

Retirement Expectations

Gábor Kézdi and Matthew D. Shapiro

First Draft June 2021; Revised July 2022

ABSTRACT

Retirement expectations—measured either as expected retirement age or the subjective probability of working until a particular age—are widely available in longitudinal studies. This chapter evaluates the performance of these measures. It finds that they are excellent predictors of retirement, though their relation with realized retirement is substantially attenuated. A wide range of variables relevant for retirement have little incremental predictive once attenuation-adjusted subjective probabilities are taken into account. Because of their wide availability and high reliability, retirement expectations are widely used for analyzing late-in-life labor supply decisions and effects of retirement.

JEL: D84, J26

KEYWORDS: Subjective expectations; Subjective probability; Retirement; Treatment effects; Elicited probabilities; Labor supply of older workers

Forthcoming in the *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw.

Acknowledgement: Gábor Kézdi was Research Associate Professor, Survey Research Center at the University of Michigan. Matthew D. Shapiro is Director and Research Professor, Survey Research Center and Lawrence R. Klein Collegiate Professor of Economics at the University of Michigan and Research Associate at the NBER. We are grateful for comments from Pamela Giustinelli and the editors and to Meifeng Yang for assistance.

Citation: Gábor Kézdi and Matthew D. Shapiro. "Retirement Expectations." In *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Amsterdam: North-Holland/Elsevier, 2023.

1. Introduction

When to retire is a key decision made by working individuals. The retirement decision is closely tied to many decisions and outcomes bearing on later life—how much to save and consume, how to spend time, where to live. Retirement is affected by health, and might affect health outcomes itself. Retirement is an important component to explicit and implicit labor contracts. Retirement is regulated by public policy and social insurance. Because of the complex, intertemporal decision-making surrounding retirement, and because it is affected by so many factors, both internal and external to the individual, modeling retirement decisions is challenging.

In this chapter, we will argue that retirement expectations are a valuable complement to both reduced-form and structural modeling of retirement. People's decision about when to retire is likely affected by a host of factors and preferences that are at best partially observable to outsiders. Moreover, even when they are observed, their effects are likely to be heterogeneous. To take into account unobservable individual heterogeneity in these factors and their effects, a substantial and growing literature uses people's reported expectations of when they would retire to answer these questions. It is important to note that retirement is something that most people think about after a certain age, so it may very well be that their own predictions are informative. We indeed show that expectations are reliable predictors of behavior and that they reflect much of what affects retirement decisions. Analysis of the expectations about when to retire also serves as a methodological template for analyzing expectations about other decisions, especially those with an intertemporal dimension.

This chapter will evaluate the value of retirement expectations. It uses a simple discrete-choice model to clarify what is being elicited in standard survey questions concerning retirement expectations. Moreover, it presents evidence on the predictive performance of retirement expectations elicited in the Health and Retirement Study (HRS), implements a machine learning procedure to assess the predictive

content of retirement expectations, and surveys the wide literature using retirement expectations for prediction and analysis.

2. Theoretical framework

To provide context for the findings of the literature and identify open questions, we first consider a general theoretical framework and what it implies for the measurement strategies of retirement age expectations. We define retirement as the transition from full-time work to not working full-time. This definition is one of the several potential definitions of retirement, but it is simple and clear, and it corresponds to the way many people think about retirement. For example, as we will see, when asked if they have a plan for retirement, more than half of the people who do have some plan would reduce hours when retiring, and less than half would stop working altogether. In addition, and perhaps for the same reason, the most widely used measure of retirement age expectations also uses this definition.

For analytic convenience, we assume that retirement is an absorbing state: Once retired, an individual is assumed not to resume full-time work.¹ They may switch back and forth between no work and part-time work. Retirement is a forward-looking decision: one weighs the benefits of working and non-working now as well as in the future. With these assumptions, retirement can be modeled as an optimal stopping problem. An individual who works full-time compares the value of continuing to work full-time to the value of retiring now. If the difference declines monotonically over time (but not necessarily smoothly), at one point in time the individual decides to stop working full-time and never resumes full-time work.

¹ Late-in-life labor supply decisions are more complex and fluid than a permanent transition from working full-time to complete retirement. For example, individuals may move from lifetime to bridge jobs, may return to work after a period of retirement, and so on. Hence, though our analytic framework is well-defined, the alternate to full-time continuing work reflects many possibilities. See Maestas (2010) for analysis of expectations of work after retirement. We return to this issue in Section 4 and Section 7.

While the main purpose of the theoretical framework discussed here is to assist in the interpretation of survey questions, the framework equally applies to modeling retirement decisions. van der Klaauw and Wolpin (2008) show how to incorporate expectations data into the moment restrictions of a dynamic programming model of household retirement and saving, and then implement the procedure using the HRS. The procedure combines unconditional working and longevity expectations with observed realizations of respondents' labor supply, health, and the other variables to increase estimation precision. van der Klaauw (2012) discusses this approach generally with an application to career decisions. See also Wiswall and Zafar (2015, 2021). In this volume, Koşar and O'Dea (2023) discuss incorporating expectations into structural models. See also Giustinelli and Shapiro (2019) for structural interpretation of probabilistic expectation questions. These structural approaches complement the reduced form approaches discussed in Section 10 of this chapter.

Consider individual i who at age t thinks about retiring. Let the set Ω_{it} include everything that goes into that decision: the individual's perceptions of the relevant state variables, their current preferences, and their beliefs about the future path of the state variables, and their own preferences. See Lumsdaine, Stock, and Wise (1992) for a discussion of embedding retirement models in a statistical framework. Ω_{it} is heterogeneous across individuals and it evolves with time. Let the binary variable w denote working full-time, and let D denote the difference of the value of continuing to work full-time and the value of retiring now, the optimal decision is

$$w_{it} = 1 \text{ iff } D_{it}(\Omega_{it}) > 0 \quad (1)$$

where w_{it} is an indicator for working full-time.

The question analyzed in this section is expectations about when that stopping would occur in the future. Time is measured in terms of age, and for convenience is assumed to be discrete with integer

values only. Individual i who is of age a looks into the future and forms beliefs about when they would retire. There are alternative ways to characterize these beliefs.

One way to characterize these beliefs is to specify a series of probabilities of whether the individual would work full-time past each subsequent age t . Importantly, at age a , the individual cannot know with certainty the relevant future set Ω_{it} , for any $t > a$, which provides the basis for each subsequent optimal decision. Instead, what they have available at age a is Ω_{ia} , which contains their preferences and perception of the state at age a as well as their beliefs about future states and preferences. Let's denote the probability that individual i would work full-time past age t as they perceive it at age a , $a < t$, as

$$p_{ia}^t = \Pr[w_{it} = 1 | \Omega_{ia}]. \quad (2)$$

From a statistical point of view, the optimal stopping problem of retirement is a duration problem. The series of the p_{ia}^t probabilities is the survival curve as of age a : they show the probability of working full-time at least to age t for each age t based on age a information. Another representation of the same problem is the probability distribution of the retirement age. Let r_{ia}^t be the joint probability of working full-time at age $t-1$ and retiring at age t ,

$$r_{ia}^t = \Pr[w_{i(t-1)} = 1 \ \& \ w_{it} = 0 | \Omega_{ia}]. \quad (3)$$

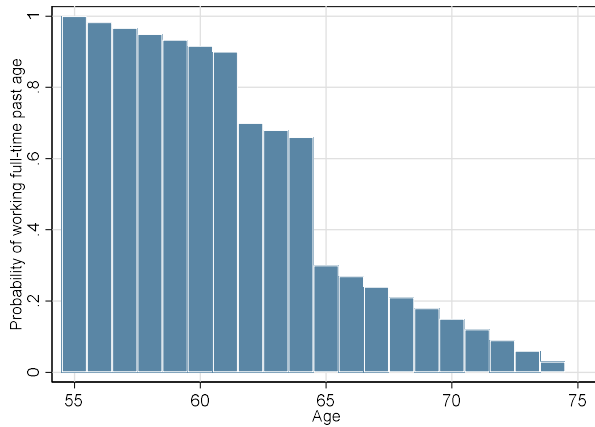
It is also the change in the survival curve,

$$r_{ia}^t = p_{ia}^{t-1} - p_{ia}^t. \quad (4)$$

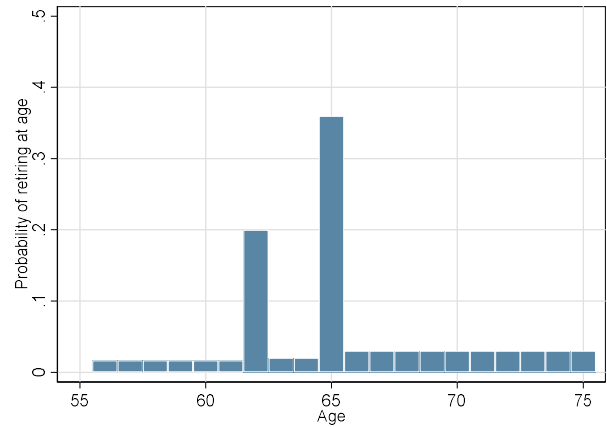
Note that the probability distribution of the retirement age is closely related to the hazard function, which is a more standard statistical representation of the same distribution. The hazard function is the probability of retiring conditional on not having retired by that age. Thus the hazard function re-normalizes the probability of the retirement age at each year of age. For our purposes, the

hazard function is less intuitive than the distribution of the unconditional probability of the retirement age, so we don't consider the hazard representation. Importantly, each of the three objects—the survival curve, the probability distribution of the retirement age or the hazard function—fully characterizes the same retirement age expectations.

Figure 1 shows a hypothetical example, though it is a realistic one for the U.S. where retirement at 62 and 65 is common. In this example, the probabilistic path to retirement is viewed from age 55 when the individual works full-time, and the individual is retired with a probability of one by age 75. The left panel is the survival curve of full-time work; the right panel is the corresponding probability of the retirement age. The survival curve shows a gradual decline in the probability of full-time work until age 62 where a larger drop occurs so that the probability of working past age 62 is 70 percent. Then some gradual decline again and an even larger drop age 65 to 30 percent. Finally, a gradual decline until age 75 when the probability of full-time work becomes zero. (Ages 62 and 65 correspond to eligibility for early claiming of Social Security and to eligibility for Medicare, respectively.) The right panel shows the probability of the retirement age at specific years of age. These probabilities are the change in the survival curve so they show a spike at ages 62 and 65, with values of 0.2 and 0.36, respectively. The gradual decline phases show up as uniform distributions within the appropriate age intervals.



The probability of working past the age threshold (survival curve of full-time work)



The probability of the retirement age

Figure 1. Two representations of the same retirement age expectations: an example

3. Measuring retirement age expectations

To characterize fully an individual’s retirement age expectations we would need to elicit their entire probability distribution of working past age thresholds or their entire probability distribution of the retirement age. However, in most survey situations it is impractical to elicit entire distributions. Instead, surveys tend to follow one of two approaches: elicit a statistic about the probability distribution of the retirement age, such as an expected value, or elicit subjective probability predictions of working past certain target ages. Neither of these approaches can completely characterize individuals’ retirement expectations. This is not a problem as long as the elicited statistic is exactly what the analyst needs. But it’s a problem when the analyst needs some other statistic or the entire distribution. In such cases, analysts need to make additional assumptions on the survival curve or the probability distribution of the retirement age, and then use interpolation and extrapolation.

Let’s return to our hypothetical example of Figure 1. Suppose that a survey elicits two points of the survival curve: the probabilities of full-time work past age 62 and 65; these are 0.7 and 0.3 here,

respectively. Knowing these two numbers would not identify the entire distribution because there are infinitely many survival curves that would fit these two points. Without a complete survival curve, we would not be able to identify the distribution of the probability of the retirement age, and thus we would not be able to compute statistics such as the expected age of retirement. In practice, we would need to make assumptions to interpolate and extrapolate from the two points to fit a survival curve and use that fitted curve for additional computations. It is an empirical question how many cut-points provide a reasonable approximation of the distribution. The choice of age 62 and 65 in the HRS (see below) is a judicious one given U.S. institutions.

Similarly, eliciting a statistic about the probability of the retirement age would not identify the probability distribution and would not identify points on the survival curve either. Candidate statistics include the expected value or the mode (the most likely retirement age). Denoting the age when individual i retires as $ageret_i$, and assuming that there is a terminal age T after which the individual would work with zero probability, the corresponding formulae are

$$E[ageret_i | \Omega_{ia}] = \sum_{t=a}^T t \times r_{ia}^t \quad (5)$$

$$\text{mode}[ageret_i | \Omega_{ia}] = \arg \max_t \{r_{ia}^t\}_{t=a}^T \quad (6)$$

In our hypothetical example in Figure 1, the expected retirement age is 65.3 and the modal age is 65. Knowing either or both of these numbers would be informative, but it would not identify the probability distributions without additional assumptions, and we would not be able to compute statistics such as the probability that the individual would work full-time past age 65.

4. Eliciting a planned or expected retirement age

Let's now turn to establishing the measurement properties of the two different elicitation strategies of subjective retirement age expectations. We start with the direct elicitation of a retirement

age and turn to subjective probability predictions after that. We analyze the Health and Retirement Study, which includes both elicitations. Given that the HRS is a longitudinal study, we will also be able to compare these elicitations to realizations.² Since its very beginning, the HRS has included questions about planning for and thinking about retirement. The first question in the sequence asks,

When you retire, do you plan to stop working altogether or reduce work hours at a particular date or age, have you not given it much thought, or what?

Multiple response categories are allowed, including stop working altogether, reducing hours, no plans to retire, or will never retire. Depending on the response, several follow-up questions elicit a retirement age; we call this the planned or expected retirement age. Not all retirement plan responses lead to an elicitation of the planned or expected retirement in the HRS before 2002.

In this chapter, we define retirement as the transition from working full-time to not working full-time. We construct a sample of HRS respondents accordingly to analyze the planned or expected retirement age. In particular, we include respondents in HRS 2002 through 2018 who were younger than 64 and worked full-time when the retirement planning questions were asked. Table 1 shows the distribution of responses to the retirement planning question (last column) and statistics of the distribution of planned or expected retirement age (previous columns).

² While the discussion in this chapter will focus on the questions in the HRS, there are similar questions in the HRS International Family of Studies including the ELSA (England), SHARE (multiple European countries), KLoSA (Korea), and LASI (India) studies. See Delavande, Lee, and Yoong (2012) for a comparison of the questions, which use questions similar to the HRS, though sometimes different ages based on national institutional arrangements. See also the Gateway to Global Aging Data (<https://g2aging.org/>). The HRS interviews both respondents and their spouses. The Federal Reserve Bank of New York's Survey of Consumer Expectation includes both probabilistic expectations (ages 62 and 67) and retirement age expectations similar to the HRS. See Federal Reserve Bank of New York (2022). The PSID uses a somewhat different questions from the HRS and its sister studies. PSID first asks, "What is the usual retirement age for people who sore with you or have the same kind of job?" [P39AGE] and then asks, "At what age do you plan to stop working?" [P40AGE]. See Panel Study of Income Dynamics Wave 41 Questionnaire (2019).

Table 1. Retirement plan and age when plans or thinks will retire

Retirement plan	Age when plans to retire or thinks will retire										Distribution of retirement plan %	
	Avg.	Median	Percent of specific age bins									total
			<62	62	63-64	65	>65	never	missing			
Stop working altogether	63.8	65	17	22	7	26	26	0	2	100	22	
Reduce hours	63.9	65	18	20	7	27	23	1	4	100	31	
No plans	66.7	65	4	8	2	24	37	1	23	100	39	
Will never stop working	-	-	0	0	0	0	0	100 ^a	0	100	4	
Other or missing	-	-	0	0	0	0	0	0	100 ^a	100	4	
Total	64.9	65	11	14	5	23	27	4	16	100	100	

Notes. ^a inferred, not elicited. Data: HRS 2002-2018, age 51-64, working full-time, retirement plans asked. N=30,155

Let's start with distribution of plans in the last column. In this sample of full-time workers aged 64 or less, 22 percent plan to stop working altogether when retired, 31 percent plan to reduce hours, and 39 percent have no plans. Four percent say they will never retire. The item nonresponse rate is 4 percent. The average planned or expected retirement age is 64.9 overall (first column), practically the same as the median of 65; it is higher among respondents with no plans. Across all plans, it is "never" or missing for 20 percent of the observations. The vast majority of "other or missing" values is among respondents with no retirement plans, and because of this high proportion, the overall item non-response rate is quite high.

As a first check on the information contained by the elicited planned or expected retirement age, we can compare the responses to the age the individuals actually retire from full-time work. Table 2 shows, for different retirement plans, whether the actual retirement age turns out to be lower, the same, or higher than expected. To allow for noisy responses with respect to the retirement window, we define the two to be equal using a 3-year window (the difference is -1, 0, or 1 years). The results show large deviations, with only 24 percent retiring at their planned or expected age, 54 percent ending up retiring younger, and 22 percent ending up retiring older. The deviations are only slightly smaller among respondents with an explicit plan, but again almost half of them end up retiring earlier. Hence, there is a

bias toward reporting a later than realized age of retirement. Perhaps more striking is the size of the variability of realizations compared to expectations. Note that this is what we would expect with a lot of uncertainty (large spread of the probability distribution of retiring at certain ages). Hence, with this large spread, a single statistic of central tendency of retirement age has relatively little information.

Table 2. Actual retirement age compared to planned/expected retirement age

Plan for retirement	Actual retirement age (%)			Total
	Lower	Same	Higher	
Stop working altogether	46	33	21	100
Reduce hours	46	23	31	100
No plans	67	19	14	100
Will never stop working	-	-	-	-
Other	-	-	-	-
Total	54	24	22	100

Notes. Actual retirement age is the year after the last year respondent worked full-time.

Same means within a 3-year bin; lower and higher mean outside this bin.

Data: HRS 2002-2018, age<65, working full-time, non-missing planned/expected retirement age. N=24,196

In addition to these measurement issues, eliciting expected retirement age has a problem of interpretation. Recall that retirement age expectations are completely characterized by either the survival curve of working or the distribution of the probability of retiring at certain ages. This measurement strategy elicits a single statistic derived from the probability distribution of retiring at certain ages. It is unclear, however, what that statistic is. It may be any statistic of central tendency, including the expected age or the modal age. Arguably, it is more likely that respondents report the mode of the distribution (the age with the highest retirement probability) than some other statistic. Bernheim (1989) indeed finds that expected retirement age is the mode based on data from the Retirement History Survey, a precursor of the HRS that was fielded between 1969 and 1979.

In summary, the combination of issues with interpretation and measurement makes the expected age measure challenging to use. Nevertheless, there are applications in which the planned or

expected retirement age is the measure of interest despite its limitations. Such applications need to address the conceptual and measurement issues.

5. Subjective retirement probability

The alternative strategy elicits the probability of working past certain age thresholds. This measurement strategy aims at eliciting points of the survival curve, so it has a conceptually clear interpretation. In this subsection, we analyze the measurement properties of this elicitation using the HRS data. Since its start in 1992, the HRS has asked its respondents about the subjective probability of working full-time past age 62 and, in a separate question, past age 65. The question is the following:

What do you think the chances are that you will be working full-time after you reach age 62 [65]?

The response scale is a percent chance on a scale of 0 to 100 (in 1992 it was a 0 to 10 scale). The HRS now includes six similar questions in total (age thresholds 62, 65, and 70, working at all and working full-time). The question asking about full-time work after ages 62 and 65 has been fielded consistently since 1994, so we focus on those two questions. Specifically, let p_{ia}^{62} denote the response of individual i , who is age a , to the question about the probability of working after age 62, and let p_{ia}^{65} denote the response to working after age 65. When referring to these two probabilities we use $t=b$ in the upper index p_{ia}^b where $b=62$ or $b=65$.

In line with our focus on retirement from full-time work, we restricted the analysis to respondents who worked full-time when asked the subjective probability question, had a nonzero survey weight, who were asked the probability question (or whose p_{ia}^{65} response we filled in), and who were 3 to 8 years younger than the age in the working probability question (54 to 59 years old for p_{ia}^{62}

and 57 to 62 years old for p_{ia}^{65}).³ For p_{ia}^{62} , there are $N=24,958$ observations for 12,872 respondents; for p_{ia}^{65} , $N=21,042$ observations for 11,122 respondents. The proportion of don't know or refused responses is below 1 percent for both questions.

While respondents make broad use of the scale 0 to 100 scale with its 101 potential responses, almost all responses are rounded to a number ending in zero or five. Additionally, there is heaping at 0, 0.5, and 1. See Bruine de Bruin, Chin, Dominitz, and van der Klaauw (2023), Manski (2023), and Giustinelli, Manski, and Molinari (2020, 2022) for discussion of these ubiquitous features of elicited probabilities.

To investigate how biased or unbiased the subjective probability responses are, we compared subjective probability predictions with actual work decisions later. For this exercise, we record whether each individual worked after the age threshold 62 or 65. When defining working full-time after age b , we included respondents who were recorded to work full-time any time between ages b and $b+5$ (changing this time window has no effect on our main conclusions). The sample sizes are smaller because we require realizations to be observed. Table 3 shows the average subjective probability predictions and the corresponding realized rates, together with the number of observations and the proportion of don't know or refused answers for these samples.

³ This focuses on the retirement expectations among those currently working also fits the structure of the questions as originally asked in the HRS. It allows us to pool over all waves of the HRS and interpret the results consistent with the theoretical framework developed above. The implementation of the HRS questions have evolved over time—reflecting both the evolution of use of probability questions and the aims to pose the questions in situations reflecting the complexity of late-in-life labor supply decisions (see also footnote 1). The HRS initially asked the P62 and P65 questions in the Current Job section (e.g., Section F of HRS 1992), so it was asked of those working. Starting in 2002, these questions and other expectations questions moved to the Expectations section (Section P). Section P also asks versions of work expectation of those not with current jobs. See Health and Retirement Study (2022). The FRBNY SCE asks the retirement expectations questions unconditionally on current labor force participation.

Table 3. Subjective probability predictions and realizations among full-time workers

	Average subjective probability prediction	Proportion working full-time	N	Percent missing
Working full-time past age 62	0.58	0.55	15,469	1%
Working full-time past age 65	0.41	0.41	12,922	1%

Notes. Percent missing is the percent of don't know or refused probability responses. Data: HRS 1994-2018. Pooled cross-section of full-time workers at the time of the elicited subjective probability prediction; whether working full-time observed 3 to 8 years after the subjective probability elicitation.

The subjective probability predictions made 3 to 8 years before the respective age thresholds are essentially unbiased. On average, full-time workers' predictions of the probability of their full-time working past the two age thresholds is close to the proportion of their full-time working past those thresholds. The last column shows that the proportion of don't know and refused answers (item-nonresponse) is negligible. Hence, the elicited retirement probability predictions are potentially high-quality measures. Importantly, the bias in predicting retirement age shown above is absent for the subjective probability elicitation.

We investigate unbiasedness further by conditioning on observable personal characteristics. We regress the difference between subjective probability predictions and realizations, $p_{ia}^b - w_{ib}$ on some important observable personal characteristics, including age, race, education, self-reported health, and the 27-item cognitive score (see e.g., Crimmins et al., 2011). All conditioning variables are measured at the time the subjective probability predictions are made. The sample used for this analysis is somewhat smaller because of missing cognitive score values. Table 4 shows the results.

Table 4. Prediction error regressions.

	(1) $p^{62} - w^{62}$	(2) $p^{65} - w^{65}$
Female	-0.010 (0.015)	0.016 (0.018)
Couple	-0.095** (0.018)	-0.044* (0.020)
Age	-0.001 (0.003)	-0.005* (0.003)
African American	-0.099** (0.023)	-0.132** (0.025)
Hispanic	-0.051 (0.031)	-0.053 (0.034)
Years of education	-0.001 (0.003)	-0.000 (0.003)
Cognitive score	-0.002 (0.007)	-0.014 (0.008)
Self-rated health very good	0.002 (0.016)	0.015 (0.019)
Self-rated health good	0.038* (0.018)	0.054** (0.021)
Self-rated health fair	0.105** (0.025)	0.080** (0.027)
Self-rated health poor	0.167** (0.054)	0.164** (0.053)
Constant	0.185 (0.157)	0.326 (0.168)
Observations	13,403	11,363
R-squared	0.015	0.011

Notes. The dependent variable is the difference between the subjective probability of working full-time at age b and a dummy variable that is one for realized working full-time at that age or later. The cognitive score variable is standardized. Robust standard errors clustered at the individual and household level are in parentheses. ** $p < 0.01$, * $p < 0.05$

The results are remarkable. Except for race and couple status, none of the standard demographics is associated with the direction of the prediction error. In fact, the coefficient estimates on years of education and cognitive score are relatively precise zeros. At the same time, subjective health is strongly associated with the prediction error: respondents in worse health are more likely to err on the positive side than respondents in better health. We can conclude that respondents' subjective

probability predictions are close to being unbiased except for negative biases among respondents who are members of a couple or are African Americans and a positive bias among respondents with low health.

Absence of biases in prediction does not imply free from any error. To investigate the error of the elicited predictions in more detail, we carried out a calibration exercise. A calibration exercise compares realizations against probability predictions at different levels of the probability prediction. In this case, it means estimating the actual proportion of respondents with $w_{ib} = 1$ for various values of p_{ia}^b . Calibration is perfect if $\Pr(w_{ib} = 1 | p_{ia}^b = p) = p, \forall p \in [0,1]$; that is among people with the same subjective probability prediction, the proportion who end up working at the target age is equal to the subjective probability prediction for all values of that subjective probability prediction.

The original response scale allows for 101 values so we need to group the responses to avoid too much noise at certain points of the calibration curve. We created bins and plotted the actual proportion of $w = 1$ observations against the average subjective probability prediction within each bin. Due to their high prevalence and potential substantive importance, we created separate bins for exact 0%, 50% and 100% responses. We put the remaining responses into bins of [1,9], [10,19], [20,29], [30,49], [51,70], [71,80], [81,90], and [91,99].

Figure 2 shows the results of the calibration exercise with 95% confidence intervals around the estimated proportion of working full-time for the two age thresholds. The standard errors allow for clustering due to repeated observations of individuals and intra-household correlations. The figure includes the linear approximation to the calibration curve and the 95 percent confidence interval around that regression (dashed line and shaded area), and a 45-degree line denoting perfect calibration (solid line).

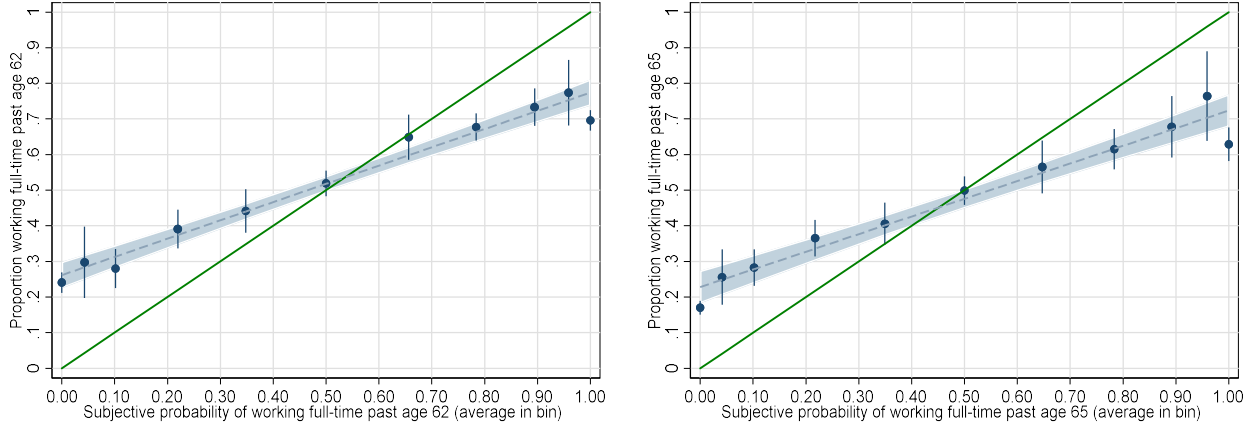


Figure 2.
Calibration curves for the subjective probability predictions p^{62} and p^{65}

The calibration curves suggest strong attenuation. These figures add several important conclusions to our previous results on unbiasedness. First, the calibration curves for the two age thresholds (62 and 65) are similar. Recall that the two samples are similar in terms of the time distance between the subjective probability predictions and actual realizations. Second, the calibration curves are flatter than the 45-degree line. Third, the calibration curves appear approximately linear except for lower values for exact 100% responses $\Pr(w_{ib} = 1 | p_{ia}^b = 1)$. Fourth, the calibration of the 50% responses is remarkably good: very close to 50% of them end up being working full-time, that is,

$$P(w_{ib} = 1 | p_{ia}^b = 0.5) \approx 0.5 .$$

What can be the reason for the attenuation? To investigate this question let's ignore the nonlinearity of the calibration curve at exact 100%, and let's consider its linear regression approximation

$$P[w_{ib} = 1 | p_{ia}^b] = E[w_{ib} | p_{ia}^b] = \alpha_b + \beta_b p_{ia}^b . \quad (7)$$

The probability limit of the OLS estimator for $\hat{\beta}_b$ is

$$\text{plim } \hat{\beta}_b = \frac{\text{Cov}(w_{ib}, p_{ia}^b)}{V(p_{ia}^b)} . \quad (8)$$

We estimate $\hat{\beta}$ to be around 0.5 for both age thresholds (see the more detailed regression results later). With a perfect 45-degree calibration line we would have $\text{plim } \hat{\beta} = 1$ and $\text{Cov}(w_{ib}, p_{ia}^b) = V(p_{ia}^b)$. It's with $\text{Cov}(w_{ib}, p_{ia}^b) < V(p_{ia}^b)$ that the linear regression approximation shows a flatter calibration curve, so $\text{plim } \hat{\beta} < 1$.

Why would we have $\text{Cov}(w_{ib}, p_{ia}^b) < V(p_{ia}^b)$? Based on simple econometrics intuition, classical additive measurement error in the subjective probability variables is an explanation. Note that measurement error cannot be entire classical here because the variable is bounded: It has to be negatively correlated with the error-free measure. Errors in responses that correspond to low error-free probabilities are more likely positive, and errors in responses that correspond to high error-free probabilities are more likely negative. That negative correlation leads to attenuation in itself due to a smaller positive $\text{Cov}(w_{ib}, p_{ia}^b)$, re-enforcing the effect of a larger $V(p_{ia}^b)$.

One source of measurement error is survey noise, which is likely present both for theoretical reasons and based on evidence. We have argued that people in this age group often think about when they would retire, and this is why eliciting their beliefs in surveys can lead to informative variables. Respondents are asked, however, about probabilities, even though most people do not carry around probabilities in their heads when they think about when they would retire. Thus, when answering the survey question they have to transform what they think into probabilities. That transformation may involve randomness so that the same beliefs may result in different probability responses.

Besides a theoretical plausibility, there is empirical evidence for survey noise in subjective probability responses and its attenuation effect. It was identified in stock market expectations from test-retest evidence by Kézdi and Willis (2011) and from differences in stated averages and averages implied by stated probabilities by Ameriks, Kézdi, Lee, and Shapiro (2020). Both papers found that taking survey

noise into account for estimation leads to stronger estimates of the associations. Kimball, Sahm, and Shapiro (2008) make the same point in the context of eliciting preference parameters in surveys.

It turns out that we can use test-retest evidence to identify survey noise with respect to retirement expectations, too, although not in terms of p^{62} or p^{65} . In 2018, in an experimental module fielded on a random subsample of its respondents to investigate their intentions to work longer, the HRS asked 753 respondents of age 64 or younger about their subjective probability of doing any work for pay after age 70. It turns out that this question has been part of the HRS core questionnaire, too, since 2012. Thus, in 2018, this same question was asked from 753 respondents twice during the same survey, separated by about 30 minutes and many other survey questions. We restrict the sample to respondents of ages 55 through 64 who worked at the time of the interview ($n=444$) and estimate a regression of one subjective probability prediction on the other. Table 5 shows these test-retest results.

Table 5. Test-retest regression of the probability of working past age 70: responses in the experimental module regressed on responses in the core survey.

	Response in experimental module
Response in core survey	0.81** (0.065)
Constant	0.07** (0.014)
Observations	444
R-squared	0.576

Data: HRS 2018 experimental module on working longer; respondents of age 55 through 64. Robust standard errors in parentheses. ** $p < 0.01$, * $p < 0.05$

While not a perfect test for p^{62} and p^{65} , this result suggests that survey noise in itself leads to a substantial attenuation in those, too, possibly as much as a third of the total attenuation. The slope coefficient estimate is approximately 0.8 for the subjective probability prediction of working at all after a 70. There is high, but not perfect, test-retest validity. Treating deviations in the responses as classical errors in variables, the 0.8 regressions coefficient in Table 5 corresponds to a signal-to-noise ratio of 4.

One would need a more complete model of retirement—including a quantification of the variance of the shocks affecting outcomes—in order to establish how much measurement error accounts for the attenuation shown in Figure 2. Nonetheless, the finding that about 20 percent of the variance of the elicited probabilities is survey response error goes a long way in explaining the attenuated response work outcomes to the probabilities of work. In the next section, we show how accounting for this attenuation is critical when using subjective probabilities to model outcomes.

Apart from any biases or errors in survey responses, a fundamental source of prediction error is—at the time of the prediction—that there is less or different information available to the individuals than at the time of the outcome, that is, $\Omega_{ia} \neq \Omega_{ib}$, $b=62$ or $b=65$ and $a < b$. If respondents integrated out all uncertainty in a rational way and formed their responses based on optimal predictions, absent survey response error the subjective probability should be an unbiased predictor of outcomes. That we see attenuation that not fully explained by survey noise is indirect evidence of departures from probabilistic thinking. For example, Ω_{ia} , the information used when making the subjective probability prediction at age a , Ω_{ia} may contain additional elements that would not be in Ω_{ib} , the information used when the actual decision is made—for example, whether one’s older peers did at the time the prediction is elicited. In general, individuals get news about their own circumstances and desires.

As noted, there is likely some evidence from the calibration curve findings of more attenuation than can be explained by survey response error. Respondents may not apply probability laws accurately, for example, they may attach disproportionately large weights to elements in the information set that are salient at the time of elicitation, but are less relevant for the actual decision or outcome. Alternatively, there may be systematic departures from systematic probabilistic thinking.⁴ To

⁴ Hudomiet, Hurd, and Rohwedder () indeed make such arguments in evaluating the predictive power of mortality expectations. They find calibration curves for mortality expectations that are very similar to those we find for retirement (see Figure 2). They, however, discount measurement response error as the sole explanation for the attenuation—pointing to biases in the probability reports including heaping at the extremes and at 50-50. More

shed more light on why the calibrations of p^{62} and p^{65} are imperfect we investigated how the slope of the calibration curve varies with observable covariates that are potentially related to the magnitude of prediction error. Note that this question is different from whether the covariates are associated with the direction of the error, which we investigated above.

Table 6 shows results from calibration curve regressions that include interactions with the covariates. The covariates are measured at the time of the elicitation of the probabilities. We show the results for p^{65} text; corresponding results p^{62} are similar. The regressions have the same binary left-hand-side variable, actually working past age 65, w_i^{65} , and the same main right-hand-side variable, the subjective probability response, p_{ia}^{65} . They are estimated on the pooled cross-sectional sample as for the calibration exercise (Figure 2): full-time workers 3 to 8 years prior to age 65. Column (1) does not include anything else and replicates the linear approximation of the calibration curve on Figure 2; column (2) includes the interaction of p_{ia}^{65} with the time between the interview and age 65 (range is 3 to 8 years); column (3) includes the interaction of p_{ia}^{65} with years of education; column (4) includes the interaction of p_{ia}^{65} with a standardized score of cognitive capacity. For easier interpretation we mean-differenced the interaction variables. Columns (2)-(4) include the interaction variable in itself as well. Their coefficients are not shown in the table for brevity. Standard errors are clustered at the household level to account for repeated individuals in the pooled cross-section and potential intra-household correlations.

generally, they point to the possibility that the distribution of subjective mortality is flatter than that of objective mortality for reasons in addition to measurement error. See their Figure 2 for an illustration of this point. We return to the issue of attenuation in Section 6.

Table 6. Calibration regressions including interactions with p^{65}

	(1)	(2)	(3)	(4)
	w^{65}	w^{65}	w^{65}	w^{65}
p^{65}	0.49** (0.018)	0.48** (0.018)	0.48** (0.018)	0.48** (0.019)
p^{65} interacted with time to age 65		-0.03** (0.007)		
p^{65} interacted with years of education			0.04** (0.007)	
p^{65} interacted with cognitive score				0.06** (0.016)
Constant	0.21** (0.009)	0.21** (0.009)	0.21** (0.009)	0.22** (0.010)
Observations	12,922	12,922	12,895	11,390
R-squared	0.14	0.14	0.14	0.14

Notes. w^{65} is whether the individual worked full-time after age 65; $w^{65}=0$ or 1. p^{65} is the subjective probability prediction of working full-time after age 65. Time to age 65 is 3 to 8 years; years of education ranges from 0 to 17; cognitive score is the standardized 27-item score combining immediate word recall (10pts), delayed word recall (10pts), counting backward from 20 (2pts), sequentially subtracting 7 from 100 (5pts). All three variables are mean-differenced. Each of columns (2)-(4) includes the interaction variable as well (coefficients not shown). All regressions are weighted by respondent survey weights. Data: Health and Retirement Study, pooled survey years 1994-2018, valid responses to p^{65} and known w^{65} , age 57-62. Standard errors are clustered at the household level. ** $p<0.01$, * $p<0.05$.

Column (1) documents the result from Figure 2 that the calibration coefficient is estimated to be 0.5, which is an attenuation by one-half of what perfect calibration would imply. The interaction results (columns 2-4) show that the slope is flatter the further away the prediction from the target age, and it is steeper the higher the level of education or the higher the cognitive score. These findings are in line with our intuition about the sources of the prediction error: The error should be smaller at shorter forecast horizons and for respondents with more education and higher cognitive capacity who are arguably better equipped for and more used to think about complex decision problems ahead of time. The estimated variation in the slope by the covariates is substantial. For example, five years closer to the target age results in a slope that is steeper by 0.13; respondents with four more years more education,

such as college versus high school have a slope that is steeper by 0.15. Thus the magnitude of the error varies systematically across individuals.

Taking the results together, we find that full-time workers' subjective probability predictions of whether they would work full-time at target ages 62 and 65 that are 3 to 8 years in the future have high quality, but they are imperfect. Subjective probability predictions are conceptually clean, and they have good measurement properties. In particular, the subjective probability predictions are essentially unbiased both overall and along important personal characteristics. While we have found some systematic heterogeneity in the direction of the prediction error by couple, race and health, but there is no such relationship with education, cognitive capacity or other demographic characteristics. In addition, the variables have negligible item nonresponse.

At the same time, the subjective probability predictions contain substantial error that makes them attenuated predictors of future retirement. The magnitude of the attenuation is about 50 percent, which means that respondents who, for example, give a 10 percentage point higher prediction of working full-time at age 65 tend to be not 10 but 20 percent more likely to be working full-time then. The magnitude of the prediction error, and thus the attenuation, is related to the forecast horizon, respondents' education and cognitive capacity in intuitive ways. Part of the attenuation is due to survey noise, which likely arises because the survey question requires a transformation of respondents' expectations into probabilities, and there appears to be random deviations in that transformation. A methodologically important finding is that the calibration curve is approximately linear, with somewhat larger prediction error for exact 100% responses.

6. Predicting retirement: subjective probability predictions and predictive analytics

Subjective retirement probability predictions can be used as either substitutes or complements to statistical models or structural models using observable characteristics to predict the retirement of

currently working people. See Koşar and O’Dea (2023) for an overview of the use of expectations in structural models. There are multiple reasons why using subjective probability predictions may be advantageous. First, subjective retirement predictions are a single variable that is cheaper to measure in a survey than the many predictor variables that may go into the decision. Second, subjective retirement predictions may include information that is not measured by other predictor variables as no dataset can measure all potential predictors. Third, subjective retirement predictions may also reflect how respondents use that information to make decisions based on their preferences and cognitive functioning. People might not behave according to the statistical or structural models used for prediction. Indeed, subjective retirement prediction may be better predictors of shifts in retirement across birth cohorts than predictive models based on other variables. The potential disadvantages of using subjective retirement predictions include its substantial prediction error that varies by the respondents’ characteristics as we documented above, and the fact that the individual subjective retirement predictions cannot properly reflect the realization of the macro environment.

To get a better understanding of how subjective retirement probability predictions perform in actual predictions, we have run a horse race between four predictive models of the actual future probability of working full-time past age 65 among people of ages 57-62 working full-time. We use the same forecast horizon of 3 to 8 years and the same sample that we used in our previous analysis. Model 1 includes the response to the subjective work probability predictions as they are ($\hat{p}_i = p_i^{65}$). Model 2 includes an adjustment to the subjective work probability predictions to re-calibrate them from the noise. In particular, Model 2 uses the rounded values of the simple OLS linear regression parameters from column (1) in Table 6, that is, $\hat{p}_i = 0.2 + 0.5 p_i^{65}$. Model 3 is a cross-validated random forest probability prediction model using a large set of predictor variables. Model 4 is the same as Model 3 except it adds the subjective probability predictions to the other predictor variables. Table 7 summarizes the details of the four predictive models.

Table 7. Details of the four predictive models

Model	Formula or method	Variables used
p^{65}	$\hat{p}_i = p_i^{65}$	p^{65}
Adjusted p^{65}	$\hat{p}_i = 0.2 + 0.5p_i^{65}$	p^{65}
RF(x)	Random Forest	x = {age, race, ethnicity, education, detailed marital status, veteran status, earnings, total household income, total household wealth, whether the respondent has a DB pension plan, a DC pension plan, total balance of the DC pension plans, total balance of the IRA accounts in the household, self-rated health status, whether the respondent has a spouse, and the age, education, work status, earnings, pensions, and self-rated health of the spouse}
RF(x, p^{65})	Random Forest	p^{65}, x

Notes. The Random Forest predictive models use the ranger package in R with default settings, treating the target variable (w^{65}) as numeric so the prediction is a probability.

We used these competing models to predict whether the individual worked full-time at or after age 65 in years 2014 through 2018. For each target year we used the variables that were measured up to three years earlier and the sample of respondents who were 57 to 62 years old then (up to four years in case there was no survey three years earlier). For example, we used data from waves 2006 through 2012 and respondents who were 57 to 62 years old in each survey wave to predict working full-time past age 65 in year 2014 (it's 2006 to 2012 and not 2011 because a few people who were 62 in 2012 turned 65 by the 2014 survey wave as some interviews are conducted in 2015).

In the language of predictive analytics (a.k.a. machine learning), years 2014 – 2018 are the five test sets, where each test set has a corresponding training set of data from three to eight years earlier, and our comparison of the various models is a five-fold cross-validation. In cross-sectional data test sets tend to be drawn randomly. The longitudinal nature of this data warrants another approach, that is, to select specific years as test sets so we can further investigate the predictive performance of the different models in different years corresponding to potentially different macro environments. To evaluate the performance of the four models we compare the predicted probabilities to the realized 0-1

variable of working full-time in the test years. The measure we use is the mean squared error (MSE) of the prediction (the square of the difference of the 0-1 realization and the predicted probability), which is also known as the Brier score (see e.g., Békés and Kézdi, 2021). The lower the MSE the better the prediction. Figure 3 shows the results. The left panel shows the average MSE across the five test years; the right panel shows the MSE for each test year separately.

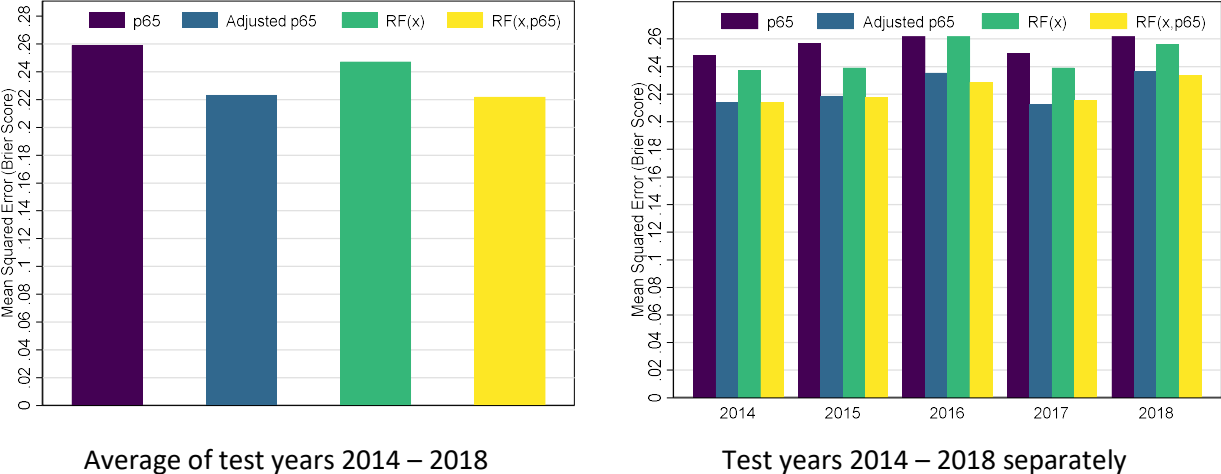


Figure 3. Predictive performance of the four models

Perhaps not surprisingly, the raw subjective probability predictions perform the worst. More surprisingly, the random forest with a rich set of predictor variables performs only slightly better. In contrast, the adjusted subjective probability predictions perform substantially better than either the unadjusted subjective probability or the random forest based on the rich set of predictors. Moreover, when the random forest prediction is based on the subjective probability plus the rich set of predictors, there is little further improvement compared to using the adjusted subjective probability predictions alone. All of this is true not only as the average over the five test years, but in each test year separately. These results are especially remarkable because the adjustment here is very simple, using the same intercept and slope coefficients for all respondents in all years. As we have shown in Table 6 above,

these parameters vary by education, cognitive score, and how far they are from the target year. The adjustment here has not used any of those differences.

Thus, we can conclude that subjective probability predictions, if adjusted in a simple way for attenuation bias, are well-performing predictors of future retirement at age 65 among full-time workers. Additional variables and more sophisticated modeling does not produce significantly better predictions. Because using adjusted subjective probability predictions is substantially less burdensome in terms of data requirements and modeling than building predictive models with many variables, this finding is important in practice.

Note that the near-sufficiency of the attenuation-adjusted retirement probability strongly supports our contention that the attenuation arises from measurement error in survey responses. If the attenuation arises instead from differences between subjective and objective probabilities as suggested by Hudomiet, Hurd, and Rohwedder (2023), other variables related to the objective probability should be helpful in predicting realized outcomes. There certainly could be omitted factors in the set x that predict retirement. Leading candidates would include macroeconomic shocks, though the year-by-year near-sufficiency of retirement probabilities suggest that they were not factors in the period studied.⁵

7. Subjective work probability predictions among non-full-time worker respondents

Before moving on to summarizing the results of the research that uses retirement age expectations, let's take a detour examining the subjective predictions of non-full-time workers. While our main focus is retirement from full-time work to some other state, the HRS asks these questions from a wider set of respondents. In particular, all respondents who worked for pay at the time of the interview are asked the p^{62} and p^{65} questions. Respondents not working for pay at the time of the

⁵ Of course, unexpected macro shocks, e.g., a dramatic change in Social Security, could lead to systematic deviations of expectations from outcomes. This would not be a violation of rational expectations, but simply an illustration of the fact that aggregated shocks do not average to zero in a cross-section.

interview are asked about the percent chance that they would work for pay at some time in the future. If their response is a nonzero percent chance, they, too, are asked the p^{62} and p^{65} questions. If their response is zero, don't know or refuse, they are skipped out of the subsequent questions. We imputed $p_{ia}^{62} = 0$ and $p_{ia}^{65} = 0$ if the response to the working-at-all question was 0%.

Table 8 shows the average subjective probability predictions and the proportion of realizations by labor force status at the time of the interview. Similar to our previous analysis, we restricted the sample to respondents who were 3 to 8 years younger than the target age when their predictions were elicited, but we include all respondents regardless of their labor force status.

Table 8. Working full-time past ages 62 and 65: Subjective probability predictions and realizations

Labor force status at the time of the subjective probability prediction	Working full-time past age 62			Working full-time past age 65		
	p^{62}	w^{62}	N	p^{65}	w^{65}	N
Works full-time	0.58	0.55	15,469	0.41	0.41	12,922
Works part-time	0.41	0.20	2,928	0.29	0.14	2,477
Unemployed	0.50	0.20	673	0.31	0.11	500
Partly retired	0.20	0.09	1,366	0.13	0.06	2,059
Retired	0.06	0.02	3,596	0.04	0.01	5,624
Disabled	0.05	0.02	1,061	0.02	0.02	1,017
Not in the labor force	0.13	0.06	1,502	0.06	0.04	1,703
Total	0.43	0.37	26,595	0.26	0.24	26,302

Notes. p^{62} and p^{65} are the average subjective probability predictions; w^{62} and w^{65} are the proportions of the corresponding realizations; p^{62} and p^{65} are measured 3 to 8 years younger than the target age. Data: HRS 1994-2018 pooled cross-section.

The first row in Table 8 repeats the information in Table 3, showing that the subjective probability predictions of full-time work past the two age thresholds are essentially unbiased among current full-time workers. At the same time, respondents in all other current labor force status tend to overpredict substantially their future full-time work chances, with two or three times as high probabilities on average. For example, respondents who work part-time tend to overpredict their probability of working full-time past each of the age thresholds, by a factor of two (0.41 instead of 0.20 for age 62; 0.29 instead of 0.14 for age 65). The same is true for the similar category of respondents who

considered themselves partially retired. The relatively few unemployed respondents in the sample tend to overpredict their likelihood of their full-time working past the age thresholds by a factor of three. Respondents in the remaining labor force status have very low full-time work chances, but they, too, tend to overpredict those low probabilities.

These results have important consequences for research uses of the subjective probability predictions. Pooling full-time workers with other groups, including part-time workers or partially retired respondents, reduces the quality of the measurement and may introduce stronger noise or bias in the results. The problem may not be very severe if the non-full-time-workers make up a small fraction of the overall sample (for example, part-time workers are only about 16 percent of the group working for pay in our sample), but the cleaner empirical strategy is to focus on full-time workers.

8. Research on the quality of retirement age expectations

The first analysis of the subjective probability predictions that we know of is by Hurd (1999). He analyzed the quality of the subjective probability predictions of working full-time past age 62 (p^{62}) in the first two waves of the HRS. His results show that p^{62} is a strong predictor of actual full-time work past age 62 two years later, it contains information beyond important variables that predict retirement, and it varies with important variables very similarly to the actual full-time work probabilities. Most studies that use the subjective probability predictions of future work to investigate substantial economic questions tend to contain some analysis of the quality of those variables, too (for example, Honig, 1996; Chan and Stevens, 2004). Their general conclusion is similar to Hurd's (1999) initial assessment and the conclusions of our own analysis: The subjective probability predictions contain a lot of information, some of which is otherwise unmeasured, and they can be used to predict actual future retirement.⁶

⁶ See Mueller and Spinnewijn (2023) for a similar favorable assessment of the predictive power of expectations of job finding and Hudomiet, Hurd, and Rohwedder (2023) for a more mixed assessment of the predictive power of expectations of health and longevity.

As for the other approach, Benitez-Silva and Dwyer (2005) use HRS data from 1992 to 2000 to examine the elicited planned or expected retirement age on a restricted sample of respondents who report to have a retirement plan. (In our sample of full-time workers, 40 percent did not have such a plan.) They test whether, conditional on observables, the expected retirement age follows a random walk, which is a test of rational expectation formation in their theoretical framework. They address measurement error in the responses by an instrumental variables strategy, and they augment it with a Heckman-type selection model to correct for selection into giving a valid answer (with longevity expectations and smoking behavior as instruments for measurement error and age of the respondent and whether the parents of the respondent reached retirement age as instruments for selection). They cannot reject the null hypothesis of planned retirement age following a random walk. They do reject a stronger, perfect foresight version, so retirement expectations appear to be rational, but they are not perfect predictors.

Papers that use this variable for investigating substantive questions tend to include some assessment of the quality of the variable (for example, Haider and Stephens, 2007). They find that the variable has a lot of useful and otherwise unmeasured information about actual future retirement, but its measurement issues require more sophisticated analysis.

Our reading of this literature reinforces the conclusion of our own analyses. Eliciting the planned or expected retirement age may be the good strategy for the substantive research question, but this variable has its issues of interpretation and high and selective item non-response. There is very valuable information in this variable, but using it requires a lot of work and additional assumptions. Subjective retirement probabilities are good although imperfect predictors that contain information not fully captured by other observables. That's great news if the research question can use this variable.

9. Potential uses of retirement age expectations

After having discussed the quality of the different approaches to elicit retirement age expectations, let's discuss how such data can be used in research and policy analysis. First and most obviously, having inexpensive and good predictions of future work probabilities for various cohorts and social groups can directly help policy analysis. For example, the U.S. Social Security Administration, and its counterparts in other countries, regularly forecast the actuarial status and financial conditions of their funds. These forecasts typically use complex models with many parts using many parameters, of which the rate at which various groups retire is only one small subset (see, e.g., SSA, 2019). Using a statistic of a single variable from available surveys, potentially adjusted in a simple way, can be used as a parameter in an appropriate model in a relatively simple way. To do so relying on our findings, the model should be about the medium run because our results establish the good properties of the (adjusted) subjective probability predictions up to 8 years from the time of the measurement of the variable.

Second, the good predictive performance of subjective probability predictions of retirement opens up the possibility to use them to evaluate the effects of various state variables and interventions on future retirement.

Third, it also allows for analyzing the effect of retirement age expectations, or deviations from such expectations, on economic behavior and outcomes. In the next section, we summarize results in the literature that use retirement expectations as a left-hand-side variable or as a right-hand-side variable. Before doing so, let's reflect on what our results imply for such applications.

We have shown that subjective probability predictions are imperfect, and the simplest way to rationalize the imperfection is as if it were a mean-zero additive measurement error to the hypothetical error-free probability prediction, with several features. This measurement error increases the variance of the error-free variable. Moreover, it is negatively correlated with the error-free variable because of its

bounded nature. We found systematic heterogeneity in the magnitude of the measurement error, most importantly with respect to education and cognitive capacity. We also found that some systematic heterogeneity in the direction of the error, by race, being in a couple, and health.

When on the left-hand-side of a regression, this measurement error does not necessarily induce bias in the coefficients of interest, but it increases estimation uncertainty. The caveat here is that the measurement error should not be correlated with the right-hand-side variable of interest. For example, the correlation of the measurement error with health can have consequences for estimating the effect of health on the probability of working past an age threshold. The potential for such correlations needs to be addressed in each application separately.

10. Research with retirement age expectations as the left-hand-side variable

What leads people to retire at certain ages has been the subject of a large literature. The most researched potential causes include health, financial incentives in the pension formula, and availability of health care (this last one in the U.S. context). In this subsection, we review the literature that attempts to estimate these effects by having the retirement age expectations as the left-hand-side variable. Having expectations on the left-hand-side instead of realized retirement allows not only for estimating effects ahead of time, but it also allows for using within-person variation in the retirement age that would be impossible to analyze using realizations. The advantage of the ability to restrict variation to within-person is that can mitigate potential biases to the effect estimates due to unmeasured and endogenous heterogeneity. Most papers use the subjective probability predictions in the HRS data, but some use other data sources and/or the direct elicitation of the expected retirement age.

McGarry (2004) uses the subjective probability predictions as a left-hand-side variable to estimate the effect of health on retirement. She uses the first two waves of the HRS, 1992 and 1994, to

examine how self-rated health and its change are associated with the subjective probability prediction of working full-time past age 62 (p^{62}) and its change. The sample consists of working people younger than 62 in 1994 (wave 2). She argues that, by focusing on working people and their expectations instead of realized retirement, one can mitigate a potential justification bias that would make people who retire for other reasons report worse or worsening health. Recall that we have found that the level of health is negatively correlated with the measurement error in p^{62} , which in this case would lead to an additional bias of cross-sectional estimates. For those reasons, we focus on her results from the specification with change in p^{62} regressed on whether the respondent reported that they had the same, better, or worse health in wave 2 than in wave 1.

The main finding is that respondents who reported deteriorating health revised their p^{62} downwards from the previous wave—by 4 percentage points (McGarry 2004, Table 6). What do the measurement properties of p^{62} that we learned in this chapter imply for this result? We don't know whether the first difference of the prediction is correlated with self-rated change in health, so we don't know if that would result in a bias. What we know is that the measurement error itself would lead to estimation uncertainty. That uncertainty is likely particularly strong here both because taking the difference magnifies the role of noise and because the sample includes the non-full-time workers with lower quality p^{62} variables. This estimation uncertainty shows up in the relatively wide 95 percent confidence interval that nevertheless does not contain zero. Hence, McGarry's result provides reasonably strong evidence for the negative effect of deteriorating health on working past age 62, but with an effect size smaller than what cross-sectional estimates would imply.

Chan and Stevens (2004) use HRS data from 1992 through 1998 to analyze the effect of retirement incentives on the subjective probability prediction p^{62} . They measure retirement incentives based on respondents' self-reported pension information, as opposed to matched objective pension plan characteristics, arguing that it is what people know about the incentives that should matter for

their behavior. They summarize incentives in a pension gain variable: how much people would gain by retiring after age 62 instead of before age 62. They estimate a moderate effect in pooled cross-sectional regressions, and a more modest effect in regressions that within-person variation only. Such within-person analysis is possible because they use longitudinal data on subjective probability predictions. Note that such within-person analysis would be impossible with data using realized retirement. With p^{62} on the left-hand-side, measurement error in this variable is unlikely to lead to biased estimates. They argue that measurement error in their causal variable is also unlikely to account for the fact that restricting to within-person variation decreases the effects to be modest. They conclude that, while financial incentives for retiring early matter, that affect may be more modest than what the previous literature suggested that analyzed realized retirement and could not condition on unobserved individual heterogeneity.

Bottazzi, Jappelli and Padula (2006) investigate the combined effect of three pension reforms in Italy in the 1990s that reduced pension benefits for some groups on savings, the expected retirement age and expectations about future pension benefits. Here we summarize their results on retirement age expectations. The authors use the Survey of Household Income and Wealth conducted by the Bank of Italy from 1989 through 2002 that elicited the expected retirement age. In a difference-in-differences analysis making use of control groups whose benefits were not affected by the reforms, they show that the pension reforms increased the expected retirement age by one to four years in various groups.

Okumura and Usui (2014) use data from the Japanese Study of Aging and Retirement (JSTAR) to estimate the effect of a gradual increase in the age eligibility of employee pensions on the expected age of pension claiming and retirement, as well as the expected pension benefits. The survey elicited the expected age of claiming. The authors document substantial item-nonresponse to both the expected pension claiming age and the expected retirement age. Their results show that, in the subsample that gave valid age responses, the increase in the age eligibility of pensions lead to a substantial increase in

the expected age of pension claiming and retirement. Our conclusion from these studies is that large policy changes tend to result in revised retirement age expectations (the Italian and Japanese studies), but people tend to be modestly responsive to the incentives in complex pension plans as those in the United States (Chan and Stevens, 2004).

The effect of health care availability on the timing of retirement is addressed by several papers in the U.S. context. Mermin et al. (2007) compare subjective retirement probability predictions among 51 to 56 year-old people across two birth cohorts: those born in 1936 – 1941 versus 1948 – 1953. They use the subjective probability predictions p^{62} and p^{65} from two waves of the HRS apart from each other by 12 years, 1992 and 2004. Their sample consists of people who worked for pay at the time of the interview. They show significantly higher reported probability of working full-time both past 62 and 65 in the younger birth cohort, both among men and women. They argue that the result is due to the fact that the younger generation plans to retire later. In an Oaxaca-type decomposition, they find that the substantial decrease in defined benefit pensions and employee-sponsored retiree health insurance can explain as much as two thirds of the increase in the likelihood of working full-time past age 62 and one third of the increase in working past 65.

Ayyagari (2019) uses HRS data to estimate the effect of the Affordable Care Act (ACA) in the USA on p^{62} . In pooled cross-sectional difference-in-differences specification, she compares p^{62} responses from before ACA to after ACA for people without employer-sponsored retiree health insurance coverage (the treatment group) to people with retiree coverage (the control group). Without retiree coverage, those who retire at age 62 would not have employer-sponsored health insurance, and they would not be qualified for Medicare until age 65. The effect of ACA here is that it could provide affordable health insurance for these years even for people who retire. People without retiree coverage are 6 percentage points less likely to work past age 62 because of the availability of health insurance through ACA (Table

1). The conclusion of these papers is that the availability of affordable health insurance in retirement is an important cause of whether people retire before age 65 in the USA.

11. Research with retirement age expectations as the right-hand-side variable

Let's start this part of the review with a note. When retirement age expectations are a right-hand-side variable of a behavioral relationship, they may be of interest even if the expectations are not unbiased predictors of retirement. As long as people behave differently if they have different expectations, learning about those effects is interesting on its own right. At the same time, aspects of measurement quality still play a role here, such as item nonresponse or measurement error due to survey noise. The former needs to be treated if it is a significant problem; the latter implies a moderate measurement error that is close to being classical.

Haider and Stephens (2007) use data from the HRS from 1992 through 2002 as well as data from its predecessor, the Retirement History Survey (RHS) from 1969 through 1977, to investigate the extent to which retirement results in a decline of consumption. More specifically, their question is whether expected retirement leads to the same reduction in food expenditures as unexpected retirement. They measure consumption by expenditures on food, and they use the elicited planned or expected retirement age and compare it to the actual retirement age to differentiate expected retirement from unexpected retirement. They find that expected retirement results in a substantially smaller reduction than unexpected retirement, which suggests that it may not be retirement itself but potentially other changes that are responsible for the decline in food expenditures with retirement. Using the elicited planned or expected retirement age brings about the problem of item-nonresponse, but the authors carry out a robustness check by transforming this variable to whether the respondent expects to retire by the subsequent survey wave and find similar results. Note that for this analysis, expected retirement age is the variable of interest; using the probability predictions instead would require functional form

assumptions to estimate that statistic. Besides its substantive results, this study highlights the methodological issues and potential solutions with using the planned or expected retirement age variable.

Romm (2015) uses data from the HRS from 1992 through 2004 to investigate the relationship between the expected time of retirement and pre-retirement wealth accumulation, and how that varies with the distribution of bargaining power between members of a couple. Later expected retirement should call for less wealth to be accumulated, *ceteris paribus*, but retirement plans are inherently endogenous to wealth accumulation through unobserved heterogeneity in preferences and constraints. The author uses the change of the subjective probability predictions p^{62} to condition on the time-invariant component of such unobserved heterogeneity and augments it with an instrumental variable strategy with self-reports of the usual retirement age at the job as an instrument. Both the non-instrumented and instrumented results indicate a strong effect in households with the husband being the sole earner, but no significant effect of married women's retirement age expectations whether or not they were the sole earners or dual earners.

Clarke et al. (2012) use the HRS data from 1998 through 2008 to investigate the effect of unexpected retirement on life satisfaction. They use the subjective probability prediction p^{62} , observed repeatedly for individuals, and they use a generalized growth mixture modeling to classify respondents into low, moderate, and high work expectation categories. They find that men with less job stability but high expectations to work full-time past age 62 experienced substantial lower life satisfaction when those expectations were not met and, instead, they retired earlier.

Another literature investigates whether and to what extent working couples coordinate retirement or take into account their spouses' resources and characteristics. Retirement expectations data have been used in some of this research. Pienta and Hayward (2002) use the first wave of HRS subjective probability predictions p^{62} and p^{65} to investigate the extent to which expectations of

members of working couples are related to their own, their spouses', and household characteristics. They find that all three matter for both husbands and wives, but they also find a stronger influence of husbands on wives' retirement expectations than vice versa.

Benitez-Silva and Dwyer (2006) extend their analysis of whether the planned or expected retirement age conforms the rational expectation hypothesis (Benitez-Silva and Dwyer, 2005, discussed above), by focusing on couples and including the spouse's expected retirement age as a conditioning variable. They find that couples do take spousal retirement age into account and they form rational expectations accordingly.

Ho and Raymo (2009) use a sample of couples in the HRS observed from 1992 through 2004 to examine how their actual retirement timing is related to their retirement age expectations. They make use of responses to a question that directly asks if couples plan to retire together. This question follows the retirement plan and planned retirement age elicitation that we discussed earlier. The authors exclude couples with one or two members reporting not to have plans for retirement, and they find that, in the remaining subsample, couples who report to plan to retire together are three times more likely to retire within 12 months than couples who do not plan to retire together.

Michaud, Van Soest, and Bissonnette (2020) estimate a structural model of the joint retirement decision of spouses that combines stated preference and retirement plans elicited in an internet supplement to the HRS. Their approach is designed to disentangle the effects of preferences and expectation, and in particular, points to husbands' misperception of wives' disutility of work.

Overall, the conclusion from these studies is that many couples plan to retire jointly, and they tend to realize these plans.

12. Using conditional subjective work probability predictions to estimate effects on retirement

Recent research on retirement age expectations as the left-hand-side variable builds on fielding conditional subjective probability questions instead of the unconditional subjective probability predictions. The idea is to control variation in the conditioning event by asking multiple questions. In practice, this amounts to asking the probability of working in one state of the world (good health, for example) followed by asking the probability of working in another state of the world (bad health). These are direct elicitations of the probabilities of potential outcomes from the same individuals. Under ideal circumstances, the difference between the two reported probabilities identifies the effect of one state versus the other (bad health versus good health) on the probability of the outcome (working), for each individual. This strategy has multiple potential uses. First, it may identify the effect without observing the various states of the world, and second, it may identify not only the average effect, but the entire distribution of the heterogeneous effects. See also Bruine de Bruin et al. (2023) and Fuster and Zafar (2023) for general discussion of the used of hypothetical and conditional expectations questions.}

Giustinelli and Shapiro (2019) define this effect as the Subjective ex ante Average Treatment Effect, abbreviated as SeaTE. It is subjective because it's stated by the individuals themselves, it is ex ante because the effect is measured without observing any change in the state variable, and it is a treatment effect at the level of the individual. They investigate the effect of bad health versus good health on the probability of working for pay in two- and four-years ahead. They field the questions as part of the fourth wave of the Vanguard Research Initiative (VRI), a longitudinal survey of Vanguard account holders, on a sample consisting of working respondents aged 57 to 81 who are in good health and were still working at the time of the survey.⁷

The questions elicit subjective probability predictions of working (unconditional), the conditioning events (bad or good health), and working conditional on health (working if in bad health,

⁷ They also report results for a nearly identical survey fielded as experimental module in the HRS 2016.

working if in good health). The SeaTE of bad health for each of individual is the difference between the two conditional probabilities. The results show a substantial average effect. The average SeaTE is -28.5 percent at a two-year horizon and -25.7 percent at a four-year horizon. There is substantial heterogeneity in the effect, with zero effect for close to 30 percent of the respondents. The authors carry out a validation exercise making use of the detailed sequence of probability questions. In particular, they compute the unconditional probability of work implied by the responses to the conditional probability questions and the probabilities of the conditioning events, and then they compare this to the directly elicited unconditional probability responses. They find a remarkable accordance with a correlation coefficient of 0.93. They also show that the subjective probabilities predict retirement decisions in the VRI panel though there are limited adverse health shocks realized.

The conditional expectations framework is helpful for controlling for endogeneity and selectivity in empirical analysis. For example, a regression of the retirement outcomes on health shocks will not produce an unbiased estimate of the treatment effect of health on retirement if health is related to unobserved taste for work. Using as simulation based on the distribution of taste for work elicited from the VRI respondents, Giustinelli and Shapiro (2019) show that such biases would emerge in estimates based on data on health and retirement outcomes.

Hudomiet, Hurd and Rohwedder (2020) report results from similar conditional probability questions fielded on the RAND American Life Panel survey on a sample of 51 to 69 year-old respondents. In particular, a series of questions asked about the probability of working past age 70 conditional on various states of the world, including good health versus bad health, a hypothetical policy change that would increase wage for workers who are over 70 years old, a large inheritance, increased longevity, and various work characteristics. Similar to Giustinelli and Shapiro (2019), they find substantial positive average effect of good health on the probability of future work (here the effect is calculated as good health versus bad health). In addition, they find a moderate average positive effect of higher wages, a

large negative effect of inheritance, a small positive effect of increased longevity, and small effects of work characteristics. Using health as the conditioning state of the world the authors carry out the validation exercise of comparing the probability of work elicited directly with the probability of work computed from the conditional probabilities and the probabilities of the conditioning events. As in Giustinelli and Shapiro (2019), they find remarkable accordance of the two. As a methodologically interesting point, Hudomiet, Hurd and Rohwedder (2020) show that the same accordance is not true if joint probabilities are elicited (for example, working and in bad health) instead of conditional probabilities (working if in bad health). They also assess whether respondents answer *ceteris paribus* in scenarios with low degrees of specification (what they call the “filling-in problem”). They find no evidence of filling-in, which supports interpreting the conditional probabilities as causal.

Conditioning versus subjectivity in general. This section’s consideration of the conditional expectation relates to a general issue in eliciting unconditional expectations: What do respondents have in mind when they provide a point response to a probabilistic question? Formally, if they are thinking probabilistically, what states are they integrating over when they give a response? Informally, what are they thinking about and what are they not thinking about? See Dominitz (1997), Bruine de Bruin et al (2023), Manski (2023), Hudomiet, Hurd, and Rohwedder (2023), and Giustinelli and Shapiro (2019) for discussion of this issue.

The issue of implicit conditioning may be important for retirement expectations. For example, there is an ambiguity of whether people’s retirement age expectations are conditional on certain circumstances, such whether they are alive at age t , whether their employer wants them to retire at a certain age, or the state of the labor market at age t . Because of limitations on the length of a survey question and the desire not to prime the answer by specifying conditions, survey questions leave such conditioning implicit. Therefore, when they interpret the responses, researchers need to be careful

about implicit assumptions about what circumstances respondents condition on and what they integrate out.

Conditional expectations questions, by being specific about conditioning events, may be a step toward addressing these ambiguities. By being explicit about a dimension, conditional probabilities are a route to grappling with the incomplete scenario problem (Manski 1999). Vignettes (Fuster and Zafar, 2023) likewise have promise for addressing this problem.

13. Conclusions

Retirement and retirement expectations are of keen interest to social scientists and policymakers. Predicting the effect of changes to policies or in circumstances requires evaluating how people's decisions would change as a result. Predicting how the retirement rate will evolve in the future requires predictions of not only how those circumstances would evolve but also how those would affect people's decisions. Retirement expectations are a key tool for analysis. This chapter demonstrates the usefulness of these measures—both by describing their statistical properties and by surveying their wide use in research and policy analysis.

Retirement expectations—measured as expected retirement age or the probability of being retired at a particular age—are among the most common expectational variables in longitudinal household surveys. Both measures have been included in the Health and Retirement Study since its inception in 1992. Similar measures are included in the international sister studies of the HRS, the Panel Study of Income Dynamics, and the Survey of Consumer Expectations.

This chapter analyzes retirement expectations in the context of a simple discrete choice model of late-in-life labor supply. It then evaluates measures using the data from the HRS. Several findings emerge. When measured in terms of expected age of retirement, expectations are a somewhat biased predictor of actual retirement age, with a tendency to retire earlier than the expected age. In contrast,

probabilistic expectations, specifically the percent chance of working full-time past the age of 62 or 65, is an unbiased predictor of actual behavior on average. The fraction of respondents who actually work beyond those ages is almost identical to the average subjective probability of working past those ages at horizons of 3 to 8 years. Additionally, the subjective probabilities are close to being conditionally unbiased predictors of retirement in the cross section. Though a few covariates (current poor health and race) have significant point estimates in a prediction error regression, the magnitude of predictable forecast errors is very small (R^2 of one or two percent).

Across individuals, however, subjective probabilities give an attenuated predictor of retirement behavior. Recalibrating the subjective probability p^b of working after age b with the linear relation $\hat{p}^b = E[w_b | p^b] = \alpha + \beta p^b$ where w_b is an indicator for working at age b robustly yields unbiased predictors of working with a value of β equal to 0.5. The value of the attenuation parameter β is remarkably consistent with what one would expect from classical measurement error where the variance of the measurement error is identified using repeated elicitations of the probabilities within a single survey, i.e., test-retest error.

The recalibrated subjective probability of retirement carries very substantial information for understanding retirement outcomes. This chapter presents a novel, machine learning (ML) model for predicting retirement. A random forest predictor of continuing to work at age 65 based on age, education, demographics, family status, earnings, pension plan, health, and spousal variables at time of elicitation of the subjective probability does less well than the recalibrated subjective probability at predicting retirement outcomes 3 to 8 years ahead. Moreover, the ML predictor has little predictive power incremental to the recalibrated subjective probability. Hence, conditional on a large set of information relevant for retirement, the recalibrated subjective probability is close to a sufficient statistic. Of course, the subjective probability will not fully account for outcomes since there is

substantial news about individuals, their workplaces, economic policy, and macroeconomic conditions that affect realizations.

Subjective probability of retirement has been widely used to study retirement outcomes and economic decisions involving late-in-life work. These uses are supported by the wide availability of retirement expectations in longitudinal studies, and by the reliability of these measures as documented in this chapter. These uses include predicting retirement (subjective probability on the “left hand side”) and analyzing retirement or retirement surprises (subjective probability on the “right hand side”). This chapter surveys a variety of these applications including the retirement income puzzle, joint labor supply decisions of spouses, and the effect of health on retirement. Additionally, this chapter discusses recent advances in eliciting and using conditional probabilities.

Because retirement is such a complex problem—involving the dynamics of income, wealth, health, joint decisions within couples, and complex interactions with private and public pension and insurance arrangements—subjective expectations are a powerful tool for analysis. Their implementation in longitudinal studies allows them to be studied as predictors of retirement and for research on the dynamics of late-in-life work. The value of retirement expectations for structural or policy analysis is underscored by the near sufficiency of probabilistic expectations for predicting retirement.

REFERENCES

- Ameriks, John, Gábor Kézdi, Minjoon Lee, and Matthew D. Shapiro. "Heterogeneity in Expectations, Risk Tolerance, and Household Stock Shares: The Attenuation Puzzle." *Journal of Business and Economic Statistics* 38.3 (2020): 633-646.
- Ayyagari, Padmaja. "Health Insurance and Early Retirement Plans: Evidence from the Affordable Care Act." *American Journal of Health Economics* 5.4 (2019): 533-560.
- Békés, Gábor and Gábor Kézdi. *Data Analysis for Business, Economics, and Policy*. Cambridge University Press, 2021.
- Benitez-Silva, Hugo and Debra S. Dwyer. "The Rationality of Retirement Expectations and the Role of New Information." *Review of Economics and Statistics* 87.3 (2005): 587-592.
- Benitez-Silva, Hugo and Debra S. Dwyer. "Expectation Formation of Older Married Couples and the Rational Expectations Hypothesis." *Labour Economics* 13.2 (2006): 191-218.
- Bernheim, B. Douglas. "The Timing of Retirement: A Comparison of Expectations and Realizations." In David A. Wise Ed., *The Economics of Aging*. Chicago: University of Chicago Press, 1989.
- Bottazzi, Renata, Tullio Jappelli, and Mario Padula. "Retirement Expectations, Pension Reforms, and Their Impact on Private Wealth Accumulation." *Journal of Public Economics* 90.12 (2006): 2187-2212.
- Bruine de Bruin, Wändi, Alycia Chin, Jeff Dominitz, Wilbert van der Klaauw. "Household surveys and probabilistic questions." In *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Amsterdam: North-Holland/Elsevier, 2023.
- Chan, Sewin and Ann Huff Stevens. "Do Changes in Pension Incentives Affect Retirement? A Longitudinal Study of Subjective Retirement Expectations." *Journal of Public Economics* 88.7-8 (2004): 1307-1333.
- Clarke, Philippa, Victor W. Marshall, and David Weir. "Unexpected Retirement from Full Time Work after Age 62: Consequences for Life Satisfaction in Older Americans." *European Journal of Ageing* 9.3 (2012): 207-219.
- Crimmins, E. M., J.K. Kim, K.M. Langa, and D.R. Weir. "Assessment of Cognition Using Surveys and Neuropsychological Assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study." *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 66.suppl_1 (2011): i162-i171.
- Delavande, Adeline, Jinkook Lee, and Joanne Yoong. "Harmonization of Cross-National Studies of Aging to the Health and Retirement Study Expectations." No. WR-861/3. RAND Corporation, 2012.
- Dominitz, Jeff. "The Role of Conditioning Information in Reports of Subjective Phenomena." Unpublished paper, California Institute of Technology, 1997.

Federal Reserve Bank of New York. Survey of Consumer Expectations Labor Market Survey. <https://www.newyorkfed.org/medialibrary/media/research/microeconomics/interactive/downloads/sc-e-labor-questionnaire.pdf> (2022).

Fuster, Andreas and Basit Zafar. "Field experiments in surveys." In *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Amsterdam: North-Holland/Elsevier, 2023

Giustinelli, Pamela, Charles F. Manski, and Francesca Molinari. "Tail and Center Rounding of Probabilistic Expectations in the Health and Retirement Study." *Journal of Econometrics* (2020), forthcoming.

Giustinelli, Pamela, Charles F. Manski, and Francesca Molinari. "Precise or Imprecise Probabilities? Evidence from Survey Response Related to Late-onset Dementia." *Journal of the European Economic Association* 20.1 (2022): 187-221.

Giustinelli, Pamela and Matthew D. Shapiro. "SeaTE: Subjective ex ante Treatment Effect of Health on Retirement." National Bureau of Economic Research Working Paper 26087 (2019).

Haider, Steven J. and Melvin Stephens Jr. "Is There a Retirement-Consumption Puzzle? Evidence Using Subjective Retirement Expectations." *The review of economics and statistics* 89.2 (2007): 247-264.

Health and Retirement Study. *Questionnaires*. <https://hrs.isr.umich.edu/documentation/questionnaires> (2022).

Ho, Jeong-Hwa and James M. Raymo. "Expectations and Realization of Joint Retirement Among Dual-Worker Couples." *Research on Aging* 31.2 (2009): 153-179.

Honig, Marjorie. "Retirement Expectations: Differences by Race, Ethnicity, and Gender." *The Gerontologist* 36.3 (1996): 373-382.

Hudomiet, Peter, Michael D. Hurd, and Susann Rohwedder. "The Causal Effects of Economic Incentives, Health and Job Characteristics on Retirement: Estimates Based on Subjective Conditional Probabilities." *Unpublished working paper presented at the AEA meetings*, January 2021 (2020).

Hudomiet, Péter, Michael D. Hurd, Susann Rohwedder. "Health expectations." In *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Amsterdam: North-Holland/Elsevier, 2023.

Hurd, Michael. "Labor Market Transitions in the HRS: Effects of the Subjective Probability of Retirement and of Pension Eligibility." In Smith JP, Willis RJ, eds. *Wealth, Work and Health: Innovations in Measurement in the Social Sciences*. Ann Arbor, MI: University of Michigan Press; 1999.

Kézdi, Gábor and Robert J. Willis. "Household Stock Market Beliefs and Learning." National Bureau of Economic Research Working Paper 17614 (2011).

Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. "Imputing Risk Tolerance from Survey Responses." *Journal of the American Statistical Association* 103.483 (2008): 1028-1038.

Koşar, Gizem and Cormac O'Dea. "Expectations data in structural micro models." In *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Amsterdam: North-Holland/Elsevier, 2023.

Lumsdaine, Robin L. , James H. Stock, and David A. Wise. "Three Models of Retirement: Computational Complexity versus Predictive Validity." In David A. Wise Ed. *Topics in the Economics of Aging*. Chicago: University of Chicago Press, 1992.

Juster, F. Thomas. "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design." *Journal of the American Statistical Association* 61.315 (1966): 658-696.

Maestas, Nicole. "Back to Work: Expectations and Realizations of Work after Retirement." *Journal of Human Resources* 4.3 (2010): 718-747.

Manski, Charles F. "Analysis of Choice Expectations in Incomplete Scenarios." *Journal of Risk and Uncertainty* 19 (1999): 49-66.

Manski, Charles F. "Steps forward, outlook." In *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Amsterdam: North-Holland/Elsevier, 2023.

McGarry, Kathleen. "Health and Retirement: Do Changes in Health Affect Retirement Expectations?" *Journal of Human Resources* 39.3 (2004): 624-648.

Michaud, Pierre-Carl, Arthur Van Soest, Luc Bissonnette. "Understanding Joint Retirement." *Journal of Economic Behavior and Organization* 173 (2020): 386-401.

Mermin, Gordon B.T., Richard W. Johnson, and Dan P. Murphy. "Why Do Boomers Plan to Work Longer?" *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 62.5 (2007): S286-S294.

Mueller, Andreas I. and Johannes Spinnewijn. "Expectations data, labor market and job search." In *Handbook of Economic Expectations*, Eds. Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Amsterdam: North-Holland/Elsevier, 2023.

Okumura, Tsunao and Emiko Usui. "The Effect of Pension Reform on Pension-Benefit Expectations and Savings Decisions in Japan." *Applied Economics* 46.14 (2014): 1677-1691.

Panel Study of Income Dynamics. *Questionnaire*.
<https://psidonline.isr.umich.edu/guide/documents.aspx> (2019).

Pienta, Amy M. and Mark D. Hayward. "Who Expects to Continue Working after Age 62? The Retirement Plans of Couples." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 57.4 (2002): S199-S208.

Romm, Aylit Tina. "The Effect of Retirement Date Expectations on Pre-Retirement Wealth Accumulation: The Role of Gender and Bargaining Power in Married US Households." *Journal of Family and Economic Issues* 36.4 (2015): 593-605.

Wiswall, Matthew and Basit Zafar. "Determinants of College Major Choice: Identification using an Information Experiment." *Review of Economic Studies* 82.2 (2015): 791-824.

Wiswall, Matthew and Basit Zafar. "Human Capital Investments and Expectations about Career and Family." *Journal of Political Economy* 129.5 (2021): 1361-1424.

van der Klaauw, Wilbert. "On the Use of Expectations Data in Estimating Structural Dynamic Choice Models." *Journal of Labor Economics* 30.3 (2012): 521-554.

van der Klaauw, Wilbert and Kenneth I. Wolpin. "Social Security and the Retirement and Savings Behavior of Low-Income Households." *Journal of econometrics* 145.1-2 (2008): 21-42.