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

Revised: June 7, 2021

The paper studies the effect of health on work among older workers using data on individuals' subjective probabilities of working to specified future horizons under alternative health states. It constructs within-person differences in work expectations across health states, interpretable as Subjective *ex ante* Treatment Effect (S*ea*TE) within potential outcomes, and document large S*ea*TE heterogeneity across individuals. Using standard discrete choice dynamic programming to derive health-contingent values of working longer, it shows that the revealed heterogeneity in taste for work biases regression-based estimates. Finally, the health-contingent working probabilities are shown to strongly predict work given realized health after two years.

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 Gábor Kézdi, Wilbert van der Klaauw, Eric French, and participants at seminars and conferences for very helpful comments and discussion.

* Department of Economics, Bocconi University, Milan 20136, Italy. Phone: +39 02 5836-3413 (office). 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. Phone: +1 734 764-8365 Fax: +1 734 763-9831. E-mail: shapiro@umich.edu.

I. Introduction

How does a treatment or state, such as poor health, affect an outcome or behavior such as retirement? In observational data, the econometrician observes the behavior of each individual or "unit" under the realized state, but not the counterfactual behavior under alternative states. Unobservability of counterfactual outcomes makes impossible inferences about causal effects at the unit level. It also makes difficult inferences about causal effects at the group level owing to selectivity and unobserved heterogeneity.

To circumvent these problems, we study the effect of health on the labor supply of healthy older workers using purposely collected data on individuals' subjective expectations of working to future horizons under alternative contemporaneous health states, along with unconditional work and health expectations. Specifically, we ask working participants in the Vanguard Research Initiative (VRI) and in the Health and Retirement Study (HRS) to report the probability on a 0-100 percent chance scale that they will still be working 2 and 4 years ahead under alternative scenarios about their health in 2 and 4 years. These measures enable us to construct within-person differences in work expectations across health states where each person is both treatment and control, thus obviating unobservability of counterfactuals and allowing for unrestricted heterogeneity across individuals.

Building on recent contributions from the survey expectations literature (see discussion of Arcidiacono, Hotz, Maurel, and Romano (2020) and Wiswall and Zafar (2021) in Section 2), we first interpret the within-person difference in subjective expectations across states as a Subjective *ex ante* Treatment Effect (S*ea*TE) at the *individual* level within textbook potential outcomes and document large S*ea*TE heterogeneity across individuals. We next use standard discrete choice dynamic programming to derive health-contingent values of working longer, which too are person-specific, and show that the heterogeneity in taste for work made observable by our approach induces a non-negligible bias in regression-based estimates of the effect of health on labor supply.

The Subjective *ex ante* Treatment Effect (S*ea*TE) of bad health on work is highly heterogeneous across individuals in both the Vanguard Research Initiative (VRI) and the Health and Retirement Study (HRS). In the VRI, S*ea*TE is zero (no effect of health on work) for almost 30 percent of initially healthy and working respondents aged 57 and higher at both 2- and 4-year horizons. A few respondents report positive S*ea*TE (more likely to work in low than in high health). The remaining 70 percent report having a strictly negative S*ea*TE, that is, a negative effect of bad health on work (median of -40 percentage points and standard deviation of 24 for the 2-year horizon; median of -30 and standard deviation of 25 for the 4-year horizon). Hence, for the median response, the *ex ante* effect of bad health on work is negative and large in magnitude. We find similar level and heterogeneity in the S*ea*TE among HRS respondents.

The problem of recovering individual-specific treatment effects and characterizing treatment-effect heterogeneity is also central to a recent and rapidly growing econometric literature that has been expanding the toolbox of causal inference methods with statistical learning and data science techniques (see Athey and Imbens (2016, 2019), Davis and Heller (2017), Chernozhukov, Demirer, Duflo, and Fernandez-Val (2020), Wager and Athey (2018), among others), stimulated by an increasing availability of large data sets in economics and related fields. This literature, which is interested in estimating causal effects from data on realized outcomes ("*ex post*" effects), has been focusing on heterogeneity with respect to observable covariates, that is, on recovering low-granularity conditional average treatment effects (CATE) or features of theirs. Instead of relying on big data and the rich observed heterogeneity that big data can afford, our *ex ante* approach relies on another increasingly available type of data: survey expectations. Moreover, our approach is completely individual-specific. Hence, our approach complements other approaches to causal inference at the granular level. By design, our approach quantifies *ex ante* effects implied by individuals' subjective expectations. In the paper, we discuss conditions under which these *ex ante* effects are informative of *ex post* effects.

We begin in Section II by reviewing the main strands of literature to which our paper contributes. These include the longstanding empirical literature studying the causal effect of health on labor supply using observational data on health and work in regression analyses or structural models and the more recent literature eliciting probabilistic expectations in economic surveys to study economic decision-making under uncertainty and expectations formation.

In Section III, we interpret S*ea*TE and its components within standard potential outcomes (POF) and discrete choice dynamic programming (DP). Within the latter, we derive expressions for the values of continued work from the working probabilities fixing health. Like the working probabilities, these measures are health contingent and completely individual specific.

In Section IV, we describe the VRI study and present descriptive results on the unconditional work probability, the unconditional health probability, the probabilities of working in high and low health, and the S*ea*TE.

In Section V, we first describe empirically the DP-based measures of continued work we derived in Section III. We show that moving from high to low health has a large, negative effect on the mean value of continued work, but the within-person correlation of this value across the two health states is substantial. Next, we simulate health and work realizations from the DP framework and show that regression-based estimates using realized health and work will be biased when there is unobserved heterogeneity in taste for work that is also correlated with health. Our approach, conversely, reveals such heterogeneity, which can be properly accounted for.

While the point of this paper is that state-contingent probabilities are of particular interest because they provide measurement for states that do not occur, it is still important to know how they predict outcomes for states that are realized. In Section VI we use the panel structure of the VRI to study how health-contingent probabilities of working predict work in realized health states. Health-contingent probabilities do strongly predict work given realization of health two years ahead.

II. Related Literature

Health and retirement. The determinants of retirement have been widely studied in economics and elsewhere (e.g. see recent reviews by Coile (2015), O'Donnell, van Doorslaer, and van Ourti (2015), Fisher, Chaffee, and Sonnega (2016), and French and Jones (2017)), and have drawn increasing attention in recent years due to the uncertain sustainability of the Social Security System and the related calls for making workers work longer (e.g. 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, horizon, expectations (e.g. see Rust and Phelan (1997), Blau and Gilleskie (2001, 2008), van der Klaauw and Wolpin (2008), Bound, Stinebrickner, and Waidmann (2010), French (2005), French and Jones (2011), and Garcia-Gomez, Galama, van Doorslaer, and Lopez-Nicolas (2017)), 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. see Grossman (1972), Bound, Schoenbaum, Stinebrickner, and Waidmann (1999), Lumsdaine and Mitchell (1999), McGarry (2004), and Blundell, Britton, Costa Dias, and French (2016)). Both retirement and health are subject to measurement issues which exacerbate the challenges of studying their relationship empirically (e.g. see Bound (1991), Dwyer and Mitchell (1999), McGarry (2004), Lindeboom and Kerkhofs (2009), and Kapteyn and Meijer (2014) on health measurement, and Gustman, Mitchell, and Steinmeier (1995), Benítez-Silva and Dwyer (2005), Gustman, Steinmeier, and Tabatabai (2010), and Maestas (2010) on concepts and measures of retirement).[1] See De Nardi, French, and Jones (2016) and French and Jones

[1] 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. see 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.

(2017) for careful reviews of the measurement and inferential challenges—in both cross-section and panel data—for estimating the effects of health and heath care.

We contribute to this literature in multiple ways. Using our method, we are able to generate individual-specific estimates of the effect of health on work/retirement and to document that the effect of bad health on older workers' prospects of working longer is negative and large on average, but also highly heterogeneous across individuals. We produce estimates that are credibly free of some of the major biases that have plagued realizations-based estimates so far, such as "justification bias" (due to retirees reporting lower health to justify their being retired) and (unobserved) heterogeneity bias.

When longitudinal data on health and labor supply realizations are available, an important strategy is to use health shocks as a source of identifying information. See 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), Blundell, Britton, Costa Dias, and French (2016), and Jones, Rice, and Zantomio (2016). Our approach shares with panel data methods the use of multiple observations per individual. In the panel, identification uses changes across time in health state and, hence, in work status. Our approach uses changes in work probabilities across alternative health states.

Subjective expectations in economic surveys. Since the early 1990s, economists have increasingly measured individuals' subjective expectations in surveys, using a 0-100 scale of percent chance. This endeavor was stimulated by the importance of subjective expectations in economic models of lifecycle behavior and human capital investment (e.g. see Hamermesh (1985) on subjective horizons and Manski (1993) on subjective returns to schooling), and by earlier empirical evidence and theoretical arguments demonstrating the greater informativeness of elicited probabilities for binary events over more commonly used "yes/no" intention measures (e.g. see Juster (1966) and Manski (1990)).[2]

Our analysis is based on two high-quality national studies, the VRI (which is representative of older US wealthholders) and the HRS (which is population representative of older US households), both of which have greatly contributed to the advancement of elicitation and analysis of stated choices through

---

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).

[2] Manski (2004, 2018), Attanasio (2009), Hurd (2009), van der Klaauw (2012), Armantier, Bruine de Bruin, Potter, Topa, van der Klaauw, and Zafar (2013), Delavande (2014), Schotter and Trevino (2014), Giustinelli and Manski (2018), and Altig, Barrero, Bloom, Davis, Meyer, and Parker (2020) trace the development of the subjective expectations literature from various perspectives. Papers by Arrondel, Calvo-Pardo, Giannitsarou, and Haliassos (2020) and Fuster, Perez-Truglia, and Zafar (2020) are examples of recent advances in this literature.

subjective expectations questions (see Hurd (2009) on HRS) and related strategic survey questions (see Ameriks, Briggs, Caplin, Shapiro, and Tonetti (2020) on VRI).

Subjective conditional expectations. Our paper advances on use of survey data on probabilistic *conditional* expectations to study *ex ante* treatment effects (Arcidiacono, Hotz, Maurel, and Romano (2020) and Wiswall and Zafar (2021)) and choice behaviors in incomplete scenarios posed by the researcher via conditioning information (e.g. Manski (1999) and Blass, Lach, and Manski (2010)). [3,4]

Methodologically, our analysis of the S*ea*TE of health on retirement in Sections IIIA and IV is closest to Arcidiacono, Hotz, Maurel, and Romano (2020), henceforth AHMR, and Wiswall and Zafar (2020), henceforth WZ. AHMR generate estimates of individual and aggregate level effects of occupation choice on future earnings of Duke undergraduates by comparing students' subjective earnings expectations across alternative scenarios of occupation choice and graduation major. WZ obtain estimates of monetary and non-monetary returns to alternative college majors among NYU undergraduates by comparing students' subjective expectations for monetary and non-monetary outcomes across alternative graduation majors. Thus, both AHMR and WZ consider generalized Roy (1951) applications, where potential treatments are alternative human capital investments (occupations and college majors respectively) and potential outcomes are future monetary consequences, and in the case of WZ also non-monetary consequences, of alternative human capital investments (own earnings, spouse's earnings, fertility, etc.).

The *ex ante* effects studied by AHMR and WZ are therefore based on students' subjective expectations for future consequences of their choice (say, earnings), conditioned on alternative choices students could make (occupations and/or majors). According to Blass, Lach, and Manski (2010)'s taxonomy, as of the time of occupation (major) choice, these expectations concern so-called *unresolvable* uncertainty. Indeed, students' choices are based on these expectations.[5] In contrast, we consider an application where the main

[3] Hudomiet, Hurd, and Rohwedder (2020), who collaborated with us in the development of the VRI and HRS modules, study the work and other domains using conditional probabilities elicited in the American Life Panel (ALP).

[4] For in-depth discussions of the role of *conditioning information* in survey elicitation of subjective data, see Dominitz (1997), Wolpin (1999), and Fischhoff, Welch, and Frederick (1999). For uses of *unconditional* choice probabilities (and in some cases also state probabilities) in econometric analysis of discrete choices, see van der Klaauw and Wolpin (2008), van der Klaauw (2012), Pantano and Zheng (2013), Stinebrickner and Stinebrickner (2014), and Wiswall and Zafar (2015). Unconditional choice probabilities may in fact be considered extreme versions of choice probabilities in incomplete scenarios in the sense of Manski (1999), where the choice scenario is left completely unspecified by the researcher.

[5] In practice, in their initial survey AHMR elicit students' expectations for earnings associated to alternative occupations before students graduate from college and select an occupation. To measure uncertainty with respect to future occupation, AHMR additionally elicit students' choice expectations over occupations. While elicitation of earnings expectations fixing occupation eliminates the uncertainty regarding future occupation, some *resolvable* uncertainty (in the sense of Blass, Lach, and Manski (2010)), concerning future earnings and other factors relevant for occupation choice, may remain at the time of the initial survey. Indeed, AHMR's respondents may update their earnings expectations and learn choice-relevant factors between the time of the initial survey and the time of choice.

choice variable (labor supply) represents the outcome of interest, and one of the main state variables (contemporaneous health) the treatment of interest. Individuals' health status at the time of their labor supply decision is obviously known to them, so there is no uncertainty about contemporaneous health as of the time of choice.[6] Recall, however, that in order to address the unobservability of individuals' labor supply in counterfactual health states, we rely on individuals' subjective expectations of working *before* the time of choice, when the value of the choice-relevant health state has not yet realized. As of the time of elicitation, individuals do face resolvable uncertainty about their health state at the time of choice. For this reason, we elicit respondents' health probabilities along with their unconditional and health-contingent working probabilities. Importantly, the health-contingent probabilities of working forming the S*ea*TE are conditioned on potential health states as of the time of choice, so they are not affected by the uncertainty individuals may perceive at the time of elicitation about their health at the time of choice.[7]

Subject expectation in choice modeling. The timing of elicitation relative to that of decision making is also important for our interpretation of the unconditional working probability, the health-contingent working probabilities underlying S*ea*TE, and the health probabilities, within the dynamic programming (DP) framework of Sections IIIB and V. While the DP framework we consider is off-the-shelf, to the best of our knowledge we are the first to investigate the interpretation and information content of related unconditional choice and state probabilities and state-contingent choice probabilities within DP, and to use these measures to recover state-contingent values of relative actions to reveal unobserved heterogeneity in individuals' valuations. Some of the papers referenced above and reviewed below have used unconditional probabilities of choice and/or state variables along with realized choices and states to aid estimation of discrete choice dynamic programming models (e.g. van der Klaauw and Wolpin (2008) and van der Klaauw (2012)). Their use of expectations, however, is to improve estimation efficiency rather than for identification. In contrast, we use the discrete choice DP framework to give a structural interpretation to the conditional probabilities.

---

[6] As already discussed, there may be uncertainty about future health. Concerns about future health may affect labor supply decisions above and beyond current health. First, health may be viewed not only as a choice-relevant state affecting individuals' labor supply decisions through state-dependent utility and productivity, but also as an outcome of labor supply. If individuals believe that working versus retiring can affect their health, when deciding whether to work or retire they will take their perceived health returns to working versus retiring into account. Second, remaining healthy years may also matter for when to retire. In either case, expectations for future health would affect labor supply decisions in addition to contemporaneous health. In this paper, we do not investigate the effect of expectations for future health on individuals' labor supply, but we believe it would be an interesting venue for future research.

[7] Alternatively, we could have retrospectively asked individuals to provide backcasts of their labor supply in unrealized health states so as to augment data on actual choices, which correspond to realized health, with individual-level measures of choices in counterfactual health states. While the *ex ante* approach to counterfactuals is not free of challenges, we believe it is cleaner than a backcasting approach where *ex post* rationalization may be a concern.

Using subjective expectations to study labor supply and its relation with health. Only a few studies to date have employed survey measures of subjective working and/or health expectations to study retirement behavior and its relationship with individuals' health. Unconditional working probabilities have been used as an outcome variable in place of or in combination with actual labor supply data to estimate regression-based ceteris paribus effects (McGarry, 2004) or structural parameters (van der Klaauw and Wolpin, 2008). Health and longevity expectations have been used to generate moment conditions to increase efficiency of structural estimates (van der Klaauw and Wolpin, 2008).

McGarry (2004) investigates the effect of health on labor supply expectations of working respondents in the HRS. 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' subjective probability of working past age 62 and its changes, finding large negative effects of health on the probability of working. The key novelty of McGarry (2004)'s 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. While we build on this idea, our innovation is to elicit the probability of work *fixing* health. By eliciting the probability of working in alternative potential health states for each individual, we can recover individual-level effects by comparing the probability of working across health states *within* individuals rather than across individuals.

van der Klaauw and Wolpin (2008) develop and estimate a rich dynamic programming model of household retirement and saving using multiple waves of the HRS. Innovating on earlier structural models of retirement, the authors combine respondents' unconditional working and longevity expectations with observed realizations of respondents' labor supply, health, and the other state variables in order to increase estimation precision. Hence, subjective expectations data are used to improve efficiency of estimates, rather than for aiding identification of structural parameters or studying particular dimensions of heterogeneity. We, instead, combine health-contingent probabilities of working and dynamic programming to construct individual-specific, health-contingent measures of the workers' value of continued work which are typically unobserved in observational data on labor supply.

Finally, a small set of studies has investigated preferences for work and retirement arrangements using stated preference methods. For instance, Kapteyn, van Soest, and Zissimopoulos (2007), van Soest and Vonkova (2014), and Ameriks, Briggs, Caplin, Lee, Shapiro, and Tonetti (2020) study preferences for full and partial retirement using hypothetical choices. Our approach advances the hypothetical choice agenda by posing the choices probabilistically, using the resulting measures to derive treatment effects at the individual level, and explicitly interpreting such measures within the mainstream frameworks of potential outcomes and dynamic programming.

III. Analytic Framework

In this section, we discuss in detail how subject expectations map into two standard analytic frameworks—potential outcomes[8] and discrete choice dynamic programming.[9]

A. Potential Outcomes Framework 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).[10] In period t, after observing the realized value of the state vector, s_{it}, the decision-maker decides whether to work or not: $d_{it} \in \{1,0\}$, where 1 denotes working, 0 not working, and i the decision maker. The state vector, s_{it}, includes the decision-maker's health and other variables discussed below. Health is also modelled as a binary variable, $h_{it} \in \{0,1\}$, where 0 denotes high health and 1 low health.

Figure 1 illustrates the setting by means of a decision tree with three time periods. Within the context of a formal DP model of labor supply, the decision tree is the extensive-form representation of the decision-maker's problem as a game against nature, where nodes are information sets and arcs are alternating decisions by nature and the agent (Rust, 1992). Here we use the tree as a unifying tool to help us illustrate the connections between the POF and DP settings as well as between the *ex post* and *ex ante* approaches.

N-nodes denote nature's decision points and A-nodes denote the agent's decision points. For simplicity we drop subscript i and display the case where health is the only state variable $(s_{it} \equiv h_{it})$. At each N-node, nature assigns a health level to the agent from the set of feasible health levels (high or low), represented as arcs exiting each N-node. In the figure, high-health arcs are labeled as H and low-health arcs as L. At each A-node, the agent optimally chooses between working and not working after learning

[8] Originating in statistics from the work of Neyman (1923), Fisher (1935), Cox (1958), and Rubin (1974), here we use POF to refer to its interpretation and developments in econometrics, as discussed for instance in Heckman and Vytlacil (1999).

[9] Eckstein and Wolpin (1989), Rust (1992, 1994), Keane and Wolpin (2009), Arcidiacono and Ellickson (2011), and others review the DP approach from various perspectives.

[10] Although AHMR's interpretation of average differences in earnings expectations across occupations as *ex ante* treatment effects is also based on POF, we think it useful to provide a formal exposition of the POF in realizations versus expectations in order to establish the mapping between the health-contingent working probabilities we elicit in our survey and the S*ea*TE of health on retirement we study. Furthermore, our exposition of POF in relation to the decision tree shown in Figure 1 helps us connect the interpretation of the state-contingent choice probabilities forming the S*ea*TE between POF and DP.

whether their health is high or low; thus, retirement is not necessarily an absorbing state. In the figure, working arcs are labeled as *W* (*d*=1) and non-working arcs as ~*W* (*d*=0). Each path through the tree yields a separate payoff (summarized in the final column). At each terminal node, the agent obtains a payoff, corresponding to a separate path through the tree.

Arcs exiting from N-nodes can be interpreted as *potential treatments* and arcs exiting from A-nodes as *potential outcomes*. At any given *t*, individual *i* is characterized by a response function, $d_{it}\left(h_{it}\right)$, which maps mutually exclusive and exhaustive treatments into outcomes. Hence, $d_{it}\left(h_{it}\right)$ is the potential labor supply outcome of person *i* at time *t* associated with health treatment h_{it}.

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

$$\Delta_{it} = d_{it}\left(1\right) - d_{it}\left(0\right) \equiv E[\Delta_{it}] + [\upsilon_{it}\left(1\right) - \upsilon_{it}\left(0\right)], \tag{1.1}$$

where: $\Delta_{it} \in \{-1, 0, 1\}$ because $d_{it}\left(h_{it}\right)$ is binary, $E[\Delta_{it}]$ represents the common "gain" (or "loss") from treatment (with the expectation taken across units), and $[\upsilon_{it}\left(1\right) - \upsilon_{it}\left(0\right)]$ represents the idiosyncratic gain (or loss) for individual *i*. Here 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*, described respectively by $h_{it} = 1$ (low health) and $h_{it} = 0$ (high health).

The variable *h* denotes potential health. We need an indicator for realized health. Define $z_{it} \in \{0, 1\}$ the realized health state of person *i* at time *t*, where 0 means high health and 1 low health as before. Then, $d_{it} \equiv d_{it}\left(z_{it}\right) \in \{0, 1\}$ is the person's *realized outcome*, that is, the labor supply *i* selects when the realized health state is z_{it}; whereas, $d_{it}\left(1 - z_{it}\right) \in \{0, 1\}$ is the *counterfactual outcome,* that is, the labor supply *i* would have selected had the realized health state been $\left(1 - z_{it}\right)$.

For example, consider an agent who is in high health and working at time *t*. This agent is hence on the top-most arc of Figure 1 (denoted as a thick solid line). At time *t*+1, this agent will be treated with either high health or low health. The individual-level treatment effect of health on labor supply at time *t*+1 for this particular agent is given by the difference in the agent's labor supply decisions across the two potential health states at *t*+1, that is, $\Delta_{it+1} = d_{it+1}\left(L\right) - d_{it+1}\left(H\right)$. Of course, this individual-level treatment effect cannot be observed because only one of the two health states is realized.

Consider the case when the agent happens to experience low health at time $t+1$ and decides not to work in that period. Then, $z_{i,t+1} = 1$ (L) is the actual treatment and $d_{i,t+1} = 0$ ($\sim W$) is the realized outcome, represented by the dashed path. The high health-working combination, (H, W), represented by the dotted path is instead counterfactual. The counterfactual outcome corresponding to the unrealized treatment is unobservable.

Treatment Effects in Realizations. Analysis with observational data on realized health and work must make assumptions to address the fact that treatment effects at the individual level are not observable. Most regression-based approaches within the POF compare outcomes *across groups* of suitably similar units rather than within units. For example, consider the Average Treatment Effect (ATE),

$$ATE_t(1-0) = E[d_{it}(1) - d_{it}(0)] = E\left[d_{it}(1)\right] - E\left[d_{it}(0)\right], \tag{1.2}$$

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

$$\begin{aligned} ATE_t(1-0) = &\left\{E\left[d_{it}(1) \mid z_{it} = 1\right]P(z_{it} = 1) + E\left[d_{it}(1) \mid z_{it} = 0\right]P(z_{it} = 0)\right\} \\ &-\left\{E\left[d_{it}(0) \mid z_{it} = 0\right]P(z_{it} = 0) + E\left[d_{it}(0) \mid z_{it} = 1\right]P(z_{it} = 1)\right\}. \end{aligned} \tag{1.3}$$

Random sampling of $\{z_{it}, d_{it}\}$ from the population distribution of realized health and labor supply asymptotically reveals all components of (1.3) except the counterfactual moments, $E\left[d_{it}(1) \mid z_{it} = 0\right]$ (mean labor supply of high-health workers in a low-health world) and $E\left[d_{it}(0) \mid z_{it} = 1\right]$ (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.

Studies based on randomized control trials (RCT) address the identification problem by randomly assigning subjects to treatments. For example, suppose that individuals are randomly assigned to either a *Control* group or to a *Treatment* group. Control individuals receive treatment $z_{it} = 0$; whereas, treated individuals receive treatment $z_{it} = 1$. In the ideal case of full randomization with perfect compliance, the mean of realized outcomes in each group can be used as a measure of the counterfactual mean for the other group, which yields point identification of the ATE parameter in (1.3).

Obviously, randomization of good and bad health across individuals is not a viable strategy to study the causal relationship between health and labor supply of interest to this paper. The bulk of the literature to date has relied on data on realized health and labor supply, so has to grapple with non-random selection of individuals into health states and unobserved heterogeneity in the gain (or loss) from treatment (bad health). For example, high-health individuals ($z_{it} = 0$) may have higher (unobserved) preference for work than low-health individuals ($z_{it} = 1$) that might persist if the former were to experience a negative health shock ($h_{it} = 1$), and *vice versa*. That is, a simple comparison of mean labor supply across high- and low-health workers, as in a linear probability model of labor supply with an intercept and a dummy for low health, will generally confound ATE with selection and heterogeneity biases. This can be seen formally in the textbook decomposition of the simple contrast of mean outcomes across treated and untreated units into ATE, selection bias, and heterogeneity bias, as follows:

$$\begin{aligned} E\left[d_{it} \mid z_{it}=1\right]-E\left[d_{it} \mid z_{it}=0\right] &= ATE_t\left(1-0\right)+\left\{E\left[\upsilon_{it}\left(0\right) \mid z_{it}=1\right]-E\left[\upsilon_{it}\left(0\right) \mid z_{it}=0\right]\right\} \\ &+[1-P\left(z_{it}=1\right)]\left\{E\left[\upsilon_{it}\left(1\right)-\upsilon_{it}\left(0\right) \mid z_{it}=1\right]-E\left[\upsilon_{it}\left(1\right)-\upsilon_{it}\left(0\right) \mid z_{it}=0\right]\right\}. \end{aligned} \tag{1.4}$$

Treatment Effects in Expectations. The approach we advance in this paper circumvents the impossibility of simultaneously observing $d_{it}\left(1\right)$ and $d_{it}\left(0\right)$ *ex post* by measuring them *ex ante*. We directly ask individuals to predict their outcome (working in this case) under specified scenarios about their treatment (health state in this case) at specific horizons.[11] We define the *Subjective ex ante Treatment Effect*,

$$SeaTE(i,t,\tau) = E_{i,t-\tau}(\Delta_{it}) \equiv P_{i,t-\tau}\left[d_{it}\left(1\right)=1\right]-P_{i,t-\tau}\left[d_{it}\left(0\right)=1\right], \tag{1.5}$$

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

We measure the *SeaTE* by eliciting the probability of the decision contingent on each health state in a survey τ periods in advance. Our approach of asking the two components of S*ea*TE separately rather than trying to elicit the casual effect directly fits the way in which respondents themselves seem to think and reason about the prediction problem (see quotation at the beginning of the paper). Moreover, this

[11] 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.

approach is natural from the perspective of economic and econometric modeling, of which conditional probabilities/expectations are essential building blocks.

The individual-level effects in (1.5) can also 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 the untreated (high health workers), given data on realized treatments; alternatively, the latter parameters may be estimated using data on health probabilities to weight the individual level effects (see also AHMR).

Under what conditions can the expression in (1.5), or features of its distribution across individuals, 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 equation (1.5) are *ceteris paribus* with respect to state variables that are relevant to the labor supply decision but are not fixed by the elicitation task. As Wolpin (1999) discusses in the context of Manski's (1999) original consideration of scenario-based expectations, there is a limit in a survey context of how completely the scenario can be specified. We make this point formally in Section IIIB, where we introduce separate notations for the state variables specified in the elicitation scenario and the state variables that are not specified in the elicitation scenario. Ultimately, and as explained more formally in Section IIIB and in the appendix appended to the text, interpretability of S*ea*TE as a causal effect *ex ante* depends on the nature of the relationship between the specified and unspecified states.

The second issue concerns the potential role of measurement error. As discussed in AHMR, 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 (classical or even non-classical) measurement error, as long as the errors have the same mean across the values of the treatment variable being contrasted. Thus, subject to the *ceteris paribus* condition discussed above and further analyzed below, in our setting the S*ea*TE of health on work at the individual level would be directly identified from the health-contingent working probabilities if the reported probabilities forming the S*ea*TE are error free, or if the two errors cancel out. If the object of interest is a feature of the distribution of S*ea*TE, for instance the AS*ea*TE, then it would suffice that the survey reports of the health-contingent working expectations have the same cross-sectional mean across the two health states underlying S*ea*TE. We discuss measurement error from rounding in Section IV.G.

A distinct but equally important question is under what conditions the *ex ante* effects addressed thus far are informative of *ex post* effects. 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.[12]

B. Dynamic Programming Interpretation of S*ea*TE

In this section, we relate the components of S*ea*TE to the individuals' decision problem within a standard framework of discrete choice dynamic programming (DP). Doing so allows interpretation of the state-contingent probabilities in terms of the individuals' optimization problem. Unlike the potential outcomes framework of the previous section, which provides useful interpretation of individual responses whether or not individuals have rational expectations, the dynamic programming framework of this section assumes individual-level rational expectations.[13]

Individual agents are represented by primitives, $u_{it}\left(s_{it}, d_{it}\right)$ and $\pi_{it}\left(s_{i,t+1} \mid s_{it}, d_{it}\right)$. As before, *i* indexes individuals and *t* time periods, with $i = 1, \ldots, N$ and $t = 0, 1, \ldots, T < \infty$. $u_{it}(s_{it}, d_{it})$ is the utility that agent *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 state vector), including health and other variables. Because health and the other state variables are generated by a Markovian stochastic process governing their evolution over time, their future values are uncertain from the viewpoint of the decision-maker. Specifically, $\pi_{it}(s_{i,t+1} \mid s_{it}, d_{it})$ is agent *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 opposed to the whole history of states and choices since the first period), as implied by the Markov-process assumption.

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

---

[12] Some weakened forms of rational expectations (e.g. respondents' subjective choice probabilities for a certain action are unbiased estimates of their objective probabilities of choosing that action), and of statistical independence of the realized treatments across the population (needed for applicability of the law of large numbers), would leave the above conclusion intact. However, aggregate shocks making treatments dependent across the population and systematic deviations from rational expectations in the form of biased expectations would generally invalidate it. See Manski (1999) for a more in-depth discussion. See also the discussion in AHMR within the context of using survey probabilities to estimate average *ex ante* effects.

[13] Recent contributions have shown that identification of dynamic discrete choice models without rational expectations is feasible (e.g. see An, Hu, and Xiao (2020)). Separate identification of agents' subjective beliefs and preferences still requires fairly strong assumptions such as time-invariant beliefs and preferences and partial knowledge of agents' subjective beliefs by the econometrician. In the case of infinite-horizon models, identification further requires exclusion restrictions. While An, Hu, and Xiao (2020)'s analysis aims at informing econometric analysis of dynamic discrete choice without expectations data, we believe that 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.

$$U_{it} = \sum_{j=0}^{T} \beta^j u_{i,t+j}\left(s_{i,t+j}, d_{i,t+j}\right), \tag{1.6}$$

where β is the discount factor. The agent behaves optimally according to the expected-utility maximizing decision rule, $\delta_{it}^*(s_{it})$, which satisfies the Bellman (1957) optimality principle. That is, at any time t and state s_{it}, δ_{it}^* is optimal also for the continuation process featuring the current state as starting point,

$$\delta_{it}^*(s_{it}) = \underset{d_{it}\in\{0,1\}}{\arg\max}\left\{u_{it}\left(s_{it}, d_{it}\right) + \beta \sum_{s_{i,t+1}} V_{i,t+1}^*\left[s_{i,t+1}, \delta_{i,t+1}^*\left(s_{i,t+1}\right)\right] \cdot \pi_{it}\left(s_{i,t+1} \mid s_{it}, d_{it}\right)\right\}, \tag{1.7}$$

where $V_{i,t+1}^*$ is the value function representing the expected present discounted value of lifetime utility from following δ_{it}^*. This expression makes transparent that $\delta_{it}^*(s_{it})$ is a deterministic function of s_{it}, given the primitives.[14]

Dynamic programming fixing future health. Our elicitation approach fixes future health states. Because health is just one element of the state vector, s_{it}, interpretation of respondents' answers within the context of the DP framework requires that the state vector be partitioned into components that are specified or fixed by the researcher in the elicitation task and those that are not specified. In general, interpretation of respondents' answers and of the derived S*ea*TE parameters depends on the relationship (or lack thereof) between the specified and unspecified components of the state vector. We therefore 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 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, the expression for the agent's utility in equation (1.6) becomes

$$U_{it} = \sum_{j=0}^{\mathrm{T}} \beta^j u_{i,t+j}\left[\left(x_{i,t+j}, y_{i,t+j}, \varepsilon_{i,t+j}\right), d_{i,t+j}\right], \tag{1.8}$$

and the related expression for the agent's optimal solution in equation (1.7) becomes

[14] 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.

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

$$\beta\sum_{\left(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1}\right)} V_{i,t+1}^{*}\left[\left(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1}\right),\delta_{i,t+1}^{*}\left(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1}\right)\right]\cdot\pi_{it}^{xy\varepsilon}\left[\left(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1}\right)\mid\left(x_{it},y_{it},\varepsilon_{it}\right),d_{it}\right], \tag{1.9}$$

where summations are implicitly taken as many times as required by the dimension of the state vector.

In our application, we specify the individual's health state so $x_{it}=h_{it}$, while leaving unspecified in y_{it} additional factors typically assumed to affect retirement decision (e.g. family and financial conditions, income, and so on).[15] 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. In particular, we assume that agents place themselves in the hypothetical situation defined by the scenario, without trying to infer why one or the other scenario might be realized (see Dominitz and Manski (1996) and Dominitz (1997)).[16]

On the other hand, y_{it} and ε_{it} are stochastic from the perspective of 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}. Absent uncertainty about factors driving choices in the future, respondents would give either a zero or one response to the elicitation task because labor supply in the future would be a deterministic function of health.

Without loss of generality because y_{it} embodies all omitted factors that could be specified, we maintain that ε_{it} is orthogonal to x_{it} and y_{it}, which implies

$$\pi_{it}^{xy\varepsilon}\left[\left(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1}\right)\mid\left(x_{it},y_{it},\varepsilon_{it}\right),d_{it}\right]=\pi_{it}^{xy}\left[\left(x_{i,t+1},y_{i,t+1}\right)\mid\left(x_{it},y_{it}\right),d_{it}\right]\pi_{it}^{\varepsilon}\left(\varepsilon_{i,t+1}\mid\varepsilon_{it},d_{it}\right).$$

[15] 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.

[16] In an intertemporal context, like the one we are considering, doing so might be easier for a respondent if the scenarios involve relatively short horizons. This design could be investigated in future surveys.

Note that ε here is unknown to both the econometrician and the individual at the time of elicitation.[17] This contrasts to the more typical setting for modeling outcome data where the respondent knows a component that is unobserved to the econometrician. Because x is fixed, equation (1.9) becomes

$$\delta_{it}^*(x_{it},y_{it},\varepsilon_{it}) = \underset{d_{it}\in\{1,0\}}{\arg\max}\ u_{it}\left[(x_{it},y_{it},\varepsilon_{it}),d_{it}\right]+$$
$$\beta\sum_{x_{i,t+1}}\left(\int_{\varepsilon_{i,t+1}}\int_{y_{i,t+1}} V_{i,t+1}^*\left[(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1}),\delta_{i,t+1}^*(x_{i,t+1},y_{i,t+1},\varepsilon_{i,t+1})\right]\cdot\pi_{it}^y(y_{i,t+1}\,|\,x_{i,t+1},y_{it},d_{it})\,dy_{i,t+1}\cdot\pi_{it}^\varepsilon(\varepsilon_{i,t+1}\,|\,\varepsilon_{it},d_{it})\,d\varepsilon_{i,t+1}\right)$$
$$\pi_{it}^x(x_{i,t+1}\,|\,x_{it},y_{it},d_{it}),$$

(1.10)

where we replace summation with integral to allow for the possibility that y and ε are continuous.

The expectation of the optimal decision $\delta_{it}^*(x_{it},y_{it},\varepsilon_{it})$ as of the time of elicitation t-1 is

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

(1.11)

where the expression for $\delta_{it}^*(x_{it},y_{it},\varepsilon_{it})$ is given in (1.10) and $P_{i,t-1}\left[\delta_{it}^*(x_{it},y_{it},\varepsilon_{it})=1\,|\,x_{it}\right]$ in the last line of expression (1.11) 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 (1.11) 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, $\pi_{it}^x(x_{i,t+1}\,|\,x_{it},y_{it},d_{it})$. 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 value of the future state.

[17] 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.

Specifically, we elicit all components of (1.11), as follows:

(i) On the right-hand side of (1.11), the probability of working fixing values of the specified state component, $P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)=1 \mid x_{it}\right]$, 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 (1.11), the unconditional working probability, $P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)=1\right]$.

Clearly, 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, $P_{i,t-1}\left[\delta_{it}^{*}=1 \mid x_{it}\right] \equiv P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)=1 \mid x_{it}\right]$, through the first two channels. This observation is key to interpretation of S*ea*TE, which is given by the difference in the state-contingent subjective probabilities of working across values of the specified state component. The 1-period-ahead S*ea*TE of health *h* on labor supply for individual *i* at time *t* is equal to,

$$\mathrm{S}ea\mathrm{TE}\left(i,t,1\right)=P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=1\right]-P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=0\right]. \tag{1.12}$$

As long as *y* may depend on *h*, in equations (1.10) and (1.11) the agent integrates 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 the S*ea*TE in equation (1.12) is that in this case the measured effect is a *total* effect. That is, it is the effect of health, operating through all of the mechanisms by which health affects labor supply. In our working illustration, it is 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.[18]

Econometric implementation: ex ante and conditional value functions. For econometric implementation it is useful to re-write the choice probabilities in (1.11) in terms of the *ex ante value function* and

[18] We are pursuing this approach in subsequent surveys. Hudomiet, Hurd, and Rohwedder (2020) investigate experimentally whether respondents tend to "fill in" unspecified aspects of the scenario, as hypothesized by Fischhoff, Welch, and Frederick (1999). They are particularly interested in unspecified aspects that, in the respondent's eye, may be related to those specified. They find no evidence of "filling-in"; for instance, when reporting the probability of working past 70 under the specified 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.

conditional value function. The DP literature makes specific additivity and orthogonality assumptions that permit estimation.

Following Arcidiacono and Ellickson (2011),[19] the *ex ante* (or *integrated*) *value function* at a generic future time *t*, $\bar{V}_{it}^{*}\left(x_{it}\right)$, is the continuation value of being in state x_{it} obtained by integrating $V_{it}\left(x_{it}, y_{it}, \varepsilon_{it}\right)$ over y_{it} and ε_{it}, that is,

$$\begin{aligned} \bar{V}_{it}^{*}\left(x_{it}\right) &= \int_{\varepsilon_{it}}\int_{y_{it}} V_{it}^{*}\left[\left(x_{it}, y_{it}, \varepsilon_{it}\right), \delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)\right]\cdot \pi_{i,t-1}^{y}\left(y_{it} \mid x_{it}, y_{i,t-1}, d_{i,t-1}\right) dy_{it}\cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it} \\ &= \int_{\varepsilon_{it}}\int_{y_{it}} \left[\left[u_{it}\left(x_{it}, y_{it}, d_{it}\right) + \varepsilon_{it}\left(d_{it}\right)\right] + \beta\sum_{x_{i,t+1}} \bar{V}_{i,t+1}^{*}\left(x_{i,t+1}\right)\pi_{it}^{x}\left(x_{i,t+1} \mid x_{it}, y_{it}, d_{it}\right)\right] \\ &\quad \cdot \pi_{i,t-1}^{y}\left(y_{it} \mid x_{it}, y_{i,t-1}, d_{i,t-1}\right) dy_{it}\cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it}. \end{aligned} \tag{1.13}$$

This formulation assumes additivity and that the residual component ε_{it} is i.i.d. across agents and time in order to deliver the standard single-crossing result for a discrete choice problem. Note that in Arcidiacono and Ellickson's setting, the econometrician is doing the integration with respect to the distribution of ε, while in ours it is the respondent, who must also carry out the integration with respect to the distribution of *y*. In the simple specification of the next subsection, the presence of *y* (or lack thereof) and the nature of its relationship with *x*, will affect the mapping between the conditional choice probabilities and the underlying value functions and, thus, the derivation of the latter from the former by inversion.

The *conditional value function* $v_{it}\left(x_{it}, y_{it}, d_{it}\right)$ is the present discounted value net of ε_{it} of choosing d_{it} and behaving optimally from period *t+1* onward, that is,

$$v_{it}\left(x_{it}, y_{it}, d_{it}\right) = u_{it}\left(x_{it}, y_{it}, d_{it}\right) + \beta\sum_{x_{i,t+1}} \bar{V}_{i,t+1}^{*}\left(x_{i,t+1}\right)\pi_{it}^{x}\left(x_{i,t+1} \mid x_{it}, y_{it}, d_{it}\right). \tag{1.14}$$

The conditional value function is a key component for forming the conditional choice probabilities that we measure and use as a basis for estimation of the parameters of the simple structural model that we specify below. Specifically, equation (1.10) can be re-written as

$$\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right) = \underset{d_{it}\in\{1,0\}}{\arg\max}\left[v_{it}\left(x_{it}, y_{it}, d_{it}\right) + \varepsilon_{it}\left(d_{it}\right)\right], \tag{1.15}$$

and the conditional choice probabilities in equation (1.11) can be re-written as

[19] Who build on Hotz and Miller (1993).

$$
\begin{aligned}
&P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it},y_{it},\varepsilon_{it}\right)=1\mid x_{it}\right] \\
&=\int_{\varepsilon_{it}}\int_{y_{it}}\delta_{it}^{*}\left(x_{it},y_{it},\varepsilon_{it}\right)\cdot\pi_{i,t-1}^{y}\left(y_{it}\mid x_{it},x_{i,t-1},y_{i,t-1},d_{i,t-1}\right)dy_{it}\cdot\pi^{\varepsilon}\left(\varepsilon_{it}\right)d\varepsilon_{it} \qquad (1.16)\\
&=\int_{\varepsilon_{it}}\int_{y_{it}}\underset{d_{it}\in\{1,0\}}{\arg\max}\left[v_{it}\left(x_{it},y_{it},d_{it}\right)+\varepsilon_{it}\left(d_{it}\right)\right]\cdot\pi_{i,t-1}^{y}\left(y_{it}\mid x_{it},x_{i,t-1},y_{i,t-1},d_{i,t-1}\right)dy_{it}\cdot\pi^{\varepsilon}\left(\varepsilon_{it}\right)d\varepsilon_{it}.
\end{aligned}
$$

Since we measure $P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it},y_{it},\varepsilon_{it}\right)=1\mid x_{it}\right]$ directly, equation (1.16) links our data to the primitives of the model.

Interpreting conditional probability responses in DP framework. We now present a simple implementation to illustrate how our data on health-contingent working probabilities can be used to derive measures of individual-specific, health-contingent valuations of continued work, using standard assumptions underlying econometric implementations of the DP framework. As above, and as common in structural models of labor supply, we treat work and health as binary. In DP models, discretization of the state space is done for tractability, that is, to avoid the curse of dimensionality problem. A parallel argument can be made for elicitation of choice probabilities in discrete scenarios. Moreover, the fact that such contingent reasoning and branching is spontaneously displayed by respondents in cognitive interviews further supports this approach.

First, consider the case where there are no unspecified state variables. As in the earlier discussion of the decision tree in Figure 1, let $\{W,\sim W\}$ be the labels for work and not work and $\{H,L\}$ for high and low health. (Recall that the indicators $d_{it}=1$ corresponds to working and $h_{it}=1$ corresponds to low health.) Because there are only four combinations of health states and labor decisions at time t, it is easy to write out the problem. Define $V_{it}\left(h_{it},d_{it}\right)$ to be the value for individual i of being in state h and making choice d at time t given expectation and optimization from t+1 onward. Then

$$
\begin{aligned}
V_{it}\left(h_{it},d_{it}\right)=(1-h_{it})&\left\{d_{it}\left(v_{it}\left(H,W\right)+\varepsilon_{it}\left(W\right)\right)+(1-d_{it})\left(v_{it}\left(H,\sim W\right)+\varepsilon_{it}\left(\sim W\right)\right)\right\} \\
+h_{it}&\left\{d_{it}\left(v_{it}\left(L,W\right)+\varepsilon_{it}\left(W\right)\right)+(1-d_{it})\left(v_{it}\left(L,\sim W\right)+\varepsilon_{it}\left(\sim W\right)\right)\right\},
\end{aligned} \qquad (1.17)
$$

where the first row refers to actions in high health and the second row in low health.

Given this standard dynamic programming formulation, maximizing $V_{it}\left(h_{it},d_{it}\right)$ yields the standard single crossing conditions for the specified health states as follows.

$$
\begin{aligned}
&\text{When } H, \\
&\quad W \text{ if } 0 \leq v_{it}\left(H,W\right) - v_{it}\left(H,\sim W\right) + \varepsilon_{it}\left(W\right) - \varepsilon_{it}\left(\sim W\right) \\
&\sim W \text{ otherwise.} \\
&\text{When } L, \\
&\quad W \text{ if } 0 \leq v_{it}\left(L,W\right) - v_{it}\left(L,\sim W\right) + \varepsilon_{it}\left(W\right) - \varepsilon_{it}\left(\sim W\right) \\
&\sim W \text{ otherwise.}
\end{aligned}
\tag{1.18}
$$

Again, as traditional in dynamic programming applications, we define objects that are the *differenced conditional value functions* and differenced residual components, where the differencing is across actions conditional on the state. Let

$$
\begin{aligned}
\tilde{v}_{it}^{H} &= v_{it}\left(H,W\right) - v_{it}\left(H,\sim W\right) \\
\tilde{v}_{it}^{L} &= v_{it}\left(L,W\right) - v_{it}\left(L,\sim W\right) \\
\tilde{\varepsilon}_{it} &= \varepsilon_{it}\left(W\right) - \varepsilon_{it}\left(\sim W\right),
\end{aligned}
\tag{1.19}
$$

where again the differencing is across working and not working. Note that because the residual components across decisions are independent of elements of the state vector, then their difference, $\tilde{\varepsilon}_{it}$, is also independent. In terms of these variables, the single cross conditions for working given health state become

$$
\begin{aligned}
\text{when } H,\ \delta_{it}^{*} &= 1 \text{ if } 0 \leq \tilde{v}_{it}^{H} + \tilde{\varepsilon}_{it} \\
&= 0 \text{ otherwise;} \\
\text{when } L,\ \delta_{it}^{*} &= 1 \text{ if } 0 \leq \tilde{v}_{it}^{L} + \tilde{\varepsilon}_{it} \\
&= 0 \text{ otherwise.}
\end{aligned}
\tag{1.20}
$$

We have reverted to the 1/0 notation for working (W) and not working ($\sim W$) to be analogous to discrete choice dynamic programming econometrics.

We now show how the health-contingent working probabilities can measure the differenced conditional value functions with arbitrary heterogeneity across individuals. At the time of elicitation, the survey respondents need to integrate out the residual component. Therefore, to analyze the conditional probabilities, we make a distributional specification for the residual uncertainty $\tilde{\varepsilon}_{it}$. Denote the cumulative distribution function of $\tilde{\varepsilon}_{it}$ as Φ. Since the differenced conditional value function (and the underlying utility functions) are only defined up to scale and location, we can take the Φ to be zero mean and unit variance without loss of generality as in the standard discrete choice model.

The survey's elicitation task maps into the discrete choice problem in (1.20). 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

$$P_{i,t-1}^{H} \equiv P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=H\right] \tag{1.21}$$

and

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

yields

$$P_{i,t-1}^{L} \equiv P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=L\right]. \tag{1.22}$$

Then (1.20) given the distributional assumption and implicitly assuming a symmetric distribution for $\tilde{\varepsilon}_{it}$ implies

$$\begin{aligned} P_{i,t-1}^{H} &= \Phi\left(\tilde{v}_{it}^{H}\right) \\ P_{i,t-1}^{L} &= \Phi\left(\tilde{v}_{it}^{L}\right). \end{aligned} \tag{1.23}$$

We invert these expressions to yield

$$\begin{aligned} \tilde{v}_{it}^{H} &= \Phi^{-1}(P_{i,t-1}^{H}) \\ \tilde{v}_{it}^{L} &= \Phi^{-1}(P_{i,t-1}^{L}). \end{aligned} \tag{1.24}$$

Given a functional form for Φ , the individually-elicited conditional probabilities yield individual-level measures of the conditional value of working versus not working in high and low health. In what follows, we will specify the distribution Φ as normal. The survey only accepts integer (percent) responses. 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 AHMR among others).

Note that this case, where there is no unspecified component of the state vector *y*, covers many cases of interest. 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. If these are all deterministic functions of health for an individual, then the value functions given in (1.24) will completely characterize decision-making. For example, say there were two possible future states: "good" where health and wage are high and "bad" where health and wage are low. Obviously, in this case, one could not distinguish observationally between the effects of health *per se* and wage.

There are cases, however, were one might want to attempt separate measurements based on *y*. As just discussed, for separate effects of *h* and *y* to be identified, the nonspecified state *y* would have to be not perfectly related to h. The appendix appended to the text shows how the conditional probabilities can be interpreted in this case.

IV. 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 this analysis.

B. 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 at the time of Survey 4 and, thus, eligible to answer the labor supply and health expectations battery;[20] (iv) who gave complete and consistent responses to the latter battery; and (v) who reported being in high health in Survey 4.[21] Table A1 in the online appendix summarizes the selection process.

[20] Some of these individuals had already retired from their career job and were working in a bridge job at the time of the survey. These individuals, too, were asked the expectations questions just described with reference to their bridge job.

[21] As fewer than 3% of respondents reported being in low health (fair or poor), we decided to exclude this small group and focus on the majority of respondents who reported being in high health (excellent, very good, or good).

The analysis sample amounts to 970 respondents aged 57 to 81, currently in high health and working. Sample size decreases to 839 respondents for the analysis of expectations with a 4 years horizon, which applies to individuals who reported a positive probability of working in 2 years. See Table A2 in the online appendix for sample statistics.

VRI respondents tend to be wealthier, more educated, and healthier than the general population. However, conditional on the sample screens (age, positive financial wealth, 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). Subsequent sections address the validity of subjective expectations.

C. S*ea*TE and Its Components

In the work-health expectations battery, eligible VRI respondents are first asked for the percent chance out of 100 that they will be working in 2 years. Next, they are asked for their self-rated health on a 5-point scale (Excellent, Very Good, Good, Fair, and Poor) and for the percent chance out of 100 that their health will be some particular state in 2 years. Finally, respondents are asked about their probability of working in the next 2 years, conditional on different health states.[22] These questions were then repeated for the 4 years 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:

[22] Both VRI and HRS respondents were already familiar with the 5-point health scale. For example, HRS respondents are asked to report their health on this scale at every wave. The 5-point health scale has been shown to be reliable in many contexts. Moreover, using this health scale follows a well-established practice in structural literature, (e.g. Blau and Gilleskie (2001, 2008), French (2005), van der Klaauw and Wolpin (2008), French and Jones (2011)), that our analysis is designed to inform. Alternatively, one could ask about work under various diagnoses (e.g. high blood pressure, cancer, in ability to lift, cognitive decline). This could be unpractical, however, especially because different conditions would be relevant for different types of work and because there are competing, multiple health risks. If respondents are heterogeneous in their use of the scale, this would not necessarily be a problem in itself, though it would be a problem if an individual comes to a subjective health assessment based on ability or willingness to work.

```
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]%

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 at 2 and 4 years 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 2-year ahead is 70.5 percent points, somewhat higher than the 65.9 percentage points mean unconditionally in 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. We discuss the S*ea*TE below.

In Figure A2 of Online Appendix A, we show box-and-whisker plots of the health-contingent working probabilities by age of the respondent at the time of the survey. The two top plots refer to the probability of working in high health, whereas the two bottom plots refer to the probability of working in low health. The plots to the left refer to the 2-year horizon, while the plots to the right refer to the 4-year horizon. Age bins 60-61, 63-64, and 65 in the two left plots are of particular interest, as a 2-year horizon from those ages implies the crossing of the early, normal, and full SS retirement ages (i.e., 62, 65, and 67), where actual labor supply displays well-known peaks. There are similar peaks for age ≤ 59, 60-61, 62, and 63-64 with the 4-years horizon. Figure A3 in Online Appendix A displays analogous box and whisker plots for the unconditional working and health probabilities.

In the left plots of Figure A2, the mean and median working expectations at 2 years feature sharp declines among the 60-61 years old (corresponding to the 62 peak), among the 63-64 years old (corresponding to the 65 peak), and among the 65 years old (corresponding to the 67 peak). Notice, however, that the mean and median working expectations do not decrease monotonically across groups of increasing age. This is consistent with increasing selectivity of the working and (high) health requirements applying to older respondents.

Moving to the 4-year ahead horizon on the right, Figure A2 reveals that the age-specific mean and median decrease sharply and steadily from the $\leq$59 and the 63-64 groups and level off (or tend to increase slightly) thereafter, again consistent with increasing selectivity of older sub-groups. The cross-sectional variance of working expectations is now fairly high in all age groups and appears higher than the cross-sectional variance of the 2-year working probabilities. This is consistent with a bigger role of heterogeneity as the forecasting horizon increases.

A comparison of the top and bottom plots by horizon reveals that the effect of a negative health shock on work is negative on average for all age groups. Indeed, the box-and-whisker plots in Figure A2 represent graphically and by age group a basic finding in Table 1: the mean of the empirical distribution of the probability of working in low health is lower than the mean of the empirical distribution of the probability of working in high health.

The effect of a negative health shock contemporaneous to the work-retirement decision has a negative effect on work for the vast majority of respondents. The average and median effects are quite large, respectively equal to -28.5 and -25 percentage points at 2 years and -25.7 and -20 at 4 years. At the same time, the large standard deviations and interquartile ranges (the first close to 28 and the second equal to 50 at both horizons) indicate that the size of the effects vary widely across respondents. Moreover, the fact that the first quartile is equal to 0 percentage points at both 2 and 4 years suggest that for 25 percent of the respondents the S*ea*TE is actually non-negative.

D. Unpacking S*ea*TE

Optimal behavior is 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 shocks 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

Does S*ea*TE vary with observed characteristics? Table 5 reports estimates from a linear regression of 2- and 4-year ahead S*ea*TEs on standard covariates capturing respondents' demographic and socioeconomic characteristics. Except for age, S*ea*TE is little predicted by covariates. Hence, most of the heterogeneity in S*ea*TE is unobserved, at least with respect to commonly available observables. Since we only observe one cross-section and the sample is limited to working respondents in high health, the age-heterogeneity likely arises from selection. The first block of Table 5 reports coefficients of age dummies. (The reference group is age less than 60, but because of the design of the VRI, most of these respondents are 59.) All the coefficients are negative, so (bad) health has a larger (depressing) effect on work for those over age 60. The difference in the effect is non-monotonic in age and only statistically significant at certain ages. Note that age is at time of survey, so the expectations refer to ages 2- or 4-years ahead. Hence, the -0.117 coefficient of the 2-year ahead S*ea*TE at age 62 reflects expectations about retirement at age 64 (or 65 depending on timing of birth). This peak disappears at the 4-year horizon at age 62, but is instead reflected in the constant. The lower S*ea*TE for the oldest groups likely reflects selection on both health and taste for work.

The coefficients of the other covariates are mainly small and statistically insignificant. Since most of the heterogeneity is unobserved, these findings imply that including commonly available covariates in observational studies will not be sufficient to control for the effects of heterogeneity.

F. Validation: Law of Total Probability

Since we ask the probability of work given health, the probability of health, and the unconditional probability of work, we are able to evaluate how well survey respondents obey the law of total probability (LTP) in their responses. The respondents are quite good at applying the LTP.

Figure 2 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, the deviations are relatively small. Therefore, the vast majority of respondents appear to understand the logic of probabilities.

The deviations from consistency of the LTP is most pronounced for respondents with self-reported unconditional working probabilities equal to 0, 50, or 100 percent. This finding is consistent with the suggestion in the literature that some respondents who give corner or 50/50 responses may be more uncertain and/or less good at probabilistic thinking (e.g. Fischhoff and Bruine de Bruin (1999), Hudomiet and Willis (2013)). Note, however, that there are significant mass points of respondents. Though not readily apparent because of the mass points at the corners of the figure, most of these respondents are getting the LTP exactly right.[23]

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 have difficulty thinking through disjunctions via contingent thinking. Much of the same research,

[23] See also Giustinelli, Manski, and Molinari (2021), who show that zero probability responses are highly informative.

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 explicating pose the contingencies.

Even so, a potential concern with our approach is that respondents may find difficult to predict outcomes in unfamiliar and/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 think 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 and/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.[24] 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 (2020) 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 (2021) 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 respondents" in Giustinelli, Manski, and Molinari (2021)'s taxonomy) and, to a lesser extent, by a lower

[24] See Giustinelli, Manski, and Molinari (2020) 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.

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. Health and Retirement Study Results

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 samples 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.[25]

Tables B3-B6 in Online Appendix B report HRS results parallel to the VRI results in Tables 1-4. On average, HRS respondents have higher probabilities of working than VRI respondents at both horizons as well as 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.

[25] The HRS analysis in this section is limited to responses that parallel the VRI sample. In work in progress with the HRS with Gabor Kezdi, we are doing a more detailed analysis that addresses response difference that are present in the HRS, but not the VRI. These include the HRS baseline includes nontrivial number of respondents in low health at baseline.

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.

V. Estimates using Discrete Choice Dynamic Programming Framework

A. DP: Eliciting Utility

We now proceed with the analysis derived from the dynamic programming specification of Section III. The elicited health-contingent probabilities yield individual-level values of working versus not working given specified health.

Table 6 shows summary statistics for these values, $\tilde{v}^H$ and $\tilde{v}^L$, obtained by inversion of the 2-year and 4-year ahead health-contingent probabilities according to equation (1.24).[26] The results in this section are qualitatively equivalent to those for the S*ea*TE in Section IV because the DP-implied values are nonlinear transformations of the health-contingent working probabilities. As expected, the mean value in high health is substantially greater than that in low health reflecting the lower value of working in low health. The conditional probability of working is reflected in the last row of the table showing the fraction who expect to work in the specified health state. For the 4-year ahead horizon, there is a substantial shift down in the willingness to work in both the high-health and low-health states.

Figure 3A shows a scatter plot of the values $\tilde{v}^H$ and $\tilde{v}^L$ for the 2-year horizon. Figure 3B shows an analogous plot for the 4-year horizon. These plots illustrate many features of the health-contingent value

[26] Note that this application of the DP model to the 4-year horizon responses analyzes them as if they were a single decision over that horizon. Given the structure of the survey where the 4-year horizon responses are not conditioned on the 2-year horizon state or decisions, this simplification is necessary. A future survey could have a multi-stage structure.

of continued work across respondents. In each of the two figures, the upper right quadrant contains the individuals who value work more than not work in both health states (where of course value is net of the residual uncertainty that will be realized at the time of the decision). The lower left quadrant has those who value work less in both states. The vast majority of individuals lie below the 45-degree line corresponding to having a lower value of work relative to not work when in low health than in high health. It is not surprising that values shift in this direction. Lower health likely decreases taste for work and the return to work. Yet, shifting in the other direction is perfectly consistent with optimization. For those above the 45-degree line, the relative attractiveness of work increases in low health. This valuation could result from need for insurance, lower value of leisure in low health, or need for income in low health. Indeed, there are a few observations in the upper left quadrant where the value of working is higher in low health than in high health. The opposite—in the lower right quadrant—is not surprisingly much more common. These represent the individuals who would quit working after a negative health shock.

There is a strong correlation between the value of work across the health states. A simple framework of summarizing is that there is a value of work in high health that is positively, but imperfectly correlated with that in low health. Consider the model of heterogeneity in taste

$$\begin{aligned}\tilde{v}_i^H &= \alpha^H + \upsilon_i^H \\ \tilde{v}_i^L &= \alpha^L + \gamma\upsilon_i^H + \upsilon_i^L,\end{aligned} \tag{1.25}$$

where α^H and α^L are the mean across individuals of the values and υ_i^H is the mean-zero heterogeneity across individuals in the value of work in the high health state. The heterogeneity in the value of work in the low health state has two components: a component correlated with that in high health, $\gamma\upsilon_i^H$, and an orthogonal component, υ_i^L. Again, from the point of view of the respondent, these components are nonstochastic. Our procedure gives a direct measurement of the LHS of equation (1.25). The orthogonal decomposition is a convenient way to summarize the observed heterogeneity.

Table 7 presents estimates of the parameters specified in equation (1.25).[27] The estimates obtained are sensible. Consider first the estimates for the 2-year ahead horizon.

- The mean utility from work shifts substantially downward when health changes from high to low. The mean is 0.97 in high health and -1.04 in low health.
- The correlation within individual of the willingness to work across health states is fairly high, but far from unity. The coefficient γ that controls this correlation is 0.71. Hence, there is persistence

[27] These estimates are from an OLS regression where the first equation just has a constant and the second equation has a constant plus the residual from the first equation. Note that this is a random coefficient model, though we do not have to specify a distribution since the values are observable.

within individuals of valuation of work across health states, implying that those with high value of work in high health carry that over into low health, but in a damped way.

For the 4-year horizon relative to the 2-year one, there is a substantial shift down in the willingness to work in the high-health state—from 0.97 to 0.48. In contrast, no shift in the willingness to work in the low-health state (-1.04 for both horizons). The estimate of the coefficient γ, controlling the correlation within individual of the willingness to work across health states, decreases slightly to 0.66.

Online Appendix Table A3 shows the same regression with the usual set of covariates. As expected, willingness to work longer declines with age. Controlling for covariates does make the coefficient γ roughly the same across the 2- and 4-year horizons. Especially for working in high health at the 2-year horizon, there is a significant, negative wealth gradient for working longer and also significant effects of spousal labor supply.

B. DP: Simulation

We can use the empirical results from the dynamic programming framework to illustrate the benefit of having data on the heterogeneity of values. Consider the estimate of a linear regression model using data on working (d_i) and health (h_i) realizations,

$$d_i = b_0 + b_1 h_i + e_i. \tag{1.26}$$

As discussed in Section III within the POF, the least squares estimate of b_1 will be an unbiased estimate of the ATE of health on work only when there is no selection or heterogeneity term in equation (1.4). Within our simple implementation of the DP framework, this requirement will fail when realized health, h_i, is correlated with the values $\tilde{v}_i^H$ or $\tilde{v}_i^L$ across individuals. Our elicitation approach is designed to render this heterogeneity observable *ex ante.*

To demonstrate how biased estimates of causal effects can emerge in data on realizations, we use our framework to construct simulated realizations of work decisions and health states. Using the DP model of Section III, equation (1.20) implies that the realized decision to work is

$$d_i = (1-h_i)\mathbf{I}[\tilde{v}_i^H + \tilde{\varepsilon}_i] + h_i\mathbf{I}[\tilde{v}_i^L + \tilde{\varepsilon}_i], \tag{1.27}$$

where $\mathbf{I}[.]$ is the indicator function, equal to 1 if the argument is positive and zero otherwise, and $d_i = 1$ if work and 0 otherwise. To simulate realizations that reflect the this heterogeneity, we use the measured health-contingent values of work ($\tilde{v}_i^H$ or $\tilde{v}_i^L$), simulated realizations of health (h_i), and the residual component ($\tilde{\varepsilon}_i$) to calculate simulated decisions according to equation (1.27). Health is modeled as {0,1}, so it is simulated using Bernoulli draws based on the health transition probability, π_i^h. As in the implementation of the DP formulation, $\tilde{\varepsilon}_i$ is simulated as standard normal.

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 individual-specific probabilities, so health transitions have the *empirical correlation* with the value of work.
3. π_i^h adjusts the individual-specific probabilities to induce a *higher correlation* between health and the value of work than is present in the VRI data.[28]

The first case implies no selection or heterogeneity, so the OLS estimate of (1.26) will yield an unbiased estimate of the average treatment effect equal to the average S*ea*TE. The second case will illustrate the extent of the bias that would be present in the VRI data. The third case magnifies the bias.

Table 8 shows estimates of the regression for simulated realization for the 2- and 4-year horizons simulated over 1000 replications for the three cases. 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 being 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 that much heterogeneity in health (most are quite healthy), so the magnitudes of the biases are fairly small. 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.

Finally, recall that Tables 7 and A3 show that there is substantial heterogeneity in the health-contingent values of work even after conditioning on standard covariates. Therefore, conditioning on such covariates in econometric applications using data on realized decisions and states, though helpful, is not likely to eliminate bias from unobserved heterogeneity.

---

[28] 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.)

VI. Relating Probabilities to Realizations

Relating state-contingent probabilities to realizations is tricky. The motivation of the approach advanced in this paper—that is, eliciting *ex ante* probabilities about state-contingent outcomes—is that in realization data only the outcome under the realized state is observed. Obviously, what state gets realized *ex post* is not randomly assigned, which leads to the inferential biases documented in the previous section. 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 counterfactual outcomes. Therefore, it is logically impossible to validate fully state-contingent probabilities and hence the S*ea*TE. For example, if our application were asking about a state that never occurs, say, a policy change that is only contemplated, there would be no scope for comparing expectations to realizations. Even if examining the predictive power of health contingent-probabilities in realization data does not produce structural estimates or constitute a test of rational expectations, it remains interesting to understand how well health-contingent expectations predict work given the realized state of health. This assessment of the information content of the contingent probabilities is important for validation their usefulness. We find significant predictive power for work outcomes at the 2-year horizon.

The panel structure of the VRI allows assessment of the predictive power of our measures. The expectations data in this paper 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 A4).

To consider how the probabilities elicited in Survey 4 predict outcomes observed in Survey 6, we compare predictions based on state-contingent and unconditional probabilities. First, we assess the predictive power of health-contingent probability given the knowledge of what health state is realized. To belabor the point made in the first paragraph of this section, 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 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 state-contingent probability measure relative to the unconditional measure, we run the prediction horse race between the two measures.

Table 9 shows these results. The dependent variable is the work outcome (1 if work, 0 if not). The predictor in column (1) is the health-contingent probability of working for the realized health state.[29] The predictor in column (2) is the unconditional probability of working.[30] The coefficient of the health-contingent probability is 0.590, precisely estimated. For the reasons just discussed, no violation of rational expectations should be inferred from the finding that the coefficient is significantly different from one. Nonetheless, there is significant predictive power based on combining the state-contingent probability of work with the realized health state. The R^2 is 21.6 percent. There are many factors other than health that affect retirement, so the R^2 can, in theory, be any value between 0 and 1. The estimates in column (2) are similar, with the R^2 being 1 percentage point lower. There are not that many bad health realizations, so the predictive power of the contingent probability is marginal.

To evaluate the incremental predictive power more directly, column (3) runs the horse race between the two probability measures. Interestingly, the health-contingent probability dominates the prediction. Health-contingent probability has a coefficient of 0.502 and the unconditional has a coefficient of 0.093. Indeed, we can test for incremental predictive power of the health-contingent versus unconditional probabilities. The unconditional probability has no incremental predictive power (the R^2 of columns (1) and (3) are the same). We formalize this inference with the Wald test, which has a value of zero and therefore a p-value of 1. Conversely, we can overwhelmingly reject the hypothesis that the health-contingent probabilities have no incremental predictive power (Wald test of 6.70 with a p-value of 0.01).

In summary, the contingent probabilities are strongly predictive of outcomes and, in a horse race, out-perform unconditional probabilities. Hence, the panel evidence shows strongly supports that the elicited probabilities have relevant informational content.

We also show results by initial age and health. The respondents were 57 to 81 when asked the contingent probability questions and therefore roughly between 59 and 83 at the point 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 still working, there may be different pattern with respect to age. Table 10 therefore interacts the predictors with age using the same specifications as

[29] The regressor uses the directly-elicited response to the contingent-probability question. Recall that there are three contingent-health states for which we elicit contingent probabilities of working (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 A5 in the Online Appendix shows tabulations of the responses by this full grid.

[30] 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 redid 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 2).

in Table 9. (The specifications also include age dummies.) The results are very similar to Table 9. Indeed, for all but the oldest two age categories, the coefficient on the health-contingent work probabilities move up. Hence, looking within age groups strengthens the results as one would expect. Of course, because of smaller number of observations per age group, the standard errors increase. Again, the health-contingent probability does better in the horse race with the unconditional probability, though the Wald test is not quite as decisive.

There is also the possibility that results vary by initial level of health. Table 11 shows results interacting with level of health (excellent, very good, and good) as of the elicitation of the state-contingent probabilities. (The regression includes age and initial health dummies. We do not interact the predictors with age to avoid an unclear proliferation of coefficients.) The results are very similar to those shown in Table 9 without the interactions.

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 A7, A8, and A9, analogues to Tables 9, 10, and 11. Compared to the 2-year ahead results, there are similarities and 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 a number of reasons: the horizon of the 4-year ahead survey is less-well aligned and, more importantly, the COVID 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.[31]

[31] 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 A4). There are a number of complications with the 4-year ahead battery, which make findings based on it more tenuous. Nonetheless, we include them both 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 ahead battery was not asked of respondents who reported zero probability of work 2-years ahead, so we do not have contingent expectations for these respondents relevant for Survey 7 outcomes even if they responded to Survey 7. Second, roughly 4-1/2 years intervened between Survey 4 and Survey 7, which could lead to under-prediction 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 of 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, which further complicates the analysis. We therefore include the 4-year ahead results for completeness, but give them less weight because of these compounding measurement problems. Finally, the health measurement is aligned with the July 2020, post-COVID work measurement. A number of individuals said their work was affected by COVID, but mainly not for reasons of own health.

VII. 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. For each person the effect is given by the difference between the individual's own estimate of the probability of working in low health versus the probability of working in high health at specified horizons. This Subjective *ex ante* Treatment Effect (S*ea*TE) gives an individual-level measurement of the treatment effect *ex ante*. Under rational expectations, absent aggregate shocks, and in absence of measurement error in survey reports of expectations (or if errors have the same cross-sectional mean across the two values of the treatment variable being compared within the S*ea*TE), the cross-sectional average of S*ea*TE gives an unbiased estimate of the standard average treatment effect (ATE). We give a formal interpretation of S*ea*TE and its components, the state-contingent choice probabilities, in the two workhorse frameworks of econometric causality: the potential outcomes framework (POF) and discrete choice dynamic programming (DP).

We document that the effect of bad health on work is highly heterogeneous across older working individuals in the Vanguard Research Initiative (VRI). The majority of respondents have negative S*ea*TE. Within this majority, there is substantial variability in the effect of health on work. Others have zero effects of health on work, some because they would always work and some because they would never work regardless of health. A very few individuals have a positive effect of bad health on work. The HRS shows similar level and heterogeneity in the S*ea*TE.

We map the conditional probabilities into a discrete choice DP formulation. The DP formulation yields empirical measures of *ex ante* values of working versus not working that are health contingent and individual specific. The DP framework yields an estimate of the individual-specific health-contingent value of work for a standard discrete choice formulation of the labor supply decision. There is a strong correlation within individuals of the value of work versus no work across health states that carries implications for interpreting the causal effect of health on work. There is negative correlation between the probability of receiving a bad health shock and the value of working, so estimates of causal effect of health on work in outcomes data will be biased.

To illustrate this possibility, we simulate realizations of health and work using our DP framework. The simulations yield the correct causal estimate of health on retirement when the heterogeneity that would normally be unobserved is taken into account. They also show that outcomes-based estimates of the ATE will be negatively biased absent accounting for correlated heterogeneity in the value of work. The bias arises because those who get negative health shocks have on average lower value of work.

We provide supporting evidence of the validity of our approach by showing that respondents are internally consistent in that their unconditional work probabilities are consistent with their health-contingent working probabilities and the law of total probability.

Finally, we use the panel structure of the VRI to compare work outcomes under realized health-states with health-contingent predictions. At 2-year ahead, we find that health-contingent probabilities explain a significant fraction of cross-individual work outcomes contingent on health realizations.

The methodology in this paper gives estimates of potential outcomes and, hence, it could be applied in a wide range of applications beyond health and retirement. More generally, the approach can be useful when treatments are difficult to manipulate experimentally or control-for econometrically, including the particularly interesting case of policies that have not yet been and might never be implemented.

Appendix

Dynamic Programming Interpretation with Correlated and Stochastic Unspecified States

We consider the interpretation of the conditional probabilities using the DP framework when there is an unspecified state y that is correlated with health. This case is distinct from the residual uncertainty ε that is additive in the value function and orthogonal to health. Extending the case in Section IIIB, 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

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

Then equation (1.23) becomes

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

where

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

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 above collapses to (1.23).

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

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Brent Meyer, and Nicholas Parker. 2020. "Surveying Business Uncertainty." Forthcoming in the *Journal of Econometrics*.

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. 2020. "Dynamic Decisions under Subjective Expectations: A structural analysis." Forthcoming in the *Journal of Econometrics*.

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.

Armantier, Olivier, Wandi Bruine de Bruin, Simon Potter, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar. 2013. "Measuring Inflation Expectations." *Annual Review of Economics* 5: 273-301.

Arrondel, Luc, Hector Calvo-Pardo, Chryssi Giannitsarou, and Michael Haliassos. 2020. Informative Social Interactions. Working Paper.

Athey, Susan and Guido W. Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *PNAS* 113(27): 7353-7360.

Athey, Susan and Guido W. Imbens. 2019. "Machine Learning Methods Economists Should Know." *Annual Reviews of Economics* 11: 685-725.

Attanasio, Orazio 2009. "Expectations and Perceptions in Developing Countries: Their Measurement and Their Use." *American Economic Review* 99(2): 87-92.

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. 2016. The Dynamic Effects of Health on the Employment of Older Workers. MRRC Working Paper 348.

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, W. and K.C. 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.

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

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2020. Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments. NBER Working Paper 24678.

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.

Cox, David R. 1958. *The Planning of Experiments.* Wiley.

Davis, Jonathan M.V. and Sara B. Heller. 2017. "Using Causal Forests to Predict Treatment Heterogeneity: An application to summer jobs." *American Economic Review: Papers & Proceedings* 107(5): 546-550.

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

Delavande, Adeline. 2014. "Probabilistic Expectations in Developing Countries." *Annual Review of Economics* 6: 1-20.

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.

Eckstein, Zvi and Kenneth I. Wolpin. 1989. "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey." *Journal of Human Resources* 24(4): 562-598.

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.

Fisher, Ronald A. 1935. *Design of Experiments.* Oliver and Boyd.

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.

Fuster, Andreas, Ricardo Perez-Truglia, Mirko Wiederholt, and Basit Zafar. 2020. "Expectations with Endogenous Information Acquisition: An Experimental Investigation." Forthcoming in the *Review of Economics and Statistics*.

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 and Charles F. Manski. 2018. "Survey Measures of Family Decision Processes for Econometric Analysis of Schooling Decisions." *Economic Inquiry* 56(1): 81-99.

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

Giustinelli, Pamela, Charles F. Manski, and Francesca Molinari. 2021. "Precise or Imprecise Probabilities? Evidence from survey response related to late-onset dementia." Forthcoming in the *Journal of the European Economic Association*.

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.

Hamermesh, Daniel S. 1985. "Expectations, Life Expectancy, and Economic Behavior." *Quarterly Journal of Economics*. 100(2): 389-408.

Heckman, James J. and Edward J. Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *PNAS* 96: 4730-4734.

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. 2020. 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.

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.

Keane, Michael and Kenneth I. Wolpin. 2009. "Empirical Applications of Discrete Choice Dynamic Programming Models." *Review of Economic Dynamics* 12: 1-22.

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. 1993. "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" In *Studies of Supply and Demand in Higher Education*, eds. Clotfelter, C.T. and M. Rothschild pp. 43-60, University of Chicago Press.

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.

Manski, Charles F. 2018. "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise." *NBER Macroeconomics Annual*, Vol. 32, No. 1, 411-471.

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.

Neyman, Jerzy. 1923. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science* 5(4): 465-480.

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.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty* 19(1-3): 171-197.

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

Roy, Andrew D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3(2): 135-146.

Rubin, Donald. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688-701.

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

Rust, John. 1994. "Structural Estimation of Markov Decision Processes." In *Handbook of Econometrics* Volume IV, Chapter 51, pp. 3081-3143, Elsevier B.V.

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.

Schotter, Andrew and Isabel Trevino. 2014. "Belief Elicitation in the Laboratory." *Annual Review of Economics* 6: 103-128.

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.

Wager, Stefan and Susan Athey. 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association* 113(523): 1228-1242.

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.

Wolpin, Kenneth I. 1999. "Commentary on 'Analysis of Choice Expectations in Incomplete Scenarios'.'' *Journal of Risk and Uncertainty* 19(1-3): 67-69.

Figure 1. Treatments and Outcomes on a Simple Health-Work Decision Tree

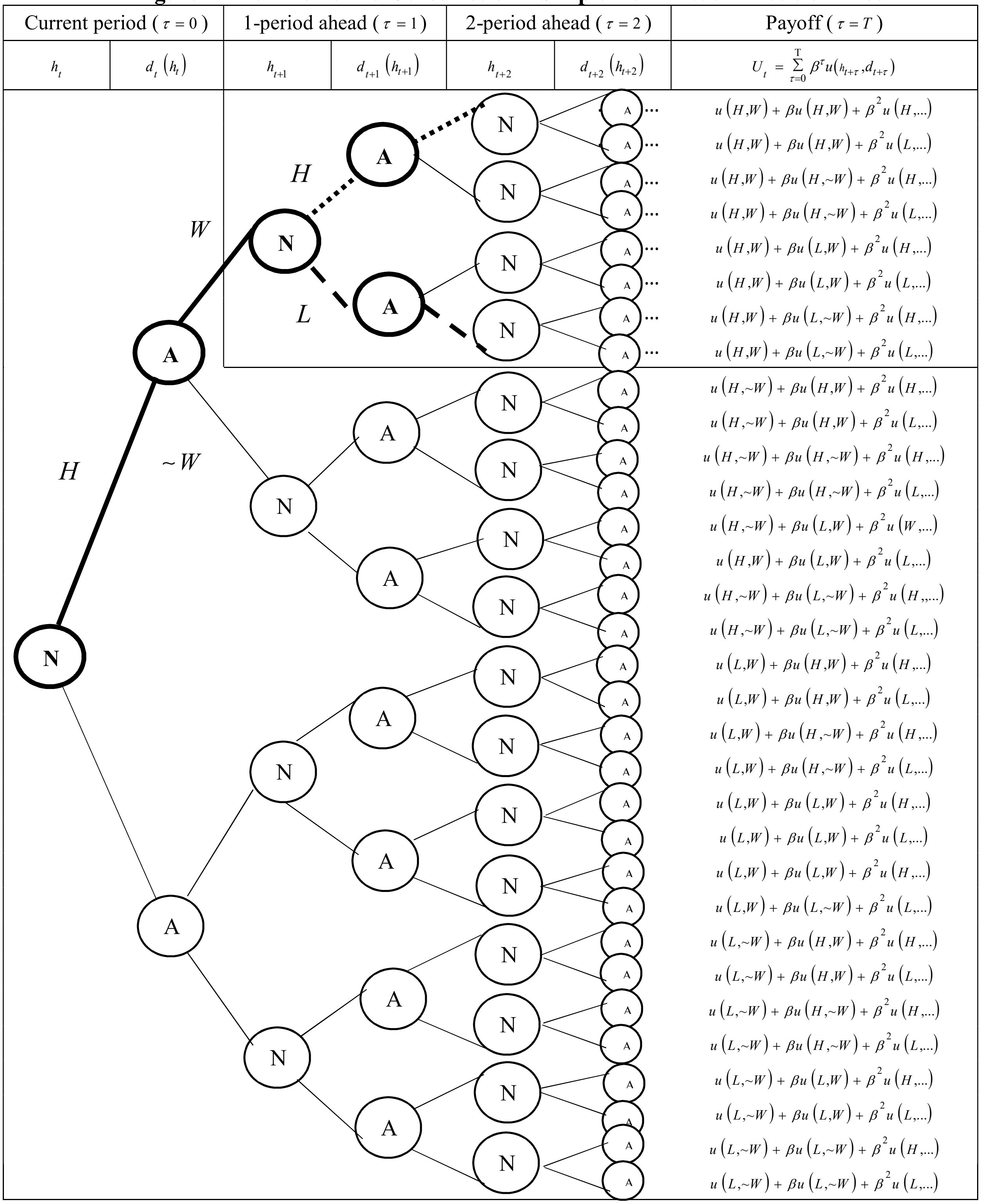
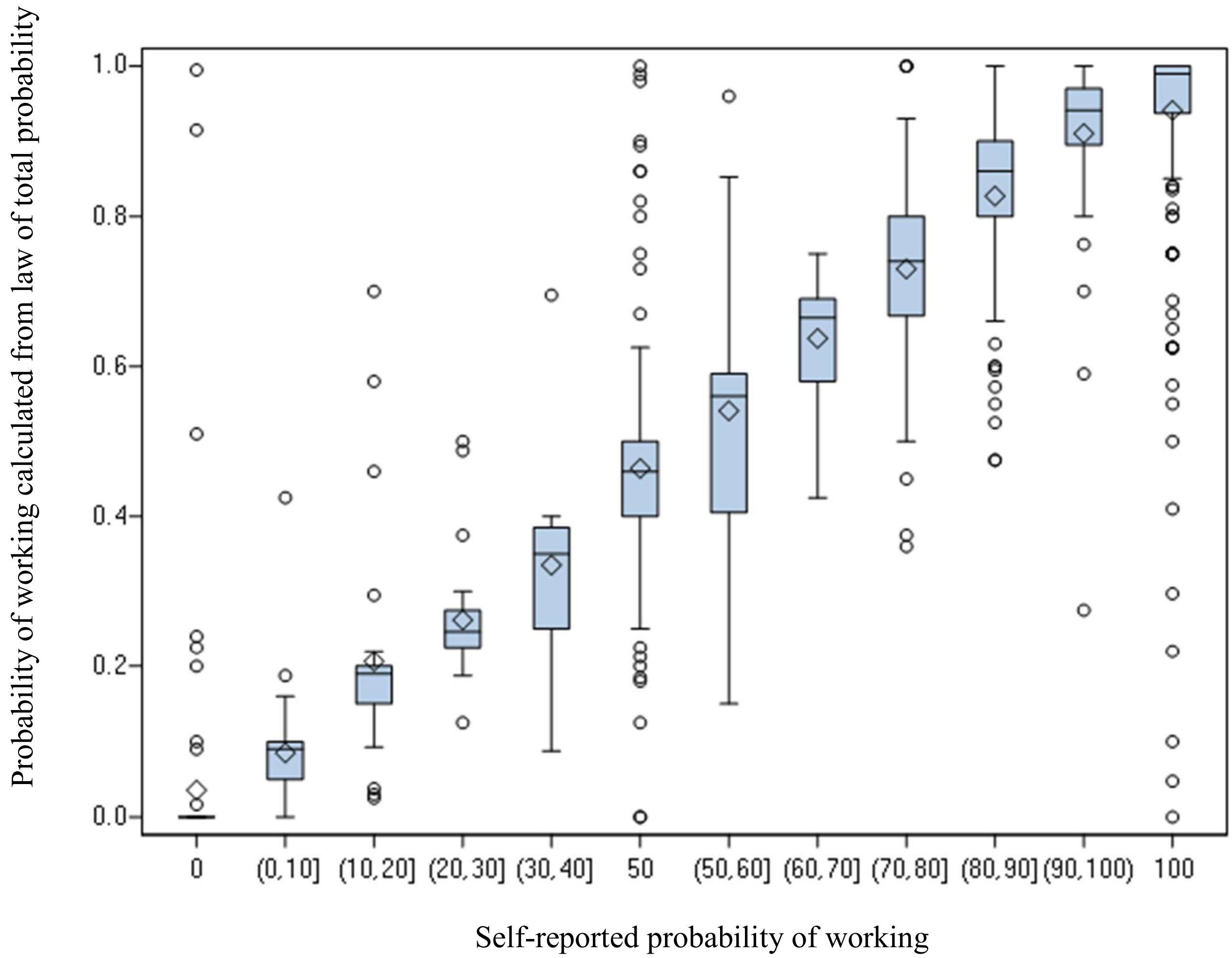


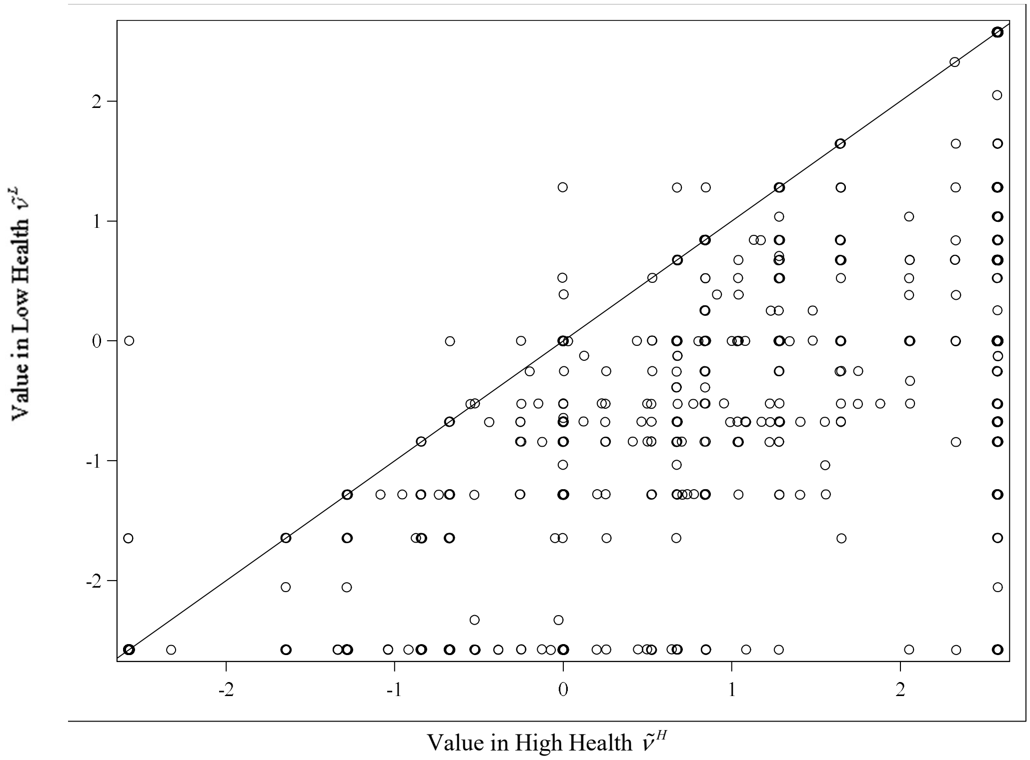
Figure 2. Are Respondents' Answers Consistent with the Law of Total Probability?



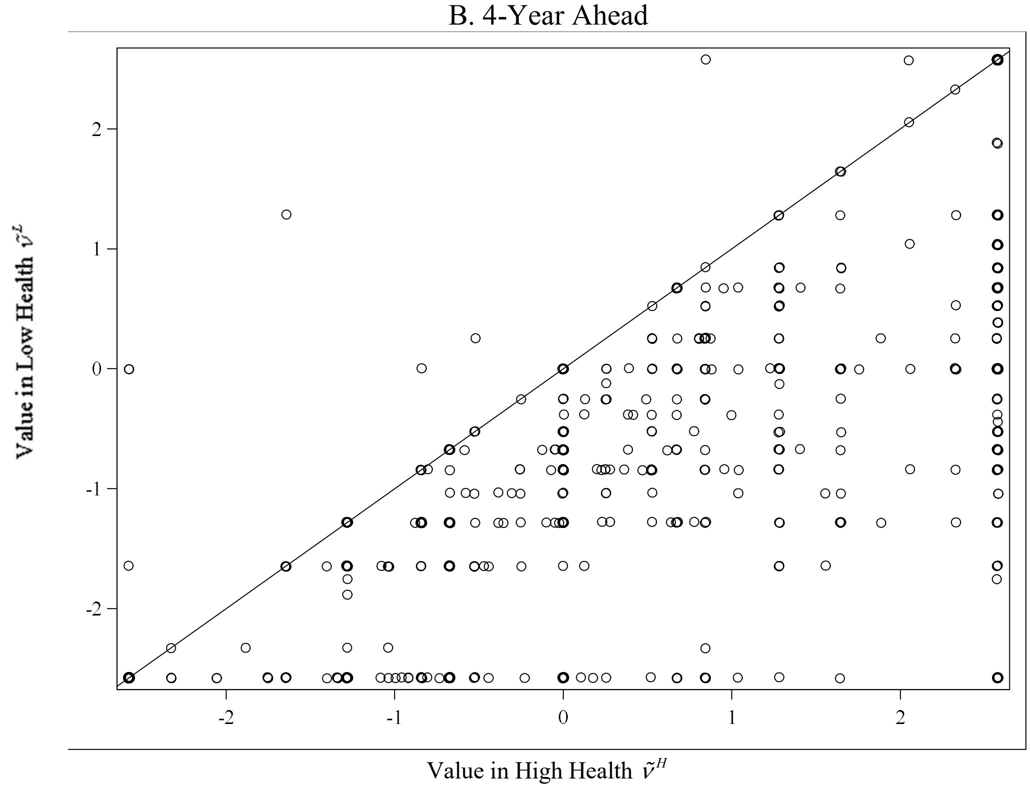
Note: Figure shows the distribution of responses for the unconditional probability of working in 2 years computed by combining the health-contingent working probabilities and health probabilities using the law of total probability (on the vertical axis) versus the self-reported reported unconditional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

Figure 3. Measured Health-Contingent Values of Working vs. Not Working

A. 2-Year Ahead



B. 4-Year Ahead



Note: Figures show the scatter plots of the differenced health-contingent values of work vs. no work, $\tilde{v}^H$ and $\tilde{v}^L$, at 2- and 4-year horizons.

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	27.9
Q25	40	5	5	50	-50
Median	80	10	40	90	-25
Q75	97.5	25	75	100	0
			4-Year Ahead		
Mean	52.7	23.5	33.0	58.7	-25.7
Std. Dev.	37	19.5	34.4	39	27.6
Q25	17	10	0	20	-50
Median	50	20	20	68	-20
Q75	90	30	50	100	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. Predictors of 2- and 4-Year Ahead S*ea*TE

	2-Year Ahead S*ea*TE	4-Year Ahead S*ea*TE
Constant	-0.150	-0.116
	(0.061)	(0.065)
Age (≤ 59 excluded)		
60-61	-0.046	-0.038
	(0.031)	(0.032)
62	-0.117	-0.055
	(0.040)	(0.041)
63-64	-0.034	-0.034
	(0.031)	(0.033)
65	-0.031	-0.111
	(0.045)	(0.051)
66-67	-0.021	0.028
	(0.037)	(0.039)
68-69	-0.120	-0.081
	(0.037)	(0.040)
70-71	-0.116	-0.088
	(0.046)	(0.049)
≥ 72	-0.088	-0.086
	(0.034)	(0.037)
Gender		
Female	0.001	-0.012
	(0.021)	(0.023)
Education		
Some college	-0.002	-0.025
	(0.044)	(0.047)
College grad	0.006	-0.010
	(0.042)	(0.044)
Other adv. degree	-0.042	-0.019
	(0.045)	(0.047)
MBA	-0.014	0.003
	(0.051)	(0.054)
JD, PhD, MD	-0.031	-0.076
	(0.050)	(0.053)
Occupation		
Operative	0.008	-0.008
	(0.025)	(0.027)
Other services	-0.020	-0.020
	(0.032)	(0.034)
Job type		
Bridge	0.008	-0.015
	(0.022)	(0.023)

Marital status		
Partnered	-0.012	-0.010
	(0.024)	(0.026)
Spouse's work status		
Working	-0.014	-0.003
	(0.023)	(0.025)
Total HH wealth		
First quintile	-0.039	-0.039
	(0.033)	(0.036)
Second quintile	-0.045	-0.084
	(0.032)	(0.034)
Third quintile	-0.020	-0.032
	(0.030)	(0.032)
Fourth quintile	-0.044	-0.044
	(0.029)	(0.031)
Replacement rate		
First quintile	-0.013	-0.022
	(0.031)	(0.033)
Second quintile	0.002	0.028
	(0.031)	(0.033)
Third quintile	-0.023	-0.023
	(0.030)	(0.032)
Fourth quintile	-0.018	-0.018
	(0.030)	(0.032)
Current salary		
First quintile	-0.039	-0.032
	(0.037)	(0.039)
Second quintile	-0.067	-0.040
	(0.034)	(0.036)
Third quintile	-0.0001	0.008
	(0.032)	(0.034)
Fourth quintile	-0.005	-0.003
	(0.030)	(0.031)
Observations	970	839
R^2	0.0484	0.0528

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.

Table 6. 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	1.80	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
Fraction positive	67.32	31.03	53.04	23.36
Observations	970	970	839	839

Note: Table shows distribution of the measured differenced health-contingent values of continued work, $\tilde{v}^H$ and $\tilde{v}^L$. See equation (1.24) of the text. When inverting the probabilities reported as 0 and 1, we recode to 0.005 and 0.995.

Table 7. Quantifying Cross-Sectional Heterogeneity in Health-Contingent Values of Work

	2-Year Ahead		4-Year Ahead	
Health state	*h=H*	*h=L*	*h=H*	*h=L*
α^h	0.97	-1.04	0.48	-1.04
	(0.05)	(0.04)	(0.06)	(0.04)
γ		0.71		0.66
		(0.02)		(0.02)
$\sigma(\upsilon^h)$	1.70	1.12	1.80	1.06
Observations	970	970	839	839

Note: Table shows mean, covariance, and variability of the measured differenced health-contingent values of continued work, $\tilde{v}^H$ and $\tilde{v}^L$, as specified in equation (1.25) of the text.

Table 8. 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)	(0.019)	(0.014)	(0.018)
Health *h*	-0.282	-0.305	-0.415	-0.254	-0.286	-0.371
	(0.040)	(0.039)	(0.039)	(0.040)	(0.035)	(0.039)
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.

Table 9. 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-year 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 10. 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	0.860		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 9.

Table 11. 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)	(0.772)	(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 9.