.

The remainder of the paper is organized as follows. In Section 2, I briefly review the related literature. In Section 3, I discuss the estimation specification that will be used for this study. Section 4 presents a description of the data set, key variables and sample definitions. Section 5 reports the estimation results, after which possible interpretations for the findings are reviewed. Conclusions and limitations are discussed in Section 6.

2. Literature Review

Ever since Herrnstein and Murray published *The Bell Curve* in 1994, there have been numerous discussions, criticism and reviews addressing the claims they made. Indeed, a full book would be required to address them all. Instead, this review will address

---

several recent studies associating behavioral outcomes with cognitive abilities that are closely related to this paper.4

Galindo-Rueda and Vignoles (2005) compared two British cohorts—one from 1958 and the other from 1970—with respect to ability and educational performance. The researchers reported a decline in the importance of ability in explaining educational performance for those born in 1970 as compared to the earlier generation born in 1958, which they tied to an economic factor. Specifically, they noted that the low-ability children with high economic status experienced the largest increases in educational attainment. Similarly, Belley and Lochner (2007) estimated the effects of cognitive ability and family income on educational attainment using two data sets, one from NLSY1979, and the second from the NLSY1997. Using a structural educational choice model, they argue for the importance of budget constraints to explain the rising influence of family income on education. As their study documented, AFQT scores were highly corrected with schooling outcomes in both 1979 and 1997. They used four quartile categorical variables to indicate AFQT levels, but did not expand their results to compare the effect of AFQT scores between the two periods. Nevertheless, in their estimation results using regressions for either high school completion or college attendance, the coefficients for the AFQT quartiles 2-4 were smaller than those for the 1979 cohorts; in comparison, when they looked at college attendance estimations, the coefficients for quartiles 2-4 were much larger than those for the 1979 cohort. The changes in the magnitude of those coefficients suggest that there might be possible changes in the role of AFQT in accounting for education decisions.

In a related study, Fryer (2010) focused on race-related gaps between labor market performance and social behavior using the same two NLSY cohorts, but also included AFQT

---

4 “Cognitive ability” and “intelligence” are considered to be interchangeable in this paper.
scores as a key control variable. Although the author does not directly discuss the role of AFQT scores, he reported the coefficients for the AFQT scores on hourly wage estimations to demonstrate substantial reductions for the 1997 cohort compared to the 1979 cohort.

In reviewing the literature, Castex and Dechter (2011) appear to be the first to definitively link the diminishing role of cognitive ability with wages. By estimating wage functions using the standard Mincer framework, they found that AFQT scores had a much smaller impact on wages for the 1997 cohort compared with the earlier cohort. Specifically, financial returns based on cognitive ability decreased by 20 to 50% (p. 4), while education contributions were found to be of greater importance in determining wage levels for the 1997 cohort. Castex and Dechter associated the decline in cognitive ability-related returns on the changing pace of technological advancements between 1979 and 1997, as well as the advance of the schooling signaling effect.

3. Empirical specification

In this section, I present the basic empirical specification for the model to be used in this paper.

The basic specification using AFQT and education for a particular type of social behavior for cohort $g$ ($g$ is either the 1979 or 1997 cohort) is represented in the following equation:

$$ Y_{gi}^{*} = X_{gi}^{t_0} B + \alpha^g AFQT + \beta^g EDUC + u_{gi} $$

where $Y_{gi}^{*}$ indicates either the level of a continuous outcome or the value of a latent variable that is positive if individual $i$ has ever engaged in a particular social behavior in cohort $g$ between $t_0 - t$. $X_{gi}^{t_0}$ represents a vector of observable individual characteristics and
family background variables in base year $t_0$. The coefficient of the AFQT score (the intelligence measure), $\alpha^{b_k}$, is of key interest to this study. \textit{EDUC} represents years of formal education. The error term is represented by $u_{g,t}^{g_k}$.

Besides (1), in order to look at the role of AFQT, we also experiment with only including AFQT or education in the estimation.

There are two main differences between the specifications that H&M used and ones employed in this paper. The first is the difference in the control variables used for estimation. H&M used the following three variables to estimate the predicted probability for each particular social behavior: AFQT scores, a standardized Duncan's Socioeconomic Index (combining parental occupational status, parental education, and family income), and standardized age.\textsuperscript{5} For this study, however, Duncan’s Socioeconomic Index was not utilized to indicate family background for the following reasons: (a) parental occupation and industry information was not reported for the 1997 cohort (thus, it was not possible to construct a reliably comparable index for both cohort years); and (b), using a single index has limitations such as restricting the coefficients of each component of the family background index to have an identical effect, and overstating the effects of IQ by using a rather crude measure of parents' socioeconomic status (e.g., Korenman & Winship, 1995; Dickens et al., 1999). Therefore, unlike H&M’s index, we used a set of variables to indicate respondents’ socioeconomic status.

The second major difference in the specifications used herein in comparison to those in H&M’s work is that we controlled for a much richer set of variables ($X_{g,t}^{b_k}$) to reduce the effects confounded by other factors. Among these independent variables, formal education

\textsuperscript{5} Duncan’s Socioeconomic Index (Duncan, 1961) is a composite of occupational prestige, income, and education, which is most commonly used as a socioeconomic status measure.
was of interest. H&M hesitated to control for education mainly because it is highly correlated with intelligence and with parents’ socioeconomic status (p. 124-125), although they examined two subsamples: a sample for high school graduates only and a sample containing college graduates with bachelor's degree only. But we introduce education as one of the independent variables along with AFQT score, following most literature after H&M.

It is important to stress that this study was not designed to examine causation. Although we incorporated measures to ensure that the two cohorts were comparable in a number of significant ways, the relationships between cognitive ability and education described in this paper should be viewed as correlations, and not as causality associations. Another limitation that should be noted is that in comparing the regression results for the two cohorts we assumed that, except for the variables that we controlled for all other unobservable characteristics would remain unchanged between these two cohorts, which of course is unlikely. However, although such an assumption is typical in cohort comparison analysis (e.g., Altonji et al., 2008), it should nonetheless be viewed as a limitation.

4. Data

This paper utilized data from the NLSY79 and NLSY97, two nationally representative samples of young men and women in the United States. When the respondents were first interviewed, they were between the ages of 14- and 22 for the 1979 cohort, and between the ages of 12–and 18 for the 1997 cohort. The original sample sizes were 12,686 and 8,984, respectively.

We restricted our sample to non-Hispanic whites and non-Hispanic blacks.\(^6\) Two principal criteria were used to select the survey years for comparing the two cohorts. First, the youth were of similar ages. Second, we used the years when the respondents were the

---

\(^6\) We dropped Hispanics from this study because H-M focused on whites and blacks in their book.
Therefore, the most frequently used sample in this study includes survey years 1979-1988 for the older cohort, and 1999-2008 for the newer cohort (although samples varied across different specifications). Table 1 displays descriptive statistics for the main variables in our main sample. Note that the sample sizes vary when looking at different social behaviors and when the specifications differ, principally due to missing values.

4.1 AFQT Scores

The variable of greatest interest for this study was cognitive ability. AFQT scores are typically used to approximate a person’s cognitive ability (for a summary of this literature, see Cawley et al., 2001). Both the NLSY79 and NLSY97 include scores from the Armed Services Vocational Aptitude Battery (ASVAB), which is a sequence of tests that cover basic math, verbal, and manual skills. Math skills are measured by scores on the Arithmetic Reasoning (AR), Numerical Operations (NO) and Mathematics Knowledge (MK) sections of the ASVAB. Verbal skills are measured by scores from the Word Knowledge (WK) and Paragraph Comprehension (PC) sections of the ASVAB. For this study, raw scores for the subject areas were standardized first, after which unadjusted AFQT scores were then calculated, representing the sum of the arithmetic reasoning score, the mathematics knowledge score, and two times the verbal composite score.

The 1997 cohort took the tests between the ages of 15 and 23, while the age range for the 1997 cohort was between 12 and 17. To address any scoring variations caused by age difference, I followed the method in Neal and Johnson (1996), I standardized the residuals from the regressions of the unadjusted AFQT scores on birth year dummies.

---

7 An example of other reasons that may cause changes in the sample size is associated with gender differences. For this study we looked at certain gender-based behavior differences (e.g., for incarceration, we examined for males only, while for low birth weight, we restricted the analysis to females only).

8 Altonji et al. (2008) described a method for making AFQT scores in NLSY79 and NLSY97 comparable. I experimented with their procedure to adjust AFQT scores, which provides qualitatively similar results as
In Table 1, note that although the average AFQT scores for other groups basically remained steady, the AFQT scores for black males increased from -0.780 to -0.460.

4.2 Social Behaviors Measures

Following H&M, we treated the social behaviors using life history measures. Specifically, between 1979-1988 for the 1979 cohort and between 1999-2008 for the 1997 cohort, if individuals had ever “engaged” in any of the following five behaviors: (a) incarceration; (b) marriage; (c) out-of-wedlock birth; (d) low birth-weight infant; or (e) poverty (according to standard federal guidelines)—they were assigned a value of 1 for that particular variable; otherwise, the variable was coded as zero.

A) Incarceration probability. In the U.S., ongoing increases in incarceration rates remain a troubling social problem. For example, the rate of incarceration in local jails in the U.S. has more than tripled since the mid-1970s (e.g., Raphael, 2009; Spelman, 2009). In addition to the direct costs of keeping a man in jail (by some estimates, about $22,000 a year),9 economists have also documented the long-term consequences of incarceration, such as negative effects on later educational attainment (Pintoff, 2006; Hjalmarsson, 2008), and employment and earnings (Holzer, 2007; Pettit and Lyons, 2007; Jung, 2010). Moreover, male incarceration has some corollary influences on the employment, health and marriage prospects of females (e.g., Charles & Luoh, 2010; Lee & Wildeman, 2011; Mechoulan, 2011).

For this study, “incarceration” was defined in the following way. An individual was classified as having been incarcerated if he had ever been in jail or spent time in either an adult or juvenile correctional institution for any arrest in any of the survey years. The percentages of incarcerated men for each cohort were 1.9% and 2.3%, respectively

---

9 See: http://www.reference.com/motif/history/cost-to-keep-someone-in-prison
(difference is insignificant). Women were dropped from this estimation because there were few women having incarceration histories in our sample.

B) Marriage.

Although the relationship between cognitive ability and marriage for youth is confounded by their educational achievement levels, getting married by age 30 is claimed by H&M to be positively associated with cognitive ability. Among all the social behaviors examined in this paper, marriage is the only one that is considered to be “desirable.” For this study, we looked at people who had ever been married within the periods of the surveys. For this study’s samples, the marriage rate diminished from 40.9% to 27.9% over the 20 years between surveys (difference significant at the 1% level), which is consistent with well-established trends showing an increase in the average age of first marriage.

C) Out-of-wedlock birth. Out-of-wedlock birth has been well documented in the literature to lead to a variety of undesirable outcomes (e.g., Willis, 1999, Colosi, 2006, Ermisch, 2009, Covington et al., 2011) The consequences include, but are not limited to, long-term negative effects on young mothers in terms of education, earnings and employment with similar outcomes for their children. If a female respondent or a male respondent’s partner/wife had given birth to any children before her/his first marriage, we assigned a 1 for this variable (0 otherwise). In our sample, the rate of out-of-wedlock childbirth had increased from 14.3% to 32.3% between the two cohorts (difference is significant at the 1% level).

---

10 H&M claimed that low IQs were associated with a disadvantage in competing for marriage partners (p. 170). H&M found that for their high school sample, higher IQ was clearly associated with higher probability of marriage by age 30, while for their college sample, IQ did not appear to have a significant effect on marriage (pages 169-172).

11 See, for example, Kuo (2005) and Brand and Davis (2011) for recent marriage trends among American youth.

12 See, for example, Wu and Wolf (2001), for a detailed discussion of consequences of out-of-wedlock birth.
D) Low birth weight. It has been well recognized that low birth weight can have negative effects on health, development and adult outcomes (e.g., Royer, 2009; Currie, 2009). Moreover, a low birth weight baby can incur substantial hospital costs (e.g., Almond et al., 2005). Following the conventional definition, if an infant weighed 5.5 pounds or below at birth, the mother was coded as having delivered a low birth weight infant. In our sample the fraction of low birth weight childbirth for the two cohorts was 1.5% and 9.7%, respectively (difference is significant at the 1% level).

E) Poverty. The consequences of poverty can be severe. For this study, if the ratio of household income to federal poverty level was less than one, the respondent was identified as living in poverty. In our sample, 16.8% of the respondents for the older cohort and 19.6% for the newer cohort were considered to be impoverished (difference is insignificant).

4.3 Other Variables

A) Formal education. Formal education is measured as completed years of schooling in the base year. As shown in Table 1, the average total schooling years (i.e., highest grade completed) for the two cohorts are 11.2 or 11.6 years, respectively (difference significant at the 1% level).

B) Family background. This study also controlled for parental education, household income, and household size in the first survey year. Parental education increased considerably, which is consistent with findings in Altonji et. al (2008). In particular, the mother’s highest grade completed rose from 11.8 years in the 1980s to 13.3 years in the 2000s, and the fathers’ highest grade completed increased from 11.8 to 13.6 years (difference is significant at the 1% level for mother’s and father’s education, respectively).

Real household income also increased substantially over the 20-year period. In our sample, 13 See Hoynes et al. (2006) for a literature survey of poverty trends and explanations. See Venkatesh (2006) for a vital and poignant study about how the poor live in a Southside Chicago neighborhood.
the average real household income in 1979 was $19,560, compared to $65,346 in 1997 (difference is significant at the 1% level).

C) Presence of parents at age 12 or 14. If a respondent at age 14 for the 1979 cohort, and at age 12 for 1997 cohort, lived in a family that did not have both a mother and father present, we counted this respondent as being raised in a single-parent household.\textsuperscript{14} Across the two cohorts, those who lived in a single-parent household increased from 32.1% to 41.8% (difference is significant at the 1% level).

Other control variables included demographic characteristics, as well as standard controlling variables such as individual characteristics (marital status, having children and part-time work status) and residence (region of residence, urban, Metropolitan Statistical Area (MSA), county-level median income, crime rates and unemployment rate).

The correlation coefficients for highest schooling years, AFQT scores and social behaviors are reported in Table 2. We note that among all the subgroups, the magnitudes of the coefficients for the AFQT scores and education do not change much across the two cohorts 0.57 and 0.66, respectively. In addition, for black males and females, the correlation coefficients between AFQT scores and marriage are positive for both cohorts. In contrast, for white males and females, marriage and AFQT scores are negatively linked (although some of them are insignificant), which confirms H&M’s finding that marriage by age 30 is positively associated with intelligence. For other social behaviors, the correlation implications are similar to what was discussed in Subsection 4.2.

5. Results

\textsuperscript{14} NLSY79 provides information for age 14, while NLSY97 provides information for age 12. Although the ages are not analogous, we did not adjust for the age difference in assessing this variable since there was no way to do so.
This section details the regression results associated with the various social behaviors examined herein for the two cohorts.\textsuperscript{15} Tables 3-9 document the three main specifications under scrutiny in this study for the whole cohort first, followed by for each race-gender group within a cohort (It should be noted, however, that other specifications returned qualitatively similar results.) For each table, in addition to the common control variables considered in the literature, panel (A) includes AFQT scores; panel (B) includes total schooling years; and panel (C) includes both AFQT scores and education. As noted, however, this study also controlled for the variables discussed in Section 4. Additionally, we report the significance of testing the equality of coefficients or predicted probabilities across the two cohorts.\textsuperscript{16}

Before determining results based solely on the behavioral measures described herein, the effects of hourly wage and annual earnings were ascertained. In other words, did our sample display the same diminishing role of cognitive ability on wages that was recently reported by Castex and Dechter (2011). Table 3 documents the Ordinary Least Squares (OLS) estimation for hourly wage, controlling for standard variables in wage equations.\textsuperscript{17} Panel (C) in Table 3 clearly shows that the effect of cognitive ability on wages is larger for the 1979 cohort in comparison to the 1997 cohort, regardless of race and gender (significant at the 1\% level). This decline is qualitatively consistent with findings reported by Castex and Dechter (2011).

In addition to wages, annual earnings were also taken into account, which Castex and Dechter (2011) did not address. Table 4 documents the OLS regressions of log annual

\textsuperscript{15} In most specifications, Chow tests reject the hypothesis that coefficients in regressions by different race-gender groups are similar.

\textsuperscript{16} Significance for comparison of coefficients or predicted probabilities is denoted by symbol † in Tables 3-9 († 10\% level, †† 5\% level and ††† 1\% level).

\textsuperscript{17} Other control variables include demographic variables, work experience, tenure, geographic variables, union membership, part-time work, and health limitations.
earnings on AFQT scores, education and other control variables. The AFQT coefficients for white females declined significantly (at the 1% level), while a similar pattern is not seen for other groups. In general, few significant changes in the coefficients of AFQT scores between the two cohorts are evident except for white females. Similarly, for most groups there are no significant changes in the education coefficient.

In the following estimation, logit regression results are provided for the social behaviors discussed above. For easier interpretation, marginal effects for logit estimations (the change in probability from a unit change in the relevant variable) are reported. Table 5 documents the probability of having been incarcerated for both white and black males. For example, as shown in panel (C), for a black male in the 1979 cohort, one unit increase in the standard deviation of AFQT score lowered the probability of having been incarcerated by 0.058 (significant at the 1% level), while for the new cohort this effect dropped to 0.040 (insignificant). As shown, changes in the effects of cognitive ability and education on the probability of incarceration between the two races were insignificant.

Table 6 features logit regression results for the probability of marriage. As shown in panel (C), the role of AFQT scores did not play much of a role across the two cohorts, except for black males. On average for black males, higher cognitive ability predicts higher probability of marriage. But we do not see similar significant associations between AFQT score and marriage among other groups.

Table 7 represents the estimation results for the probability of out-of-wedlock births. Across the two cohorts, the role of AFQT score in predicting the likelihood of out-of-wedlock fertility became larger only for white males. In contrast, the effects of education were shown to be more important for both black males and black females.

---

18 We dropped female observations from these estimations due to the low percentage of females who had been incarcerated.
Table 8 documents the regression results for the probability of females having a low-birth-weight infant (had they given birth). Neither cognitive ability nor education was shown to have large significant effects for white females in both cohorts. It is noticeable, however, that for black females, education had higher predictive power on low-birth-weight childbirth for the 1997 cohort, the higher AFQT scores, the lower the probability of giving birth low-birth-weight infants.

Table 9 reports the estimation results for poverty. One change is worth noting for white males: the effect of AFQT scores on the likelihood of being in poverty changed from negative in the 1980s to positive in the 2000s. For black males, the comparative effect of education on poverty changed into insignificant. But the role of education became more important for black males.

To summarize, in general this study did not document a pattern of substantial decline in the predictive ability of cognitive aptitude for social behaviors in comparison to what was reported by Castex and Dechter (2011) with respect to hourly wage. Moreover, when compared with the role of AFQT scores, education levels were shown to exert a greater impact across the two cohorts for the investigated behaviors. For instance, education had a positive effect on marriage for the new cohorts. In comparison, the predictive power of education declined across both cohorts for out-of-wedlock childbirth.

6. Discussion and Conclusion

In conclusion, the findings discussed herein provide no support for the hypothesis that the use of cognitive ability as measured by AFQT scores in predicting an array of social behaviors differed considerately across the two cohorts. There are a number of reasons to explain why the role of AFQT scores did not decline for social behaviors as it did for hourly wage indicators. One obvious interpretation is that measuring cognitive ability is
inherently problematic (e.g., Heckman and Vytlacil, 2001). If this limitation was valid, however, it would represent an insurmountable stumbling block to meaningful research, which clearly has not been the case as evidenced by a significant number of studies that have used cognitive ability as a research variable. Another possibility is related to the reasons that Castex and Dechter (2011) suggested for the declining importance of the role of AFQT scores in predicting wages. The researchers claimed that changes in the pace of technological progress, as well as changes in employee learning speed, might account for the decline in the effect of AFQT scores on wages in the 2000s. If these reasons are valid, they should not be expected to affect social behaviors in the same way they influence wage-earning potential.

To summarize, this paper compared the role of intelligence as measured by AFQT scores on social behaviors for two cohorts—one first surveyed in 1979 and the other in 1997. By assessing logistic estimations for each social behavior for the older cohort and the newer cohort, and then comparing the coefficients of AFQT scores and education, we did not identify a clearly-decreased role of AFQT scores in predicting social behavior in comparison to what was reported for hourly wage effects by Castex and Dechter (2011). This paper finds that, in general, the role of cognitive ability (measured as AFQT scores) has not substantially changed over the past 20 years (although there have been increases or reductions). In contrast, the predictive power of formal schooling has been found to have larger changes in the past 20 years. In closing, several limitations must be noted. First, this investigation should not be considered a strict causal analysis. The effects of unobservable characteristics, such as non-cognitive ability, are certainly possible; this study, however,
lacked the methodological tools to assess this problem.\footnote{For example, with respect to non-cognitive ability, although for the older cohort the measures for non-cognitive ability are available, there are no comparable measures for the new cohort.} Second, as noted earlier, the variables for the two cohorts are not wholly comparable. Therefore, some of the variables had to be omitted from the comparisons. Future research studies could consider extending the present study by incorporating an examination of unobservable characteristics into the analysis.

References


### Tables

#### Table 1. Statistical Description

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White male</td>
<td>Black male</td>
</tr>
<tr>
<td>AFQT (T_0)</td>
<td>0.490 (0.87)</td>
<td>-0.780 (0.89)</td>
</tr>
<tr>
<td>Highest grade completed((T_0))</td>
<td>11.155 (1.77)</td>
<td>10.891 (1.76)</td>
</tr>
<tr>
<td>Diploma or GED (T_0)</td>
<td>0.490 (0.50)</td>
<td>0.440 (0.50)</td>
</tr>
<tr>
<td>Incarceration ever (T)</td>
<td>0.012 (0.11)</td>
<td>0.094 (0.29)</td>
</tr>
<tr>
<td>Married ever (T)</td>
<td>0.439 (0.496)</td>
<td>0.241 (0.428)</td>
</tr>
<tr>
<td>Out of wedlock birth ever (T)</td>
<td>0.064 (0.245)</td>
<td>0.271 (0.449)</td>
</tr>
<tr>
<td>Low birth weight ever (T)</td>
<td>0.026 (0.16)</td>
<td>0.050 (0.22)</td>
</tr>
<tr>
<td>Poverty ever (T)</td>
<td>0.079 (0.271)</td>
<td>0.220 (0.414)</td>
</tr>
</tbody>
</table>

| Female \(T_0\) | 0.478 (0.50) | 0.498 (0.50) |
| Black \(T_0\) | 0.258 (0.44) | 0.281 (0.45) |
| Age \(T_0\) | 18.447 (2.13) | 18.713 (2.03) | 18.365 (2.11) | 18.958 (1.81) | 18.366 (2.05) | 18.093 (1.42) | 18.137 (1.44) | 18.184 (1.40) | 18.208 (1.41) | 18.101 (1.40) |
| Real wage \(T_0\) | 10.986 (9.17) | 10.361 (11.32) | 8.814 (8.68) | 9.669 (11.27) | 9.911 (10.67) | 11.203 (36.26) | 15.343 (18.08) | 10.496 (39.07) | 9.714 (25.76) | 12.438 (87.56) |
| Log real family income \(T_0\) | 9.839 (0.62) | 8.999 (0.89) | 9.758 (0.66) | 9.146 (0.84) | 9.519 (0.87) | 10.918 (1.00) | 10.082 (1.52) | 10.941 (0.91) | 10.137 (1.36) | 10.700 (1.18) |
| Mother’s education(years) \(T_0\) | 12.100 (2.23) | 10.759 (2.72) | 12.116 (2.18) | 11.338 (2.40) | 11.756 (2.37) | 13.553 (2.52) | 12.636 (2.28) | 13.580 (2.51) | 12.579 (2.06) | 13.271 (2.44) |
| Father’s education(years) \(T_0\) | 12.323 (3.11) | 9.867 (3.40) | 12.389 (3.03) | 10.845 (3.66) | 11.787 (3.31) | 13.827 (2.89) | 12.448 (2.50) | 13.781 (2.76) | 12.940 (2.12) | 13.599 (2.76) |

N | 1748 (954) | 2223 (968) | 5893 (1160) | 465 (3308) |

Source: NLSY79 and NLSY97 data. Notes: All statistics are weighted with the NLSY cross-sectional weights. Standard deviations are reported in parenthesis. \(T_0\) indicates 1979 or 1999 for the two cohorts, respectively and \(t\) for 1988 or 2008.
Table 2. Correlation Coefficients of AFQT, Education and Social Behaviors for 1979 & 1997 Cohorts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
<td>AFQT</td>
<td>Education</td>
</tr>
<tr>
<td>White male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.656***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incarceration</td>
<td>-0.169***</td>
<td>-0.157***</td>
<td></td>
<td>-0.154***</td>
<td>-0.218***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marriage</td>
<td>-0.050*</td>
<td>-0.096***</td>
<td>-0.014</td>
<td>-0.076***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-of-wedlock birth</td>
<td>-0.186***</td>
<td>-0.250***</td>
<td></td>
<td>-0.277***</td>
<td>-0.358***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low birth weight</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td>-0.263***</td>
<td>-0.217***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.575***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incarceration</td>
<td>-0.213***</td>
<td>-0.200***</td>
<td></td>
<td>-0.221***</td>
<td>-0.292***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marriage</td>
<td>0.092***</td>
<td>0.096***</td>
<td>0.072*</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-of-wedlock birth</td>
<td>-0.210***</td>
<td>-0.228***</td>
<td></td>
<td>-0.252***</td>
<td>-0.318***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low birth weight</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td>-0.263***</td>
<td>-0.298***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.619***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incarceration</td>
<td>-0.040</td>
<td>-0.041*</td>
<td></td>
<td>-0.052*</td>
<td>-0.089***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marriage</td>
<td>-0.066**</td>
<td>-0.151***</td>
<td></td>
<td>-0.055*</td>
<td>-0.111***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-of-wedlock birth</td>
<td>-0.297***</td>
<td>-0.337***</td>
<td></td>
<td>-0.330***</td>
<td>-0.502***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low birth weight</td>
<td>-0.075***</td>
<td>-0.085***</td>
<td></td>
<td>-0.060**</td>
<td>-0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td>-0.290***</td>
<td>-0.258***</td>
<td></td>
<td>-0.223***</td>
<td>-0.201***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.608***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incarceration</td>
<td>-0.074**</td>
<td>-0.079**</td>
<td></td>
<td>-0.050</td>
<td>-0.110***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marriage</td>
<td>0.080**</td>
<td>0.022</td>
<td>0.088**</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-of-wedlock birth</td>
<td>-0.260***</td>
<td>-0.321***</td>
<td></td>
<td>-0.340***</td>
<td>-0.378***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low birth weight</td>
<td>-0.048</td>
<td>-0.079**</td>
<td>-0.033</td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td>-0.400***</td>
<td>-0.410***</td>
<td></td>
<td>-0.419***</td>
<td>-0.374***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: NLSY79 and NLSY97 data.
Notes: ***Significant at 1% level; **significant at 5% level; *significant at 10% level.
### Table 3. OLS Regression Results (Coefficients) on Log Wage for 1979 & 1997 Cohorts

<table>
<thead>
<tr>
<th>Variables</th>
<th>1979 cohort</th>
<th>1997 cohort</th>
<th>Comparison across cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) AFQT</td>
<td>(2) Education</td>
<td>(3) N R²</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.166***</td>
<td>5,324</td>
<td>0.113***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td></td>
<td>(0.011)</td>
</tr>
<tr>
<td>White male</td>
<td>0.134***</td>
<td>1,748</td>
<td>0.080***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td></td>
<td>(0.019)</td>
</tr>
<tr>
<td>Black male</td>
<td>0.155***</td>
<td>954</td>
<td>0.054*</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>White female</td>
<td>0.186***</td>
<td>1,687</td>
<td>0.146***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td>(0.020)</td>
</tr>
<tr>
<td>Black female</td>
<td>0.155***</td>
<td>954</td>
<td>0.054*</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.074***</td>
<td>5,523</td>
<td>0.102***</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td>(0.009)</td>
</tr>
<tr>
<td>White male</td>
<td>0.089***</td>
<td>1,830</td>
<td>0.096***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
<td>(0.018)</td>
</tr>
<tr>
<td>Black male</td>
<td>0.099***</td>
<td>985</td>
<td>0.064***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td>(0.021)</td>
</tr>
<tr>
<td>White female</td>
<td>0.037***</td>
<td>1,755</td>
<td>0.111***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td></td>
<td>(0.019)</td>
</tr>
<tr>
<td>Black female</td>
<td>0.099***</td>
<td>985</td>
<td>0.064***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td>(0.021)</td>
</tr>
<tr>
<td>(C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.147***</td>
<td>0.041***</td>
<td>5,323</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>0.106***</td>
<td>0.062***</td>
<td>1,748</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>0.115***</td>
<td>0.072***</td>
<td>954</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.016)</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>0.185***</td>
<td>0.003</td>
<td>1,686</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>0.115***</td>
<td>0.072***</td>
<td>954</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.016)</td>
<td></td>
</tr>
</tbody>
</table>

Source: NLSY79 and NLSY97 data.
Notes: 1) All statistics are weighted with the NLSY provided cross-sectional weights. Standard errors in parenthesis.
2) Control variables that are not reported include work experience, tenure at current job, union status, part-time indicator, family background (parental education, parental employment in the year before the initial survey, family household income in the first initial survey year), presence of parents at age 12 or 14 (only mother present, only father present, or neither mother or father), current individual characteristics (marital status, dummy for having children, and part-time work) and residence (region of residence, urban, MSA, county-level log median income, county-level crime rates and county-level unemployment rate).
3) Symbol † indicates significance using Wald test for the equivalence of coefficients across the two cohorts. * for significance of coefficients. †††, ***Significant at 1% level; ††, **significant at 5% level; †, * significant at 10% level.
Table 4. OLS Regression Results (Coefficients) on Log Annual Earnings for 1979 & 1997 Cohorts

<table>
<thead>
<tr>
<th>Variables</th>
<th>1979 cohort</th>
<th>1997 cohort</th>
<th>Comparison across cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>AFQT</td>
<td>Education</td>
<td>N</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.290***</td>
<td>5.278</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>0.204***</td>
<td>1.791</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>0.353***</td>
<td>969</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>0.172</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>0.186***</td>
<td>1.687</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>0.227***</td>
<td>909</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>0.176</td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.146***</td>
<td>5.467</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>0.164</td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>0.153***</td>
<td>1.873</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>0.233***</td>
<td>1.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>0.155</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>0.063**</td>
<td>1.688</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>0.072***</td>
<td>953</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>0.117</td>
<td></td>
</tr>
<tr>
<td>(C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.254***</td>
<td>0.085***</td>
<td>5.277</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.014)</td>
<td>0.197</td>
</tr>
<tr>
<td>White male</td>
<td>0.152***</td>
<td>0.115***</td>
<td>1.791</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.019)</td>
<td>0.164</td>
</tr>
<tr>
<td>Black male</td>
<td>0.279***</td>
<td>0.145***</td>
<td>969</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.029)</td>
<td>0.193</td>
</tr>
<tr>
<td>White female</td>
<td>0.302***</td>
<td>0.016</td>
<td>1.632</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.030)</td>
<td>0.120</td>
</tr>
<tr>
<td>Black female</td>
<td>0.209***</td>
<td>0.034*</td>
<td>935</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.018)</td>
<td>0.179</td>
</tr>
</tbody>
</table>

Source: NLSY79 and NLSY97 data.

Notes: 1) All statistics are weighted with the NLSY provided cross-sectional weights. Standard errors in parenthesis.
2) Control variables that are not reported include work experience, tenure at current job, union status, part-time indicator, family background (parental education, parental employment in the year before the initial survey, family household income in the first initial survey year), presence of parents at age 12 or 14 (only mother present, only father present, or neither mother or father), current individual characteristics (marital status, having children, and part-time dummies) and residence (region of residence, urban, MSA, county-level log median income, county-level crime rates and county-level unemployment rate).
3) Symbol † indicates significance using Wald test for the equivalence of coefficients across the two cohorts. * for significance of coefficients. †† † , ***Significant at 1% level; † † , **significant at 5% level; † , * significant at 10% level.
Table 5. Logistic Regression Results (Marginal Effects) on Incarceration Probability for 1979 & 1997 Cohorts

<table>
<thead>
<tr>
<th>Variables</th>
<th>1979 cohort</th>
<th></th>
<th></th>
<th>1997 cohort</th>
<th></th>
<th></th>
<th>Comparison across cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) AFQT</td>
<td>(2) Education</td>
<td>N Pseudo-R²</td>
<td>(3)</td>
<td>(4) AFQT</td>
<td>(5) Education</td>
<td>N Pseudo-R²</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>-0.010***</td>
<td>3,353</td>
<td></td>
<td>-0.011***</td>
<td>2,414</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>0.242</td>
<td></td>
<td>(0.002)</td>
<td>0.231</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.004***</td>
<td>2,108</td>
<td></td>
<td>-0.006***</td>
<td>1,677</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>0.250</td>
<td></td>
<td>(0.002)</td>
<td>0.267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>-0.071***</td>
<td>1,219</td>
<td></td>
<td>-0.055***</td>
<td>702</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>0.136</td>
<td></td>
<td>(0.010)</td>
<td>0.195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>-0.007***</td>
<td>2,414</td>
<td></td>
<td>-0.010***</td>
<td>2,887</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>0.231</td>
<td></td>
<td>(0.003)</td>
<td>0.188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.004***</td>
<td>2,247</td>
<td></td>
<td>-0.005**</td>
<td>1,956</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>0.229</td>
<td></td>
<td>(0.002)</td>
<td>0.189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>-0.033***</td>
<td>1,282</td>
<td></td>
<td>-0.049***</td>
<td>891</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>0.104</td>
<td></td>
<td>(0.010)</td>
<td>0.142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>-0.007***</td>
<td>-0.001***</td>
<td>3,558</td>
<td>-0.009***</td>
<td>2,395</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>0.224</td>
<td>(0.002)</td>
<td>0.240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.003**</td>
<td>-0.002***</td>
<td>2,107</td>
<td>-0.006***</td>
<td>1,669</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>0.278</td>
<td>(0.002)</td>
<td>0.268</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>-0.058***</td>
<td>-0.020***</td>
<td>1,219</td>
<td>-0.040***</td>
<td>693</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.006)</td>
<td>0.151</td>
<td>(0.008)</td>
<td>0.217</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: NLSY79 and NLSY97 data.
Notes: 1) All statistics are weighted with the NLSY provided cross-sectional weights. Standard errors in parenthesis. Only males are included.
2) Control variables that are not reported include family background (parental education, parental employment in the year before the initial survey, family household income in the first initial survey year), presence of parents at age 12 or 14 (only mother present, only father present, or neither mother or father), current individual characteristics (marital status, dummy for having children, and part-time work) and residence (region of residence, urban, MSA, county-level log median income, county-level crime rates and county-level unemployment rate).
3) Symbol ††† indicates significance of test for the equivalence of predicted probabilities across the two cohorts.
* for significance of marginal effects in logit estimation. ††† , **Significant at 1% level; †† , **significant at 5% level; † , * significant at 10% level.
Note: None of the differences of marginal effects across cohorts are significant.
Table 6. Logistic Regression Results (Marginal Effects) on Marriage for 1979 & 1997 Cohorts

<table>
<thead>
<tr>
<th>Variables</th>
<th>1979 cohort</th>
<th></th>
<th></th>
<th>1997 cohort</th>
<th></th>
<th></th>
<th>Comparison across cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) AFQT</td>
<td>(2) Education</td>
<td>N Pseudo-R²</td>
<td>(4) AFQT</td>
<td>(5) Education</td>
<td>N Pseudo-R²</td>
<td>(7) AFQT</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.007</td>
<td>2,743</td>
<td>(0.014)</td>
<td>0.021*</td>
<td>3,294</td>
<td>†</td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.019</td>
<td>1,035</td>
<td>(0.019)</td>
<td>0.028</td>
<td>1,239</td>
<td>†</td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>0.021</td>
<td>1,066</td>
<td>(0.025)</td>
<td>0.086***</td>
<td>333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.004</td>
<td>1,000</td>
<td>(0.027)</td>
<td>-0.011</td>
<td>1,160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>0.012</td>
<td>312</td>
<td>(0.054)</td>
<td>0.039*</td>
<td>192</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.009</td>
<td>2,912</td>
<td>(0.012)</td>
<td>0.022**</td>
<td>3,874</td>
<td>††</td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.004</td>
<td>1,110</td>
<td>(0.017)</td>
<td>0.022</td>
<td>1,452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>0.052**</td>
<td>1,120</td>
<td>(0.026)</td>
<td>0.051**</td>
<td>533</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.023</td>
<td>1,052</td>
<td>(0.015)</td>
<td>0.022</td>
<td>1,342</td>
<td>††</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>-0.067</td>
<td>321</td>
<td>(0.047)</td>
<td>0.020</td>
<td>547</td>
<td>††</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.006</td>
<td>-0.004</td>
<td>2,743</td>
<td>0.014</td>
<td>0.022*</td>
<td>3,272</td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.018</td>
<td>-0.000</td>
<td>1,035</td>
<td>0.025</td>
<td>0.014</td>
<td>1,232</td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>-0.002</td>
<td>0.051**</td>
<td>336</td>
<td>0.077***</td>
<td>0.019</td>
<td>1,487</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.002</td>
<td>-0.007</td>
<td>1,000</td>
<td>-0.023</td>
<td>0.040*</td>
<td>1,527</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>0.032</td>
<td>-0.071</td>
<td>312</td>
<td>0.040*</td>
<td>-0.002</td>
<td>1,461</td>
<td>†</td>
</tr>
</tbody>
</table>

Source: NLSY 79 and NLSY97 data.

Notes: 1) All statistics are weighted with the NLSY provided cross-sectional weights. Standard errors in parenthesis.
2) Only those who are not enrolled in school are included.
3) Control variables that are not reported include family background (parental education, parental employment in the year before the initial survey, family household income in the first initial survey year), presence of parents at age 12 or 14 (only mother present, only father present, or neither mother or father), current individual characteristics (dummy for having children and part-time work) and residence (region of residence, urban, MSA, county-level log median income, county-level crime rates and county-level unemployment rate).
4) Symbol † indicates significance of test for the equivalence of predicted probabilities across the two cohorts.
5) * for significance of marginal effects in logit estimation. †††, ***Significant at 1% level; ††, **significant at 5% level; †, * significant at 10% level.
Table 7. Logistic Regression Results (Marginal Effects) on Out-of-Wed Birth for 1979 & 1997 Cohorts

<table>
<thead>
<tr>
<th>Variables</th>
<th>1979 cohort</th>
<th></th>
<th>1997 cohort</th>
<th></th>
<th>Comparison across cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>AFQT</td>
<td>Education</td>
<td>N</td>
<td>Pseudo-R²</td>
<td>AFQT</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.052***</td>
<td>6,816</td>
<td>(0.006)</td>
<td>0.240</td>
<td>-0.090***</td>
</tr>
<tr>
<td>White male</td>
<td>-0.021***</td>
<td>2,106</td>
<td>(0.005)</td>
<td>0.140</td>
<td>-0.050***</td>
</tr>
<tr>
<td>Black male</td>
<td>-0.112***</td>
<td>1,243</td>
<td>(0.025)</td>
<td>0.136</td>
<td>-0.116***</td>
</tr>
<tr>
<td>White female</td>
<td>-0.061***</td>
<td>2,175</td>
<td>(0.010)</td>
<td>0.189</td>
<td>-0.106***</td>
</tr>
<tr>
<td>Black female</td>
<td>-0.090***</td>
<td>1,292</td>
<td>(0.023)</td>
<td>0.202</td>
<td>-0.153***</td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.012***</td>
<td>7,199</td>
<td>(0.004)</td>
<td>0.226</td>
<td>-0.041***</td>
</tr>
<tr>
<td>White male</td>
<td>-0.011**</td>
<td>2,246</td>
<td>(0.005)</td>
<td>0.127</td>
<td>-0.021***</td>
</tr>
<tr>
<td>Black male</td>
<td>-0.016</td>
<td>1,309</td>
<td>(0.015)</td>
<td>0.124</td>
<td>-0.074***</td>
</tr>
<tr>
<td>White female</td>
<td>-0.003</td>
<td>2,301</td>
<td>(0.005)</td>
<td>0.169</td>
<td>-0.049***</td>
</tr>
<tr>
<td>Black female</td>
<td>-0.024</td>
<td>1,343</td>
<td>(0.018)</td>
<td>0.196</td>
<td>-0.051*</td>
</tr>
<tr>
<td>(C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.051***</td>
<td>6,816</td>
<td>(0.006)</td>
<td>0.240</td>
<td>-0.086***</td>
</tr>
<tr>
<td>White male</td>
<td>-0.018***</td>
<td>2,106</td>
<td>(0.005)</td>
<td>0.142</td>
<td>-0.046***</td>
</tr>
<tr>
<td>Black male</td>
<td>-0.114***</td>
<td>1,243</td>
<td>(0.026)</td>
<td>0.136</td>
<td>-0.097***</td>
</tr>
<tr>
<td>White female</td>
<td>-0.064***</td>
<td>2,175</td>
<td>(0.010)</td>
<td>0.190</td>
<td>-0.106***</td>
</tr>
<tr>
<td>Black female</td>
<td>-0.083***</td>
<td>1,292</td>
<td>(0.025)</td>
<td>0.203</td>
<td>-0.169***</td>
</tr>
</tbody>
</table>

Source: NLSY 79 and NLSY97 data.
Notes: 1) All statistics are weighted with the NLSY provided cross-sectional weights. Standard errors are reported in parenthesis.
2) Control variables that are not reported include family background (parental education, parental employment in the year before the initial survey, family household income in the first initial survey year), presence of parents at age 12 or 14 (only mother present, only father present, or neither mother or father), current individual characteristics (marital status, and part-time work) and residence (region of residence, urban, MSA, county-level log median income, county-level crime rates and county-level unemployment rate).
3) Symbol † indicates significance of test for the equivalence of predicted probabilities across the two cohorts. * for significance of marginal effects in logit estimation. †††, ***Significant at 1% level; ††, **significant at 5% level; †, * significant at 10% level.
<table>
<thead>
<tr>
<th>Variables</th>
<th>1979 cohort</th>
<th></th>
<th>1997 cohort</th>
<th></th>
<th>Comparison across cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1) AFQT</td>
<td>(2) Education</td>
<td>(3) N</td>
<td>Pseudo-R²</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>-0.008</td>
<td></td>
<td></td>
<td>1,937</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.007</td>
<td></td>
<td></td>
<td>1,039</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>-0.013</td>
<td></td>
<td></td>
<td>851</td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td>(C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>-0.009**</td>
<td></td>
<td></td>
<td>2,023</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.006</td>
<td></td>
<td></td>
<td>1,094</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>-0.015***</td>
<td></td>
<td></td>
<td>878</td>
<td></td>
</tr>
<tr>
<td>(C)</td>
<td></td>
<td>(D)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>-0.006</td>
<td></td>
<td></td>
<td>1,937</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.004</td>
<td></td>
<td></td>
<td>1,039</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>-0.008</td>
<td></td>
<td></td>
<td>851</td>
<td></td>
</tr>
</tbody>
</table>

Source: NLSY 79 and NLSY97 data.

Notes: 1) All statistics are weighted with the NLSY provided cross-sectional weights. Standard errors in parenthesis.

Only women who have giving birth are included.

2) Control variables that are not reported include family background (parental education, parental employment in the year before the initial survey, family household income in the first initial survey year), presence of parents at age 12 or 14 (only mother present, only father present, or neither mother or father), current individual characteristics (marital status and part-time work) and residence (region of residence, urban, MSA, county-level log median income, county-level crime rates and county-level unemployment rate).

3) Symbol †† indicates significance of test for the equivalence of predicted probabilities across the two cohorts.

* for significance of marginal effects in logit estimation. ‡‡‡ , ***Significant at 1% level; †† , **significant at 5% level; † , * significant at 10% level.
Table 9. Logistic Regression Results (Marginal Effects) on Poverty for 1979 & 1997 Cohorts

<table>
<thead>
<tr>
<th>Variables</th>
<th>1979 cohort</th>
<th>1997 cohort</th>
<th>Comparison across cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>AFQT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo-R²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.065***</td>
<td>6.816</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.055***</td>
<td>2.106</td>
<td>0.056***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>-0.087***</td>
<td>1.243</td>
<td>-0.050**</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.055***</td>
<td>2.175</td>
<td>-0.022*</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>-0.059***</td>
<td>1.292</td>
<td>-0.111***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.030***</td>
<td>7.199</td>
<td>-0.065***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.030**</td>
<td>2.246</td>
<td>-0.038**</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>-0.047***</td>
<td>1.309</td>
<td>-0.070***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.021</td>
<td>2.301</td>
<td>-0.084***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>-0.024*</td>
<td>1.343</td>
<td>-0.051***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.059***</td>
<td>-0.017*</td>
<td>6.816</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>White male</td>
<td>-0.047**</td>
<td>-0.020</td>
<td>2.106</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>Black male</td>
<td>-0.077***</td>
<td>-0.025**</td>
<td>1.243</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>White female</td>
<td>-0.053***</td>
<td>-0.006</td>
<td>2.175</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Black female</td>
<td>-0.050***</td>
<td>-0.026</td>
<td>1.292</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.016)</td>
<td></td>
</tr>
</tbody>
</table>

Source: NLSY 79 and NLSY97 data.

Notes: 1) All statistics are weighted with the NLSY provided cross-sectional weights. Standard errors in parenthesis.
2) Control variables that are not reported include family background (parental education, parental employment in the year before the initial survey, family household income in the first initial survey year), presence of parents at age 12 or 14 (only mother present, only father present, or neither mother or father), current individual characteristics (marital status, dummy for having children, part-time work) and residence (region of residence, urban, MSA, county-level log median income, county-level crime rates and county-level unemployment rate).
3) Symbol †† indicates significance of test for the equivalence of predicted probabilities across the two cohorts. * for significance of marginal effects in logit estimation. †††, ***Significant at 1% level; ††, **significant at 5% level; †, * significant at 10% level.