S*ea*TE: Subjective *ex ante* Treatment Effect of Health on Retirement

Pamela Giustinelli*
Bocconi University, IGIER, and LEAP

Matthew D. Shapiro**
University of Michigan and NBER

First Draft: July 11, 2019

Current Draft: January 2, 2023

The paper studies the effect of health on work among older workers by eliciting 2- and 4-year-ahead subjective probabilities of working under alternative health states. These measures predict work outcomes. Person-specific differences in working probabilities across health states can be interpreted as Subjective *ex ante* Treatment Effects (S*ea*TE) in a potential outcomes framework and as taste for work within a discrete choice dynamic programming framework. There is substantial heterogeneity in expectations of work conditional on health. The paper shows how heterogeneity in taste for work correlated with health can bias regression estimates the effect of health on retirement.

Keywords: Subjective probabilities, treatment- or state-contingent expectations, *ex ante*, individual-level treatment effects, unobserved heterogeneity, retirement, health.

JEL Codes: C21, C83, D84, J26.

Acknowledgements. This research uses data from the Vanguard Research Initiative (VRI) that was developed by a research team under a program project grant from the National Institute on Aging P01-AG026571. The Vanguard Group Inc. supported the data collection of the VRI. Vanguard's Client Insight Group and IPSOS SA were responsible for implementing the VRI survey and provided substantial input into its design. John Ameriks, Andrew Caplin, and Matthew D. Shapiro are co-principal investigators of the VRI. The design of the VRI benefited from the collaboration and assistance of Joseph Briggs, Wandi Bruine de Bruin, Alycia Chin, Mi Luo, Minjoon Lee, Brooke Helppie McFall, Ann Rodgers, and Christopher Tonetti as part of the program project, from Annette Bonner (Vanguard), and Wendy O'Connell (IPSOS SA). This project uses the VRI survey that was designed by Ameriks, Briggs, Caplin, Lee, Shapiro, and Tonetti. The conditional probability battery used in this paper, which was included in VRI Survey 4 and a module of the HRS 2016, was designed in collaboration with Michael Hurd, Peter Hudomiet, Gabor Kezdi, Susann Rohwedder, and Robert Willis. For documentation of the VRI, including a dynamic link to the survey instrument, see http://ebp-projects.isr.umich.edu/VRI/. For documentation of the HRS, see HRS 2016 experimental module 5 at https://hrs.isr.umich.edu/documentation/questionnaires. This research was supported by a grant from the U.S. Social Security Administration (SSA), funded as part of the Retirement Research Consortium (RRC). The findings and conclusions expressed are solely those of the author(s) and do not represent the views of SSA. We thank Feiya Shao, Ann Rodgers, and Dawn Zinsser for their assistance. We are grateful to Tom Crossley, Martin Fankhauser, Gábor Kézdi, Wilbert van der Klaauw, Eric French, and participants at many seminars and conferences for very helpful comments and discussion.

* Department of Economics, Bocconi University, Milan 20136, Italy. E-mail: pamela.giustinelli@unibocconi.it.
** Department of Economics and Survey Research Center, University of Michigan, Ann Arbor, MI 48106-1248, United States of America and NBER. E-mail: shapiro@umich.edu.

The willingness of older individuals to work later in life, especially as a consequence of economic, social, policy, and health shocks, is of key importance to economic performance. The uncertain sustainability of social security systems in the U.S. and elsewhere has drawn increasing attention to the issue of whether older workers would be able and willing to work longer,[1] while generating a resurgence of interest in the longstanding quest to quantify the effect of negative health transitions on the labor supply of older workers.[2] These issues are brought into sharp resolve by the effects on late-in-life labor force participation arising from the combined health and economic shocks of the COVID-19 pandemic. Assessing the effect of such shocks, especially those that do not recur regularly, is a challenge.

This paper introduces an approach to addressing this challenge. It uses survey modules implemented in the Vanguard Research Initiative (VRI) and the Health and Retirement Study (HRS) where respondents report the probability that they will still be working 2 and 4 years ahead, both unconditionally and under scenarios about their health in 2 and 4 years. We also elicit respondents' expectations about their health at the same horizons. We subsequently observe their actual health and work realizations.

This approach complements the econometric approach to studying the effect of health on retirement that relies on realizations of health and labor supply. An advantage of our approach over the latter is that it generates data on individuals on the effect of conditioning events such as bad health whether or not the conditioning event occurs. The approach, of course, depends on the validity of the elicited subjective expectations. We address validity both by examining the internal consistency of elicited expectations and by demonstrating their ability to predict realizations.

Health may operate through multiple mechanisms, including preferences, productivity, financial incentives, expectations, and longevity. Existing evidence points to a negative effect of bad health on employment, but there is little consensus on the magnitude of the effect. Blundell, Britton, Costa Dias, and French (2021) show that regression estimates of the total effect of health on labor supply near retirement are highly sensitive to controlling for initial conditions related to health and employment, pointing to a likely importance of persistent unobserved heterogeneity in taste for work and/or productivity that may be correlated with health. Elicited subjective expectations rely on individuals to take account of these multiple and hard-to-observe factors. We provide formal interpretations of the health-contingent working probabilities within potential outcomes (POF) and discrete choice dynamic programming (DP).[3]

[1] Coile, Milligan, and Wise (2018), Coile (2018), and Berger, Lopez-Garcia, Maestas, and Mullen (2021).
[2] Coile (2015), French and Jones (2017), and Blundell, Britton, Costa Dias, and French (2021).
[3] See Juster (1966) and Manski (1990, 1999) for early ideas on elicitation and interpretation of choice probabilities, and van der Klaauw (2012), Pantano and Zheng (2013), Stinebrickner and Stinebrickner (2014), Arcidiacono, Hotz, Maurel, and Romano (2020), and Wiswall and Zafar (2021) for recent advances.

The first interpretation within POF motivates the use of elicited health-contingent working probabilities to construct person-specific differences in work expectations across health states, interpretable as Subjective *ex ante* Treatment Effects (S*ea*TE), and to characterize the heterogeneity of effects across individuals. The second interpretation within DP clarifies the mapping between elicited health-contingent working probabilities and latent health-contingent values of continuing to work versus not. They contain information on individual unobserved taste for work. This approach enables us to assess the extent of the bias in regression estimates of the overall effect of health on labor supply implied by failing to account for unobserved heterogeneity in taste for work that is correlated with health.

The paper finds a large average effect of health on retirement, but also notable heterogeneity in the effect, in the sample of working VRI respondents aged 57 and higher. The S*ea*TE of bad health on work, i.e., the difference between the probability of working in bad health versus in good health, averages -28.5 percentage points 2 years ahead and -25.7 percentage points 4 years ahead. These averages mask substantial heterogeneity in both S*ea*TE and the underlying health-contingent working probabilities. S*ea*TE is zero (no effect of health on work) for almost 30 percent of respondents at both horizons. A few respondents report probabilities implying a positive S*ea*TE (more likely to work in low than in high health). The remaining 70 percent have a strictly negative S*ea*TE. We find similar results in the HRS.

The paper is organized as follows. Section I reviews the relevant literature. Section II interprets the health-contingent working probabilities within POF and DP. Section III describes the VRI study and presents descriptive results on the elicited probabilities, S*ea*TE, and its heterogeneity. It also provides evidence on the internal validity of the elicited probabilities. Additionally, it replicates the analysis in the HRS. Section IV uses the VRI panel structure to link the elicited expectations to realizations. Section V presents a simulation exercise that illustrates the potential for bias in regression estimates of the effect of health on retirement based on the distributional information in the elicited subjected probabilities.

I. Related Literature

Health and retirement with data on realizations. The determinants of retirement have been widely studied in economics and elsewhere (e.g., see reviews by Coile (2015), O'Donnell, van Doorslaer, and van Ourti (2015), De Nardi, French, and Jones (2016), Fisher, Chaffee, and Sonnega (2016), and French and Jones (2017)), and have drawn increasing attention in recent years due to the uncertain sustainability of social security systems and related calls for making workers work longer (e.g., Coile, Milligan, and Wise (2018), Coile (2018), and Berger, Lopez-Garcia, Maestas, and Mullen (2021)).

The role of health has been subject to much debate, owing to its centrality to workers' ability of working longer and to the difficulties of unpacking the health-retirement nexus. Because health might operate through a variety of mechanisms such as preferences, productivity, financial incentives, expectations, and horizon (e.g., Rust and Phelan (1997), Blau and Gilleskie (2001, 2008), French (2005), van der Klaauw and Wolpin (2008), Bound, Stinebrickner, and Waidmann (2010), French and Jones (2011), Garcia-Gomez, Galama, van Doorslaer, and Lopez-Nicolas (2017), and Blundell, Britton, Costa Dias, and French (2021)), the sign of the relationship is theoretically ambiguous. Additionally, health might affect labor supply in a variety of forms such as expected trajectory vs. unexpected shocks, earlier vs. later changes, types of conditions (e.g., Grossman (1972), Bound, Schoenbaum, Stinebrickner, and Waidmann (1999), Lumsdaine and Mitchell (1999), McGarry (2004), and Blundell, Britton, Costa Dias, and French (2021)). Both retirement and health are subject to measurement issues which exacerbate the challenges of studying their relationship empirically (e.g., Bound (1991), Dwyer and Mitchell (1999), Lindeboom and Kerkhofs (2009), Datta Gupta and Larsen (2010), Kapteyn and Meijer (2014), and Blundell, Britton, Costa Dias, and French (2021) on health measurements, and Gustman, Mitchell, and Steinmeier (1995), McGarry (2004), Benítez-Silva and Dwyer (2005), Gustman, Steinmeier, and Tabatabai (2010), and Maestas (2010) on concepts and measures of work and retirement).[4]

Using panel data on realized health and labor supply from the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study of Aging (ELSA), the recent analysis of Blundell, Britton, Costa Dias, and French (2021) investigates two main issues and their implications for inference. The first is how health is measured and modelled. The second is the potential relevance of permanent individual heterogeneity in unobserved preferences, productivity, and health. Using regression equations linking work and health derived from a life-cycle labor supply model, they find estimates of the total effect of health on labor supply near retirement to be overall robust to alternative choices of the health variables (e.g., subjective, objective, or combinations of theirs), but highly sensitive to controlling for initial conditions related to health and employment. This suggests that potential biases stemming from classical measurement error ("attenuation bias") and from misreporting of health *ex post* to justify retirement ("justification bias") are

[4] The magnitude of the relationship is hard to quantify empirically as health and work are jointly determined and might feed dynamically into each other. This has prompted researchers to investigate the potential effect of retirement on health (e.g., Rohwedder and Willis (2010), Coe and Zamarro (2011), and Behncke (2012)). This paper does not address this feedback, but it could be potentially addressed using our approach. That is, one could allow health to be an outcome of labor supply, rather than just a state of nature affecting workers' labor supply decisions through state-dependent utility, productivity (wage), etc. To investigate the effect of labor supply on subsequent health using subjective conditional expectations one would need to elicit respondents' subjective expectations for future health conditional on alternative labor supply decisions (e.g., working versus not working).

empirically small or offsetting each other. Whereas the presence of persistent unobserved heterogeneity in taste for work and/or productivity related to health is likely more consequential for inference.

By deploying our novel data on subjective expectations about work, health, and work given health, our analysis strongly complements this line of work, while shedding further light on issues it raises. First, we provide direct evidence of substantial individual-level heterogeneity in work attachment across healthy older workers and in their expected response to changes in health. Second, we show that the latent values of continued work underlying the health-contingent working probabilities are correlated within individuals across health states, suggesting that indeed they capture individual-specific preference for work. At the same time, we show that the cross-sectional heterogeneity in individual-specific preference for work, which is typically unobserved in realizations data, induces a bias in regression estimates of the overall effect of health on labor supply due to its correlation with the health transitions. We do so while avoiding the justification bias since, as first noted by McGarry (2004), using data on work expectations enables one to focus on workers.

While our approach employs data on subjective expectations, it features clear connections with approaches employing hypothetical choice data. Since choice probabilities under incomplete scenarios are essentially probabilistic measures of stated preferences (e.g., Manski (1990, 1999) and Blass, Lach, and Manski (2010)), our approach is closely related to the literature investigating preferences for work and retirement arrangements using hypothetical choice methods (e.g., Kapteyn, van Soest, and Zissimopoulos (2007), van Soest and Vonkova (2014), and Ameriks, Briggs, Caplin, Lee, Shapiro, and Tonetti (2020)). We contribute to this literature by eliciting the hypothetical choices probabilistically, and then interpreting them by using potential outcomes and dynamic programming frameworks.

At the same time, our approach shares with panel data methods the use of multiple observations per individual, enabling a within-individual analysis. In a panel, identification uses changes in health across time periods leading to changes in work status (e.g., Bound, Schoenbaum, Stinebrickner, and Waidmann (1999), Disney, Emmerson, and Wakefield (2006), van der Klaauw and Wolpin (2008), McGeary (2009), Cai (2010), Garcia-Gomez (2011), Maurer, Klein and Vella (2011), Datta Gupta, Kleinjans, and Larsen (2015), and Jones, Rice, and Zantomio (2016)). Our approach uses changes in work probabilities across alternative health scenarios. An advantage of our approach is that the state-contingent choice probabilities can be elicited in a single survey, while observation of a sufficient number of health transitions ("shocks") and related work transitions may require a long panel.[5,6] Another advantage is the complete flexibility in how to specify the relevant state. Given our interest in the effect of overall health on retirement and limited survey time, and in keeping with the tradition of structural models of retirement, we have decided to specify the state in terms of overall health. Alternatively, one could easily pose health scenarios describing specific health conditions. Note however that Blundell, Britton, Costa Dias, and French (2021) show that accounting for only selected health conditions can lead to substantial underestimation of the overall effect of health on labor supply near retirement.

Of course, like any other method, our approach is not free of potential limitations. Our choice of specifying the scenarios in terms of overall health, and on the standard 5-point scale (Excellent, Very Good, Good, Fair, Poor) typically used to elicit self-reported health in surveys, is subject to the criticism that different people may interpret the concept and/or scale differently. Moreover, some respondents may not hold rational expectations about their labor supply, or their expectations reports may feature measurement error. We address these issues in the next sections.

Health and retirement with data on expectations. Only a few papers have employed survey measures of subjective expectations to study retirement behavior and its relationship with individuals' health.

McGarry (2004) investigates the effect of health on labor supply expectations of working respondents in the early HRS waves. Using a regression analysis across individuals, the paper explores the roles of a variety of health measures (e.g., contemporaneous, lagged, and changes in self-reported health, diagnosed health conditions, and subjective longevity expectations) on respondents' unconditional probability of working past age 62 and its changes between the first and second HRS waves. The key innovation of the analysis is to replace actual labor supply with unconditional working expectations as a dependent variable so as to focus on working respondents and avoid justification bias in self-reported health among retirees. She finds negative average effects of health on the probability of working, larger than those previously documented in the literature using labor supply realizations. We build on and extend McGarry's idea by directly eliciting health-contingent working probabilities, instead of unconditional ones, and comparing them across hypothetical health states within respondents, rather than across respondents with different realized health states. Thus, our approach enables us to characterize the individual-level heterogeneity in older workers' attachment to work in different health states and in their response to health changes.

[5] Whenever possible, papers employing longitudinal data seek to exploit unexpected health shocks (e.g., unexpected accidents or insurgence of specific conditions), as credible sources of identifying information. However, fully exogenous shocks are rare.
[6] While in our survey we did not ask respondents to condition their 4-year-ahead expectations on the 2-year-ahead state or decision, one could imagine designing a survey with a multi-stage structure. Such a survey would generate expectations measures featuring individual-level panel variation both across states and across time.

van der Klaauw and Wolpin (2008) develop and estimate a dynamic programming model of household retirement and saving, using multiple waves of the HRS. Innovating on earlier structural models of retirement, they combine respondents' unconditional work and longevity expectations with observed realizations of respondents' labor supply, health, and the other state variables to increase estimation precision. Specifically, survey expectations are used to generate extra moment conditions for matching, in addition to those sufficient for identification, with the goal of improving efficiency of estimates. Our analysis does not entail the specification and estimation of a structural dynamic model. Instead, we use the DP framework to provide a formal interpretation of the health-contingent working probabilities as containing information about unobserved heterogeneity in taste for work, and to motivate their use in the simulation exercise where we assess the bias in regression estimates of the effect of health on work induced by failing to account for such unobserved heterogeneity.

Hudomiet, Hurd, and Rohwedder (2021) elicit subjective probabilities of working past age 70 from working respondents in the American Life Panel (ALP), under alternative scenarios about health, wealth, and other factors that may affect retirement. Like us, they compare health-contingent working probabilities across health states and find an average S*ea*TE of -23.4 percent chance points, very similar to those we obtain in the VRI (-28.5 and -25.7 at 2 and 4 years) and in the HRS (-26.8 and -26.9 at 2 and 4 years), notwithstanding some differences in the elicitation horizons and respondent characteristics.

II. Analytic Framework

A. POF Interpretation of S*ea*TE

We consider a simple potential outcomes framework (POF), with a binary treatment (the health state) and a binary outcome (the labor supply decision).[7] In period t, after observing the realized value of the state vector, s_{it}, the decision-maker decides whether to work or not, $d_{it} \in \{0,1\}$, where 1 denotes working, 0 not working, and i the decision maker. The state vector, s_{it}, includes the decision-maker's health, $h_{it} \in \{0,1\}$, where 1 denotes low health and 0 high health, and other variables discussed below.

Within-person differences in potential outcomes across pairs of hypothetical treatments yield *individual-level treatment effects* of the form,

$$\Delta_{it} = d_{it}(1) - d_{it}(0) \equiv E[\Delta_{it}] + [\upsilon_{it}(1) - \upsilon_{it}(0)], \quad (1)$$

[7] Arcidiacono, Hotz, Maurel, and Romano (2020)'s interpretation of average differences in students' earnings expectations across hypothetical occupations as *ex ante* treatment effects is also based on potential outcomes. Our presentation of the potential outcomes framework in this paper helps make explicit the value of data on conditional probabilities.

where $\Delta_{it} \in \{-1,0,1\}$, $E[\Delta_{it}]$ is the common "gain" (or "loss") from treatment (with the expectation taken across units), $[\upsilon_{it}(1)-\upsilon_{it}(0)]$ is the idiosyncratic gain (or loss) for individual *i*, and being treated corresponds to experiencing a negative health shock contemporaneous to the time of the decision. Recovering this effect entails the evaluation and comparison of the labor supply decisions that person *i* would make in two mutually exclusive and alternative states of the world at time *t*, $h_{it}=1$ (low health) and $h_{it}=0$ (high health).

Treatment Effects in Realizations. Define $z_{it} \in \{0,1\}$ to be the realized health state of person *i* at time *t*. As before, 0 means high health and 1 low health. Then, $d_{it} \equiv d_{it}(z_{it}) \in \{0,1\}$ is the person's *realized outcome*, whereas $d_{it}(1-z_{it}) \in \{0,1\}$ is the *counterfactual outcome*.

Obviously, the individual-level treatment effect in (1) is unobservable by construction because one of the two health states is necessarily counterfactual. Most implementations of POF compare outcomes *across groups* of suitably similar units rather than within units. For example, consider the Average Treatment Effect (ATE),

$$ATE_t(1-0) = E[d_{it}(1)-d_{it}(0)] = E\left[d_{it}(1)\right] - E\left[d_{it}(0)\right], \tag{2}$$

where $E\left[d_{it}(h)\right] \equiv P\left[d_{it}(h)=1\right]$ denotes the mean of the population labor supply distribution at time *t* if everyone were to be treated with health level *h*, conditioning variables are omitted for simplicity, and $E[\upsilon_{it}(1)] = E[\upsilon_{it}(0)] = 0$. Decomposition of the two means into realized and counterfactual components yields,

$$\begin{aligned} ATE_t(1-0) = &\left\{E\left[d_{it}(1)\,|\,z_{it}=1\right]P(z_{it}=1) + E\left[d_{it}(1)\,|\,z_{it}=0\right]P(z_{it}=0)\right\} \\ &-\left\{E\left[d_{it}(0)\,|\,z_{it}=0\right]P(z_{it}=0) + E\left[d_{it}(0)\,|\,z_{it}=1\right]P(z_{it}=1)\right\}. \end{aligned} \tag{3}$$

Random sampling of $\{z_{it}, d_{it}\}$ from the population distribution of realized health and labor supply asymptotically reveals all components of (3) except the counterfactual moments, $E\left[d_{it}(1)\,|\,z_{it}=0\right]$ (mean labor supply of high-health workers in a low-health world) and $E\left[d_{it}(0)\,|\,z_{it}=1\right]$ (mean labor supply of low-health workers in a high-health world). Thus, without additional assumptions on these counterfactual moments, the ATE parameter is not (point) identified.

With perfect compliance to treatment assignment, randomization of good and bad health across individuals would yield identification of ATE in (3). Randomization is not feasible with health realizations. Hence, the literature using observational data on health and labor supply needs to grapple with non-random selection of individuals into health states and with unobserved heterogeneity in the gain (or loss) from treatment (i.e., bad health).[8] For example, high-health individuals may have higher (unobserved) preference for work than low-health individuals that might persist if the former were to experience a negative health transition ($h_{it} = 1$), and *vice versa*. In this case, a simple comparison of mean labor supply across high- and low-health workers will generally confound ATE with selection and heterogeneity terms, as follows:

$$\begin{aligned} E\left[d_{it} \mid z_{it}=1\right]-E\left[d_{it} \mid z_{it}=0\right] = ATE_t(1-0) + \left\{E\left[\upsilon_{it}(0) \mid z_{it}=1\right]-E\left[\upsilon_{it}(0) \mid z_{it}=0\right]\right\} \\ +[1-P(z_{it}=1)]\left\{E\left[\upsilon_{it}(1)-\upsilon_{it}(0) \mid z_{it}=1\right]-E\left[\upsilon_{it}(1)-\upsilon_{it}(0) \mid z_{it}=0\right]\right\}. \end{aligned} \tag{4}$$

Treatment Effects in Expectations. In this paper, we measure potential outcomes *ex ante* by directly asking individuals to predict their labor supply in alternative scenarios about their health at specific horizons.[9] We define the *Subjective ex ante Treatment Effect*,

$$SeaTE(i,t,\tau) = E_{i,t-\tau}(\Delta_{it}) \equiv P_{i,t-\tau}\left[d_{it}(1)=1\right] - P_{i,t-\tau}\left[d_{it}(0)=1\right], \tag{5}$$

as the individual-level expectation at $t-\tau$ of the individual-level treatment effect at *t*, Δ_{it}.

We elicit the two components of S*ea*TE separately, as this approach fits the way in which respondents themselves seem to think and reason about the prediction problem (see evidence and discussion in Subsection III.G). Moreover, this approach is natural from the perspective of economic and econometric modeling, of which conditional probabilities are essential building blocks.

The individual-level effects in (5) can be aggregated across individuals to generate subjective *ex ante* versions of popular group-level parameters; for example, the Average Subjective *ex ante* Treatment Effect (AS*ea*TE), $E\left[SeaTE(i,t,\tau)\right]$, where the expectation is taken across individuals. In a similar fashion, the average S*ea*TE may be also separately computed among the treated (low health workers) and among the untreated (high health workers), using realized treatments or elicited health probabilities to weight the individual level effects (see Arcidiacono, Hotz, Maurel, and Romano (2020)).

[8] An exception is Stephens and Toohey (2022), who exploit a randomized trial aimed at reducing coronary heart disease mortality risk in the United States to evaluate the causal effect of health on workers' labor market outcomes.
[9] Following Manski (1999), a scenario can be formalized as a function assigning a potential choice set and environment to each member of the population. Hence, it is interpretable as a treatment policy or program. In our application, we focus on specification or fixing of specific features of the choice environment (a state variable) and leave the choice set unspecified. We assume that the latter consists of the two alternative options of working vs. not working.

An important question is under what conditions the expression in (5), or features of its cross-sectional distribution, may be interpreted as subjective *ex ante* causal effects. The answer depends on two main issues. The first issue concerns whether or not survey reports of the two health-contingent probabilities of working in (5) are *ceteris paribus* with respect to state variables that are relevant to the labor supply decision but are not specified in the elicitation task. We address this point formally in Subsection II.B, where we introduce separate notation for the state variables specified in the elicitation scenario and those not specified. Ultimately, interpretability of S*ea*TE as a causal effect *ex ante* depends on the nature of the relationship between the specified and unspecified states.[10]

The second issue concerns measurement error. As discussed by Arcidiacono, Hotz, Maurel, and Romano (2020), the *ex ante* treatment effects are identified directly from the subjective expectations data as long as survey expectations are not ridden by measurement error. Moreover, the average *ex ante* effects would be still identified in presence of measurement error of any form as long as the errors have the same mean across the values of the treatment variable being contrasted. We further discuss measurement error from rounding in Subsection III.G.

A distinct but equally important question is under what conditions the *ex ante* effects introduced in this section are informative of *ex post* effects. Here the answer depends on the nature of individuals' expectations. If at the time of elicitation individuals have rational expectations about the choice they would make in each state, and in the absence of unanticipated aggregate shocks, the AS*ea*TE coincides with the mean (*ex post*) effect of treatment on outcome (see also discussions in Manski (1999) and Arcidiacono, Hotz, Maurel, and Romano (2020)).[11]

---

[10] Hudomiet, Hurd, and Rohwedder (2021) investigate experimentally whether respondents tend to "fill in" unspecified aspects of the scenario, as hypothesized by Fischhoff, Welch, and Frederick (1999). They are interested in unspecified aspects that, in the respondent's eye, may be related to those specified. They find no evidence of "filling-in"; e.g., when reporting the probability of working past 70 under the scenario that they will earn a high wage, respondents do not seem to assume that they will be in high health. This evidence is consistent with a *ceteris paribus* interpretation of the state-contingent probabilities.

[11] Some weakened forms of rational expectations (e.g., respondents' subjective probabilities of working are unbiased estimates of their objective probabilities of working, which we show hold empirically for the VRI respondents under 70 but not over 70), and of statistical independence of realized health across the population, would leave the conclusion intact. However, aggregate shocks making negative health transitions dependent across the population and systematic deviations from rational expectations due to biased beliefs would generally invalidate it.

B. DP Interpretation of S*ea*TE

We now relate the components of S*ea*TE to the individuals' decision problem within a standard discrete choice dynamic programming (DP) framework in order to provide an interpretation of the health-contingent working probabilities in terms of the individuals' optimization problem. This framework, in addition to giving a structural interpretation to the responses, allows us to make explicit how they are probabilistic.[12]

Individuals are characterized by primitives, $u_{it}\left(s_{it},d_{it}\right)$ and $\pi_{it}\left(s_{i,t+1}\,|\,s_{it},d_{it}\right)$, with $i=1,\ldots,N$ and $t=0,1,\ldots,T<\infty$. $u_{it}(s_{it},d_{it})$ is the utility that person *i* derives in period *t* from choosing labor supply $d_{it}\in\{0,1\}$, given the realizations of the state variables collected in s_{it}. The expression $\pi_{it}(s_{i,t+1}\,|\,s_{it},d_{it})$ is person *i*'s subjective probability over the states' realizations in the next period (*t+1*), conditional on the agent's information set in the current period. The latter is summarized by the realized state and decision at *t*, as implied by the assumption that health and the other state variables are generated by a Markovian process.

With additively time-separable utility, the person's utility functional at *t* is given by

$$U_{it}=\sum_{j=0}^{T}\beta^{j}u_{i,t+j}\left(s_{i,t+j},d_{i,t+j}\right), \tag{6}$$

where β is the discount factor. The person behaves optimally according to the expected utility maximizing decision rule, $\delta_{it}^{*}\left(s_{it}\right)$, which satisfies the Bellman (1957) optimality principle. Formally,

$$\delta_{it}^{*}\left(s_{it}\right)=\underset{d_{it}\in\{0,1\}}{\arg\max}\left\{u_{it}\left(s_{it},d_{it}\right)+\beta\sum_{s_{i,t+1}}V_{i,t+1}^{*}\left[s_{i,t+1},\delta_{i,t+1}^{*}\left(s_{i,t+1}\right)\right]\cdot\pi_{it}\left(s_{i,t+1}\,|\,s_{it},d_{it}\right)\right\}, \tag{7}$$

where $V_{i,t+1}^{*}$ is the value function representing the expected present discounted value of lifetime utility from following $\delta_{it}^{*}\left(s_{it}\right)$, which is a deterministic function of s_{it} given the primitives.[13]

Dynamic programming fixing future health. We partition the state vector into variables fixed in the elicitation task and variables not fixed in the elicitation task. We further partition the latter into variables

[12] As traditional in the DP literature, in this subsection we maintain rational expectations. Recent contributions have begun to show that identification of dynamic discrete choice models without rational expectations is feasible (e.g., An, Hu, and Xiao (2021)). Bridging the emerging literature on dynamic discrete choice modelling without rational expectations with the fast-expanding literature on survey expectations is a promising venue for future research.

[13] Following Rust (1992) and the traditions of the DP literature, at this point we specify this dynamic program at a high level of abstraction including leaving constraints implicit.

that the researcher could reasonably fix in the elicitation task, if they decided to do so, and variables capturing any residual uncertainty of the agent at the time of elicitation about aspects of the choice environment that might affect their future decision. Formally, $s_{it} = (x_{it}, y_{it}, \varepsilon_{it})$, where x_{it} denotes the *specified* component of s_{it}, y_{it} denotes the *unspecified* component of s_{it}, and ε_{it} denotes the *residual* component of s_{it}. Under this partition,

$$U_{it} = \sum_{j=0}^{\mathrm{T}} \beta^j u_{i,t+j} \left[\left(x_{i,t+j}, y_{i,t+j}, \varepsilon_{i,t+j} \right), d_{i,t+j} \right]. \tag{8}$$

In our survey, we specify the individual's health state so $x_{it} = h_{it}$, while leaving unspecified in y_{it} additional factors that may affect retirement decision (e.g., family and financial conditions, income, and so on).[14] Because x_{it} is fixed in the elicitation task, it is no longer stochastic to the respondent at the time of elicitation. In this context, the variation in health is experimental; because we are fixing health, the estimates can have a causal interpretation. We assume that respondents place themselves in the hypothetical situation defined by the scenario, without trying to infer why one or the other scenario might be realized (see also discussions in Dominitz and Manski (1996) and Dominitz (1997)). On the other hand, y_{it} and ε_{it} are stochastic at the time of elicitation. We assume that respondents hold subjective distributions for the unspecified components of the choice environment at time t and allow them to express any uncertainty they might have about future decision, $\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)$ due to the uncertainty they might perceive about y_{it} and ε_{it}.

Without loss of generality because y_{it} embodies all omitted factors that could be specified, we maintain that ε_{it} is orthogonal to both x_{it} and y_{it}. This and the fact that x is fixed imply

[14] Specified scenarios are generally *incomplete* (Manski, 1999). An incomplete scenario can be thought of and formalized as a collection of scenarios, each sharing some common feature (the specified components). In our application, scenarios have in common a specified health level and a horizon length. Furthermore, the elicitation tasks implicitly condition on being alive. Likewise, the Markov transitions are implicitly conditional on being alive. Using the standard convention that utility when dead is normalized to be zero, conditioning on being alive is natural and has no effect on the optimization problem. A full model would, of course, need to account for mortality risk.

$$\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})=\underset{d_{it}\in\{1,0\}}{\arg\max}\; u_{it}\left[(x_{it},y_{it},\varepsilon_{it}),d_{it}\right]+$$

$$\beta\sum_{x_{i,t+1}}\left(\int_{\varepsilon_{i,t+1}}\int_{y_{i,t+1}}V_{i,t+1}^{*}\left[(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1}),\delta_{i,t+1}^{*}(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1})\right]\cdot\pi_{it}^{y}(y_{i,t+1}\mid x_{i,t+1},y_{it},d_{it})\,dy_{i,t+1}\cdot\pi_{it}^{\varepsilon}(\varepsilon_{i,t+1}\mid\varepsilon_{it},d_{it})\,d\varepsilon_{i,t+1}\right)$$

$$\pi_{it}^{x}(x_{i,t+1}\mid x_{it},y_{it},d_{it}),$$

(9)

where we replace summation with integral to allow for the possibility that y and ε are continuous. Here ε is unknown to both the econometrician and the individual at the time of elicitation.[15] This is in contrast to the more typical setting for modeling outcome data where the respondent knows a component that is unobserved to the econometrician.

The expectation of the optimal decision $\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})$ as of elicitation is

$$\begin{aligned}
&P_{i,t-1}\left[\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})=1\right]\\
&=\sum_{x_{it}}\left[\int_{\varepsilon_{it}}\int_{y_{it}}\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})\cdot\pi_{i,t-1}^{y}(y_{it}\mid x_{it},x_{i,t-1},y_{i,t-1},d_{i,t-1})\,dy_{it}\cdot\pi_{i,t-1}^{\varepsilon}(\varepsilon_{it}\mid\varepsilon_{i,t-1},d_{i,t-1})\,d\varepsilon_{it}\right]\cdot\pi_{i,t-1}^{x}(x_{it}\mid x_{i,t-1},y_{i,t-1},d_{i,t-1})\\
&=\sum_{x_{it}}P_{i,t-1}\left[\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})=1\mid x_{it}\right]\cdot\pi_{i,t-1}^{x}(x_{it}\mid x_{i,t-1},y_{i,t-1},d_{i,t-1}),
\end{aligned}\tag{10}$$

where $P_{i,t-1}\left[\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})=1\mid x_{it}\right]$ is the individual's expected optimal decision in period *t*, obtained by integrating $\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})$ with respect to the distributions of ε_{it} and y_{it}, and by evaluating the resulting function at a particular realization of x_{it} specified by the elicitation task.

Consider the implications of (10) for a survey response. From the viewpoint of a respondent at time *t*-1, the optimal choice at time *t*, $\delta_{it}^{*}(x_{it},y_{it},\varepsilon_{it})$, is a random variable, as it is a function of random variables x_{it}, y_{it}, and ε_{it}. As the elicitation scenario fixes the value of x_{it}, the uncertainty associated to the stochastic process for x_{it} gets partialed out into the transition probabilities, $\pi_{it}^{x}(x_{i,t+1}\mid x_{it},y_{it},d_{it})$. On the other hand, uncertainty may remain about y_{it} and ε_{it}. For this reason, we allow respondents to report their expected choice probabilistically, expressed as their subjective probability of working contingent on the specified health state. Absent uncertainty about factors driving choices in the future, respondents would give either a zero or one response because labor supply in the future would be a deterministic function of health.

[15] Here we treat ε as a scalar. Each process of the choice environment could feature its own residual component, e.g., one in the agent's utility, one in the wage process, and so on.

We elicit all components of (10) as follows:

(i) On the right-hand side of (10), the probability of working fixing the state, $P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)=1 \mid x_{it}\right]$, and the probability of the specified state, $\pi_{i,t-1}^{x}\left(x_{it} \mid x_{i,t-1}, y_{i,t-1}, d_{i,t-1}\right)$, with $x_{it} \equiv h_{it}$.

(ii) On the left-hand side of (10), the unconditional working probability, $P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)=1\right]$.

Health can affect the unconditional probability of working through three channels. The first channel is preference (i.e., agent's utility). The second is the mechanism or mechanisms represented by the unspecified component, y_{it}, (e.g., wage or productivity). The third is uncertainty (i.e., agent's subjective belief about the stochastic process governing health). On the other hand, health only affects the health-contingent working probabilities through the first two channels.

The survey's elicitation task maps into the quantities described. Specifically, the question

```
If your health is excellent [very good/good] two years from now, what
are the chances that you will be working for pay?
```

yields

$$P_{i,t-1}^{H} \equiv P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=0\right] \tag{11}$$

and

```
If your health is fair [poor] two years from now, what are the chances
that you will be working for pay?
```

yields

$$P_{i,t-1}^{L} \equiv P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=1\right]. \tag{12}$$

SeaTE. We can now link these elicitations to S*ea*TE. The 1-period-ahead S*ea*TE of health on work for individual i at time t is $\mathrm{S}ea\mathrm{TE}(i,t,1)=P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=1\right]-P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=0\right]$. As long as y may depend on h, in equation (10) the individual should integrate over the future values of y_{it} that are consistent with the values of h_{it} fixed in the elicitation task. The main implication for interpretation of S*ea*TE is that in general it measures a *total* effect: the effect of health operating through all the mechanisms by which health affects labor supply. In our working illustration, it would be the effect of health on labor supply through both utility and productivity. One could decompose the effect of health that operates through wages versus other factors by specifying both wages and health in the elicitation task.

III. Eliciting Conditional Probabilities: Survey and Basic Results

A. The Vanguard Research Initiative (VRI)

The VRI is a longitudinal survey-administrative linked dataset on older wealthholders, who are account holders at the Vanguard Group. At the time of the initial survey wave in 2013, recruited respondents were aged 55 and above, web-survey eligible, and had at least $10,000 in financial assets at Vanguard. As of 2015, four surveys were completed by a panel of about 3,000+ VRI respondents, with each survey focusing on a different aspect of retirement decision-making. Our analysis is based mainly on Survey 4 (Labor), while Survey 1 (Wealth), Survey 2 (Long-term Care), and Survey 3 (Transfers) provide relevant covariates. Additionally, we use realized health and work in 2017, collected in Survey 6, to validate our 2-year ahead probability measures elicited in Survey 4.

Survey 4 begins by asking whether an individual is working. If so, it gets facts about the current job and establishes if it is the career job (Current job battery). If yes, it gets information about whether the individual is searching for another job (On-the-job search battery). If not, it gets information about the career job, separation from it, and subsequent search (Career job, Separation, and Career-to-bridge search batteries). If not working, there is a similar sequence starting with information about last job. This sequence establishes information about career job, bridge job (if relevant), and the transitions and search. Respondents who were working in either a career job or bridge job at the time of Survey 4 were asked a series of questions regarding their labor supply and health expectations (described below) that are the key inputs to our analysis.

B. Analytic Sample

We select our sample from respondents who meet the following criteria: (i) who have taken the first 4 surveys of the VRI; (ii) who were working in either their career or bridge job at the time of Survey 4 and, thus, eligible to answer the labor supply and health expectations battery; (iv) who gave complete and consistent responses to the latter battery; and (v) who reported being in high health in Survey 4.[16] Table A1 in the online appendix summarizes the selection process. The analytic sample amounts to 970 respondents aged 57 to 81, currently working, and in high health. For the analysis of expectations with a 4-year horizon, the sample decreases to 839 respondents because it excludes those who reported a zero percent chance of working 2 years ahead. See Table A2 in the online appendix for sample statistics.

[16] This filter was applied *ex post*, as fewer than 3% of respondents reported being in low health (fair or poor).

VRI respondents tend to be wealthier, more educated, and healthier than the general population. However, conditional on the VRI sample screens (age, positive financial wealth, and internet access), they are broadly similar to those from the HRS and the Survey of Consumer Finances (SCF) (Ameriks, Caplin, Lee, Shapiro and Tonetti, 2015).

C. S*ea*TE and Its Components

In the work-health expectations battery, eligible VRI respondents are first asked the percent chance out of 100 that they will be working in 2 years. Next, they are asked their self-rated health on a 5-point scale (Excellent, Very Good, Good, Fair, and Poor), and the percent chance out of 100 that their health will be equal to specific states in 2 years. Finally, respondents are asked about their probability of working in the next 2 years if their health is equal to specific states.[17] These questions were then repeated for the 4-year horizon. Hence, as common in the analysis of data from biennial surveys (HRS, PSID, for examples), when relating these answers to the analytic framework above, we will treat a time period as two years.

To economize on the number of questions, the expectations module uses three partitions of the 5-point scale of self-rated health. The partition of future health states used in the questions depends on the current level of health reported by the respondent. Figure A1 of Online Appendix A shows the partitions used for each level of initial health. Note that they map uniquely to the high (excellent/very good/good) and low (fair/poor) dichotomous classification used in this paper. Online Appendix C gives the full survey module including how the partitioning affects questions.

For example, consider a respondent who reported being in good health at Survey 4. This respondent is asked the following sequence of questions for the 2-year horizon:

```
1) What are the chances that you will be working 2 years from now? [fill-in
   box]%

2) What are the chances that your health will be fair or poor 2 years from now?
   [fill-in box]%
```

[17] Both VRI and HRS respondents were already familiar with the 5-point health scale that has been shown to be reliable in many contexts and is routinely used in the structural literature we reviewed earlier. Alternatively, one could ask about work under various diagnoses (e.g., high blood pressure, cancer, inability to lift, cognitive decline). This could be impractical, however, especially because different conditions would be relevant for different types of work and because there are competing, multiple health risks. Moreover, Blundell, Britton, Costa Dias, and French (2021) show that accounting for only selected health conditions can lead to substantial underestimation of the overall effect of health on labor supply near retirement. If respondents are heterogeneous in their use of the scale, this would not necessarily be a problem for our approach, though it would be a problem if respondents come to a subjective health assessment based on their ability or willingness to work.

3) `What are the chances that your health will be very good or excellent 2 years from now? [fill-in box]%`

4) `If your health is very good or excellent 2 years from now, what are the chances that you will be working for pay? [fill-in box]%`

5) `If your health is good 2 years from now, what are the chances that you will be working for pay? [fill-in box]%`

6) `If your health is fair or poor 2 years from now, what are the chances that you will be working for pay? [fill-in box]%`

Table 1 shows the empirical distributions of our main survey measures: the unconditional probability of working, the unconditional probability of low health, the probability of working in low health, the probability of working in high health, and the S*ea*TE. For streamlined exposition and to match with the binary specification of the DP problem, we combine the two answers for the high health state (E/VG/G) into a single high state (*H*) for the statistics in this section. (See Appendix Tables A5 and A6 for statistics with the underlying detail.) For each measure, it reports the mean, standard deviation, first quartile, median, and third quartile of the empirical distribution. The figures shown in the top panel refer to the 2-year ahead expectations, while those in the bottom panel refer to the 4-year ahead expectations. Respondents' working expectations 2 and 4 years ahead are very heterogeneous and span the whole support of the 0-100 percent chance scale. The median belief of 80 percentage points in the top panel is quite high. This figure decreases to 50 percentage points at the 4-year horizon.

Health expectations are relatively high and less heterogeneous than work expectations. The mean of the distribution of respondents' 2-year-ahead subjective probability of entering low (fair or poor) health is 16.6 percentage points; at the 4-year horizon, the mean is 23.5 percentage points.

The next two columns show the empirical distributions of the working probabilities in low health and in high health. Consider first the percent chance of working in high health. Its mean probability 2 years ahead is 70.5 percentage points, somewhat higher than the 65.9 percentage points mean unconditionally in the first column. The median displays a similar pattern. The relative similarity between reports of unconditional working probabilities and of working probabilities in high health results from the high and relatively undispersed probability of remaining in high health.

Having respondents entertain a scenario of low health lowers substantially their self-reported working expectations at both horizons. For example, in the 2-year horizon the median of the distribution of the health-contingent working probabilities drops from 90 to 40 percentage points between high and low health. Similarly, the mean drops from 71 to 42 percentage points. In the 4-year case, the median drops from 68 to 20 and the mean from 59 to 33.

Online Appendix A presents and discusses results by age group. A noteworthy result is that working probabilities track focal retirement ages.

D. Unpacking S*ea*TE

We now discuss the responses in terms of S*ea*TE. Optimal behavior can be consistent with negative, zero, or positive S*ea*TE. Negative S*ea*TE is the leading case corresponding to a lower probability of working in low health owing to health-contingent disutility of work or productivity. Table 2 shows that approximately 70 percent of respondents have expectations consistent with a lower probability of working in low health. Most of the remaining respondents have zero S*ea*TE, which means that they have the same probability of working regardless of health. A few respondents have positive S*ea*TE, which is a logical possibility, for example, due to valuing leisure less or income more in low health. These fractions are similar across the 2- and 4-year horizons.

There are three ways to have zero S*ea*TE: never work regardless of health, always work regardless of health, or work with the same likelihood regardless of health. Table 3 shows that almost a third of these respondents expect to never work in 2 years. Another 47 percent of respondents expect to always work, while the remaining 21 percent are interior. The fractions who expect to never work and always work flip at 4 years as the tendency to retire regardless of one's state of health increases.

Table 4 focuses on the size of S*ea*TE among respondents with negative S*ea*TE using the same format as Table 1. Among this majority group where negative health transitions are expected to reduce work, there remains considerable heterogeneity in the size of the effect of health on work.

E. Observed Heterogeneity in S*ea*TE

Is S*ea*TE related to observed respondent characteristics? Table 5 reports estimates of linear regressions of 2- and 4-year ahead S*ea*TEs on a vector of covariates capturing respondents' background characteristics. Table A2 in the Online Appendix reports the summary statistics for the covariates.

The covariates are dummy variables for respondents' age (60-61, 62, 63-64, 65, 66-67, 68-69, 70-71, ≥ 72 vs. ≤ 59); gender (female vs. male); educational attainment (some college, college graduate, other advanced degree, MBA, JD/PhD/MD vs. high school or less); occupation (operative, other services vs. management and professional); job type (bridge vs. career); marital status (partnered vs. not); spouse's work status (working vs. not, for partnered respondents only); total household wealth (quintiles 1-4 vs. 5); replacement rate (quintiles 1-4 vs. 5); current salary (quintiles 1-4 vs. 5); and work status as of Survey 1 (completely retired vs. not). Since all our respondents were working at Survey 4, the latter dummy distinguishes between respondents who had previously retired but returned to work by Survey 4 and respondents who worked continuously. Time-varying characteristics are measured as of Survey 4. Thus, the reference group is single male respondents under 60, with a high school diploma or less, working in a bridge job of a management or professional type, in the top quintiles of household wealth, replacement rate, and salary, and working at Survey 1. The average 2-year S*ea*TE in the reference group is -15 points on the 0-100 percent chance scale and the average 4-year S*ea*TE is -11 points.

The S*ea*TE of bad health on work decreases (meaning a negative and larger in magnitude effect) non-monotonically with age, *ceteris paribus*. Relative to the under-60, respondents aged 62, 68-69, 70-71, and 72+ have an economically and statistically significantly lower S*ea*TE of 10 to 12 points 2 years ahead. Similarly, respondents aged 65, 68-69, 70-71, and 72+ have a lower S*ea*TE of 8 to 11 points 4 years ahead. Thus, older respondents and respondents whose S*ea*TE horizon corresponds to statutory retirement ages report negative and larger in magnitude perceived effects of a downward transition from high to low health on work. Since our sample is limited to a single cross-section of working respondents in high health, the observed heterogeneity in S*ea*TE by age likely arises from selection on both health and taste for work, especially for the oldest groups.

Over 9.5 percent of respondents in the analytic sample (see Table A2), all of whom are by construction working at the point of the elicitation of the conditional probabilities in Survey 4, reported being completely retired in Survey 1, which was fielded 2 years earlier. This "unretirement" is a large and highly significant indicator of differences in S*ea*TE across respondents, especially 2 years ahead. As reported in the last line of Table 5, on average, respondents who were completely retired by Survey 1 but were observed to work at Survey 4 have an 10.8-point higher S*ea*TE 2 years ahead (statistically significant), and a 3.9-point higher S*ea*TE 4 years ahead (not significant), than their counterparts who are observed to work in both waves. That is, the S*ea*TE of respondents who are observed to unretire between VRI waves 1 and 4 is still negative on average, but smaller in magnitude than the average S*ea*TE of respondents who are observed to work continuously.

Not surprisingly, respondents who are observed to unretire in between waves tend to be older on average. They are more likely to work in bridge jobs, and correspondingly earn substantially lower earnings as they work less hours/weeks. Their health is essentially the same as those who worked continuously.[18] While we do not know or directly observe these respondents' reasons for returning to work after retiring, since they are working bridge or part-time jobs, it suggests they unretired because they missed some aspect of working. Consequently, a taste or need for work makes them less sensitive to health shocks. It is not surprising that this effect diminishes at the 4-year horizon, when most expect to be retired regardless of health.

Table 5 reveals further heterogeneity in S*ea*TE. Respondents with a current salary in the bottom two quintiles of the distribution have a 7 to 8 percentage point lower S*ea*TE 2 years ahead (statistically significant). Respondents with a household wealth in the second quintile have also an 8 percentage point lower S*ea*TE, this time 4 years ahead. The estimated differences for the other groups are usually smaller in magnitude and not statistically significant.

While S*ea*TE correlates sensibly with selected covariates, most of its variation across VRI respondents reflects unobserved heterogeneity in differential work attachment across health states. Thus, conditioning on commonly available covariates in observational studies is unlikely to account for all unobserved heterogeneity and its distorting effects on inference, an issue we study in Section VI.

F. Internal Consistency: Law of Total Probability

Since we elicit the probability of work given health, the probability of health, and the unconditional probability of work, we can evaluate the coherence of responses with the law of total probability (LTP). VRI respondents are quite good at applying the LTP.

Figure 1 gives plots of the reported unconditional probability of work versus that implied by the LTP for the 2-year horizon using box and whiskers plots for various bins. (The survey did not ask the unconditional probability of work for the 4-year horizon, so we can only do this exercise for the 2-year horizon.) A large majority of the observations lies very close to the 45-degree line, corresponding to the case in which the self-reported probability and the calculated one are equal to each other. The correlation between the two measures is 0.928. For those responses that deviate, most deviations are relatively small. Therefore, the vast majority of respondents appear to understand the logic of probabilities.

The deviations from consistency of the LTP appear most pronounced among respondents with self-reported unconditional working probabilities equal to 0, 50, or 100 percent.[19] Note, however, that there

[18] The regression controls for all these differences.

[19] This finding is in principle consistent with the suggestion in the literature that some respondents who give 50/50 or corner responses may be more uncertain and/or less good at probabilistic thinking (e.g., Fischhoff and Bruine de Bruin (1999) and Hudomiet and Willis (2013)). However, using HRS data, Giustinelli, Manski, and Molinari (2022b) find that 0-percent

are significant mass points of respondents at the corners of the figure and, though not readily apparent because of the mass points, most of these respondents are getting the LTP exactly right.

G. Credibility of Measuring Expectations for Hypothetical Scenarios

How reliable is the elicitation of expectations, particularly under hypothetical scenarios? Any intertemporal economic choice requires individuals to think about the future. The perspective of the literature on subjective expectations is that it is possible to elicit useful measures that can be used for economic analysis. We showed in the previous subsection that our respondents are quite adept at working with conditional probabilities. In this subsection, we consider additional issues related to elicitation.

Do Respondents Think Conditionally? Cognitive Interviews. At the outset of this project, we conducted cognitive interviews in the Fall 2011 and Spring 2012 with a research team of economists led by psychologist Wandi Bruine de Bruin. A group of respondents drawn from the Health and Retirement Study (HRS)'s pre-test sample was encouraged to think aloud while answering percent chance questions about their future labor supply and other events, like those asked in the expectations section of the HRS (section P) and used in this paper. The following is a representative quotation for the cognitive interviews:

> I was thinking about my health, the way things are going in the economy, I don't know if it's going to really pick up [...] There might be a chance of me working and there might be a chance that there won't be much work when I'm that age. If I'm in good health [...] Well, I have no retirement, so if I am working, it's going to have to be later than 65 if my health is good where I can work.

When thinking aloud, many respondents made similar explicit reference to the uncertainty surrounding the events they were asked to predict and addressed the question by going through chains of contingent hypothetical (or "if") reasoning of the type displayed in the reported text. In the question asking respondents their subjective probability of working past 62, many of them thought of the probability of working past 62 if in good health or if in bad health, which motivated the approach of this paper.

Are the Contingencies Salient? Decision-making under uncertainty typically involves disjunctions of possible states: either one state will occur, or another. Although many HRS pre-test respondents who answered our cognitive interviews displayed a systematic and spontaneous use of contingent thinking without probes by the interviewers, research in psychology and economics has found that laypeople may

---

responses reflect, if anything, less rounding and imprecision than interior responses, while 50-percent responses reflect no differential imprecision, though somewhat more rounding.

have difficulty thinking through disjunctions via contingent thinking. Much of the same research, however, has also found that when the decision problem is framed in a way that makes the contingencies explicit, quality of choice or probabilistic judgments improve substantially (e.g., see Shafir (1994), Martinez-Marquina, Niederle, and Vespa (2019), and Esponda and Vespa (2019), among others). Our survey questions explicitly pose the contingencies.

Even so, a potential concern with our approach is that respondents may find it difficult to predict outcomes in unfamiliar or low-probability scenarios. While this may be a fair concern in general, we do not believe it to be a serious one in our context. Working respondents in the VRI and HRS are accustomed to thinking about retirement planning and related issues involving their future health, aging, working conditions, work prospects, personal finances, and the like. While VRI respondents tend to be healthier and wealthier than the general population of the same age, thus facing a lower probability of entering a bad health state over the next period on average, they are still older individuals likely to have had direct personal experience or indirect ones through family, friends, and colleagues, with health issues and aging circumstances.

Rounding in Elicited Probabilities. Probabilities elicited on a 0-100 percent chance scale often display heaping at 0, 50, 100, and other multiples of 10 and 5 percent, pointing to rounding of reports, a non-classical form of measurement error.[20] Especially since the S*ea*TE is the difference between two elicited probabilities, there may be concerns that rounding introduces non-classical measurement error. Recent work in the survey expectations literature has shown that the prevalence of rounding depends on features of the survey mode, elicitation format, question wording, and expectation domain. For instance, using all expectations collected in the HRS between 2002 and 2014, Giustinelli, Manski, and Molinari (2022a) find less rounding among survey reports of subjective probabilities within the domain of personal finances (including the probability of working past specified ages), relative to those belonging to the domains of personal health and general economic conditions. In the domain of long-term care utilization and insurance purchase behaviors with uncertain dementia state, Giustinelli, Manski, and Molinari (2022b) find that survey reports of state-contingent choice probabilities feature substantially less rounding or approximating than unconditional choice probabilities. This decrease in the incidence of rounded or approximated reports is driven by a lower proportion of respondents perceiving belief ambiguity when answering the state-contingent questions than when answering the unconditional questions ("imprecise-probability

[20] See Giustinelli, Manski, and Molinari (2022a) for an extensive analysis of rounding of probabilistic expectations in the HRS, including a review of the relevant literature and discussions on the nature of measurement error resulting from rounding and its possible interpretations within the context of survey reports of subjective probabilities.

respondents” in Giustinelli, Manski, and Molinari (2022b)’s taxonomy) and, to a lesser extent, by a lower proportion of precise-probability respondents who round their reports when answering state-contingent questions than when answering unconditional questions.

Other Checks on Validity within Survey. The VRI project also implemented a number of procedures and checks on the validity of the responses and the understanding of the respondents: (1) The survey firm conducted chats with pilot respondents. These chats did not reveal problems with respondents’ comprehension of expectations questions or the even more complex strategic survey questions (SSQ) analyzed by Ameriks, Briggs, Caplin, Shapiro, and Tonetti (2020). (2) There was no item-level nonresponse on the conditional probability questions for those who entered the battery. (3) Only 5.6 percent of respondents gave inconsistent answers to our probabilistic questions. (4) The free-response question at the conclusion of the survey did not yield complaints about these questions. Finally, the responses have good internal validity in that they closely obey the law of total probability.

H. Replication in the Health and Retirement Study

We do a parallel analysis of results in this section using data from an experimental module of the 2016 administration of the Health and Retirement Study (HRS), where we fielded the same battery of expectations questions as in the VRI. We summarize the main findings here and describe the analysis in greater detail in Online Appendix B. For parallelism with the main analysis based on VRI data, we focus on a sample of 483 HRS respondents who, in addition to taking the module, were working and in high health at the time of the survey and gave complete and consistent (or close to consistent) responses to the expectations battery. The VRI respondents, who are sampled from a population with significant financial wealth, are older, healthier at the same age, more educated, and more affluent than the HRS respondents. Hence, the results are not meant to be directly comparable, but rather to demonstrate the applicability of the approach in different populations.

Tables B3-B7 in Online Appendix B report HRS results parallel to the VRI results in Tables 1-5. On average, HRS respondents have higher probabilities of working than VRI respondents at both horizons, both unconditionally and conditional on either health state. They also have higher average probability of entering low health, although the difference is not large, especially at 4 years.

The HRS sample has more zero and positive S*ea*TE respondents than the VRI sample and fewer negative S*ea*TE respondents, although the differences are quite small (3 percent more zero S*ea*TE at 2 years, 0.4 percent more positive S*ea*TE at 2 years, 1.6 percent more positive S*ea*TE at 4 years). This similarity between HRS and VRI responses obtains despite there being evidence that HRS respondents have more difficulty in answering the questions (e.g., more violations of the law of total probability and more inconsistent responses). See Online Appendix B.

Even though the proportion of zero S*ea*TE respondents is only marginally higher in the HRS than in the VRI, their composition in terms of the underlying conditional probabilities looks quite different. In particular, the relative size of the never-work group is much smaller in the HRS than in the VRI, whereas the relative size of the always-work and maybe-work groups is larger.

Among negative S*ea*TE respondents, the distribution of S*ea*TE is remarkably similar in the VRI and HRS samples. Hence, the estimated effect of health on work is quite similar despite the difference in the samples and in responses.

As in the VRI sample, S*ea*TE varies non-monotonically with age. The oldest age group, which in the HRS sample is made of respondents aged 63-64,[21] has a significantly lower S*ea*TE (i.e., higher in absolute value) than the reference group of the under 60 by about 11 points at 2 years and 7 points at 4 years. Different from the VRI sample, in the HRS sample S*ea*TE varies also by educational attainment. More educated respondents have a lower S*ea*TE (i.e., higher in absolute value) than respondents with a high school diploma or less. At 2 years, the difference is large (-9 points) and statistically significant only among respondents with an advanced degree beyond college. Whereas at 4 years, all education groups (some college, college graduate, other advanced degree) have a significantly lower S*ea*TE (-8 points) than respondents in the lowest education group (high school or less).

[21] This is because our experimental module was fielded among the younger subgroup (under 65) of a random subsample of HRS respondents. The older subgroup (65 and over) was used for a companion module eliciting unconditional and dementia-contingent subjective probabilities of long-term care outcomes, analyzed by Giustinelli, Manski, and Molinari (2022b).

IV. Panel Validation: Relating Expectations to Realizations

We use the panel structure of the VRI to assess whether the conditional probabilities, combined with health outcome, can predict work outcomes. Note that an ideal comparison of state-contingent expectations to realizations would require observing everyone in every state *ex post*, which is not possible due to unobservability of counterfactuals. While this is the very reason that motivated us to elicit and analyze *ex ante* state-contingent expectations, it also makes it impossible to validate fully state-contingent probabilities and, hence, the S*ea*TE. More generally, if one were to use our approach to elicit subjective expectations for an outcome corresponding to a state that never occurs, say, a policy change that is only contemplated, there would be no scope for comparing expectations to realizations. It is nevertheless both interesting and important to assess how well health-contingent expectations predict work given realized health to gauge their usefulness. We find significant predictive power for work outcomes 2 years ahead.

Our expectations data are from Survey 4 (fielded in late 2015). VRI Survey 6 has realizations that roughly match the 2-year horizon. Of the 970 respondents in Survey 4 who completed the 2-year-ahead expectations battery, 584 responded to Survey 6 (fielded in early 2018). There is no evidence of selective non-response to Survey 6 conditional on age, probability of working, and probability of low health (see Online Appendix Table A3).

To consider how the probabilities elicited in Survey 4 predict outcomes observed in Survey 6, we compare predictions based on health-contingent and unconditional work probabilities. First, we assess the predictive power of the health-contingent working probability given knowledge of what health state is realized. As mentioned earlier, this prediction exercise is different from analyzing the S*ea*TE because, by construction, realization data reflect only one health state per individual. Second, we consider the standard approach of predicting actual work using the unconditional probability of working, which averages over the potential realizations of the health state via the health probabilities. Third, to assess whether there is predictive value added in the health-contingent probabilities relative to the unconditional ones, we perform a prediction horse race between the two.

Table 6 shows these results. The dependent variable is realized labor supply (1 if work, 0 if not). The predictor in column (1) is the health-contingent probability of working under the health scenario corresponding to the health state realized *ex post*.[22] The predictor in column (2) is the unconditional

[22] The regressor uses the directly elicited response to the health-contingent working probability question. Recall that there are three health states for which we elicit state-contingent working probabilities (see Figure A1). Using the direct response avoids having to aggregate into the two-way classification (*H*/*L* health) that is convenient for the discrete choice specification used above. Thus, the prediction exercise uses the elicited health-contingent work probabilities directly with the detail used in eliciting them. Table A4 in the Online Appendix shows tabulations of the responses by this full grid.

probability of working.[23] The coefficient of the health-contingent probability is 0.590, precisely estimated. The R^2 is 21.6 percent. The estimates in column (2) are similar, with the R^2 being 1 percentage point lower. There are not that many negative health transitions, so the predictive power of the health-contingent work probability is just marginally better than the predictive power of the unconditional work probability. Taking these specifications as tests of rational expectations leads to a rejection of the null of rationality (slope coefficient equal to one). Nonetheless, the predictive power of both the health-contingent and unconditional work probabilities of VRI respondents are remarkably highly predictive of respondents' realized labor supply.[24]

To evaluate the incremental predictive power of the health-contingent work probabilities over the unconditional work probabilities more directly, column (3) presents a horse race between the two. The health-contingent work probability dominates the prediction, with a coefficient of 0.502 versus 0.093 of the unconditional work probability. We test for incremental predictive power of the health-contingent versus unconditional probabilities using a Wald test. The unconditional probability has no incremental predictive power between columns (1) and (3) (Wald test of 0 and p-value of 1). Conversely, we can overwhelmingly reject the hypothesis that the health-contingent probabilities have no incremental predictive power between columns (2) and (3) (Wald test of 6.70 with a p-value of 0.01).

In sum, the state-contingent probabilities are strongly predictive of outcomes and, in a horse race with the unconditional probabilities, they outperform the latter. Hence, our panel evidence provides strong support about the information content of survey-elicited probabilities, especially the state-contingent ones.

We further investigate whether the predictive power of our survey expectations vary by key initial characteristics, namely age and health. The respondents' age ranged between 57 and 81 at Survey 4 and therefore roughly between 59 and 83 at the time of realization. The nature of retirement decisions varies substantially over this age range. Because the VRI is a short panel and the work probabilities were only elicited from individuals who were working as of Survey 4, there may be different patterns with respect

[23] We use the measure of the unconditional probability that is constructed from the health-contingent probabilities of working and the health probabilities using the law of total probability. We re-estimated the regression presented in this section using the directly elicited measure of unconditional probability. The results are very similar, which is as expected given that the vast majority of responses respect the law of total probability (see Figure 1).

[24] The specifications in (2) corresponds to a classical regression test of rationality of work expectations. The null hypothesis of rational expectations (RE) requires that the regression intercept equals 0 and the slope equals 1 (under no aggregate shocks), or just the latter (under aggregate shocks). Because in our application the outcome variable is binary, the regression test exhausts the main implications of rational expectations, which would not be true if the outcome variable were continuous. See discussions in D'Haultfoeuille et al. (2021) and Crossley et al. (2021), who have developed new tests of rational expectations that take into account the information on higher moments (beyond the first) contained in survey expectations about continuous variables whenever these expectations are elicited as multiple points on the respondent's subjective belief distribution.

to age. Table 7 uses the same specifications as Table 6 while interacting the subjective probabilities used as main predictors with age. (The specifications also include age dummies.)

The overall results are similar to those in Table 6, while revealing interesting heterogeneity by age. For all but the oldest two age groups (70-71 and 72+) the coefficient on the health-contingent work probability increases visibly, ranging between 0.639 and 0.860 across age groups for the health-contingent work probabilities, and between 0.637 and 0.896 for the unconditional work probabilities. For selected age brackets (62, 68-69 and, for the unconditional work probabilities, also 65), the null hypothesis of a unit slope cannot be rejected at conventional significance levels.[25] Hence, looking within age groups strengthens the results. Of course, because of the smaller number of observations available per group, the standard errors tend to increase. Again, the health-contingent work probability does better in the horse race with the unconditional probability.

Table 8 shows results for specifications where the subjective work probabilities are interacted with respondents' initial health (excellent, very good, or good) at Survey 4. (The regression includes age and initial health dummies. We do not interact the predictors with age to avoid a proliferation of coefficients.) Overall, the results are similar to the earlier ones. Subjective work probabilities are highly predictive of actual labor supply for all groups, though respondents in excellent health have a larger coefficient than the other groups (0.740 vs. 0.501 and 0.569 for the health-contingent work probabilities and 0.751 vs. 0.515 and 0.584 for the unconditional work probabilities). The null hypothesis of unity is rejected for all groups.

The VRI panel also allows for evaluation of the 4-year ahead forecasts. The results paralleling those for 2-year ahead are reported in the online appendix Tables A6, A7, and A8, analogues to Tables 6, 7, and 8. Compared to the 2-year ahead results, there are similarities as well as differences. The coefficients of the predictors (columns 1 and 2 of the tables) are quite similar across horizons, though not surprisingly, the R^2 is lower at the longer horizon. In the horse race regression, neither measure drives out the other, the point estimate of the coefficient on the unconditional probability is substantially higher than that of the health-contingent probability. While acknowledging that the health-contingent results are not as good at 4 years as 2 years, the 4-year ahead forecasts are problematic for various reasons: the horizon of the 4-

[25] D'Haultfoeuille et al. (2021) and Crossley et al. (2021), too, uncover interesting dimensions of heterogeneity in rationality of expectations. Using data on earnings expectations collected in the Labor Market Module of the New York Fed's Survey of Consumer Expectations (SCE), D'Haultfoeuille et al. (2021) find that the rational expectations hypothesis is more realistic for college-educated respondents and less realistic for less educated ones. Using data on earnings expectations collected in the Berea Panel Study (BPS), Crossley et al. (2021) find that earnings expectations become more accurate as students progress through college, and especially after they leave college. The analysis almost always rejects the null of rational expectations based on the data collected in the in-school period but not in the post-college period.

year ahead survey is less-well aligned and, more importantly, the COVID-19 crisis occurred during the interval between the elicitation of the probabilities and the realizations. Hence, we report the results in the online appendix with suitable cautions about their interpretation.[26]

V. Unobserved Heterogeneity in Taste for Work and Its Consequences for Inference

We now investigate how the presence of unobserved heterogeneity in taste for work that may be correlated with individuals' health would affect regression estimates of the causal effect of health on labor supply in realizations data. Consider the linear regression of work on health,

$$d_i = b_0 + b_1 h_i + e_i, \tag{13}$$

where $d_i \in \{0,1\}$, $h_i \in \{0,1\}$, and the time subscript and conditioning on exogenous covariates are omitted for simplicity. As above, d=1 indicates working and h=1 indicates bad health. As discussed in Subsection II.A, the least squares estimate of b_1 will be an unbiased estimate of ATE of health on work as long as there is no selection or heterogeneity in equation (4). This is unlikely to hold in observational data. For example, individuals may differ in their health-contingent taste for work and the latter may be correlated with individuals' health.[27] Our elicitation approach is designed to render this unobserved heterogeneity observable *ex ante.*

Building on the DP framework of Subsection II.B, the actual decision to work can be expressed as

$$d_i = (1-h_i)\mathbf{I}[\tilde{v}_i^H + \tilde{\varepsilon}_i] + h_i\mathbf{I}[\tilde{v}_i^L + \tilde{\varepsilon}_i], \tag{14}$$

where $\mathbf{I}[.]$ is the indicator function, equal to 1 if the argument is positive and 0 otherwise; $\tilde{v}_i^h = v_i(h,1) - v_i(h,0)$ is the difference in the health-contingent value of continuing to work ($d_i = 1$)

[26] Of the 839 respondents in Survey 4 who completed the 4-year ahead battery, 397 responded to Survey 7 (fielded beginning July 2020). Again, there is no evidence of selective attrition conditional on observables (see Table A3). There are multiple complications with the 4-year battery, which make findings based on it more tenuous. Nonetheless, we include them for completeness and because they are an unusual opportunity to compare realizations to state-contingent expectations over multiple periods. First, recall that the 4-year battery was not asked of respondents who reported zero probability of working 2 years ahead, so we do not have expectations for these respondents relevant for Survey 7 outcomes. Second, roughly 4-1/2 years intervened between Survey 4 and Survey 7, which could lead to underprediction of not working because of the passage of time if we used 4-and-1/2-year-ahead outcomes. Third, there was an unexpected aggregate shock, COVID-19, which led to more-than-expected labor market exit. Because of the pandemic, Survey 7 asked about employment retrospectively as of January 2020. We use those retrospective answers for work outcomes for the 4-year ahead exercise because they better align with the horizon and abstract from the pandemic shock. Doing so, however, may introduce recall or justification bias in the outcome variable. Moreover, the health measurement is aligned with the July 2020, post-COVID work measurement. A number of respondents said their work was affected by COVID-19, but mainly not for reasons related to own health.

[27] This argument will generally apply even after conditioning on exogenous covariates, which one may include to address selection on observables. Furthermore, while availability of panel data on realized health and work would enable one to address endogeneity caused by time-invariant unobserved heterogeneity via first differencing or fixed effects, the approach would require observability of a sufficient number of health transitions, it would limit the econometrician's conditioning ability to time-varying covariates, and it would not address endogeneity from time-varying unobserved heterogeneity.

versus not ($d_i = 0$) in health state h as of the time of elicitation, with $h \in \{H, L\} \equiv \{0,1\}$; and $\tilde{\varepsilon}_i = \varepsilon_i(1) - \varepsilon_i(0)$. We derive our measures of health-contingent valuations of continued work, $\tilde{v}_i^H$ and $\tilde{v}_i^L$, by inversion of the elicited health-contingent work probabilities, P_i^H and P_i^L, of equation (10). See the Appendix [to main text (see below), not the online appendix] for a full derivation of the simulated environment based on the framework in Section II. The Appendix also presents the parametric specification for implementing the simulation.[28]

Before proceeding with the simulation, in Table 9 we show main features of the empirical distributions of $\tilde{v}_i^H$ and $\tilde{v}_i^L$ across respondents, obtained from the 2-year and 4-year ahead health-contingent subjective probabilities. The patterns are qualitatively equivalent to those observed for the S*ea*TE in Section III, as the DP-implied values are nonlinear transformations of the elicited health-contingent work probabilities. As expected, the mean value in high health (0.97 at 2 years and 0.48 at 4 years) is substantially greater than that in low health (-0.35 at 2 years and -0.72 at 4 years), reflecting the lower value of working in low health. Notwithstanding, there is a high correlation across health states in respondents' values of work: 0.73 at 2 years and 0.74 at 4 years.

To simulate work outcomes according to equation (14), and reflecting unobserved heterogeneity in taste for work, we use the health-contingent values of work ($\tilde{v}_i^H$ or $\tilde{v}_i^L$) we derived from the elicited health-contingent work probabilities along with draws for health (h_i) and the residual component ($\tilde{\varepsilon}_i$). Specifically, we simulate health using Bernoulli draws from the subjective health transition probability we elicited in our survey, π_i^h, and the residual component from a standard normal.[29]

We consider three cases for correlation of the health transition probability with the value of work:

1. π_i^h is fixed at the sample mean, so health transitions are *uncorrelated* with the value of work.
2. π_i^h is the elicited probability, so health transitions have the *empirical correlation* with the value of work.

[28] We implement the simplest setting without unspecified state variables, y, as it encompasses many relevant cases. The value function may shift for multiple reasons given health. Preferences for work versus leisure may be a function of health; wages may be a function of health; medical costs may be a function of health, which for the simulation are collapsed into a single dimension via $\tilde{\varepsilon}_i$.

[29] We could have simulated labor supply decisions directly via Bernoulli draws from the elicited health-contingent work probability corresponding to the simulated health draw, thus avoiding distributional assumptions on the residual component. However, we think it useful to implement the simulation in a manner that is coherent with the DP framework of Subsection II.B and therefore explicitly acknowledges the presence of the residual component. Our procedure also allows us easily to consider the alterative cases shutting down and magnifying the correlation of taste for work with health.

3. π_i^h adjusts the elicited probability to induce a *higher correlation* between health and the value of work than is present in the VRI data.[30]

The first case implies no selection or heterogeneity, so OLS estimation of (14) will yield an unbiased estimate of the average treatment effect equal to the average S*ea*TE. The second case illustrates the extent of the bias that would be present in the VRI data. The third case magnifies the bias.

Table 10 shows estimates of the regression for each of the three cases (uncorrelated, empirical correlation, and higher correlation) for the 2-year horizon (columns 1-3) and for the 4-year horizon (columns 4-6), with realizations simulated over 1000 replications.

In the *uncorrelated* cases, the estimated coefficient of health is unbiased and therefore equals the average S*ea*TE in Table 1.

The *empirical* cases yield a biased estimate because of the positive correlated heterogeneity in value of work and health transitions in the VRI. There is a slight positive correlation between the value of work and the probability of remaining in high health. The sign of this correlation is not surprising because individuals in situations with attractive jobs (e.g., high socioeconomic status) are also likely to have better health. The estimated coefficient of health is larger in absolute value than the causal effect because those who get bad health shocks disproportionately have lower value of work. The VRI respondents do not have much heterogeneity in health (most are quite healthy), so the magnitude of the bias is modest. Even so, the bias is nontrivial, overstating by 10% the health-related job transitions relative to the causal effect.

In other samples with more heterogeneity in health, this bias would be even more important as illustrated by the *higher correlation* case.

VI. Conclusion

In this paper, we study the effect of health on work among healthy older workers using data on individuals' subjective probabilities of working to specified future horizons under alternative health states. We call this expectations-based effect S*ea*TE, for Subjective *ex ante* Treatment Effect. We document that the S*ea*TE of a negative health transition on labor supply is on average negative but highly heterogeneous in magnitude—and even in sign—across respondents in two U.S. national studies, the Vanguard Research Initiative (VRI) and the Health and Retirement Study (HRS). Consistent with the empirical evidence from realizations data on health and work which point to an overall negative average effect of bad health on

[30] π_i^h is adjusted by subtracting 0.1 from individuals in the bottom quintile of $\tilde{v}_i^H$ and 0.05 from those in the second quintile of $\tilde{v}_i^H$ and by adding 0.075 to the top two quintiles of $\tilde{v}_i^H$. (The latter quintiles are combined as they have a common $\tilde{v}_i^H$ corresponding to individuals who gave a 100% change of working when in high health.)

labor supply, the majority of VRI respondents (close to 70%) report a lower probability of working in low than in high health (negative S*ea*TE). Within this majority the magnitude of the effect varies widely across respondents. Moreover, a large fraction of respondents (close to 30%) has health-contingent probabilities of working implying a zero effect of health on labor supply, some because they would always work and some because they would never work regardless of health. A few individuals have a positive effect of bad health on work. We document similar level and heterogeneity in S*ea*TE among HRS respondents.

We further show that the documented heterogeneity in S*ea*TE correlates sensibly with selected covariates, including respondent age (especially when age plus the elicitation horizon corresponds to a customary retirement age), whether the respondent is observed to unretire between survey waves, and current salary or wealth. Notwithstanding, a substantial amount of heterogeneity in S*ea*TE cannot be explained by observable respondent characteristics, suggesting that most of its cross-sectional variation reflects persistent unobserved heterogeneity in differential work attachment across health states.

Using a discrete choice DP framework, we show that the health-contingent working probabilities embed information about unobserved heterogeneity in taste for work across workers and we investigate empirically the potentially distorting effects on inference implied by a failure to account for such unobserved heterogeneity when it is correlated with health. In our data, there is a negative correlation between the health-contingent values of working underlying the health-contingent working probabilities and the probability of a negative health transition. We show that this correlation induces a nontrivial bias in regression estimates of the overall effect of health on work.

We validate our measures and, hence, our proposed approach. We show that the health-contingent and unconditional work expectations are both highly internally consistent in that they respect the law of total probability. Moreover, they are strongly predictive of work realizations after 2 years. At the same time, it is important to note that a full validation of S*ea*TE is made impossible—at least at the individual level—by the very problem S*ea*TE aims to overcome, that is, the unobservability of counterfactual outcomes. A full validation of S*ea*TE may be possible for its mean, or another distributional feature, in a setting where the corresponding feature of the distribution of counterfactual outcomes can be reliably estimated, for instance, within the context of a randomized control trial.

Looking beyond this paper, because our methodology gives estimates of potential outcomes, it could be used in a wide range of applications. More generally, and very much in the spirit of stated preference methods, the approach may be useful when treatments are difficult to manipulate experimentally or control-for econometrically, especially the particularly interesting case in which one would like to study behavioral responses to policies that have not yet been and might never be implemented.

Appendix

(i) Deriving the health-contingent values of work from the health-contingent work probabilities

Building on the DP framework of Section II.B, we now derive measures of health-contingent valuations of working versus not, $\tilde{v}_i^H$ and $\tilde{v}_i^L$, from the elicited health-contingent work probabilities, P_i^H and P_i^L.

Following Arcidiacono and Ellickson (2011), we define the *conditional value function* as the present discounted value (net of ε_{it}) of choosing d_{it} and behaving optimally from *t+1* onward,

$$v_{it}\left(h_{it}, y_{it}, d_{it}\right) = u_{it}\left(h_{it}, y_{it}, d_{it}\right) + \beta \sum_{x_{i,t+1}} \bar{V}_{i,t+1}^{*}\left(h_{i,t+1}\right) \pi_{it}^{h}\left(h_{i,t+1} \mid h_{it}, y_{it}, d_{it}\right).$$

We further define the *ex ante* (or *integrated*) *value function* as the continuation value of being in health state h_{it} after integrating over the unspecified state, y_{it}, and the residual component, ε_{it},

$$\begin{aligned}
\bar{V}_{it}^{*}\left(h_{it}\right) &= \int_{\varepsilon_{it}} \int_{y_{it}} V_{it}^{*}\left[\left(h_{it}, y_{it}, \varepsilon_{it}\right), \delta_{it}^{*}\left(h_{it}, y_{it}, \varepsilon_{it}\right)\right] \cdot \pi_{i,t-1}^{y}\left(y_{it} \mid h_{it}, y_{i,t-1}, d_{i,t-1}\right) dy_{it} \cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it} \\
&= \int_{\varepsilon_{it}} \int_{y_{it}} \left[\left[u_{it}\left(h_{it}, y_{it}, d_{it}\right) + \varepsilon_{it}\left(d_{it}\right)\right] + \beta \sum_{x_{i,t+1}} \bar{V}_{i,t+1}^{*}\left(h_{i,t+1}\right) \pi_{it}^{h}\left(h_{i,t+1} \mid h_{it}, y_{it}, d_{it}\right) \right] \\
&\quad \cdot \pi_{i,t-1}^{y}\left(y_{it} \mid h_{it}, y_{i,t-1}, d_{i,t-1}\right) dy_{it} \cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it},
\end{aligned}$$

where ε_{it} is assumed to enter the utility additively and to be i.i.d. across individuals and time, thus yielding the standard single-crossing result for a discrete choice problem.[31]

Equation (10) can then be re-written as

$$\begin{aligned}
& P_{i,t-1}\left[\delta_{it}^{*}\left(h_{it}, y_{it}, \varepsilon_{it}\right) = 1 \mid h_{it}\right] \\
&= \int_{\varepsilon_{it}} \int_{y_{it}} \delta_{it}^{*}\left(h_{it}, y_{it}, \varepsilon_{it}\right) \cdot \pi_{i,t-1}^{y}\left(y_{it} \mid h_{it}, h_{i,t-1}, y_{i,t-1}, d_{i,t-1}\right) dy_{it} \cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it} \qquad (15) \\
&= \int_{\varepsilon_{it}} \int_{y_{it}} \underset{d_{it} \in \{1,0\}}{\arg\max} \left[v_{it}\left(h_{it}, y_{it}, d_{it}\right) + \varepsilon_{it}\left(d_{it}\right)\right] \cdot \pi_{i,t-1}^{y}\left(y_{it} \mid h_{it}, h_{i,t-1}, y_{i,t-1}, d_{i,t-1}\right) dy_{it} \cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it},
\end{aligned}$$

which directly links the elicited health-contingent probabilities to the primitives of the DP framework.

We begin with the case without unspecified state variables, *y*, as it is simpler while encompassing many relevant cases. The value for individual *i* of being in state $h_{it} \in \{0,1\}$ and making choice $d_{it} \in \{0,1\}$ at time *t* given expectation and optimization from *t*+1 onward is

[31] In standard DP settings for realizations data, the econometrician is doing the integration with respect to the distribution of ε, while in ours it is the respondent, who must additionally carry out the integration with respect to the distribution of *y*. Note that the i.i.d. assumptions are standard in the DP literature, and would likely be invoked by those using these subjective probabilities in structural modeling. That said, they play no role in our analysis that features a single-period realization. Any common shock would be absorbed in the constant in the simulated regression.

$$
\begin{aligned}
V_{it}\left(h_{it},d_{it}\right) = &(1-h_{it})\left\{d_{it}\left(v_{it}\left(0,1\right)+\varepsilon_{it}\left(1\right)\right)+(1-d_{it})\left(v_{it}\left(0,0\right)+\varepsilon_{it}\left(0\right)\right)\right\} \\
&+h_{it}\left\{d_{it}\left(v_{it}\left(1,1\right)+\varepsilon_{it}\left(1\right)\right)+(1-d_{it})\left(v_{it}\left(1,0\right)+\varepsilon_{it}\left(0\right)\right)\right\},
\end{aligned}
\tag{16}
$$

where the first row refers to actions in high health ($h_{it} = 0$) and the second row in low health ($h_{it} = 1$). Maximization of $V_{it}\left(h_{it},d_{it}\right)$ yields the standard single crossing conditions, as follows:

$$
\begin{aligned}
&\text{when } h_{it} = 0, \\
&\qquad \delta_{it}^{*} = 1 \quad \text{if } 0 \le \tilde{v}_{it}^{H} + \tilde{\varepsilon}_{it} \\
&\qquad\quad\; = 0 \quad \text{otherwise;} \\
&\text{when } h_{it} = 1, \\
&\qquad \delta_{it}^{*} = 1 \quad \text{if } 0 \le \tilde{v}_{it}^{L} + \tilde{\varepsilon}_{it} \\
&\qquad\quad\; = 0 \quad \text{otherwise,}
\end{aligned}
\tag{17}
$$

where

$$
\begin{aligned}
\tilde{v}_{it}^{H} &= v_{it}\left(0,1\right) - v_{it}\left(0,0\right) \\
\tilde{v}_{it}^{L} &= v_{it}\left(1,1\right) - v_{it}\left(1,0\right) \\
\tilde{\varepsilon}_{it} &= \varepsilon_{it}\left(1\right) - \varepsilon_{it}\left(0\right).
\end{aligned}
\tag{18}
$$

In the simulation exercises, we assume that $\tilde{\varepsilon}_{it}$ is distributed as a standard normal with cdf Φ implies,

$$
\begin{aligned}
P_{i,t-1}^{H} &= \Phi\left(\tilde{v}_{it}^{H}\right) \\
P_{i,t-1}^{L} &= \Phi\left(\tilde{v}_{it}^{L}\right),
\end{aligned}
\tag{19}
$$

which, by inversion, yield the values of working versus not working in high and low health,

$$
\begin{aligned}
\tilde{v}_{it}^{H} &= \Phi^{-1}(P_{i,t-1}^{H}) \\
\tilde{v}_{it}^{L} &= \Phi^{-1}(P_{i,t-1}^{L}).
\end{aligned}
\tag{20}
$$

To perform the inversion in the simulation exercise, we recode responses of probability zero as 0.005 and of probability one as 0.995, following an established practice in the literature (e.g., Blass, Lach, and Manski (2010), Wiswall and Zafar (2015), and Arcidiacono, Hotz, Maurel, and Romano (2020)).Table 9 shows the empirical distribution of $\tilde{v}^{H}$ and the $\tilde{v}^{L}$.

(ii) DP Interpretation with Correlated and Stochastic Unspecified States

We now consider the interpretation of the health-contingent work probabilities using the DP framework when there is an unspecified state y that is correlated with health. Extending the case in Section II.B, suppose that the unspecified state is also binary. To model correlation with health, assume it can take on two values (y^{+H}, y^{-H}), if health is high, and potentially two different values (y^{+L}, y^{-L}), if health is low. Let the probability of y given health to be

$$\begin{aligned} P\left(Y^{+H} \mid H\right) &= \pi^{+H} \\ P\left(Y^{-H} \mid H\right) &= 1-\pi^{+H} \\ P\left(Y^{+L} \mid L\right) &= \pi^{+L} \\ P\left(Y^{-L} \mid L\right) &= 1-\pi^{+L} \end{aligned} . \tag{21}$$

Then equation (19) becomes

$$\begin{aligned} P_{i,t-1}^{H} &= \Phi\left(\tilde{v}_{it}^{+H}\right)\pi^{+H} + \Phi\left(\tilde{v}_{it}^{-H}\right)\left(1-\pi^{+H}\right) \\ P_{i,t-1}^{L} &= \Phi\left(\tilde{v}_{it}^{+L}\right)\pi^{+L} + \Phi\left(\tilde{v}_{it}^{-L}\right)\left(1-\pi^{+L}\right) \end{aligned} , \tag{22}$$

where

$$\begin{aligned} \tilde{v}_{it}^{+H} &= v_{it}\left(H, Y^{+H}, W\right) - v_{it}\left(H, Y^{+H}, \sim W\right) \\ \tilde{v}_{it}^{-H} &= v_{it}\left(H, Y^{-H}, W\right) - v_{it}\left(H, Y^{-H}, \sim W\right) \\ \tilde{v}_{it}^{+L} &= v_{it}\left(H, Y^{+L}, W\right) - v_{it}\left(H, Y^{+L}, \sim W\right) \\ \tilde{v}_{it}^{-L} &= v_{it}\left(H, Y^{-L}, W\right) - v_{it}\left(H, Y^{-L}, \sim W\right) \end{aligned} , \tag{23}$$

that is, the differenced conditional value functions under the four possible combinations of health and the unspecified state. Hence, the health-contingent probability of working, $P_{i,t-1}^{h}$, is the weighted average of the probability of working given health and the unspecified state $\left(\Phi\left(\tilde{v}_{it}^{+h}\right), \Phi\left(\tilde{v}_{it}^{-h}\right)\right)$, with weights equal to the probabilities of the unspecified state given health $\left(\pi^{+h}, \pi^{-h}\right)$.

Note that the presence of y does not necessarily cause the complication given above. For example, consider the leading case for studying health and retirement that has the wage a function of health. If wage is the only state affecting retirement that is a function of health, then the model in the main text applies. In terms of the notation of the appendix, the probabilities of the unspecified state give health, $\left(\pi^{+h}, \pi^{-h}\right)$, are degenerate corners, so the expression (22) collapses to (19).

The complication in interpretation discussed here would, however, arise if health shifts the utility function independently of wage (e.g., taste heterogeneity) and the probabilities $\left(\pi^{+h}, \pi^{-h}\right)$ are not corners. The conditional probability approach can still be used in this setting, but one would need to elicit the conditional probabilities of working fixing all combinations of health and wage.

References

Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D. Shapiro, and Christopher Tonetti. 2020. "Long-Term Care Utility and Late-in-Life Saving." *Journal of Political Economy* 128: 2375-2451.

Ameriks, John, Andrew Caplin, Minjoon Lee, Matthew D. Shapiro, and Christopher Tonetti. 2015. The Wealth of Wealthholders. NBER Working Paper 20972.

Ameriks, John, Joseph Briggs, Andrew Caplin, Minjoon Lee, Matthew D. Shapiro, and Christopher Tonetti. 2020. "Older Americans Would Work Longer If Jobs Were Flexible." *American Economic Journal: Macroeconomics* 12(1): 174-209.

An, Yonghong, Yingyao Hu, and Ruli Xiao. 2021. "Dynamic Decisions under Subjective Expectations: A structural analysis." *Journal of Econometrics* 222(1), Part C: 645-675.

Arcidiacono, Peter and Paul B. Ellickson. 2011. "Practical Methods for Estimation of Dynamic Discrete Choice Models." *Annual Reviews of Economics* 3: 363-394.

Arcidiacono, Peter, Joseph V. Hotz, Arnaud Maurel, and Teresa Romano. 2020. "*Ex Ante* Returns and Occupational Choice." *Journal of Political Economy* 128(12): 4475-4522.

Behncke, Stefanie. 2012. "Does Retirement Trigger Ill Health?" *Health Economics* 21(3): 282-300.

Bellman, Richard E. 1957. *Dynamic Programming*. Princeton University Press.

Benítez-Silva, Hugo and Debra S. Dwyer. 2005. "The Rationality of Retirement Expectations and the Role of New Information." *Review of Economics and Statistics* 87(3): 587-592.

Berger, Benjamin, Italo Lopez-Garcia, Nicole Maestas, and Kathleen Mullen. 2021. "The Link between Health and Working Longer: Differences in Work Capacity." In Can't Work, Can't Retire: America's Aging Workforce. Eds. Lisa Berkman and Beth C. Truesdale.

Blau, David M. and Donna B. Gilleskie. 2001. "The Effect of Health on Employment Transitions of Older Men." *In* Worker Wellbeing in a Changing Labor Market, vol. 20, 35-65, Elsevier.

Blau, David M. and Donna B. Gilleskie. 2008. "The Role of Retiree Health Insurance in the Employment Behavior of Older Men." *International Economic Review* 49(2): 475-514.

Blass, Asher A., Saul Lach, and Charles F. Manski. 2010. "Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability." *International Economic Review* 51(2): 421-440.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French. 2021. "The Impact of Health on Labor Supply Near Retirement" Forthcoming in the *Journal of Human Resources*.

Bound, John. 1991. "Self-Reported Versus Objective Measures of Health in Retirement." *Journal of Human Resources* 26(1): 106-138.

Bound, John, Michael Schoenbaum, Todd Stinebrickner, and Timothy Waidmann. 1999. "The Dynamic Effects of Health on the Labor Force Transitions of Older Workers." *Labour Economics* 6(2): 179-202.

Bound, John, Todd Stinebrickner, and Timothy Waidmann. 2010. "Health, Economic Resources and the Work Decisions of Older Men." *Journal of Econometrics* 156: 106-129.

Bruine de Bruin, Wandi and Katherine G. Carman. 2018. "Measuring Subjective Probabilities: The effect of response mode on the use of focal responses, validity, and respondents' evaluations." *Risk Analysis* 31(10): 2128-2143.

Bruine de Bruin, Wandi, Alycia Chin, Jeff Dominitz, and Wilbert van der Klaauw. 2023. Household Surveys and Probabilistic Questions, in R. Bachmann, G. Topa, and W. van der Klaauw eds., 'Handbook of Economic Expectations', Elsevier, chapter 1, pp. 3-31.

Cai, Lixin. 2010. "The Relationship Between Health and Labour Force Participation: Evidence from a Panel Data Simultaneous Equation Model." *Labour Economics* 17: 77–90.

Coe, Norma B. and Gema Zamarro. 2011. "Retirement Effects on Health in Europe." *Journal of Health Economics* 30(1): 77-86.

Coile, Courteny C. 2015. "Economic Determinants of Workers' Retirement Decisions." *Journal of Economic Surveys* 29(4): 830-853.

Coile, Courteny C. 2018. Working Longer in the United States: Trends and Explanations in "Social Security Programs and Retirement around the World: Working Longer", pp. 299-324, Coile, C.C. and K. Milligan and D.A. Wise eds., University of Chicago Press.

Coile, Courteny C. and Kevin Milligan and David A. Wise. 2018. Introduction to "Social Security Programs and Retirement around the World: Working Longer" in "Social Security Programs and Retirement around the World: Working Longer", pp. 1-32, Coile, C.C. and K. Milligan and D.A. Wise eds., University of Chicago Press.

Crossley, Thomas and Yifan Gong and Todd Stinebrickner and Ralph Stinebrickner. 2021. Examining Income Expectations in the College and Early Post-college Periods: New Distributional Tests of Rational Expectations. Working Paper 01, Institute for Fiscal Studies.

Currie, Janet and Brigitte C. Madrian. 1999. Health, Health Insurance and the Labor Market in the Handbook of Labor Economics, Vol. 3, Part C, 3309-3416.

Datta Gupta, Nabanita and Mona Larsen. 2010. "The Impact of Health on Individual Retirement Plans: Self-Reported Versus Diagnostic Measures." *Health Economics* 19: 792-813.

Datta Gupta, Nabanita, Kristin J. Kleinjans, and Mona Larsen. 2015. "The Effect of a Severe Health Shock on Work Behavior: Evidence from Different Health Care Regimes." *Social Science & Medicine* 136-137: 44-51.

De Nardi, Mariacristina, Eric French, and John B. Jones. 2016. "Savings After Retirement: A survey." *Annual Review of Economics* 8(1): 177-204.

D'Haultfoeuille, Xavier and Christophe Gaillac and Arnaud Maurel. 2021. "Rationalizing Rational Expectations: Characterizations and Tests." *Quantitative Economics* 12(3): 817-842.

Disney, Richard, Carl Emmerson, and Matthew Wakefield. 2006. "Ill Health and Retirement in Britain: A Panel Data-Based Analysis." *Journal of Health Economics* 25: 621-649.

Dominitz, Jeff. 1997. The Role of Conditioning Information in Reports of Subjective Phenomena. Working Paper.

Dominitz, Jeff and Charles F. Manski. 1996. "Student Expectations of the Returns to Schooling." *Journal of Human Resources* 31(1): 1-26.

Dwyer, Debra and Olivia Mitchell. 1999. "Health Problems as Determinants of Retirement: Are Self-Rated Measures Endogenous?" *Journal of Health Economics* 18(2): 173-193.

Esponda, Ignacio and Emanuel Vespa. 2019. Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory. Working Paper.

Fischhoff, Baruch and Wandi Bruine de Bruin. 1999. "Fifty-Fifty = 50%?" *Journal of Behavioral Decision Making* 12: 149-163.

Fischhoff, Baruch, Ned Welch, and Shane Frederick. 1999. "Construal Processes in Preference Assessment." *Journal of Risk and Uncertainty* 19(1-3): 139-164.

Fisher, Gwenith G., Dorey S. Chaffee, and Amanda Sonnega. 2016. "Retirement Timing: A Review and Recommendations for Future Research." *Work, Aging and Retirement* 2(2): 230-261.

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

French, Eric and John B. Jones. 2011. "The Effects of Health Insurance and Self-Insurance on Retirement Behavior." *Econometrica* 79(3): 693-732.

French, Eric and John B. Jones. 2017. "Health, Health Insurance and Retirement: A Survey." *Annual Reviews of Economics* 9: 383-409.

Garcia-Gomez, Pilar. 2011. "Institutions, Health Shocks and Labour Market Outcomes Across Europe." *Journal of Health Economics* 30: 200-213.

Garcia-Gomez, Pilar, Titus J. Galama, Eddy van Doorslaer, and Angel Lopez-Nicolas. 2017. Interactions Between Financial Incentives and Health in the Early Retirement Decision. Tinbergen Institute Discussion Paper 44.

Giustinelli, Pamela, Charles F. Manski, and Francesca Molinari. 2022a. "Tail and Center Rounding of Probabilistic Expectations in the Health and Retirement Study." *Journal of the Econometrics* 231(1): 265-281.

Giustinelli, Pamela, Charles F. Manski, and Francesca Molinari. 2022b. "Precise or Imprecise Probabilities? Evidence from survey response related to late-onset dementia." *Journal of the European Economic Association* 20(1): 187-221.

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

Gustman, Alan L., Olivia S. Mitchell, and Thomas L. Steinmeier. 1995. "Retirement Measures in the Health and Retirement Study." *Journal of Human Resources* 30: S57-S83.

Gustman, Alan L., Thomas L. Steinmeier, and Nahid Tabatabai. 2010. *Pensions in the Health and Retirement Study*. Harvard University Press.

Hotz, Joseph V. and Robert A. Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies* 60(3): 497-529.

Hudomiet, Peter and Robert J. Willis. 2013. "Estimating Second Order Probability Beliefs from Subjective Survival Data." *Decision Analysis* 10(2): 152-170.

Hudomiet, Peter, Michael D. Hurd, and Susann Rohwedder. 2021. The Causal Effects of Wages, Wealth, and Health on Retirement: Estimates Based on Subjective Conditional Probabilities. To be presented in the special session on "Using Subjective Expectations of Individuals in Studies of Decision Making" of the ASSA 2021 virtual conference, January 4, 2021.

Hudomiet, Peter, Michael D. Hurd, and Susann Rohwedder. 2023. Mortality and Health Expectations, in R. Bachmann, G. Topa, and W. van der Klaauw eds., 'Handbook of Economic Expectations', Elsevier, chapter 8, pp. 225-259.

Hurd, Michael D. 2009. "Subjective Probabilities in Household Surveys." *Annual Review of Economics* 1: 543-562.

Jones, Andrew M., Nigel Rice, and Francesca Zantomio. 2016. "Acute Health Shocks and Labour Market Outcomes." University of York, HEDG Working Paper 04.

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

Kapteyn, Arie and Erik Meijer. 2014. "A Comparison of Different Measures of Health and their Relation to Labor Force Transitions at Older Ages." In *NBER Book Discoveries in the Economics of Aging*, Chicago University Press, chapter 3, pp. 115-150.

Kapteyn, Arie, Arthur van Soest, and Julie Zissimopoulos. 2007. "Using Stated Preferences Data to Analyze Preferences for Full and Partial Retirement." IZA Discussion Paper 2785.

Kezdi, Gabor and Matthew D. Shapiro. 2023. Retirement Expectations, in R. Bachmann, G. Topa, and W. van der Klaauw eds., 'Handbook of Economic Expectations', Elsevier, chapter 10, pp. 293-320.

Lindeboom, Maarten and Marcel Kerkhofs. 2009. "Health and Work of the Elderly Subjective Health Measures, Reporting Errors and the Endogenous Relationship between Health and Work." *Journal of Applied Econometrics* 24(6): 1024-1046.

Lumsdaine, Robin and Olivia S. Mitchell. 1999. "New Developments in the Economic Analysis of Retirement." In *Handbook of Labor Economics* Volume 3, Chapter 50, pp. 3261-3307, Elsevier B.V.

Maestas, Nicole. 2010. "Back to Work: Expectations and Realizations of Work After Retirement." *Journal of Human Resources* 45(3): 718-747.

Manski, Charles F. 1990. "The Use of Intentions Data to Predict Behavior: A Best-Case Analysis." *Journal of the American Statistical Association* 85(412): 934-940.

Manski, Charles F. 1999. "Analysis of Choice Expectations in Incomplete Scenarios." *Journal of Risk and Uncertainty* 19(1-3): 49-66.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72: 1329-1376.

Martinez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa. 2019. "Failures in Contingent Reasoning: The Role of Uncertainty." *American Economic Review* 109(10): 3437-3474.

Maurer, Jurgen, Roger Klein, and Francis Vella. 2011. "Subjective Health Assessments and Active Labor Market Participation of Older Men: Evidence from a Semiparametric Binary Choice Model With Nonadditive Correlated Individual-Specific Effects." *Review of Economics and Statistics* 93(3): 764–774.

McGarry, Kathleen. 2004. "Health and Retirement: Do Changes in Health Affect Retirement Expectations?" *Journal of Human Resources* 39(3): 624-648.

McGeary, Kerry A. 2009. "How Do Health Shocks Influence Retirement Decisions?" *Review of Economics of the Household* 7: 307-321.

O'Donnell, Owen, Eddy van Doorslaer, and Tom van Ourti. 2015. Health and inequality. In *Handbook of Income*, Volume 2B, eds. Atkinson, A.B. and F.J. Bourguignon, Chapter 18, pp. 1419-1533, Elsevier B.V.

Pantano, Juan and Yu Zheng. 2013. Using Subjective Expectations Data to Allow for Unobserved Heterogeneity in Hotz-Miller Estimation Strategies. Working Paper.

Rohwedder, Susan and Robert J. Willis. 2010. "Mental Retirement." *Journal of Economic Perspectives* 24(1): 119-138.

Rust, John. 1992. Do People Behave According to Bellman's Principle of Optimality? Working Paper E-92-10, Hoover Institution, Stanford University.

Rust, John and Christopher Phelan. 1997. "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets." *Econometrica* 65(4): 781-831.

Shafir, Eldar. 1994. "Uncertainty and the Difficulty of Thinking Through Disjunctions." *Cognition* 50: 403-430.

Stephens, Melvin Jr. and Desmond Toohey. 2022. "The Impact of Health on Labor Market Outcomes: Evidence from a Large-Scale Health Experiment" *American Economic Journal: Applied Economics* 14(3): 367-399.

Stinebrickner, Todd and Ralph Stinebrickner. 2014, "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout" *Review of Economic Studies* 81(1): 426-472.

van der Klaauw, Wilbert. 2012. "On the Use of Expectations Data in Estimating Structural Dynamic Models." *Journal of Labor Economics* 30(3): 521-554.

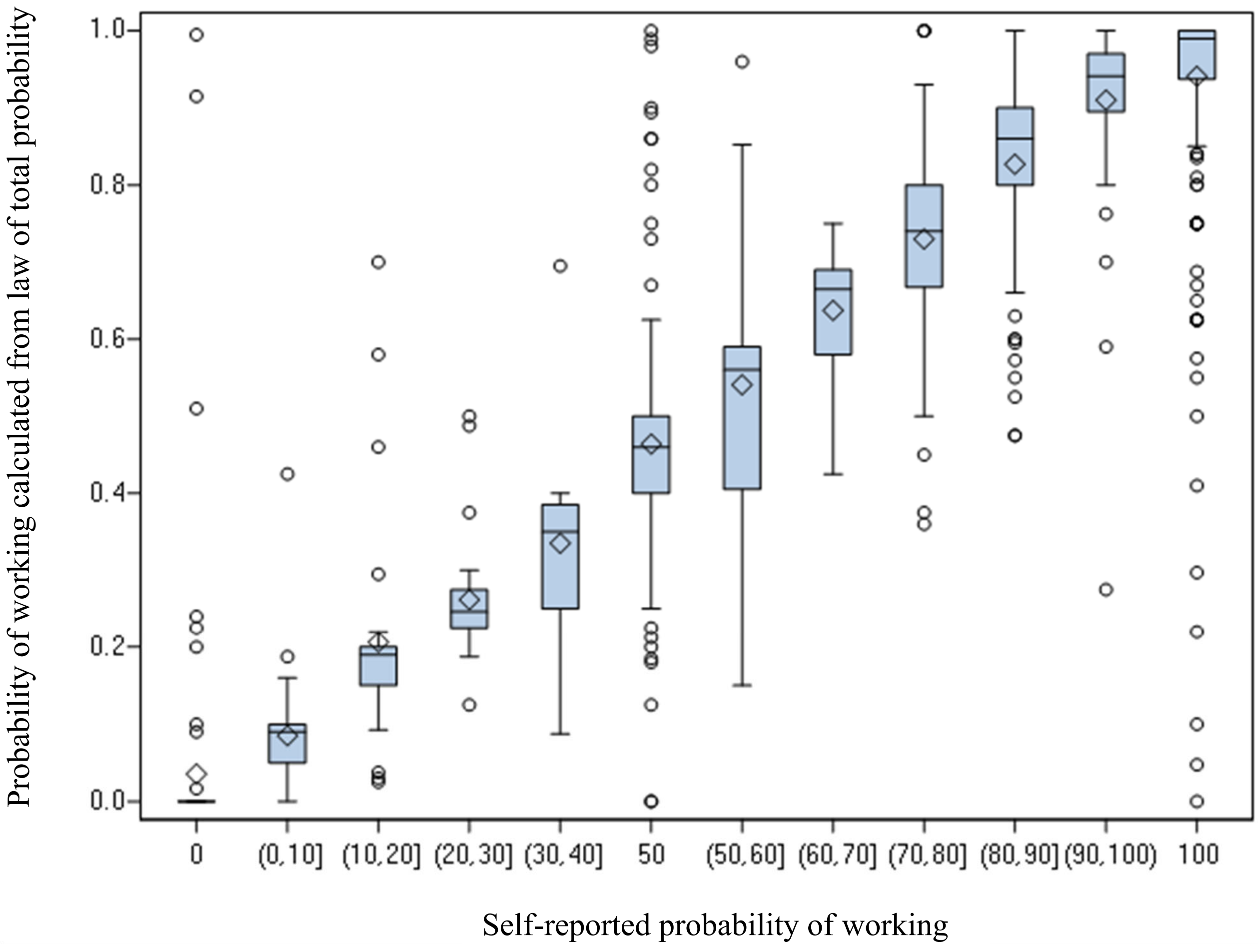
van der Klaauw, Wilbert and Kenneth I. Wolpin. 2008. "Social Security and the Retirement and Savings Behavior of Low-Income Households." *Journal of Econometrics* 145: 21-42.

van Soest, Arthur and Hana Vonkova. 2014. "How Sensitive Are Retirement Decisions to Financial Incentives? A Stated Preference Analysis." *Journal of Applied Econometrics* 29(2): 246-264.

Wiswall, Matthew and Basit Zafar. 2015. "Determinants of College Major Choice: Identification using an Information Experiment." *Review of Economic Studies* 82(2): 791-824.

Wiswall, Matthew and Basit Zafar. 2021. "Human Capital Investments and Expectations about Career and Family." *Journal of Political Economy* 129(5): 1361-1424.

Figure 1. Are Respondents' Answers Consistent with the Law of Total Probability?



Note: Figure shows the distribution of responses for the unconditional probability of working in 2 years computed by combining the health-contingent working probabilities and health probabilities using the law of total probability (on the vertical axis) versus the self-reported reported unconditional probability of working in 2 years (on the horizontal axis). The correlation between the two measures is 0.928.

Table 1. Percent Chance of Working, Health, and Health-Contingent Working

	Working	Low Health	Working in Low Health	Working in High Health	S*ea*TE
			2-Year Ahead		
Mean	65.9	16.6	41.9	70.5	-28.5
Std. Dev.	35.3	16.5	36.1	36.0	27.9
Q25	40.0	5.0	5.0	50.0	-50.0
Median	80.1	10.0	40.0	90.0	-25.0
Q75	97.5	25.0	75.0	100.0	0.0
			4-Year Ahead		
Mean	52.7	23.5	33.0	58.7	-25.7
Std. Dev.	37.0	19.5	34.4	39.0	27.6
Q25	17.0	10.0	0.0	20.0	-50.0
Median	50.0	20.0	20.0	68.0	-20.0
Q75	90.0	30.0	50.0	100.0	0.0

Note: Sample size is 970 for the 2-year ahead sub-sample and 839 for the 4-year ahead sub-sample. Table shows mean, standard deviation, first quartile (Q25), median, and third quartile (Q75) across respondents for each reported probability (measured as a percent chance between 0 and 100) and for S*ea*TE. Probability of working is calculated from the law of total probability using elicited probabilities of health and of work fixing health (see text for discussion). S*ea*TE is the different between the probability of working in low versus high health.

Table 2. S*ea*TE: Negative, Zero, or Positive (fraction of responses, percent)

	2-Year Ahead	4-Year Ahead
Negative S*ea*TE	70.31	70.80
Zero S*ea*TE	28.45	28.25
Positive S*ea*TE	1.24	0.95
Observations	970	839

Note: Tables shows the fraction of respondents with negative S*ea*TE (lower chance of working in low health than in high health), zero S*ea*TE (same chance of working across low and high health), and positive S*ea*TE (higher chance of working in low health than in high health).

Table 3. Unpacking Zero S*ea*TE (fraction of responses, percent)

	2-Year Ahead	4-Year Ahead
Never work	31.88	41.35
Always work	47.10	34.18
Maybe work	21.01	24.47
Observations	276	237

Note: Table shows distribution of responses among respondents who give the same probability of working in high and low health. In both health states, never-work respondents have zero probability of working, always-work respondents have probability one of working, and maybe-work respondents have interior probability of working.

Table 4. Unpacking Negative S*ea*TE (percent chance)

	2-Year Ahead	4-Year Ahead
Mean	-40.9	-36.8
Std. Dev.	24.1	25.1
Q25	-50	-50
Median	-40	-30
Q75	-20	-15
Observations	682	594

Note: Table reports same statistics as Table 1 for the subset of respondents who have lower probability of working in low health than in high health.

Table 5. Indicators of 2- and 4-Year Ahead S*ea*TE

	2-Year Ahead S*ea*TE	**4-Year Ahead S*ea*TE**
Constant	-0.154	-0.119
	(0.061)	(0.065)
Age at S4 (≤ 59 excluded)		
60-61	-0.042	-0.037
	(0.031)	(0.032)
62	-0.113	-0.054
	(0.040)	(0.041)
63-64	-0.034	-0.033
	(0.031)	(0.033)
65	-0.029	-0.110
	(0.045)	(0.051)
66-67	-0.017	0.029
	(0.037)	(0.039)
68-69	-0.124	-0.080
	(0.037)	(0.040)
70-71	-0.120	-0.090
	(0.046)	(0.049)
≥ 72	-0.095	-0.089
	(0.034)	(0.037)
Gender (Male excluded)		
Female	0.004	-0.011
	(0.021)	(0.023)
Education (≤ HS excluded)		
Some college	-0.000	-0.025
	(0.044)	(0.047)
College grad	0.014	-0.007
	(0.042)	(0.044)
Other adv. degree	-0.038	-0.017
	(0.045)	(0.047)
MBA	-0.007	0.005
	(0.051)	(0.054)
JD, PhD, MD	-0.018	-0.071
	(0.049)	(0.053)
Occupation at S4 (Management and professional excluded)		
Operative	0.009	-0.007
	(0.025)	(0.027)
Other services	-0.020	-0.019
	(0.032)	(0.034)
Job type at S4 (Career excluded)		
Bridge	0.004	-0.017
	(0.022)	(0.023)

Marital status at S4		
Partnered	-0.011	-0.009
	(0.024)	(0.026)
Spouse's work status at S4		
Working	-0.012	-0.003
	(0.023)	(0.024)
Total HH wealth at S4		
First quintile	-0.035	-0.037
	(0.033)	(0.036)
Second quintile	-0.039	-0.082
	(0.032)	(0.035)
Third quintile	-0.019	-0.031
	(0.030)	(0.032)
Fourth quintile	-0.041	-0.043
	(0.029)	(0.031)
Replacement rate at S4		
First quintile	-0.022	-0.025
	(0.031)	(0.034)
Second quintile	0.002	0.027
	(0.031)	(0.033)
Third quintile	-0.032	-0.025
	(0.030)	(0.032)
Fourth quintile	-0.024	-0.020
	(0.030)	(0.032)
Current salary at S4		
First quintile	-0.077	-0.044
	(0.038)	(0.041)
Second quintile	-0.072	-0.042
	(0.034)	(0.036)
Third quintile	-0.002	0.007
	(0.031)	(0.034)
Fourth quintile	-0.009	-0.005
	(0.030)	(0.032)
Work Status at S1		
Completely retired	0.108	0.039
	(0.036)	(0.041)
Observations	970	839
R^2	0.0575	0.0539

Note: OLS estimates of mean linear regressions of 2-year and 4-year ahead S*ea*TE on covariates. Standard errors reported in parenthesis under the corresponding point estimate. S1 stands for Survey 1 and S4 stands for Survey 4.

Table 6. Predicting Work 2-Year Ahead: Health-Contingent versus Unconditional Probabilities

	(1)	(2)	(3)
Constant	0.301	0.322	0.301
	(0.037)	(0.036)	(0.037)
Health-contingent work probability	0.590		0.502
	(0.047)		(0.191)
Unconditional work probability		0.595	0.093
		(0.048)	(0.197)
Observations	584	584	584
R^2	0.216	0.207	0.216
Test for no incremental predictive power of:			
Unconditional work probability (3 vs. 1), $\chi^2(1)$ [p-value]			0.00 [1.00]
Health-contingent probability (3 vs. 2), $\chi^2(1)$ [p-value]			6.70 [0.01]

Note: The table shows how well health-contingent and unconditional probabilities of working predict realized work 2 years ahead. The dependent variable is 1 if the person actually works 2-year ahead and zero otherwise. Specification (1) uses as predictor the health-contingent probability of working for the health state that was actually realized 2 years ahead. Specification (2) uses as predictor the unconditional probability of working constructed from the health-contingent probabilities of working and the health probabilities according the law of total probability. Specification (3) uses both predictors. See notes to Table A5 for summary statistics. Standard errors in parenthesis. Last rows report tests for no incremental predictive power.

Table 7. Predicting Work 2-Year Ahead: Health-Contingent versus Unconditional Probabilities, with Age Interactions

	(1)	(2)	(3)
Constant	0.443	0.447	0.445
	(0.075)	(0.074)	(0.075)
Health-contingent work probability			
≤ 59	0.730		0.823
	(0.135)		(0.495)
60-61	0.672		1.398
	(0.135)		(0.652)
62	0.857		0.474
	(0.201)		(0.695)
63-64	0.653		1.272
	(0.124)		(0.595)
65	0.659		-0.662
	(0.192)		(0.551)
66-67	0.639		0.804
	(0.155)		(0.498)
68-69	0.784		0.373
	(0.149)		(0.810)
70-71	0.100		0.057
	(0.202)		(0.520)
≥ 72	0.406		0.053
	(0.102)		(0.528)
Unconditional work probability			
≤ 59		0.695	-0.097
		(0.136)	(0.495)
60-61		0.637	-0.756
		(0.138)	(0.664)
62		0.896	0.420
		(0.213)	(0.730)
63-64		0.618	-0.641
		(0.126)	(0.602)
65		0.828	1.479
		(0.203)	(0.579)
66-67		0.641	-0.1889
		(0.169)	(0.541)
68-69		0.832	0.443
		(0.158)	(0.859)
70-71		0.107	0.051
		(0.218)	(0.560)
≥ 72		0.438	0.383
		(0.110)	(0.562)
Observations	584	584	584
R^2	0.261	0.253	0.275
Test for no incremental predictive power of:			
Unconditional work probability (3 vs. 1), $\chi^2(9)$ [p-value]			11.28 [0.26]
Health-contingent probability (3 vs. 2), $\chi^2(9)$ [p-value]			17.72 [0.04]

Note: Regressor interacted with age dummies. Regressions include age dummies (not reported). See also note to Table 6.

Table 8. Predicting Work 2-Year Ahead: Health-Contingent versus Unconditional Probabilities, with Initial Health Interactions

	(1)	(2)	(3)
Constant	0.219	0.243	0.222
	(0.079)	(0.077)	(0.079)
Health-contingent work probability			
Good	0.501		0.424
	(0.104)		(0.298)
Very Good	0.569		0.489
	(0.067)		(0.309)
Excellent	0.740		0.498
	(0.087)		(0.438)
Unconditional work probability			
Good		0.515	0.089
		(0.111)	(0.317)
Very Good		0.584	0.086
		(0.069)	(0.321)
Excellent		0.751	0.252
		(0.089)	(0.446)
Observations	584	584	584
R^2	0.248	0.240	0.248
Test for no incremental predictive power of:			
Unconditional work probability (3 vs. 1), $\chi^2(3)$ [p-value]			0.00 [1.00]
Health-contingent probability (3 vs. 2), $\chi^2(3)$ [p-value]			6.21 [0.10]

Note: Regressor interacted with initial health dummies. Regressions include age and initial health dummies (not reported). See also note to Table 6.

Table 9. Differenced Values of Working vs. Not Working in High and Low Health

	2-Year Ahead		4-Year Ahead	
	$\tilde{v}^H$	$\tilde{v}^L$	$\tilde{v}^H$	$\tilde{v}^L$
Mean	0.97	-0.35	0.48	-0.72
Std. Dev.	1.70	1.65	1.79	1.60
Q25	0	-1.64	-0.84	-2.58
Median	1.28	-0.25	0.47	-0.84
Q75	2.58	0.67	2.58	0
Correlation	0.73		0.75	

Note: Table shows distribution of the measured differenced health-contingent values of continued work, $\tilde{v}^H$ and $\tilde{v}^L$. See text and Appendix (i) for details of calculations. Observations: 970 (2-year ahead) and 839 (4-year ahead).

Table 10. Relationship between Health and Work with Simulated Realizations

	2-Year Ahead			4-Year Ahead		
	Uncorrelated	Empirical	Higher correlation	Uncorrelated	Empirical	Higher correlation
Constant	0.703	0.709	0.730	0.586	0.594	0.621
	(0.011)	(0.011)	(0.011)	(0.014)	(0.014)	(0.014)
Health *h*	-0.282	-0.305	-0.415	-0.254	-0.286	-0.371
	(0.040)	(0.039)	(0.040)	(0.038)	(0.035)	(0.036)
SEE	0.463	0.461	0.448	0.488	0.485	0.474
Obs.	970	970	970	839	839	839

Note: Table reports mean values from the 1000 replications. Uncorrelated case has health transition probability fixed at sample mean of the health probabilities elicited in the survey. Empirical case uses individual-specific health probabilities elicited in the survey. Highly correlated case has stronger correlation of health transition probability and value of work as described in text. The left-hand-side variable is the simulated realized decision to work (d). The right-hand-side variable is simulated realized health state (h). h=1 is low health and d=1 is work.