The Causal Effects of Wages, Wealth, and Health on Retirement: Estimates Based on Subjective Conditional Probabilities*

Péter Hudomiet (RAND)

Michael D. Hurd (RAND, NBER, NETSPAR)

Susann Rohwedder (RAND, NETSPAR)

December 2020

ABSTRACT

This paper uses subjective conditional probabilities to study the causal effects of health, wage rate, wealth, and longevity on retirement. We fielded a survey in the RAND American Life Panel that asked about the subjective unconditional probabilities of working after ages 62, 65, and 70, and about the subjective probabilities of working after age 70 under varying conditions. For example, we asked about the probability of working after age 70 conditional on being in good health and conditional on being in bad health. The difference between the two conditional probabilities provides the subjective causal effect of health, or other factors, on retirement. We carried out a number of internal and external consistency checks of the reported probabilities, such as testing if they satisfy basic probability laws and comparing them to conditional outcomes, and found the quality of the data to be high. Then we found that the subjective causal effects on retirement of alterations in the wage rate, of large windfall gains in wealth, and of health are substantial. In contrast, the subjective causal effect of longevity is small. The estimated elasticities were typically larger for workers. Finally we fitted a flexible retirement model on the subjective expectations as well as on actual labor supply to estimate the elasticities of these factors on labor supply at various ages between 50 and 80. The model estimates were closely comparable for both types of data. We conclude that subjective conditional probabilities provide a rich framework and they can be a valuable addition to revealed preference data in estimating causal effects of economic incentives on labor supply.

Key words: subjective causal effects, expectations, retirement, health, economic incentives

JEL codes: J26, D84

* This research was supported by the National Institute on Aging [P01 AG008291] and the Alfred P. Sloan Foundation [G-2016-7199).

Introduction

Because of increasing life expectancies, it has been frequently argued that it would be desirable if people retired at later ages (Maestas and Zissimopoulos, 2010; Turner, 2011). Longer working lives would improve older individuals' financial security and it would potentially relieve some of the financial pressures on entitlement programs such as Social Security, Medicare and Medicaid in the U.S. But to develop policies that would encourage later retirement requires quantification of the effects of those policies both in the population and in subpopulations. Estimating the causal effects of policies on retirement is challenging because it is difficult to find convincing exogeneous variation in those factors that policy would manipulate. Therefore most of the literature uses observational statistical methods, microsimulation models or structural econometric models that by nature rely on strong assumptions.

This paper develops an alternative strategy to study causality. We propose to estimate the *subjective causal effect* of economic incentives and health on retirement using novel survey questions. We ask individuals about the probabilities that they would work after age 70 conditional on varying future health, wage rate, wealth levels, and mortality scenarios, and we use the change in their stated subjective probabilities of retirement to quantify the causal effects. This conditional probability approach provides an alternative to the use of quasi-experiments or randomized controlled trials to study causality, and if successful, it could be applied across a wide range of topics in social science.

A large literature has focused on understanding how economic incentives, health, and other socio-economic factors affect retirement. See recent summaries by Blundell, French & Tetlow (2016) and Fisher, Chaffee & Sonnega (2016). Poor health is one of the strongest predictors of early retirement (Bound, Stinebrickner & Waidmann, 2010; Blundell, Britton, Costa Dias & French, 2017; French, 2005; French & Bailey, 2017; Hudomiet, Hurd, Rohwedder & Willis, 2018; McGarry 2004; Rice, Lang, Henley & Melzer, 2011; van Rijn, Robroek, Brouwer & Burdorf. 2014; Vandenberghe, 2019). Though the correlation between health and retirement is well established, it is not known how much of the correlation is due to a causal mechanism. It is difficult to estimate the causal effects of health on retirement, because it is challenging, if not impossible, to find exogeneous variation in health. Moreover, individuals with different levels of health likely differ in many important, unobserved characteristics, such as productivity, discount rate or job characteristics, which also influence retirement decisions.

The literature on the effect of economic incentives, such as wealth and earnings, on retirement is also large (Blundell, French & Tetlow, 2016; Bound, Cullen, Nichols & Schmidt, 2004; Daminato & Padula, 2020; French, 2005; Gruber & Wise, 2004; Gustman & Steinmeier, 2005; Hanappi & Nagl, 2019; Maestas et al, 2017; Maestas, Mullen & Strand, 2013). Though most of the literature uses non-experimental methods, there is limited quasi-experimental evidence on the effects of these factors. Brown, Coile, & Weisbenner (2010) used inheritance receipt to instrument wealth in a retirement model and found large effects of wealth on retirement. Sevak (2002) used fluctuations on the stock market to create an exogenous variation in financial wealth, and also found strong wealth effects. Imbens, Rubin & Sacerdote (2001) found that lottery winners tended to reduce their labor supply particularly when they were close to retirement. These approaches are interesting and encouraging, but they also rely on untestable assumptions, and it is less clear whether the local treatment effects identified by these papers generalize to the population.

Subjective *unconditional* probabilities, such as the probability of working after age 65, have been used in a wide variety of research, such as the effect of health on retirement, or forecasts of labor force participation (Baldini, Mazzaferro, Onofri, 2019; Bissonnette & van Soest, 2010; Manski, 2004; Hurd, 2009). There is little research, however, on *subjective conditional probabilities*, such as the probabilities of working after age 70 conditional on being in good health. By asking multiple questions with varying conditions, subjective conditional probabilities can be used to recover the *subjective* causal effects of the conditions. Like subjective unconditional probabilities in general, subjective conditional probabilities have the desirable property that they are properly scaled so that it may be possible to compare average responses to the corresponding actual outcomes. Such a comparison is not possible with qualitatively stated responses (Manski, 2004).

Subjective conditional probabilities are related to stated preferences, with the important difference that the probability format of subjective conditional probabilities offers individuals the opportunity to express their uncertainty in quantifiable ways (i.e. using the probability scale). Stated choices typically require the respondent to choose one or the other, which, in some plausible scenarios, can lead to mis-prediction (Manski,2004 and Juster, 1966).

In the RAND American Life Panel (ALP) we asked about the subjective unconditional probabilities of working at ages 62, 65, and 70. Then we asked a series of questions about the subjective probabilities that individuals would work after age 70, varying one-by-one their future health, income, wealth, and longevity.

An alternative is to query about joint probabilities such as the probability of good health *and* working after age 70. We hypothesized that respondents would find it easier to form and report conditional probabilities than joint probabilities, The latter is cognitively more demanding, and we had observed in an earlier survey that few respondents were able to compute the joint probability of two events in a relatively simple example. To find out whether our hypothesis was correct, we also queried some joint subjective probabilities along with the conditional probabilities.

We present indicators of the validity of the subjective probability measures: we examine how closely the subjective expectations compare to actual retirement behavior, to find if the responses are internally consistent; and we compare randomized question formats. We find that subjective unconditional and conditional probabilities show strong consistency with each other and with observed labor supply. For example, actual labor supply as observed in survey data varies by reported health similarly to responses to the corresponding subjective conditional probabilities. We also find that the responses to the conditional and unconditional probabilities satisfy the law of total probability to a great precision, which we use as a test of internal consistency. This stands in contrast to responses to joint probability questions which we found to be largely inconsistent with unconditional probabilities. Apparently respondents find it difficult to form joint probabilities. Breaking up joint events into single conditions increases the quality of answers.

We estimate the subjective causal effects of economic incentives and of health on working after age 70 by estimating the means of the subjective conditional probabilities. Our estimates of the subjective causal effects of health, income, and wealth on retirement are large, while the effect of longevity on retirement is considerably smaller. We find notable differences by current labor market status, and smaller differences by income, gender, race, and education.

We develop and estimate a joint model of (1) actual labor supply as observed in survey data, (2) the unconditional probabilities, and (3) all the conditional probabilities. The model permits us to estimate the effect of health, wealth, earnings, and longevity on labor supply, either as measured by revealed preference data or as measured by the subjective expectations, from age 50 to age 80.

Our results complement and expand on the limited prior studies of subjective conditional probabilities. Dominitz & Manski (1996) used conditional probabilities to estimate the subjective causal effect of schooling on wages. Zafar (2011) studied the subjective effects of major choice on future wages. Giustinelli & Shapiro (2019) studied the effects of health on retirement using data from an experimental module fielded in the Health and Retirement Study (HRS). Like them we find that conditional probabilities are promising tools to study causality, especially when experimental methodologies are not feasible or ethical to conduct.

Subjective conditional probabilities provide a rich framework to evaluate policies and provide evidence that, in some ways, is superior to natural experiments and randomized controlled trials. First, natural experiments and randomized controlled trials are often not feasible or just not available for social scientists. Instead, subjective conditional probabilities can be asked on a wide range of topics. Second, even when they are available, the population subject to the experiment or trial is often not the population that is the target of a policy. For example, an older cohort may have generated data about retirement but a policy change will be applied to a younger cohort. To predict the effect requires the assumption that younger and older cohorts behave similarly. Subjective conditional expectation questions can be directly asked from the target population, those younger individuals who are not yet retired. Third, subjective conditional probabilities provide information on possibly heterogeneous behavioral effects across individuals rather than a single "average treatment effect" such as is produced by a controlled trial (Deaton, 2010). In fact the framework permits estimating subjective causal effects at the individual level, which is typically not possible from revealed preference data. The individual effects can be related to other, observable data to find which types respond to the policy. Fourth, it is far cheaper to collect and analyze data on subjective conditional probabilities than designing and implementing randomized controlled trials.

Subjective causal effects: Theory

Suppose we are interested how factor X affects the retirement decisions of workers. X could indicate that the worker has good or bad health, has various levels of financial assets, or his or her job has certain wage rate or other characteristics.

We ask whether and by how much different values of X would change the fraction of the population working after age *t*, for example, how lowering the income tax of older workers would change labor supply after age 70.

The causal effect of the factor on the employment rate after age *t*, $\Delta(t)$, is given by:

$$\Delta(t) = F\left(W(t) \mid X = 1\right) - F(W(t) \mid X = 0), \tag{0.1}$$

where $F\left(W(t)|X=1\right)$ denotes the fraction of the labor force that would be working after age *t* if all workers had $X=1$, $F\left(W(t)|X=0\right)$ denotes the same fraction without X. This paper primarily focuses on labor supply at age $t=70$.

The fundamental identification problem of revealed preference data is that individuals can only be observed in one state, that is, when X is either 0 or 1 but not both. Unless the factor is randomly assigned, the observed difference between the set of individuals with and without X may not yield the causal effect of X.

Our approach is to replace $F\left(W(t)|X=j\right)$ with the average of the subjective conditional probabilities that individuals would work under the two hypothetical states:[1]

$$\Delta^S(t)=\frac{1}{N}\sum_{i=1}^{N}\Delta_i^S(t), \tag{0.2}$$

$$\Delta_i^S(t)=\Pr_i\left(W(t)|X=1\right)-\Pr_i\left(W(t)|X=0\right), \tag{0.3}$$

where $\Pr_i\left(W(t)|X=1\right)$ denotes individual *i*'s subjective conditional probability of working after age *t* if the factor applies, $\Pr_i\left(W(t)|X=0\right)$ denotes his or her subjective conditional probability of working without the factor, $\Delta_i^S(t)$ is the subjective causal effect on individual *i*, and $\Delta^S(t)$ is the average subjective causal effect of the factor.

The methodology only requires collecting data on subjective conditional probabilities, $\Pr_i\left(W(t)|X=j\right)$, one for each value of *j*. Because all individuals in the sample are asked about their work expectations if the factor is turned on and in the counterfactual state in which the factor is turned off, the fundamental problem of identification is avoided. This is a major advantage of this approach.

The basic version of the methodology can be used to identify the subjective causal effects of X at a particular age, such as age 70. Data on subjective conditional probabilities can also enter a parametric retirement model together with data on subjective unconditional probabilities and the actual labor supply of individuals. Such a model requires additional assumptions about functional forms, the relationship between actual and subjective labor supply, etc., but it allows estimating the effects of X on broader outcomes. We present estimates of such a model in the Results section.

[1] If one of the scenarios coincides with the status quo then there would be only one hypothetical scenario.

A related stated preference approach would ask individuals if they would work in the two scenarios (yes or no). Then the average of the yes/no answers might also identify the subjective effect of the factor. However, as pointed out by Manski (2004) and Juster (1966), that favorable outcome might not obtain. For example, if all individuals have a 45% probability of working past age 70, all may say they would not choose that outcome; yet in realization 45% would have. The main advantage of our approach is that it allows individuals to express their uncertainty about their future choices. For example, no one can know for sure how they would choose their labor force status in the future without fully observing all other important, stochastic factors that may affect their future choices.

The value of conditional expectations depends on whether survey answers represent beliefs that individuals use to forecast their own future behavior and situation, and whether their forecasts are accurate, at least on average. If individuals report their subjective expectations with error, but the error has a zero mean in the population, the average subjective causal effects, $\Delta^{S}(t)$, will still provide an unbiased estimate of $\Delta(t)$. If the error term has a non-zero mean, however, the causal effects may be biased. To learn about the properties of the expectation questions, we implemented a number of internal and external validity tests, which will be discussed in the Results section.

A potential concern with hypothetical questions, such as the subjective conditional expectations, is the "fill-in problem": do individuals provide *ceteris paribus* answers, or do they fill-in different unspecified future conditions? For example, when we ask about retirement probabilities conditional on a low future income, some individuals may infer that their (unspecified) future health would also be lower or that a low future income would accompany a job with reduced demands or harsh working conditions. In this paper we test the fill-in problem by randomly assigning individuals alternative question wordings, in which we either leave any other conditions unspecified (potentially suffering from the fill-in problem), or we explicitly specify some of these other conditions.

Data

The RAND American Life Panel

We designed and fielded a survey on the RAND American Life Panel (ALP). The ALP is an ongoing Internet panel survey with a sample of about 6,000 respondents over age 18, operated and maintained at RAND. It covers the U.S. population age 18 and over. The majority of the panel members have their own Internet access. RAND has ensured Internet access for the remaining panel members by providing a laptop or an Internet service subscription or both. Accordingly the sample does not suffer from selection due to a lack of Internet access. Post-stratification weights are provided so that after weighting, the ALP approximates the distributions of age, sex, ethnicity, education, and income in the Current Population Survey. About twice a month, respondents receive an email request to visit the ALP website to complete questionnaires that typically take no more than 30 minutes to finish. Respondents are paid an incentive of about $20 per 30 minutes of survey time, and pro-rated accordingly for shorter surveys. Response rates are typically between 75% and 85% of the enrolled panel members, depending on the topic, the time of year, and how long a survey is kept in the field. The turn-around time between questionnaire design and the fielding of a survey is short, facilitating rapid responses to new events or insights. Thus, surveys can be operated at high frequency, reducing the risk of missing events or the effects on households.

The ALP has conducted a large number of longitudinal surveys of its respondents, so that over time it has accumulated data on a wide range of covariates. For example, ALP respondents have been asked about their financial knowledge, their retirement planning, and hypothetical questions designed to reveal parameters such as risk aversion. They have been given the HRS survey instrument in modules one at a time over an extended period, so that we have responses to the HRS health queries, income and asset data and to the HRS cognitive battery. These data can be linked to the data collected in any other ALP survey such as ours.

The survey on subjective conditional probabilities

We fielded our survey in late 2017 and early 2018 on a sample of individuals over the age of 50 in the ALP.[2] Our analytic sample consists of the 2,183 people who: (1) answered some of the basic questions in the survey (labor force status, self-assessed health, personality questions); and (2) had some attachment to the labor force. The latter requirement was due to this study's focus on work and retirement. A person was considered to have some attachment to the labor force if any of the following was true:

1. He or she worked at the time of the survey.
2. He or she worked for at least five of the previous 15 years (if younger than 65).
3. He or she worked for at least five years after the age of 50 (if older than 65).

Those who did not meet any of these criteria were skipped out of the survey after a screener module.

[2] The survey #487 is available at https://alpdata.rand.org/index.php?page=data&p=showsurvey&syid=487.

Our primary analytic sample consists of 1,691 individuals aged 50-69. The conditional retirement probabilities were only asked in this age range, because these questions ask about the probabilities of working after age 70. The survey included an additional 492 individuals of age 70-80. This older sample is used in models of actual labor force status.

Table A1 in the online appendix shows basic characteristics of the sample. The sample is balanced in gender, age, and race, but it is more educated than the general U.S. population. This is a feature of the ALP in general. The sample is also more attached to the labor market due to the screener questions. For example, the fraction of 50-69-year-olds who are currently working is 68.7% in the ALP sample vs. 57.9% in the 2016 HRS (the latter not shown in the table). The sample is fairly diverse in subjective health, labor force status, and income. The younger (50-69) and older (70-80) samples are similar to each other, but the older sample members are more likely to be white and to be retired.

The survey queried individuals about their expectations to work after ages 62, 65, and 70. After an introductory sentence about probability questions, the survey asked:

> *What are the chances that you will be doing any work for pay after you reach age 62?*

with similar questions for target ages 65 and 70. We refer to these questions as *unconditional probabilities*, and we denote them by P62, P65, and P70. The survey also asked similar questions about working full-time:

> *What are the chances that you will working full-time after you reach age [62/65/70]?*

The conditional probability questions were similar to the unconditional ones, except that they specified certain hypothetical events. For example, to study the effect on labor supply of excellent , very good or good health (which for simplicity we refer to as "good health") we used the following question:

> *Suppose when you reach age 70 your health is excellent, very good or good. In that case what are the chances that you will be doing any work for pay after you reach age 70?*

A similar question was asked about P70 conditioning on future fair or poor health (which for simplicity we refer to as "bad health).[3] The survey also asked about the probability of the conditioning event:

[3] The health conditions (excellent, very good, good, fair, poor) were the same as the answer options to the subjective health question that asked individuals to rate their health on this five-point same scale.

What are the chances that your health will be excellent, very good or good at age 70?

We also elicited the joint probability as follows:

And what are the chances that both will happen: At age 70 your health will be excellent, very good or good, and you will be doing any work for pay after you reach age 70?

We wanted to know whether question ordering mattered, so we randomized the order: one group was first asked about the conditional probability of working at age 70 and then the joint probability; the order was reversed for the second group. We found that question order made little difference, and so we will not discuss the ordering further.

After the section on health and retirement conditional probabilities, we asked a series of other probability questions that conditioned on varying values of wealth, wages, and longevity. These questions used randomized question formats to study the fill-in problem; they will be reviewed in the results section.

All probability questions in the survey allowed respondents to answer “Don’t know” or simply to skip the question without providing any answers. Tables A2 and A3 in the online appendix show that very few people skipped the questions, but the ratio of “Don’t know” answers was non-negligible, sometimes reaching 10%. However, almost all people (98.3%) answered at least one probability question. In the descriptive part of the paper we use non-missing observations when showing the means of various variables. When we estimate the retirement model that combines data on labor supply from revealed preferences and from subjective expectations, we use multiple imputation to fill in the missing values.

Results

We present the results in three steps. First, we discuss a number of internal and external consistency checks of the unconditional, conditional, and joint probabilities. Second, we show descriptive results on the subjective causal effects of health, wage, and wealth on working after age 70. Third, we fit a retirement model on revealed preference and subjective data, which we use to estimate the effect of the conditions on actual and subjective labor supply at ages between 50 and 80.

Validation of subjective unconditional, conditional, and joint probabilities

Unconditional probabilities

We first compared the unconditional subjective retirement probabilities with actual labor supply. Table 1 shows such comparisons on six outcomes: working after age 62, 65, and 70, and working full-time after these ages. The subjective versions of these outcomes are the unconditional probabilities discussed earlier. The actual versions were created by averaging working and full-time working statuses among those sample members whose ages were exactly 62, 65, and 70. Because of conditioning on single-year ages, the sample sizes for the measures of actual labor supply are relatively small. The left two columns of Table 1 shows results based on the ALP sample. The right two columns replicate the analysis on a larger, independent sample: the 2012-2016 waves of the HRS (the waves that asked about P70). Not all measures are available in all samples. The HRS only asks about the probabilities of working *full-time* after ages 62 and 65, but not about doing any work after these ages. The ALP did not ask about work hours from self-employed workers, and so it is not possible to estimate the actual fraction of full-time workers in that sample.

Table 1 shows that both the observed and subjective probabilities of working fall with age in the ALP and HRS samples, as workers retire. Labor supply in the ALP somewhat exceed that in the HRS, because of the screener questions in the ALP. The ALP-HRS differential is smallest at age 70 when about two thirds of each sample are retired already.

The actual and subjective probabilities of working after age 65 are very close in the ALP (54.6% vs. 56.9). The differences are slightly larger at ages 62 (79% vs. 71%) and at age 70 (34.4% vs. 28.4%), though these differences are not statistically significant at 5% due to small sample sizes for the estimates based on observed labor supply. In the HRS (right two columns in Table 1), the actual and subjective probabilities of any work after 70 are very close (12.1% vs. 11.3%), and the actual and subjective probabilities of working full-time after ages 62/65/70 in the HRS are scattered around each other.

There can be discrepancies between the observed and subjective averages for many reasons, such as biases in the subjective measures, small sample sizes, cohort trends in labor supply, or differences in how respondents interpret "*after you reach age [62/65/70]?*" in the unconditional probability questions. Nevertheless, it is reassuring to see in Table 1 that the actual and subjective values are not systematically different.

The six panels of Figure A1 in the online appendix shows additional consistency checks of the unconditional probabilities. They show age patterns in the probabilities in the ALP and HRS samples in the total sample and by labor market status. All three unconditional probabilities have a largely flat age profile in both samples, which is expected in a steady state model with rational expectations. The figures also show that work expectations strongly vary by labor market status: they are basically zero among those who are out of the labor force, reflecting the actual relatively infrequent return to the labor market at these ages. Work expectations among workers slightly increase with age, which is expected if there is heterogeneity in retirement probabilities and those with higher retirement expectations retire earlier.

Conditional probabilities

As a consistency check of the conditional probabilities, we first compare the percent of the population working by health and the projected percent from the subjective conditional probabilities (Table 2). The first column shows the percent working at ages 68-72 aggregated from HRS waves 2006-2014 stratified by the actual health of individuals at those observed ages. For example, among those whose health was excellent, very good or good, 32.2% were working. The next column shows similar percentages from our ALP survey (36.8%). The last column shows the average subjective conditional probability: conditional on health being excellent, very good or good, the average probability is 35.4%. The actual and subjective probabilities of work conditional on fair or poor health are also similar in the three cases: the value based on observed labor supply in HRS is 14.3%, and 12% in the ALP, while the subjective value is 16%. In steady-state, where successive cohorts reach retirement age with similar expectations and similar determinants of retirement we would expect the averages in the three columns to be similar, and, indeed, they are.

Second, we investigate if conditional probabilities are consistent with the reported unconditional probabilities in the ALP, and with the law of total probability:

$$\Pr_i\left(W(70)\right)=\Pr_i\left(W(70)\,|\,H=G\right)\Pr_i\left(H=G\right)+\Pr_i\left(W(70)\,|\,H=B\right)\left(1-\Pr_i\left(H=G\right)\right), \tag{0.4}$$

where $\Pr_i\left(W(70)\right)$ indicates individual i's reported unconditional probability of working after age 70, $H=G$ indicates the event that health is good at age 70, and $H=B$ indicates bad health.

All terms in (0.4) are available in the survey, and we check if the reports satisfy the equation.

Table 3 shows that the average values of P70 calculated from the conditional probabilities are remarkably close to the unconditional probabilities in the total sample (29.1% vs. 28.6%). Moreover, the values are consistent in all subsamples stratified by current health.

The consistency of P70 based on the conditional probabilities is found throughout the distribution of P70, not just in the mean. Panel A of Figure 1 compares the cumulative distribution functions of the unconditional P70 with the P70 derived from the conditional probabilities. The two distribution functions closely track each other. Panel B of the figure shows the nonparametric regression of P70 from the conditional probability on the unconditional P70. The P70 value calculated from conditional probabilities line up very closely with the unconditional P70 below the value of 50%. At values of P70 greater than 50%, the calculated values fall short but the discrepancies are mostly minor.

Joint probabilities

To inspect the consistency of the reported joint probabilities with the unconditional probability, we use the equation

$$\Pr_i\left(W(70)\right)=\Pr_i\left(W(70),H=G\right)+\Pr_i\left(W(70),H=B\right). \tag{0.5}$$

The two joint probabilities on the right hand side of (0.5) are also available in the ALP survey. Table 3 shows that the joint probabilities imply P70 values that are substantially larger than the reported unconditional P70. The discrepancy is consistently large in all subgroups stratified by current health. Panel A of Figure A2 in the online appendix show a similarly large discrepancies in the cumulative distribution functions of the two variables. For example, some 20% of the implied values of P70 based on joint probabilities are 100% or more. Panel B shows the non-parametric regression of the two variables. The values of P70 calculated from the joint probabilities are substantially greater than the unconditional P70 at all parts of its distribution.

To investigate the reason for the discrepancy Table A4 in the appendix shows the components that are used in the calculation of P70 from the conditional and joint subjective probabilities. We found that respondents in all health categories do not distinguish between the conditional and joint probabilities. It seems people do not understand joint probabilities or they are not able to express accurately joint probabilities. We have found this lack of understanding joint probabilities in prior work (Hudomiet et al, 2018) where we tested individuals' knowledge of various laws of probability. An implication is that if an analysis needs joint probabilities, it is best to ask respondents about conditional and marginal probabilities and then compute the joint probability.

Overall, the unconditional and the conditional probabilities showed favorable characteristics: The reported values were consistent with each other and with values based on revealed preference data in the ALP and in the HRS. At the same time we found that joint probabilities are not reported accurately in the ALP, and so we do not use them further in this project.

The effects of health, income, wealth and longevity, and the fill-in problem

The effect of health

The subjective causal effect of good versus bad health is defined as the difference between the subjective probabilities of working conditional on good, and conditional on bad health:

$$\Delta_i^{health}\left(70\right) = \Pr_i\left(W(70) \mid H = G\right) - \Pr_i\left(W(70) \mid H = B\right). \tag{0.6}$$

Table 4 shows the unconditional P70 values stratified by current self-assessed health (measured at the time of the survey at ages 50-69); and compares them to P70 values that condition on future health (at age 70). The table is restricted to 1,349 individuals who answered all questions. The unconditional probability (column 1) varies by current health, as expected, but the main division is between those whose self-assessed health is excellent, very good or good versus those whose health is fair or poor. Except for those in "poor" health, the conditional probabilities (columns 3 and 4) vary less with current health, which indicates that conditioning works as intended: current health has a reduced effect on future labor force status if future health is controlled. There is some residual variation with current health which is to be expected because the conditioning is relatively coarse; thus, someone with current excellent health may expect to be toward the top of the "health good or better" band at age 70 whereas someone with current good health may expect to be toward the bottom of the band. Under the condition "health good or better" (column 3) the increase in P70 relative to unconditional P70 among those whose actual health is good or better is small, on the order of 6 percentage points (ppts). indicating that those individuals already put a fairly high probability on their health at age 70 being "good or better." That is, the conditioning contained only a modest amount of news so that respondents only modestly updated their probabilities. Among those whose initial health was fair or poor the gains are about 12 ppts, indicating that the conditioning contained considerable news. When the conditioning is to bad health (health fair or poor) the reductions in P70 are large among those initially in good or better health (12 to 15 ppts) again reflecting the large amount of news, and small among those already in fair or poor health, reflecting the small amount of news.

The subjective causal effect of health is large and, except for those initially in poor health, it does not vary much with current health. The overall effect, 19 ppts, is substantial compared with the current labor force participation rate in the older population: among those 70-74 the participation rate was 19.7%.

Table 5 shows our preferred causal effect estimate of health (and the to-be-discussed other conditions) on working after age 70 for 1,018 working and 412 non-working individuals who answered the two conditional probability questions.[4] The effect of health (good versus bad) on working after age 70 is 23.4 ppts among workers and 11.2 ppts among non-workers. Good health, thus, would more than double labor supply after age 70.

Table A5 in the online appendix shows linear regressions of the subjective causal effect on a series of demographic and economic predictors. Recall that we measured the subjective causal effect on the individual level, and hence running this regression is feasible. We found that the strongest predictor was labor force status: The causal effect of health on labor supply is largest for workers, independently of full or part time status, smallest for the retired, and it is in between for those who are not working and are not retired. The effect may be the smallest for the retired, because many of them do not plan to work under any circumstances. The causal effect is also somewhat smaller for wealthier households. Other covariates are not related strongly to the outcome variable.

The effect of income

In addition to eliciting the subjective response to income changes, we used these questions to test the fill-in problem by randomizing three versions of the questions regarding the effect of earnings on retirement. The first version used our preferred wording:

> ***Version 1:*** *Suppose that Congress changed the tax system in a way that all workers above age 70 would bring home 20% more in wages compared to what they currently make.*
>
> *In this case, what are the chances that you would be doing any work for pay after you reach age 70?*

The objective of the wording of this question was to encourage the respondent to think that the demands of the job would not increase, and that no unspecified treatment of taxes would come into play.

It is possible that some individuals implicitly condition on (fill-in) being in good health when they answer the Version 1 question, even though the question does not intend to condition on health. To test this,

[4] The total sample size in Table 5 (1,430) exceeds the sample size in Table 4 (1,349), because Table 5 does not restrict the sample to observations with non-missing health and unconditional P70 values. The estimated causal effects are similar in the two tables.

the conditional statement in Version 2 further specified that individuals' health would be good at age 70:

> ***Version 2:*** *Suppose that Congress changed the tax system in a way that all workers above age 70 would bring home 20% more in wages compared to what they currently make. Suppose further that when you reach age 70, your health would be excellent, very good or good.*

Similar responses to the Version 1 and Version 2 questions would suggest that Version 1 suffers from the fill-in problem (that some individuals assume good health). Conversely, if individuals provide higher probabilities of working when their health is explicitly conditioned to be good (Version 2), then we may conclude that Version 1 does not suffer from the fill-in problem.

Version 3 used a more compact wording:

> ***Version 3:*** *Now imagine that you earned 20% more than you do now…*

While short and simple, this wording leaves open any fill-in. For example, individuals may assume that they make more money, because they are in better health or because they got a much better job than what they currently have. Furthermore, Version 3 could be interpreted as specifying that earnings would be 20% higher at the current age, and possibly at all future ages.

We followed each version with a corresponding version concerning pay reductions of 20%. For example, Version 1 was followed by

> *Now suppose instead that Congress changed the tax system in a way that all workers above age 70 would bring home 20% less in wages compared to what they currently make.*

Each individual was assigned to the same version number in the wage decrease follow-up as in the wage increase question, and also in other experiments involving wealth and longevity.

The subjective causal effect of 20% wage increase would be

$$\Delta_i^{wage,*}(70) = \Pr_i\left(W(70) \mid \Delta y = 20\%\right) - \Pr_i\left(W(70) \mid \Delta y = 0\%\right) \tag{0.7}$$

where $\Pr_i\left(W(70) \mid \Delta y = 20\%\right)$ is the probability of working after age 70 if wages increase by 20%, and $\Pr_i\left(W(70) \mid \Delta y = 0\%\right)$ is the probability that the wage rate remains the same. The counterfactual conditional probability, however, is not available in the survey. Instead, we use the conditional probability with respect to a 20% wage cut, and define the subjective causal effect as

$$\Delta_i^{wage}(70) = \frac{\Pr_i\left(W(70)\,|\,\Delta y = 20\%\right) - \Pr_i\left(W(70)\,|\,\Delta y = -20\%\right)}{2} \;, \tag{0.8}$$

We divide this difference by 2, so that the estimated effect corresponds to a 20% (rather than 40%) change in wages.

Two alternative subjective causal effects can be defined:

$$\Delta_i^{wage,2}(70) = \Pr_i\left(W(70)\,|\,\Delta y = 20\%\right) - \Pr_i\left(W(70)\right) \tag{0.9}$$

$$\Delta_i^{wage,3}(70) = \Pr_i\left(W(70)\right) - \Pr_i\left(W(70)\,|\,\Delta y = -20\%\right) \tag{0.10}$$

(0.9) and (0.10) use slightly stronger assumptions than (0.8), but they allow comparing the effects of positive and negative shocks in wages. They assume that the unconditional $\Pr_i\left(W(70)\right)$ probability is equivalent with the counterfactual conditional probability if the scenario is *not* implemented (i.e. $\Pr_i\left(W(70)\,|\,\Delta y = 0\%\right)$). This would be true if people assign a 0% chance that Congress would adopt such tax changes in the future. This may be a reasonable assumption, given that there is no current discussion about such tax changes. But our preferred method is (0.8), because it is valid under milder conditions.

Table 5 shows our preferred causal effect estimate of wages based on the Version 1 wording. Among workers, a 20% increase in wages would increase labor supply at age 70 by 10.8 percentage points, from 31.2% to 42.1%. This is a very large effect, though it is less than half as much as the effect of good health. Similar to the health effect, the causal effect of a 20% wage increase is less strong among non-working individuals: this change would increase labor supply age 70 in this group at by 7 percentage points from 13.6% to 20.6%.

Table A6 in the online appendix investigates the fill-in problem by comparing the three wording versions. Version 2, which specified being in good health at 70, increased P70 under both conditions (i.e. if wages go up or down by 20%), and it also led to a slightly larger causal effects compared to the preferred version (11.5 v.s. 9.4 ppts). We interpret this as evidence that most individuals did not fill-in good health when answering the preferred unspecified question format (version 1). Version 3, which used a compact though somewhat ambiguous wording, lead to very different response patterns, and a substantially muted subjective causal effect (3 ppts). Overall, the preferred Version 1 wording appears superior to the other two versions.

So far we estimated the subjective causal effects of income by subtracting the reported probabilities conditional on a wage gain and on a wage loss. We did not directly ask about the probability conditional on *no change* in wages. But under the assumption that the reported unconditional probability of working past age 70 corresponds to "no change" in wages, we can compare the effects of wage gains to wage losses. Table A6 shows that the preferred Version 1 implies fairly symmetric effects, with a slightly larger response to wage losses: The effect of a 20% wage gain is 8.3 ppts, and the effect of a 20% wage cut is 10.4 ppts.

The effect of wealth

We randomized three versions of the questions regarding the effect of wealth on retirement:

> ***Version 1:*** *Now please think about your situation today, including your current health and financial situation. Suppose you were to inherit $500,000.*
>
> *In this case, what are the chances that you would be doing any work for pay after you reach age 70?*

This version explicitly directs the respondent toward a ceteris paribus interpretation. This is reinforced by specifying that the wealth shock is the result of an inheritance, rather than by, say, past saving which may cause the respondent to think of higher earnings.

> ***Version 2:*** *Suppose you were to inherit $500,000.*

This version does not restrict fill-in about (unspecified) aspects of the financial situation.

> ***Version 3:*** *Suppose you had $500,000 more in financial assets than you do today.*

The subjective causal effect of $500,000 of wealth is

$$\Delta_i^{wealth}(70) = \Pr_i\left(W(70) \mid \Delta a = \$500,000\right) - \Pr_i\left(W(70) \mid \Delta a = \$0\right). \tag{0.11}$$

The counterfactual conditional probability (i.e. working after 70 *without* inheriting $500k, $\Pr_i\left(W(70) \mid \Delta a = \$0\right)$) is not available in the survey, and we replace it with the unconditional probability, $\Pr_i\left(W(70)\right)$. Therefore, we assume that individuals either do not currently expect to inherit $500,000, or if they do, they interpreted our questions as receiving an *additional* $500,000. The question wording suggests such a ceteris paribus interpretation. Some individuals in the sample, however, may have expected a large inheritance and at the same time they did not interpret the conditional probability question as a ceteris paribus change in wealth. The subjective causal effects for these individuals would be biased toward zero, but we expect this bias to be small.

Table A7 in the online appendix shows that the three randomized question wording are very similar. The mean P70s are close, and the mean causal effects are all between -12.3 and -14.3 ppts. We conclude that respondents interpreted all three questions similarly. Because of simplicity, we suggest Version 2 in applications studying the causal effect of wealth.

Table 5 shows the causal effects of wealth on labor supply among workers and non-workers. Because of the similarities of the three question versions, we use all three in this table. Among workers, an additional $500k wealth would decrease labor supply at age 70 by 17.3 percentage points, from 35.7% to 18.4%. This is a very large effect, more than the effect of a 20% wage cut, and almost as much as the effect of health. Similar to the health and wage effects, the causal effect of wealth is also less strong among non-working individuals. Half million dollars would decrease labor supply in this group by only 3.1 ppts. We note that this effect is limited by the low baseline labor supply in this group, which is only 10.5%.

The effect of longevity

If people expect to live longer, they may need to retire at a later age. We randomized three versions of the questions regarding the effect of longevity on retirement.

> ***Version 1:*** *Now imagine that scientists discover a new medicine that adds an extra ten years to your life, and those would be 10 healthy years. All other aspects of your life would be unchanged…*
>
> ***Version 2:*** *Now imagine that scientists discover a new medicine that adds an extra ten years to your life, but all other aspects of your life would be unchanged.*
>
> ***Version 3:*** *Now imagine that scientists discover a new medicine that adds an extra ten years to your life.*

Version 1 and Version 2 again direct respondents toward a ceteris paribus interpretation, while Version 3 offers the simplest wording. Version 1 further specifies “good health” to test if people filled this condition in.

The subjective causal effect of longevity is

$$\Delta_i^{longevity}(70) = \Pr_i\left(W(70) \mid \Delta l = 10\right) - \Pr_i\left(W(70) \mid \Delta l = 0\right). \tag{0.12}$$

The counterfactual conditional probability (i.e. working after 70 *without* the discovery of this new drug) is not available in the survey, and is, again, replaced by the unconditional probability, $\Pr_i\left(W(70)\right)$. The assumption is that individuals either do not currently expect the discovery of such a drug, or if they do, they interpreted the condition as providing an *additional* 10 years of life compared to their current expectations. We believe this is a reasonable assumption.

Table A8 shows the performance of the three versions. There is little difference between Versions 2 & 3, and we conclude that adding the extra causal language "*but all other aspects of your life would be unchanged*" does not change the interpretation of the question much. Thus the fill-in problem seems to be unimportant in this case. Specifying that the added years are healthy years (Version 1), however, increases the effect of P70 significantly (9.1 ppts vs. 2.9 and 3.6 ppts in Versions 2 & 3). We can think of two explanations. The first is that if the added years are healthy years it is likely that health would be good at age 70, so we are observing a health effect on work due to fill-in. The second is that if the added years are healthy people will want more wealth to spend in the healthy state, and so will work longer. Either way, we Versions 2 & 3 appear better to estimate the pure effects of longevity and we prefer those.

Table 5 shows that the causal effects of longevity on labor supply, based on Versions 2 & 3, is only 2ppts among workers. This is considerably smaller than the effects of health, wages, and wealth. Longevity does not seem to be on the mind of workers when they consider retirement. The estimates effects, however, are somewhat larger among non-workers: 10 years of additional longevity would increase labor supply at 70 by 6 ppts, from 11.8% to 17.8%. Non-working individuals may more actively think about the possibility of running out of wealth due to longer than expected longevity.

A joint model of labor supply based on revealed preference data and on subjective expectations

So far we only analyzed the subjective causal effect of the different factors on labor supply at age 70. We have not considered the subjective effects at other ages, or the effects on actual labor supply as observed in revealed preference data. The main reason was data limitation: the causal effect estimates are based on the subjective conditional probabilities that only asked about subjective labor supply at age 70. This section shows that by introducing additional assumptions one can estimate the effect of the conditions more generally.

We propose a joint model of actual labor supply, the unconditional retirement probabilities, and the conditional retirement probabilities. The model is based on a duration model, in which the length of working life is the "duration" and retirement is the "event." The model allows the subjective expected labor supply to differ from the actual labor supply by adding a to-be-estimated deviation-term. And we allow the various factors to shift the retirement hazards up or down compared to the baseline scenario.

Methods

The baseline retirement hazard at age *t* is modeled as piecewise Gompertz,

$$h_i^*(t)=\begin{cases}\lambda_i\eta\exp(\eta t) & \text{if} \quad t\le 62\\ G_{62}\lambda_i\eta\exp(\eta t) & \text{if} \quad 62<t\le 65\\ G_{65}\lambda_i\eta\exp(\eta t) & \text{if} \quad 65<t\end{cases} \tag{0.13}$$

λ_i is the shape and η is the scale parameter of the hazard function. This model allows the retirement hazard to change slope at ages 62 and 65 (by G_{62} and G_{65}), when many workers retire. This functional form fits the labor supply observed in the ALP data well, as we will show below.[5]

The shape parameter λ_i varies across individuals,

$$\ln(\lambda_i)=\beta' x_i+u_i. \tag{0.14}$$

x_i indicates observable covariates, and u_i is unobserved heterogeneity. Standard retirement models do not include an unobserved heterogeneity term, because it is not identified from retirement data alone. The subjective expectations data, however, provide additional information about individuals retirement chances, which help identify unobserved heterogeneity in the joint model.

Let $S_i^*(t\,|\,s)$ denote the probability of surviving on the labor market (i.e. not retiring) from age *s* to age *t*. The online appendix shows that a transformation of this survival probability can be written as

$$\begin{aligned}\ln\left(-\ln S_i^*(t\,|\,s)\right)&=T(t,s)+\ln\lambda_i\\ &=T(t,s)+\beta' x_i+u_i\end{aligned}. \tag{0.15}$$

$T(t,s)$ is a function that captures the age pattern in retirement and λ_i shifts this curve up or down. The term $T(t,s)$ depends on parameters η, G_{62}, and G_{65}, but it does not depend on the λ_i, or the x_i-s or

[5] We considered an alternative model, in which a mass of workers retire at ages 62 and 65 (i.e. the survival function jumps at 62/65). Such a model fits the labor supply data similarly well, but it is hard to build jumps into subjective retirement models in a compact way.

u_i as is shown in the online appendix. In this piecewise Gompertz model, thus, the log-log transformation of the survival function has a regression-like representation, with the important difference that the constant term of the regression is replaced by a non-linear function, $T(t,s)$.[6]

Next we assume that the conditions, such as a 20% wage increase, can shift the survival curve up and down similarly to the observable x_i-s. That is, a condition $C=c$

$$\begin{aligned} \ln\left(-\ln S_i^*(t\,|\,s,C=c)\right) &= \ln\left(-\ln S_i^*(t\,|\,s)\right)+\gamma_c' x_i \\ &= T(t,s)+\beta' x_i+\gamma_c' x_i+u_i \end{aligned} \quad (0.16)$$

The effect of the conditions, $\gamma_c' x_i$, thus, may vary by observable characteristics. In this paper we assume that the conditions cannot affect $T(t,s)$, the baseline age-pattern in retirement.[7]

Next, the subjective unconditional probabilities of working after age *t* for someone who is working at age *s* is

$$\begin{aligned} \ln\left(-\ln S_i^{subj}(t\,|\,s)\right) &= \ln\left(-\ln S_i^*(t\,|\,s)\right)+\delta' x_i+m_{it} \\ &= T(t,s)+\beta' x_i+\delta' x_i+u_i+m_{it} \end{aligned} \quad (0.17)$$

$\delta' x_i$, captures the deviation of the expected labor supply from that observed in the revealed preference data of older cohorts, and it is allowed to vary with observables; m_{it} is a mean-zero measurement error term, specific to each probability question (working after age 62/65/70).

Finally, the conditional probability answers are

$$\begin{aligned} \ln\left(-\ln S_i^{subj}(t\,|\,s,C=c)\right) &= \ln\left(-\ln S_i^*(t\,|\,s,C=c)\right)+\delta' x_i+m_{itc} \\ &= T(t,s)+\beta' x_i+\delta' x_i+\gamma_c' x_i+u_i+m_{itc} \end{aligned} \quad (0.18)$$

The model assumes that the conditions affect the subjective and the actual retirement in the same way, by shifting the log-log survival curve by $\gamma_c' x$. That is, subjective probabilities may deviate from actual labor supply ($\delta' x_i$), but this deviation affects all conditional and unconditional probabilities in the same

[6] Many other survival models have similar representations, but different models have different $T(t,s)$ terms, or in some cases they require different transformations.

[7] This is a strong assumption, because some conditions may have different effects on retirement at younger or older ages, but the ALP data does not allow us to estimate such heterogeneity, because all conditional probabilities referred to a single age, 70. It would be an interesting extension to collect data on conditional probability questions at other ages, and incorporate those responses into this model.

way, on average. This is the strongest assumption in this model, but it is needed to identify the effects of the factors on retirement in the revealed preference data. The assumption essentially means that the conditional probability questions are valid in the sense that the reported changes in subjective probabilities approximate the true causal effects of the conditions on labor supply. The assumption may or may not hold, but it cannot be tested using the ALP data alone. This identifying assumption is more likely to hold if individuals are asked about important and salient life decisions, such as the effect of wealth on retirement. And it is more likely to be violated if a hypothetical condition covers a sensitive topic or if the condition is very difficult to imagine for respondents.

Finally, the model has a few other technical assumptions:

1. All errors terms (u_i, m_{it}, m_{itc}) are normally distributed with zero means.
2. All measurement error terms are uncorrelated except the ones that ask about the same target age (70). We assumed that the unconditional and conditional probability questions about working after 70 has a shared measurement error term, $m_{i70,shared}$, which is included in all questions, and additional uncorrelated measurement error terms.[8]
3. Missing probability answers are missing at random. This assumption is not very restrictive, because the actual labor supply is observed for all individuals, and almost all respondents (98.3%) answered at least one subjective probability question. The missing data is filled-in using multiple imputation.
4. We only use the randomized wording-versions of the conditional probability questions that were identified to work well in the previous section: all three wealth questions, Version 1 of the wage questions, and Versions 2&3 of the longevity questions. Other responses were considered missing, and they were imputed using multiple imputation.
5. Some individuals provide 0% or 100% answers to some probability questions, for which the log-log transformation is undefined. We assume that these answers were not exact but rounded: 0% responses referred to a true probability between 0% and 0.5%, and 100% responses refer to true probabilities in the interval 99.5%-100%. Then we treated these observations as partially missing, and imputed them using multiple imputation from these intervals.

[8] This assumption significantly improved the fit of the model at younger ages in joint models that incorporated many conditional probability questions simultaneously. Without this assumption the model primarily focused on fitting the data at age 70, because most questions asked about that age.

6. A person is considered retired according to the revealed preference data obtained in the ALP if he or she does not work, and for the multiple choice labor market status question in the ALP reported a status of retired, disabled, or homemaker. The working expectations of these individuals were close to 0% (See Figure A1 in the appendix), while the working expectations of other non-workers (such as the unemployed) were significantly higher.
7. $S_i^{subj}\left(t \mid s\right)$, above, refers to the work expectations of individuals who are not yet retired. We assumed that retired individuals' work expectations are shifted by and additional term $\delta_r ' x_i$.[9]
8. We model the effect of each condition compared to the baseline (unconditional) scenario.
9. The youngest age in our data is 50, but a few individuals are already retired at that age. We modeled the probability of working at age 50 as a probit:

$$S_i^*\left(50\right) = \Phi\left(\alpha ' x_i\right) \tag{0.19}$$

We estimated this model using the Markov Chain Monte Carlo method, which allows incorporating many equations and random variables, and makes multiple imputation relatively straightforward. The estimation was carried out in Matlab. We tested the estimation model on simulated data, and as Table A9 in the appendix shows the model accurately recovered the true model parameters.

Results

We first show results of the model without any demographic covariates. The output of the estimation model is shown in Table A10 in the appendix. Figure 2 shows the fit of the model on actual labor supply. The blue line shows the model's prediction, and the red circles indicate the fraction of the sample who are not yet retired in single age bins running from 50 to 80. The model estimates with standard errors are also available in the top left corner of Table 6.

5% of the sample is predicted to be retired already at age 50. The survival curve (i.e. the fraction who remain on the labor market) slowly decreases between age 50 and 62, the decline accelerates first at 62, then further at 65. The fraction of retired individuals continues to rise with age, but the rate of increase slows after age 70. Overall, 22% of the sample is predicted to be retired at age 62, 36% at age 65, 67% at

[9] The ALP survey asked about retirement expectations from both workers and non-workers. Most retired individuals gave very low probabilities of working in the future. Retirement, thus, is an absorbing state for most, but not all workers. The small, cross-sectional ALP sample does not allow us to separately model inflows and outflows from retirement. So we decided to model retirement as an absorbing state, but allow expectations to differ between working and retired individuals. The fitted value of our model was similar if we simply ignored the expectations of retired individuals.

age 70, and 79% at age 75. The fit of the model on actual labor supply is good, as the red circles scatter around the blue line. This was expected, though far from assured, because the joint model was fit on nine subjective variables (three unconditional and six conditional) and only on a single measure from the revealed preference data. Thus, the weight of the subjective information was higher than the weight on the revealed preference information in the estimates. It is reassuring that the model still fitted the observed labor supply well.

Figure 3 and the bottom left corner of Table 6 show the predicted survival-in-the-labor-market probabilities from age 50 to other ages based both on the revealed preference data and on the subjective probabilities. The actual and subjective labor supplies are close to one another, though the subjective probabilities are somewhat below the probabilities obtained from revealed preference data. This means that individuals, on average, expect to retire slightly earlier than workers are actually observed to do. This deviation of the expectations from actual retirement behavior among older individuals, however, is small. For example, the survival probabilities in the labor market from age 50 to 62 are 81.8% in the revealed preference data and 77.6% based on the subjective expectations. The corresponding numbers until age 75 are 22.2% and 17.8%.

Figure 4 and Table 6 and Table **7** show the effect of the conditions on estimates of labor supply at various ages based on revealed preference data and the subjective expectations. Baseline labor supply at age 65 is 63.9%. If all workers had good health, labor supply would be 70.2%, or 6.3 ppts higher. If all workers had bad health (i.e. fair or poor health), labor supply would be 46.3%, or 17.6 ppts lower. Moving the entire labor force from bad to good health would increase labor supply at age 65 from 46.3% to 70.2%, or by 23.9 ppts. These are very large effects. The effects on actual and subjective labor supply are similar, but this is due to the modeling assumptions. The effects at other ages are also similar in percentage point terms, though they are typically smaller, especially at the oldest ages. For example, labor supply at age 75 would be 27% and 10.4% in the two health scenarios, which is only a 16.6 ppts difference. Note that though the percentage point differences are smallest, the percentage differences are largest at the old ages, because baseline labor supply is low at old ages.

The wealth effects are also large. If all individuals in the U.S. got $500k, labor supply at age 65 would be reduced from 63.9% to 41.1%, or by 22.8 ppts. Similar to the health effects, we find that the effect of wealth on labor supply at older ages is smaller in percentage point terms, but larger in percentage terms.

The effect of the wage rate on labor supply is also large. A 20% increase in the wage rate after 70 would increase labor supply to 71.6% (by 7.7 ppts), and a 20% cut in wages would decrease it to 47.4% (by 16.5 ppts). We thus find a considerably larger effects of a wage cut compared to a wage increase on labor supply at age 65. The effects are more symmetric at older ages.

The effect of longevity on labor supply is positive, but small, in line with the estimates in the previous section.

We finally explored heterogeneity by population subgroups. To allow for maximum flexibility, we re-estimated the joint model by gender, race, and education. Table 8 shows labor supply at age 65 in these groups under the baseline scenario, and under three hypothetical scenarios: good health, higher wealth, and higher wages. The table shows evidence that the effects of the conditions are largest in subgroups with lowest baseline labor supply. Though female labor supply is below that of men, female labor supply is more sensitive to the three conditions compared to men. The pattern is similar by education, though the effect of wages is similar across the education groups. Labor supply among racial groups do not vary systematically.

Discussion and Conclusions

We had several objectives. First, we wanted to produce evidence on the validity of subjective conditional probabilities. The value of conditional expectations depends on whether survey answers represent beliefs that individuals use to forecast their own future behavior and situation, and to make current decisions that will impact the future. We found conditional probabilities to be very consistent with unconditional probabilities: the calculated P70 based on the conditional probabilities and the unconditional health probabilities align well with the unconditional P70. We found conditional probabilities to line up very closely with actual conditional outcomes measures in the large HRS. We also found conditional probabilities to vary in systematic and meaningful ways with other variables: P70 increases when health is specified to be better, when wages are higher, and when longevity increases; P70 declines with a windfall wealth gain.

Second, we wanted to explore the fill-in problem. Mostly we found that responses were similar irrespective of specifying *ceteris paribus* (in common language), but that they changed substantially when adding an important condition such as good health. For example, the subjective causal effects of a wealth shock were similar across three alternatives that all used causal language in three different ways. When we randomized alternative versions of changes in the wage rate and mortality, we found that additional specification of about health significantly changed response patterns. We conclude that, while the fill-in problem is a potential concern, in our applications any evidence for its empirical importance is limited. Subjective conditional expectations questions should ideally be asked in a way that implies a *ceteris paribus* interpretation.

Our third objective was, conditional on favorable outcomes for the first two goals, to quantify causal effects on labor supply based on subjective conditional probabilities. We found that the subjective causal effect of health was large among both workers and non-workers. For example, labor supply at age 70 would be more than twice as large if all workers had good health vs. if all had fair or poor health. The effect of financial wealth was found to be large among workers, but less so among non-workers. The effect of a tax change that changed take home pay of workers above age 70 would also have a large effect of labor supply: A 20% increase in wages would increase labor supply at age 70 by about 10 ppts. Finally, we found that effect of longevity on labor supply is positive, but small. Longevity does not seem to be on the mind of workers when they consider retirement. An implication is that health and financial preparedness are the most important determinants of retirement, while general improvements in longevity are unlikely to prompt individuals to change their retirement plans. We find that the casual effects tended to be larger for those who worked at the time of the survey compared to those who did not, possibly because the latter group included individuals who already left the labor market and faced large hurdles of re-entry.

The conditional probability questions in the ALP survey all asked about labor supply at age 70. Our fourth objective was to develop a joint model of actual and subjective labor supply that allows estimating the effect of the conditions over the entire age range between 50 and 80. This model had to rely on additional assumptions, such as how the conditions shift the labor market hazard rates. The prediction of this model generally agreed with the descriptive results. We showed, however, that the effect of the conditions, expressed as a percentage of baseline labor supply, strongly increase with age, because baseline labor supply shrinks with age. This pattern lines up with retirement elasticity estimates from structural models, such as those reported in Blundell et al. (2016).

An interesting extension of our work could be to collect data on conditional probability questions about work after ages other than 70 and test if the patterns agree with the predictions of the fitted duration model. Another extension could be to expand the scope of the conditional probability questions. Our survey only asked about the effects of a $500k wealth shock, and 20% changes in the wage rate. It would be interesting to ask questions using additional values and test if the implied causal effect estimates are different.

An additional important future research topic concerns whether individuals use subjective conditional probabilities in making current decisions, whether their survey responses represent those beliefs and whether their survey responses are good forecasts of their own future behavior. Subjective conditional probabilities, such as stated preferences in general, are subject to the problem that the measures may differ from actual outcomes. Despite this potential drawback the technique solves the problem of unobserved tastes and characteristics that may undermine causal estimations based on revealed preference data.

References

Baldini, M., Mazzaferro, C. & Onofri, P. (2019). Pension expectations, reforms and macroeconomic downturn in Italy. What can microdata tell us? *Applied Economics*, 51(13), 1396-1410.

Bissonnette, L. & van Soest, A. H. O. (2010). Retirement Expectations, Preferences, and Decisions. *Netspar Panel Paper*; 06/2010-18. NETSPAR.

Blundell, R., French, E. & Tetlow, G. (2016). Retirement Incentives and Labor Supply. in: Piggott, John & Woodland, Alan (ed.), Handbook of the Economics of Population Aging, Elsevier, 457-566.

Blundell, R., Britton, J., Costa Dias, M., & French, E. (2017). The impact of health on labour supply near retirement. IFS Working Papers W17/18, Institute for Fiscal Studies.

Bound, J., Cullen, J., Nichols, A., & Schmidt, L. (2004). The welfare implications of changing disability insurance benefit generosity. Journal of Public Economics. 88, 2487-2514.

Bound, J. Stinebrickner, T. & Waidmann, T. (2010). Health, economic resources and the work decisions of older men. Journal of Econometrics, 156(1), 106-129.

Brown, J. R., Coile, C. C. & Weisbenner, S. J. (2010). The Effect of Inheritance Receipt on Retirement. *The Review of Economics and Statistics*. 92(2), 425-434.

Daminato, C. & Padula, M. (2020). "The Life-Cycle Effects of Pension Reforms: A Structural Approach," CSEF Working Papers 585, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.

Deaton, A. S. (2010). "Instruments, Randomization and Learning About Development." Journal of Economic Literature 48: 424-455.

Dominitz, J. & Manski, C. (1996): "Eliciting Student Expectations of the Returns to Schooling. *Journal of Human Resources*, 31: 1-26.

French, E. (2005). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour. *Review of Economic Studies* 72 (2):395-427.

French, E. & Bailey, J. J. (2017). Health, Health Insurance, and Retirement: A Survey. *Annual Review of Economics* 2017 9:1, 383-409.

French, E. & Jones, J.B. (2011). The Effects of Health Insurance and Self-Insurance on Retirement Behavior. Econometrica, 79(3), 693-732.

Gruber J. and Wise D.A. (eds.) (2004). Social security programs and retirement around the world: Micro-estimation. Chicago, IL: University of Chicago Press.

Giustinelli, P. & Shapiro, M. D. (2019). "SeaTE: Subjective ex ante Treatment Effect of Health on Retirement," NBER Working Papers 26087, National Bureau of Economic Research, Inc.

Gustman, A.L. & Steinmeier, T.L. (2005). The social security early entitlement age in a structural model of retirement and wealth. Journal of Public Economics, 89(2-3), 441-463.

Hanappi, T. & Nagl, W. (2019). "Retirement Behaviour in Austria: Effects of Incentives on Old-Age Labour Supply." *Fiscal Studies*, 40(3): 301-328.

Hudomiet P., Hurd, M. & Rohwedder, S. (2018). Measuring Probability Numeracy. *RAND Labor and Population Working Paper*, WR-1270.

Hudomiet P., Hurd, M. Rohwedder, S. & Willis, R. J. (2018). "The Effect of Physical and Cognitive Decline at Older Ages on Job Mismatch and Retirement," NBER Working Papers 25229, National Bureau of Economic Research, Inc.

Hurd, M. (2009). "Subjective Probabilities in Household Surveys." *Annual Review of Economics,* 1: 543-562. PMC3106311.

Imbens, G. W., Rubin D. B. & Sacerdote, B. I. (2001). Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players. *American Economic Review*, 91(4), 778-794.

Juster, F. (1966). Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design. *Journal of the American Statistical Association*, 61 (315):658-696.

Maestas, N., Mullen, K., Powell, D., von Wachter, T., & Wenger, J. (2017). Working Conditions in the United States - Results of the 2015 American Working Conditions Survey. *RAND Working Paper*, RR-2014-APSF. Santa Monica, CA: RAND Corporation.

Maestas, N., Mullen, K., & Strand, A. (2013). Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *American Economic Review*, 103(5), 1797-1829.

Maestas, N., & Zissimopoulos, J. (2010). How Longer Work Lives Ease the Crunch of Population Aging. *Journal of Economic Perspectives*, 24 (1):139-60.

Manski, C. (2004). Measuring Expectations. *Econometrica* 72 (5):1329-1376.

McGarry, K. (2004). Health and retirement: do changes in health affect retirement expectations? *Journal of Human Resources* 39 (3):624-648.

Rice, N. E, Lang, I. A., Henley, W, & Melzer, D. (2010). Common health predictors of early retirement: findings from the English Longitudinal Study of Ageing. *Age and Ageing* 40 (1):54-61.

van Rijn, R. M., Robroek, S. J. W., Brouwer, S., & Burdorf. A. (2014). Influence of poor health on exit from paid employment: a systematic review. *Occupational and Environmental Medicine* 71 (4):295-301.

Sevak, P. (2002). Wealth Shocks and Retirement Timing: Evidence from the Nineties. Michigan Retirement Research Center Working Papers wp027, University of Michigan.

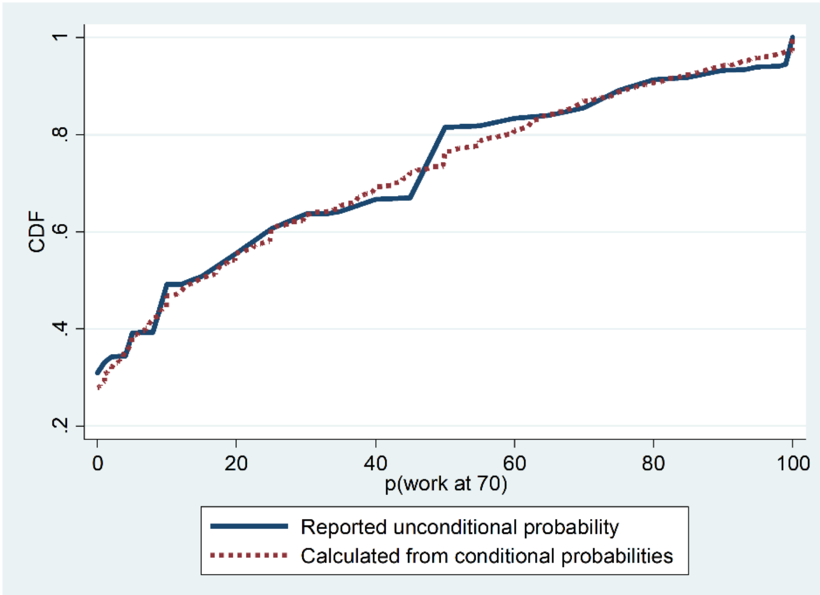
Turner, J. A. (2011). Longevity Policy: Facing Up to Longevity Issues Affecting Social Security, Pensions, and Older Workers. Books from Upjohn Press, W.E. Upjohn Institute for Employment Research.

Vandenberghe, V. (2019). "Health, Cognition and Work Capacity Beyond the Age of 50," GLO Discussion Paper Series 295, Global Labor Organization (GLO).

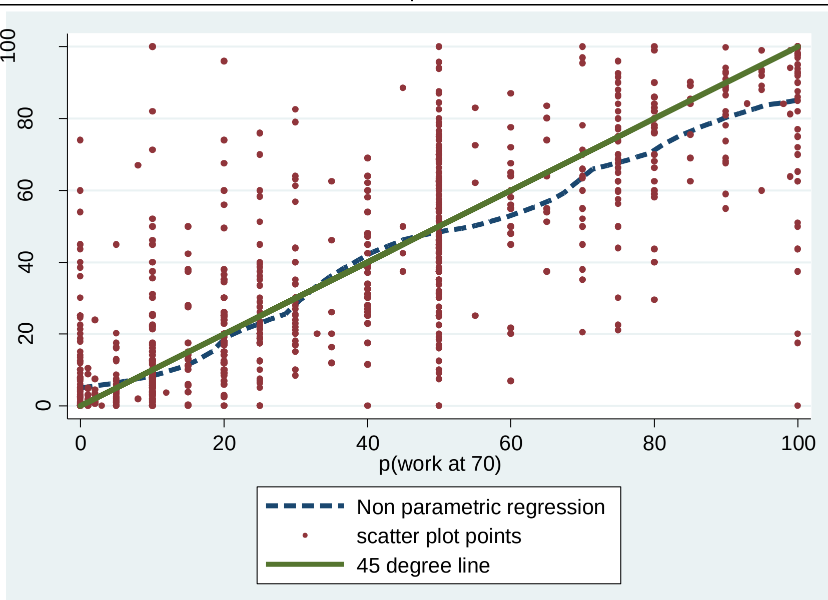
Zafar, B. (2011). "How Do College Students Form Expectations?" *Journal of Labor Economics*, 29(2): 301-348.

Figures and Tables

Figure 1. Comparison of reported unconditional P70 and P70 calculated from conditional probabilities



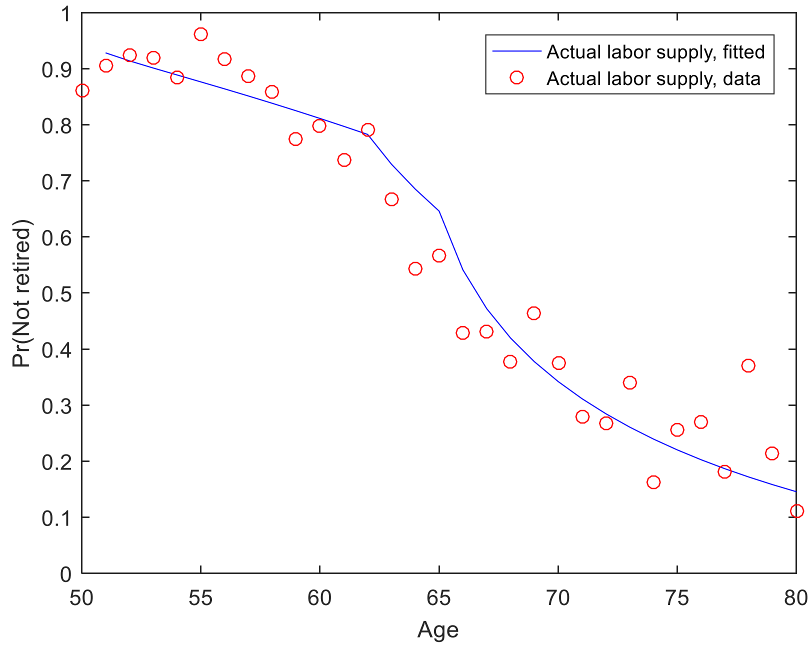
Panel A: The c.d.f.-s of reported and calculated P70.



Panel B: Non-parametric regression of calculated P70 on reported P70

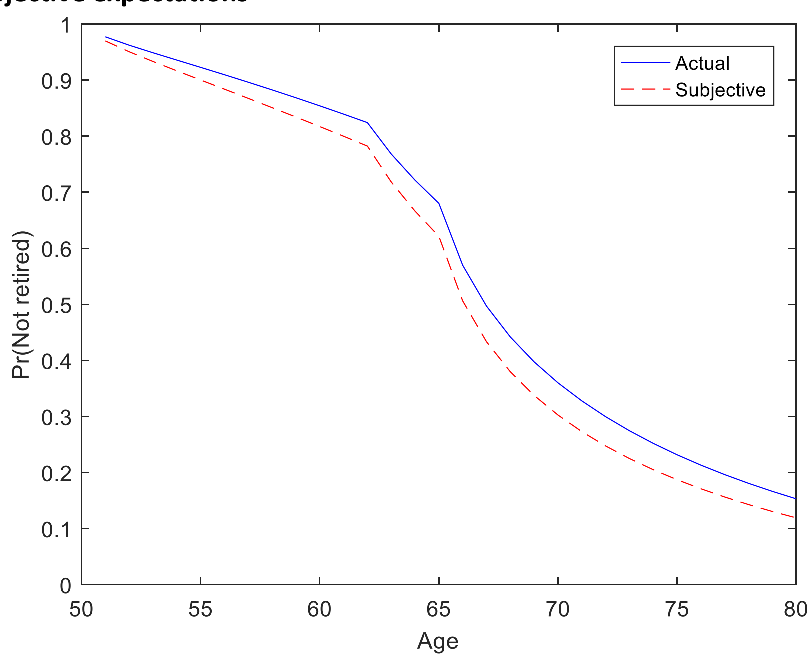
* ALP, Age 50-69, unweighted. Let Pr(W70|G) and Pr(W70|B) denote the reported probability of working past age 70 conditional on being in good or bad health at age 70, and Pr(G) denotes the probability of being in good health at age 70. The formula to calculate the probability of working past age 70 from conditional probabilities was: $P70^{cond}$ = Pr(W70|G) Pr(G) + Pr(W70|B)(1- Pr(G)).

Figure 2. Labor supply by age: comparison of data and model fit



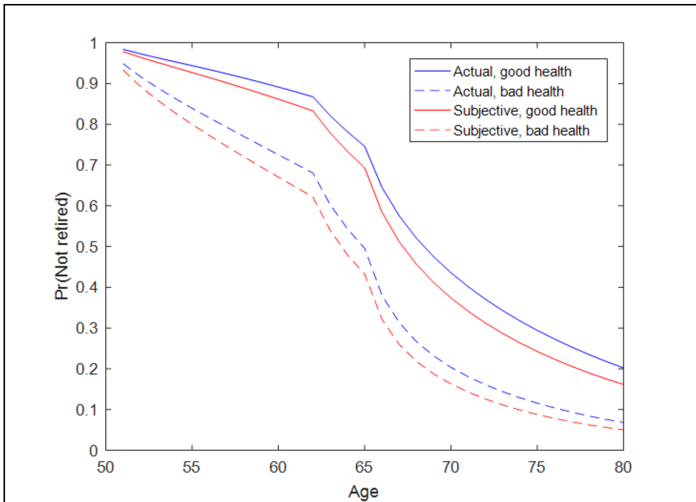
* ALP, Age 50-80, unweighted. The red circles show the fraction of individuals not yet retired in single age-bins. The blue line shows the fitted value of the model.

Figure 3. Survival curves of remaining in the labor market based on revealed preference data and on subjective expectations

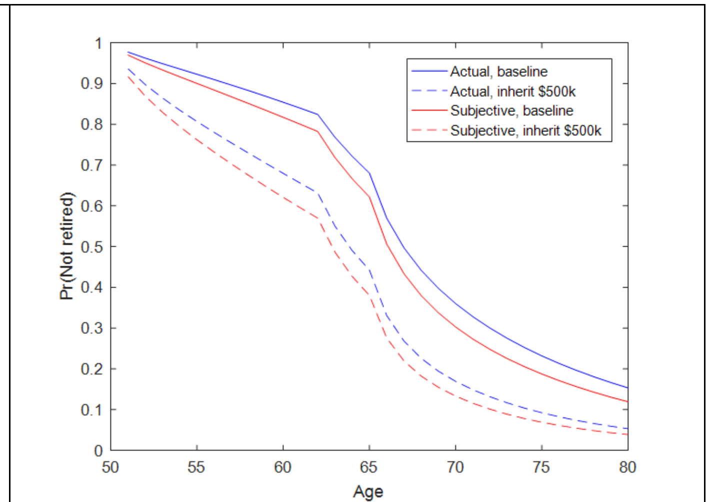


* ALP, Age 50-80, unweighted. The lines show actual and subjective probabilities of not being retired conditional on not being retired at age 50.

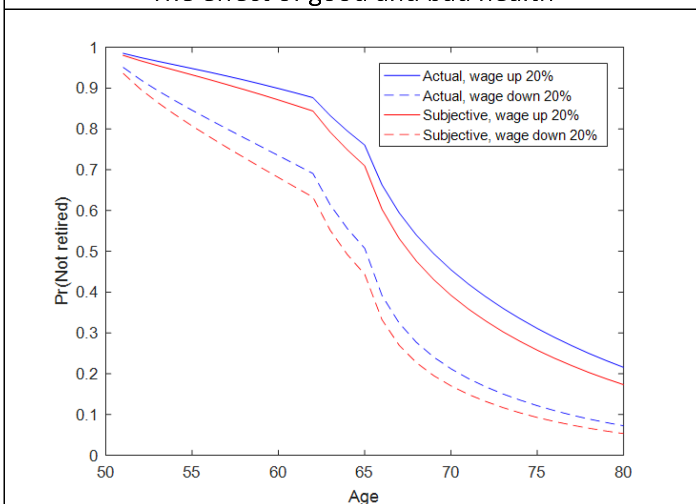
Figure 4. The effect of health, wealth, wages, and longevity on the labor market survival curves based on revealed preference data and subjective expectations



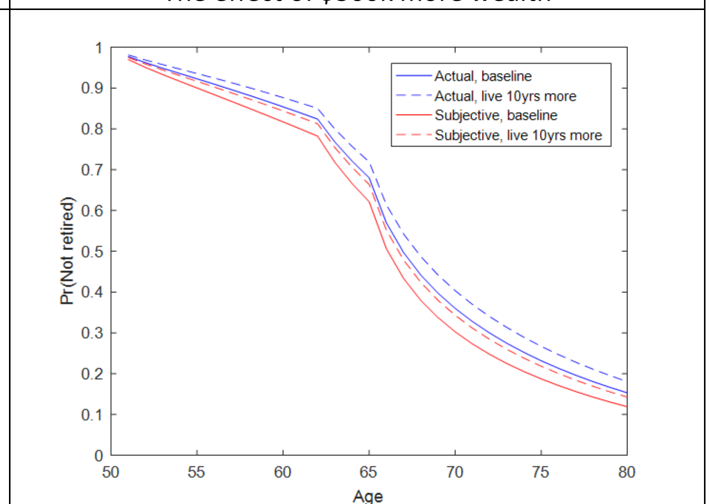
The effect of good and bad health



The effect of $500k more wealth



The effect of 20% wage increase and cut



The effect of 10 more years of longevity

Table 1. Actual and subjective probabilities of working at various ages, HRS and ALP

	ALP 2018		HRS 2012-2016	
	N	mean	N	mean
Works for pay, age 62	105	0.790	1,937	0.541
Prob work after 62	*913*	*0.710*	-	-
Works full-time, age 62	-	-	1,943	0.395
Prob work full-time after 62	*920*	*0.648*	*19,150*	*0.473*
Works for pay, age 65	90	0.546	1,678	0.430
Prob work after 65	*1,164*	*0.569*	-	-
Works full-time, age 65	-	-	1,681	0.257
Prob work full-time after 65	*1,179*	*0.488*	*24,389*	*0.314*
Works for pay, age 70	88	0.344	1,244	0.302
Prob work after 70	*1,463*	*0.284*	*31,105*	*0.205*
Works full-time, age 70	-	-	1,245	0.121
Prob work full-time after 70	*1,434*	*0.210*	*31,052*	*0.113*

* Italics indicate subjective expectations of working after ages 62/65/70 either in full-time or in any jobs. Non-italics indicate actual labor force status at age 62/65/70. Weighted statistics.

Table 2. Actual and subjective probabilities of working conditional on health, HRS and ALP

Health (actual or conditional)	Percent working, age 68-72 HRS	Percent working, age 68-72 ALP	Subjective conditional probability of working, ALP, age 50-69
good or better	32.2	36.8	35.4
fair or poor	14.3	12.0	16.0
Difference	17.9	24.8	19.4
N	13,845	375	1,431

* The actual probabilities of working measure the fraction of the HRS or ALP samples age 68-72 who are doing any work for pay, by their current health. The HRS values are based on the 2006-2014 waves; the ALP values are from wave 487. Weighted statistics.

Table 3. Mean subjective probabilities of working past age 70: comparing the unconditional reports with values created from conditional or joint probabilities, by current health

	N	Probability of working past age 70		
		Reported unconditional probability	Calculated from conditional probabilities	Calculated from joint probabilities
Current health	[1]	[2]	[3]	[4]
Excellent	168	30.2	33.9	54.4
Very good	499	32.5	31.3	55.2
Good	361	28.8	28	53.2
Fair	123	21.3	19.5	34.3
Poor	41	7.8	7.3	13.6
Total	1192	29.1	28.6	50.9

* ALP, Age 50-69, unweighted statistics. See notes under Figure 1 of the text for definitions.

Table 4. Mean subjective probabilities of working past age 70, unconditional and conditional on health at age 70, by current health,

	N	Probability of working past age 70			Subjective causal effect
		Reported unconditional	Conditional on health good or better	Conditional on health fair or poor	[3]-[4]
Current health	[1]	[2]	[3]	[4]	[5]
Excellent	184	29.5	36.9	16.3	20.6
Very good	556	32.3	36.6	17.0	19.6
Good	414	28.2	34.9	16.6	18.2
Fair	146	20.7	32.3	13.9	18.3
Poor	49	10.2	21.7	7.2	14.6
Total	1349	28.6	35.1	16.1	19.0

* ALP, Age 50-69, unweighted. "Health good or better" conditions on health being excellent, very good or good. Current self-rated health is measured at ages 50-69.

Table 5. The subjective causal effect of health, wealth, wage, and longevity on working past age 70

	N	Probability of working past age 70			
		No condition	With condition	Subjective causal effect, ([3] - [2])	Standard error of [4]
Panel A: Workers	[1]	[2]	[3]	[4]	[5]
Health: good or better	1018	19.8	43.2	23.4	[0.8]
Wealth: $500k more	1007	35.7	18.4	-17.3	[0.9]
Wage: 20% more	338	31.2	42.1	10.8	[0.8]
Longevity: 10 more years	662	34.1	36.1	2.0	[1.0]
Panel B: Non-workers					
Health: good or better	412	5.3	16.5	11.2	[1.0]
Wage: 20% more	138	13.6	20.6	7.0	[1.0]
Longevity: 10 more years	253	11.8	17.8	6.0	[1.5]
Wealth: $500k more	397	10.5	7.4	-3.1	[1.0]

* ALP, Age 50-69, unweighted.

Table 6. The predicted effects of health and wealth on labor supply

	Baseline		Effect of conditions					
			Good health		Bad health		500k more wealth	
Labor supply at 50	0.950	0.013	-	-	-	-	-	-
Labor supply at 62	0.777	0.010	0.042	0.004	-0.138	0.010	-0.184	0.013
Labor supply at 65	0.639	0.012	0.063	0.006	-0.176	0.009	-0.228	0.012
Labor supply at 70	0.333	0.012	0.073	0.007	-0.148	0.007	-0.181	0.009
Labor supply at 75	0.211	0.015	0.059	0.007	-0.107	0.008	-0.128	0.009
Survive 50-62 actual	0.818	0.014	0.044	0.004	-0.145	0.009	-0.194	0.013
Survive 50-62 subjective	0.776	0.017	0.051	0.005	-0.161	0.009	-0.214	0.013
Survive 50-65 actual	0.673	0.016	0.066	0.006	-0.186	0.010	-0.240	0.013
Survive 50-65 subjective	0.614	0.014	0.072	0.006	-0.191	0.009	-0.243	0.013
Survive 50-70 actual	0.351	0.014	0.076	0.008	-0.156	0.008	-0.190	0.010
Survive 50-70 subjective	0.293	0.008	0.072	0.007	-0.139	0.007	-0.167	0.009
Survive 50-75 actual	0.222	0.016	0.063	0.007	-0.113	0.008	-0.135	0.010
Survive 50-75 subjective	0.178	0.009	0.055	0.006	-0.095	0.006	-0.112	0.007

Table 7. The predicted effects of wages and longevity on labor supply

	Baseline		Effect of conditions					
			Wage up 20%		Wage down 20%		Live 10 more years	
Labor supply at 50	0.950	0.013	-	-	-	-	-	-
Labor supply at 62	0.777	0.010	0.051	0.006	-0.128	0.015	0.025	0.006
Labor supply at 65	0.639	0.012	0.077	0.009	-0.165	0.016	0.037	0.008
Labor supply at 70	0.333	0.012	0.091	0.012	-0.141	0.012	0.042	0.010
Labor supply at 75	0.211	0.015	0.075	0.011	-0.102	0.010	0.033	0.008
Survive 50-62 actual	0.818	0.014	0.053	0.006	-0.135	0.015	0.027	0.006
Survive 50-62 subjective	0.776	0.017	0.062	0.008	-0.150	0.016	0.031	0.007
Survive 50-65 actual	0.673	0.016	0.081	0.009	-0.174	0.017	0.039	0.009
Survive 50-65 subjective	0.614	0.014	0.088	0.010	-0.179	0.017	0.042	0.010
Survive 50-70 actual	0.351	0.014	0.096	0.013	-0.148	0.013	0.044	0.010
Survive 50-70 subjective	0.293	0.008	0.090	0.012	-0.132	0.011	0.041	0.010
Survive 50-75 actual	0.222	0.016	0.079	0.011	-0.107	0.010	0.035	0.009
Survive 50-75 subjective	0.178	0.009	0.070	0.010	-0.090	0.008	0.031	0.007

Table 8. The predicted effects of health, wealth, and wages on labor supply by subgroups

	Baseline		Effect of conditions					
			Good health		500k more wealth		Wage up 20%	
Labor supply at 65								
All	0.639	0.012	0.063	0.006	-0.228	0.012	0.077	0.009
Men	0.659	0.017	0.054	0.008	-0.222	0.017	0.043	0.013
Women	0.623	0.018	0.069	0.008	-0.232	0.019	0.104	0.013
Non-Hispanic white	0.644	0.013	0.056	0.006	-0.238	0.013	0.068	0.010
Non-Hispanic Black	0.633	0.049	0.071	0.027	-0.159	0.054	0.102	0.037
Non-Hispanic other race	0.672	0.051	0.083	0.041	-0.084	0.055	0.080	0.043
Hispanic	0.600	0.044	0.084	0.023	-0.190	0.050	0.081	0.035
High school or less	0.512	0.039	0.077	0.018	-0.287	0.038	0.066	0.032
Some college	0.640	0.020	0.067	0.009	-0.237	0.023	0.072	0.013
College	0.682	0.016	0.053	0.007	-0.206	0.016	0.077	0.013
Survive 50-65 subjective								
All	0.614	0.014	0.072	0.006	-0.243	0.013	0.088	0.010
Men	0.637	0.020	0.061	0.009	-0.236	0.017	0.049	0.014
Women	0.606	0.020	0.083	0.009	-0.256	0.019	0.126	0.015
Non-Hispanic white	0.628	0.015	0.061	0.007	-0.249	0.014	0.075	0.011
Non-Hispanic Black	0.603	0.060	0.107	0.032	-0.209	0.054	0.156	0.039
Non-Hispanic other race	0.517	0.058	0.103	0.048	-0.091	0.055	0.099	0.051
Hispanic	0.474	0.041	0.118	0.026	-0.217	0.047	0.115	0.046
High school or less	0.555	0.040	0.093	0.019	-0.326	0.042	0.079	0.038
Some college	0.591	0.025	0.080	0.011	-0.253	0.024	0.086	0.016
College	0.656	0.019	0.061	0.009	-0.224	0.015	0.089	0.015

Online Appendix

Derivation of the survival function

The hazard of retirement at age t is

$$h_i^*(t)=\begin{cases}\lambda_i\eta\exp(\eta t) & if \quad t\le 62\\ G_{62}\lambda_i\eta\exp(\eta t) & if \quad 62<t\le 65\\ G_{65}\lambda_i\eta\exp(\eta t) & if \quad 65<t\end{cases}. \tag{0.20}$$

The probability of surviving on the labor market from age *s* to age *t* is

$$S_i^*(t\,|\,s)=\begin{cases}\exp\left(-\lambda_i H(t\,|\,s)\right) & if \quad t\le 62, s\le 62\\ S_i^*(62\,|\,s)\exp\left(-G_{62}\lambda_i H(t\,|\,62)\right) & if \quad 62<t\le 65, s\le 62\\ S_i^*(65\,|\,s)\exp\left(-G_{65}\lambda_i H(t\,|\,65)\right) & if \quad 65<t, s\le 62\\ \exp\left(-G_{62}\lambda_i H(t\,|\,s)\right) & if \quad 62<t\le 65, 62<s\le 65\\ S_i^*(65\,|\,s)\exp\left(-G_{65}\lambda_i H(t\,|\,65)\right) & if \quad 65<t, 62<s\le 65\\ \exp\left(-G_{65}\lambda_i H(t\,|\,s)\right) & if \quad 65<t, 65<s\end{cases}, \tag{0.21}$$

where $H(t\,|\,s)$ is defined as

$$H(t\,|\,s)\equiv\exp(\mu t)-\exp(\mu s)\ . \tag{0.22}$$

After minor transformations (0.21) becomes

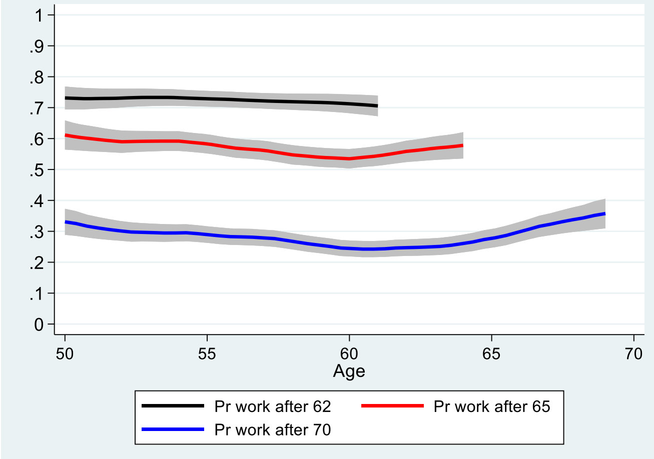
$$\ln\left(-\ln S_i^*(t\,|\,s)\right)=T(t,s)+\ln\lambda_i\ , \tag{0.23}$$

where

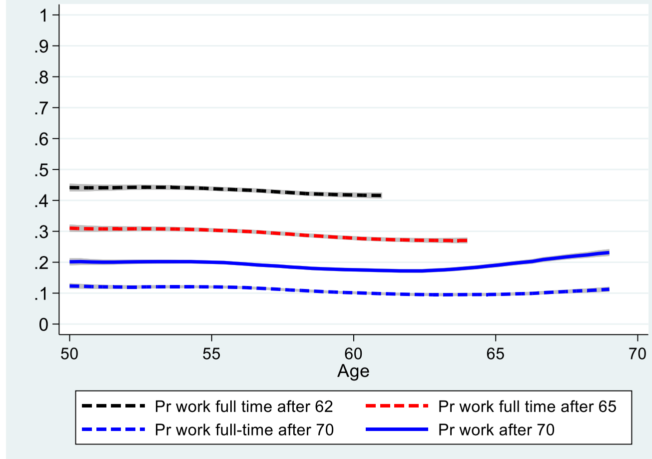
$$T(t,s)=\begin{cases}\ln\left[H(t\,|\,s)\right] & if \quad t\le 62, s\le 62\\ \ln\left[H(62\,|\,s)+G_{62}H(t\,|\,62)\right] & if \quad 62<t\le 65, s\le 62\\ \ln\left[H(62\,|\,s)+G_{62}H(65\,|\,62)+G_{65}H(t\,|\,65)\right] & if \quad 65<t, s\le 62\\ \ln\left[G_{62}H(t\,|\,s)\right] & if \quad 62<t\le 65, 62<s\le 65\\ \ln\left[G_{62}H(65\,|\,s)+G_{65}H(t\,|\,65)\right] & if \quad 65<t, 62<s\le 65\\ \ln\left[G_{65}H(t\,|\,s)\right] & if \quad 65<t, 65<s\end{cases}. \tag{0.24}$$

Additional Figures and Tables

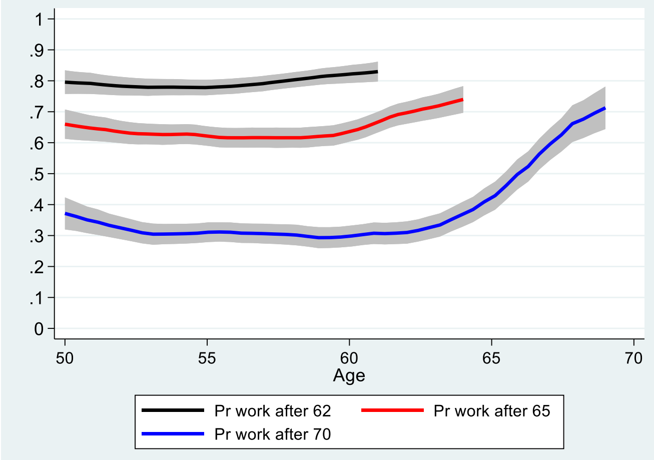
Figure A1. Age patterns in retirement expectations by labor force status, ALP and HRS



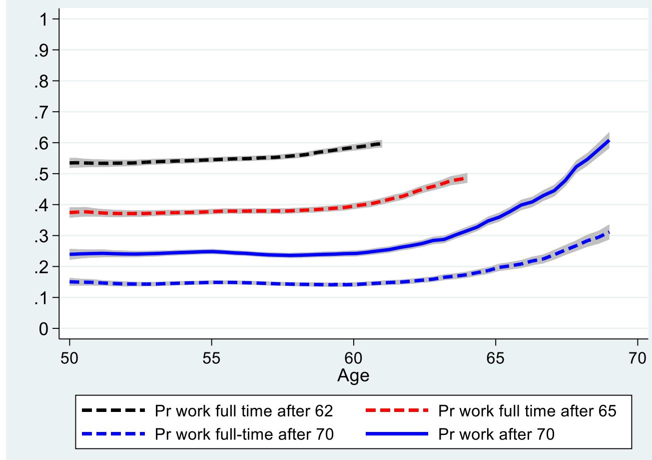
Panel A: Total sample, ALP 2018



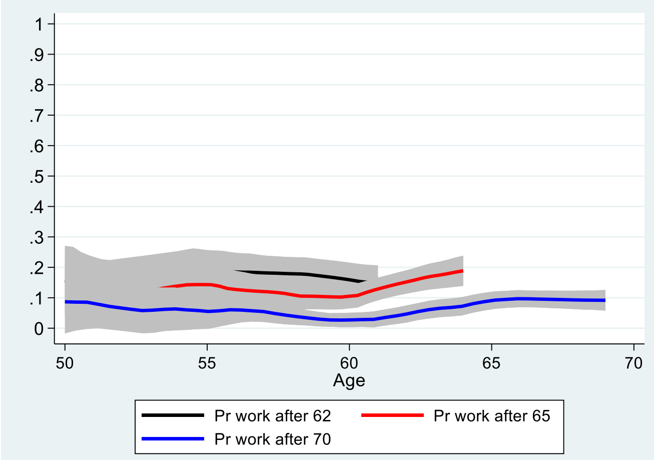
Panel B: Total sample, HRS 2012-2016



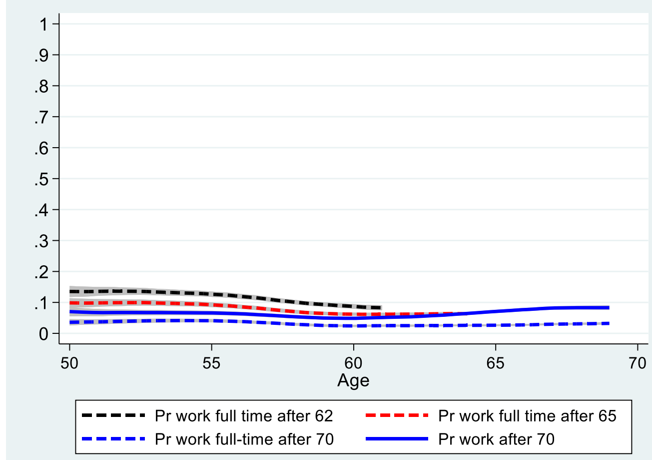
Panel C: In the labor force, ALP 2018



Panel D: In the labor force, HRS 2012-2016



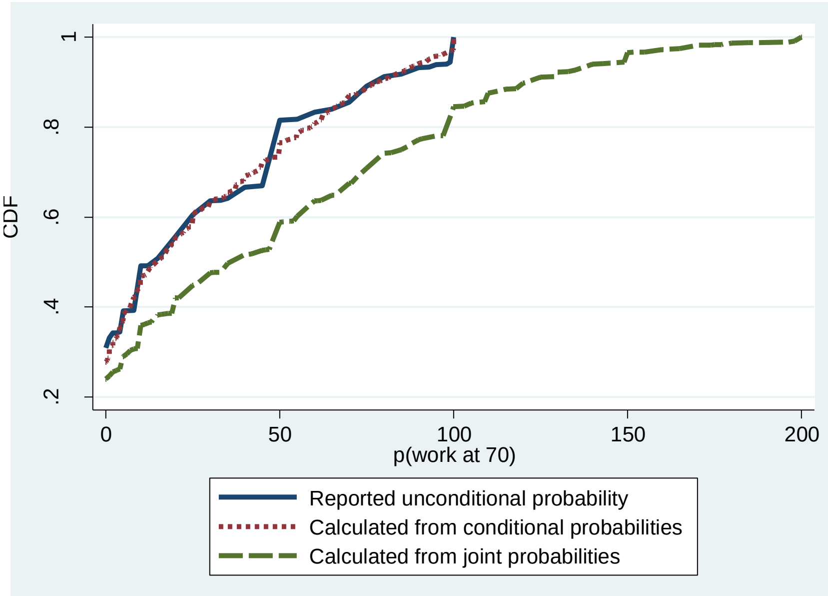
Panel E: Out of the labor force, ALP 2018



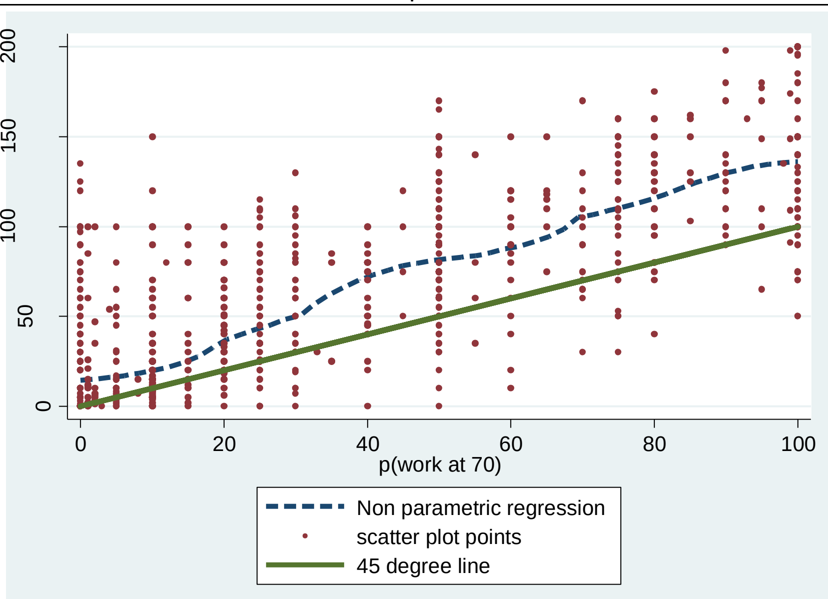
Panel F: Out of the labor force, HRS 2012-2016

* Kernel-weighted local polynomial smoothing. The solid lines show expectations of doing any work for pay after age X, the dashed lines show expectations of working full-time after age X.

Figure A2. Comparison of reported unconditional P70 and P70 calculated from conditional and joint probabilities



Panel A: The c.d.f.-s of reported and calculated P70.



Panel B: Non-parametric regression of P70 calculated from joint probabilities on reported P70

* ALP, Age 50-69, unweighted. Let Pr(W70|G) and Pr(W70|B) denote the reported probability of working past age 70 conditional on being in good or bad health at age 70, Pr(W70 & G) and Pr(W70 & B) denote the joint probabilities of working and good or bad health, and Pr(G) denotes the probability of being in good health at age 70. The formula to calculate the probability of working past age 70 from conditional probabilities was: $P70^{cond}$ = Pr(W70|G) Pr(G) + Pr(W70|B)(1- Pr(G)). Let The formula for the calculated P70 based on joint probabilities was: $P70^{joint}$ = Pr(W70 & G) + Pr(W70 & B).

Table A1. Summary statistics

	Age 50-69			Age 70-80		
	N	**Mean**	**Sd**	**N**	**Mean**	**Sd**
Age	1,691	60.0	5.4	492	73.6	2.9
Female	1,691	0.550	0.498	492	0.498	0.501
Less than high school	1,691	0.020	0.138	492	0.016	0.127
High school	1,691	0.132	0.339	492	0.110	0.313
Some college	1,691	0.355	0.479	492	0.364	0.482
College or more	1,691	0.493	0.500	492	0.510	0.500
White non-Hispanic	1,691	0.792	0.406	492	0.907	0.291
Black non-Hispanic	1,691	0.083	0.276	492	0.030	0.172
Other non-Hispanic	1,691	0.037	0.188	492	0.018	0.134
Hispanic	1,691	0.088	0.284	492	0.045	0.207
Married/Partnered	1,691	0.636	0.481	492	0.575	0.495
Divorced/Separated	1,691	0.200	0.400	492	0.201	0.401
Widowed	1,691	0.052	0.222	492	0.171	0.377
Never Married	1,691	0.111	0.314	492	0.053	0.224
Current Health: Excellent	1,689	0.125	0.331	492	0.081	0.274
Current Health: Very good	1,689	0.396	0.489	492	0.437	0.497
Current Health: Good	1,689	0.316	0.465	492	0.325	0.469
Current Health: Fair	1,689	0.121	0.327	492	0.124	0.330
Current Health: Poor	1,689	0.042	0.201	492	0.033	0.178
Works for pay	1,691	0.687	0.464	492	0.260	0.439
Retired	1,691	0.212	0.409	492	0.715	0.452
Disabled	1,691	0.044	0.206	492	0.002	0.045
Homemaker	1,691	0.014	0.116	492	0.004	0.064
Unemployed	1,691	0.021	0.144	492	0.004	0.064
Other labor force status	1,691	0.022	0.146	492	0.014	0.119

* Unweighted statistics. Labor force status is based on a multiple choice question, and the categories were defined sequentially. For example, "homemaker" means that the person does not work for pay, did not mention retirement or disability status, but mentioned being a homemaker.

Table A2. Answer patterns to the retirement probability questions

N = 1,691	**Valid answer**	**Mean**	**Missing because of age**	**Missing this version**	**Don't know**	**Skip**
Any work after 62	913	0.722	710	0	63	5
Any work after 65	1164	0.570	417	0	103	7
Any work after 70	1463	0.285	0	0	221	7
P70 if health good	1518	0.358	0	0	164	9
P70 if health bad	1503	0.153	0	0	174	14
P70 if inherit $500k, all else equal (V1)	530	0.158	0	1125	34	2
P70 if inherit $500k (V2)	530	0.135	0	1118	42	1
P70 if $500k more assets (V3)	511	0.157	0	1139	41	0
P70 if wage up 20% at 70, health good (V1)	502	0.468	0	1125	62	2
P70 if wage up 20% at 70 (V2)	493	0.360	0	1118	80	0
P70 if wage up 20% (V3)	474	0.336	0	1139	78	0
P70 if wage down 20% at 70, health good (V1)	503	0.209	0	1125	59	4
P70 if wage down 20% at 70 (V2)	510	0.163	0	1118	62	1
P70 if wage down 20% (V3)	483	0.256	0	1139	69	0
P70 if 10 more health years, all else equal (V1)	515	0.393	0	1125	49	2
P70 if 10 more years, all else equal (V2)	511	0.307	0	1118	61	1
P70 if 10 more years (V3)	491	0.324	0	1139	60	1

* Age 50-69. Unweighted statistics. The conditional probability questions used randomized question formats indicated by V1, V2, and V3.

Table A3. Distribution of the number of valid answers to the nine probability questions

# of valid probability answers	N	%
0	28	1.7
1	20	1.2
2	27	1.6
3	32	1.9
4	38	2.3
5	68	4.0
6	100	5.9
7	366	21.6
8	318	18.8
9	694	41.0
Total	1,691	100.0

* Age 50-69. Unweighted statistics.

Table A4. Subjective probabilities: working at age 70 conditional on health at age 70; working at age 70 and health at age 70, by current health, ALP, ages 50-69, unweighted

	N	Pr(G)	Pr(W70|G)	Pr(W70 & G)	Pr(W70|B)	Pr(W70 & B)
Current health	[1]	[2]	[3]	[4]	[5]	[6]
Excellent	171	78.6	37.7	38.1	17	16.7
Very good	511	71.6	37	37.1	17	18.4
Good	372	57.9	35.2	32.9	17	19.6
Fair	129	22	31.8	16.2	14.1	16.9
Poor	43	6.6	22	5.2	6.1	7.8
Total	1226	60.9	35.5	32.6	16.3	18

* Pr(G) denotes the probability of being in good, very good or excellent health at age 70. Pr(W70|G) and Pr(W70|B) denote the probability of working past age 70 conditional on being in good or bad health at age 70. Pr(W70 & G) and Pr(W70 & B) denote the joint probabilities of working and good or bad health.

Table A5. Linear regression model of the subjective causal effect of health on retirement

	coef.	s.e.
Age 50-54	ref.	
Age 55-59	-2.563	(1.835)
Age 60-64	2.553	(1.863)
Age 65-69	6.164***	(2.153)
Female	1.117	(1.377)
High school or less	ref.	
Some college	3.925*	(2.050)
College or more	4.277*	(2.257)
White non-Hispanic	ref.	
Black non-Hispanic	0.975	(2.508)
Hispanic	2.681	(2.393)
Other race	5.770*	(3.462)
Married/Partnered	-2.714*	(1.453)
Health: Excellent	ref.	
Health: Very good	-1.091	(1.967)
Health: Good	-1.481	(2.063)
Health: Fair	0.355	(2.613)
Health: Poor	-0.266	(3.689)
Probability Numeracy	1.182	(1.143)
Number Series Score	0.00268	(0.0292)
Family Income, $1,000	-0.0273***	(0.0105)
Retired (and not working)	ref.	
Other not working	13.68***	(2.746)
Working full time	18.49***	(1.947)
Working part time	18.80***	(2.239)
Self-Employment	5.881***	(2.012)
Current/Last Job Characteristic Score: Cognitive	0.0721**	(0.0303)
Current/Last Job Characteristic Score: Physical	0.0328	(0.0208)
Current/Last Job Characteristic Score: Social	-0.0166	(0.0304)
Constant	-2.742	(16.61)
R-squared	0.108	
N	1410	

* ALP, Age 50-69, unweighted.

Table A6. Mean subjective probabilities of working past age 70, unconditional and conditional on higher/lower wage

	N	Probability of working past age 70			
		Unconditional	Conditional on 20% wage increase	Conditional on 20% wage cut	Subjective causal effect of 20% wage, ([3] - [4])*0.5
Randomized versions	[1]	[2]	[3]	[4]	[5]
V1: wage change at age 70	437	27.4	35.7	17.0	9.4
V2: wage change at 70 and health good at 70	436	30.6	45.1	22.1	11.5
V3: wage change immediately	423	27.8	32.2	26.3	3.0
All	1296	28.6	37.7	21.7	8.0

* ALP, Age 50-69, unweighted statistics. Condition in version 1: Congress changes the tax system so that workers above age 70 make 20% more/less. Condition in version 2 is the same, plus that the individual's health is good, very good or excellent at age 70. The condition in version 3: The wage of the person is 20% more/less than today.

Table A7. Mean subjective probabilities of working past age 70, unconditional and conditional on higher wealth

	N	Probability of working past age 70		
		Unconditional	Conditional on $500k more in wealth	Subjective causal effect, ([3] - [2])
Randomized versions	[1]	[2]	[3]	[4]
V1: $500k inheritance, all else equal	471	30.3	16.1	-14.2
V2: $500k inheritance	477	27.4	14.1	-13.3
V3: $500k more assets	457	28.2	15.9	-12.3
All	1405	28.6	15.3	-13.3

* ALP, Age 50-69, unweighted statistics. Version 1 and 2 specify an inheritance of $500k, while version 3 specifies an increase of $500k of financial assets (source not specified). Version 1 further specifies that health and financial situation would not change except for the inheritance.

Table A8. Mean subjective probabilities of working past age 70, unconditional and conditional on longevity

	N	Probability of working past age 70		
		Unconditional	Conditional on 10 more years of life	Subjective causal effect, ([3] - [2])
Randomized versions	[1]	[2]	[3]	[4]
V1: 10 more health years, all else equal	471	30.3	39.4	9.1
V2: 10 more years, all else equal	477	27.4	30.3	2.9
V3: 10 more years	457	28.2	31.8	3.6
All	1405	28.6	33.9	5.3

* ALP, Age 50-69, unweighted statistics. Version 1 specifies 10 extra years in good health, and that all other aspects of life would be unchanged. Version 2 adds 10 years of life but makes no statement about health; all other aspects of life would be unchanged. Version 3 adds 10 years of life but makes no statements about other aspects of life.

Table A9. The output of the model on simulated data

	Param.	Coef.	S.e.
Basic parameters			
Scale parameter	1.5	1.570	0.071
Log shape constant (actual)	-4.0	-4.116	0.113
Shape deviation (subjective)	0.5	0.472	0.030
Shape deviation (retired)	1.0	0.979	0.048
Slope change at 62	1.0	1.013	0.048
Slope change at 65	2.0	2.003	0.071
Normal inverse of labor supply at age 50	1.5	1.482	0.031
Effects of conditions			
Good health	-0.5	-0.511	0.015
Bad health	1.0	1.010	0.016
Inherit 500k	1.5	1.495	0.018
Wage up by 20%	-0.5	-0.493	0.015
Wage down by 20%	1.0	1.012	0.016
Longevity 10 more years	-0.1	-0.100	0.015
Standard deviations of errors			
Unobserved heterogeneity	2.0	2.011	0.018
Noise in pw62	2.0	1.992	0.022
Noise in pw65	1.5	1.507	0.016
Noise in pw70	1.0	1.004	0.010
Noise in pw70_GoodHealth	1.0	1.012	0.009
Noise in pw70_BadHealth	1.0	1.000	0.011
Noise in pw70_500k	1.0	0.999	0.012
Noise in pw70_WageUp20%	1.0	1.003	0.009
Noise in pw70_WageDown20%	1.0	1.003	0.011
Noise in pw70_10MoreYears	1.0	0.998	0.010
Shared noise term for Pw70 questions	0.5	0.495	0.029

* We used a simulated sample size of 20,000. The estimation model was carried out by MCMC using 1,000,000 simulation draws, from which the first 100,000 was discarded as burn-in, and the rest of the 900,000 were used for inference.

Table A10. The output of the model on ALP data

	Coef.	S.e.
Basic parameters		
Scale parameter	1.251	0.320
Log shape constant (actual)	-4.131	0.543
Shape deviation (subjective)	0.408	0.109
Shape deviation (retired)	0.657	0.219
Slope change at 62	1.416	0.209
Slope change at 65	2.723	0.317
Normal inverse of labor supply at age 50	1.658	0.124
Effects of conditions		
Good health	-0.503	0.046
Bad health	1.224	0.053
Inherit 500k	1.566	0.073
Wage up by 20%	-0.625	0.078
Wage down by 20%	1.149	0.104
Longevity 10 more years	-0.291	0.067
Standard deviations of errors		
Unobserved heterogeneity	2.337	0.086
Noise in pw62	2.543	0.122
Noise in pw65	1.886	0.090
Noise in pw70	0.955	0.035
Noise in pw70_GoodHealth	1.190	0.036
Noise in pw70_BadHealth	1.278	0.042
Noise in pw70_500k	1.975	0.062
Noise in pw70_WageUp20%	1.337	0.063
Noise in pw70_WageDown20%	1.643	0.088
Noise in pw70_10MoreYears	1.659	0.055
Shared noise term for Pw70 questions	1.313	0.087

* The estimation model was carried out by MCMC using 1,000,000 simulation draws, from which the first 100,000 was discarded as burn-in, and the rest of the 900,000 were used for inference.