NBER WORKING PAPER SERIES

*SEA*TE: SUBJECTIVE *EX ANTE* TREATMENT EFFECT OF HEALTH ON RETIREMENT

Pamela Giustinelli
Matthew D. Shapiro

Working Paper 26087
http://www.nber.org/papers/w26087

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2019

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 investigator 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 Survey 4 of the VRI 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

*Sea*TE: Subjective *ex ante* Treatment Effect of Health on Retirement
Pamela Giustinelli and Matthew D. Shapiro
NBER Working Paper No. 26087
July 2019
JEL No. C21,C83,D84,J26

ABSTRACT

The Subjective *ex ante* Treatment Effect is the difference between the probabilities of an outcome conditional on a treatment. The *Sea*TE yields *ex ante* causal effects at the individual level. The paper gives an interpretation in two workhorse econometric frameworks: potential outcomes and dynamic programming. It finds large effect heterogeneity of health on work in two surveys of older workers, the VRI and the HRS. It shows how reduced-form estimates of health on work are biased when there is unobserved heterogeneity in taste for work. Using the VRI's panel structure, it validates the elicited conditional probabilities of work given health.

Pamela Giustinelli
Department of Economics
Bocconi University
Milan, Italy
pamela.giustinelli@gmail.com

Matthew D. Shapiro
Department of Economics
University of Michigan
611 Tappan St
Ann Arbor, MI 48109-1220
and NBER
shapiro@umich.edu

A data appendix is available at http://www.nber.org/data-appendix/w26087

"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."* (57 years old working man explaining his answer to a question asking the percent chance that he will be working past age 62.)

I. Introduction

How does a treatment, such as poor health, affect a behavior, such as retirement? In behavioral data, by which we mean realized outcomes or decisions of economic agents, the econometrician observes only the behavior conditional on the treatment, but not the counterfactual behavior absent the treatment. When there is unobserved heterogeneity across individuals, this inherent feature of behavioral data makes difficult inferences about causal effects. Rather than relying on behavioral data, the paper uses subjective expectations of a behavior or outcome conditional on values of the treatment to elicit *ex ante* treatment effects. This strategy allows each individual to be both treatment and control, thereby obviating the unobserved counterfactual problem and allowing for completely unrestricted heterogeneity across individuals. This approach yields an individual-level estimate of the treatment effect. Thus, this paper provides a strategy for quantifying person-specific treatment effects and for characterizing the distribution of causal effects across the population.

We implement our approach by asking older workers participating in the Vanguard Research Initiative (VRI) to report the conditional probability on a 0-100 percent chance scale that they will be working to specified horizons under alternative scenarios about their future health. Using these data, we generate individual and aggregate level estimates of the *Subjective ex ante Treatment Effect*, or S*ea*TE, of health on retirement, given by the difference between respondents' probabilities of working in low versus high health.

Using the panel structure of the VRI, we show that elicited conditional probabilities strongly predict realized work given realized health two-years ahead. Hence, behavioral outcomes support the validity of the conditional probabilities. Additionally, we elicit the same conditional probability measures in the Health and Retirement Study (HRS) to apply the approach to a population-representative sample.

The Subjective *ex ante* Treatment Effect (S*ea*TE) of health on work is highly heterogeneous across individuals. S*ea*TE is zero (no effect of health on work) for almost 30 percent of 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 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 health on work is large.

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 behavioral 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 to estimate causal effects. Relative to the existing literature, our method shows enormous heterogeneity in the treatment effect of health on work. We also document that the individual-level treatment effect is correlated with baseline willingness to work.

In Section III, we interpret S*ea*TE using two mainstream frameworks of econometric causality: the potential outcomes framework (POF) and dynamic programming (DP). Starting from the conditional working probabilities and employing standard econometric assumptions for dynamic discrete choice models, we further derive expressions for the values of continued work defined in the DP framework. Like the conditional 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 conditional probabilities of working in high and low health, and the S*ea*TE.

In Section V, we present the measures of continued work we derived in Section III and we estimate a simple econometric model of health and retirement based on these measures. The model implies 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. We use a simulation of the simple structural model of labor supply to show that reduced-form results using realized health and work will be biased when there is heterogeneity in taste for work that is also correlated with health.

We take two distinct approaches for validating the conditional probability approach. In the last part of Section V, we show that the conditional probability-based measures of VRI respondents match quite well realized work status, both unconditionally and conditional on realized health.

Additionally, in Section VI we replicate our analysis of S*ea*TE in the Health and Retirement Study (HRS).

II. Related Literature

Literature on 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)). The role of health has been subject to much debate owing to the difficulties of unpacking the health-work nexus.

The sign of the relationship is theoretically ambiguous. 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)). It 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)).[1]

Furthermore, both retirement and health are subject to several measurement issues which exacerbate the challenges of studying their relationship (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, Gustman, Steinmeier, and Tabatabai 2010, and Maestas 2010 on concepts and measures of retirement).

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 document that the effect is highly heterogeneous. We produce estimates that are credibly free of major biases that have plagued realizations-based estimates, such as justification bias and heterogeneity bias.

[1] The magnitude of the relationship is hard to quantify empirically as health and work are jointly determined and tend to 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.

Literature on subjective expectations. 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 (e.g., Hamermesh (1985) builds the case for measurement of perceived horizons or longevity expectations) 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 (see Juster 1966 and Manski 1990).

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), Carroll (2017), and Giustinelli and Manski (2018) trace the development of the subjective expectations literature from various perspectives. Papers by Arrondel, Calvo-Pardo, Giannitsarou and Haliassos (2017), Bordalo, Gennaioli, Ma, and Shleifer (2017), and Fuster, Perez-Truglia and Zafar (2017) are recent advances in this literature.

This paper advances on use of probabilistic expectations data to predict choice behavior in incomplete scenarios posed by the researcher (Manski 1999). See also Dominitz (1997), Wolpin (1999), and Blass, Lach, and Manski (2010). van der Klaauw and Wolpin (2008), van der Klaauw (2012), and Pantano and Zheng (2013) use unconditional expectations to aid in estimating dynamic programming models. In this paper, we use expectations about future actions conditional on a specified future state to quantify the causal effect of health on retirement and to provide an interpretation of the effect within two mainstream frameworks of econometric causality, the potential outcome framework (POF) and dynamic programming (DP).

Methodologically, our paper is closest to recent works by Arcidiacono, Hotz, Maurel, and Romano (2017), henceforth AHMR, and Wiswall and Zafar (2016), henceforth WZ. AHMR generate estimates of individual and aggregate level effects of occupation choice on earnings of Duke undergraduates by comparing students' subjective conditional 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 conditional expectations for monetary and non-monetary outcomes across alternative graduation majors. Hence, these papers consider a modeling framework à la Roy (Roy 1951), where potential treatments are alternative human capital investments (occupation choices and college majors) and potential outcomes are the monetary and/or non-monetary consequences of making alternative investments (earnings, marriage market outcomes, etc.).[2] We, instead, focus on a class of models of intertemporal decision-making amenable to dynamic programming treatment, where potential treatments are alternative states of nature (individuals' health) and potential outcomes (working vs. not) are feasible actions that agents can take after learning the realized states.

Using subjective expectations to study labor supply and its relation with health. Within the labor supply literature, 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 ceteris paribus effects (McGarry 2004) or structural parameters (van der Klaauw and Wolpin 2008). In turn, health and longevity expectations have been used to generate moment conditions for structural estimation (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, 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 innovation of the analysis is to replace actual labor supply with unconditional working expectations as a dependent variable in order to focus on working respondents and avoid justification bias in self-reported health among retirees. Our contribution is to introduce the use of conditional probabilities.[3]

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 estimates precision. We, on the other hand, combine

[2] These studies build on earlier work by Dominitz and Manski (1996) who elicit and analyze subjective earnings distributions of a sample of Wisconsin high school students and college undergraduates under alternative scenarios for future schooling.
[3] Hudomiet, Hurd, and Rohwedder (2019), who collaborated with us in the development of the VRI and HRS modules, are studying the work and other domains using conditional probabilities elicited in the HRS and the ALP.

conditional working probabilities and dynamic programming to derive individual-specific, health-contingent measures of value of continued work that is unobserved in behavioral data.

Data on unconditional choice expectations, (or on choice expectations elicited under a single scenario), can be used to perform unconditional (or conditional) prediction of population behavior. Combined with data on choice realizations, they can be further used to improve estimation efficiency (as in van der Klaauw and Wolpin (2008) and van der Klaauw (2012)). Moreover, its greater variation over realized choices can sometimes help address specific issues (as in McGarry 2004). Identification of treatment effects or structural parameters and extrapolation to alternative scenarios, on the other hand, requires elicitation of conditional choice expectations under multiple alternative scenarios, or a combination of unconditional and conditional choice expectations (see Wolpin 1999). This paper advances that agenda.

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 (forthcoming) study preferences for full and partial retirement using hypothetical choices. Our approach advances the hypothetical choice agenda by posing the choices probabilistically.

III. Analytic Framework

A major advantage of our probability-based approach over the traditional approach based on realizations data is that it enables us to circumvent the logical impossibility of observing individuals' counterfactual labor supply behavior corresponding to the health states that individuals do not experience. This is achieved by asking respondents to predict their labor supply behavior under *all* health states that they might experience. We additionally elicit respondents' subjective expectations of experiencing those health states. Hence, for each respondent our data encompass probabilistic *ex ante* predictions of the respondent's health and labor supply outcomes which *ex post* will be either realized (actual) or counterfactual.

An important related advantage of this data structure is that it naturally allows for within-person comparisons of (expected) labor supply behavior across alternative health states. These comparisons yield person-specific effect of health on labor supply and allow for unrestricted effect heterogeneity across individuals. This effect may be interpreted as a causal partial effect (as opposed to a non-causal or total one) under specific conditions spelled out below. These conditions involve the relationship (or lack thereof) between the state variables whose values are set by the researcher in the elicitation scenario and, thus, are kept fixed by the respondent when reporting their labor supply expectations and those variables whose values are not specified by the researcher and, thus, might not be kept fixed by the respondent.

The *ex ante* approach and measures that we advance in this paper can be employed to study the causal relationship between health and retirement—or any other state and behavior—within two mainstream microeconometric frameworks, potential outcomes (POF)[4] and dynamic programming (DP).[5] Empirical causal analyses have been overwhelmingly implemented using data on realizations of the relevant variables: realized states and choices in DP and realized treatments and outcomes in the POF. In this section, we demonstrate how to use *ex ante* measures of predicted hypotheticals to identify and estimate causal parameters of interest in both a POF and DP framework.

A. Potential Outcomes Framework Interpretation of S*ea*TE

We consider a simple setting where labor supply is modelled as a binary variable. 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 and 0 not working. 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 1 denotes low health and 0 high 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 (see 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.

[4] Originating in statistics from the work of Neyman (1923), Rubin (1974, 1976), Holland (1986), and their collaborators, here we use POF to refer to its interpretation and developments in econometrics, as discussed for instance in Heckman (2001, 2005, 2008, 2010) and Manski (1995, 2007).
[5] Eckstein and Wolpin (1989), Rust (1992, 1994), Keane and Wolpin (2009), and Arcidiacono and Ellickson (2011), and others review the DP approach from various perspectives.

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 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, $d_{it}(h_{it})$, which maps mutually exclusive and exhaustive treatments into outcomes. Hence, $d_{it}(h_{it})$ 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,

$$\Delta_{it} = d_{it}(1) - d_{it}(0), \tag{1.1}$$

with $\Delta_{it} \in \{-1, 0, 1\}$. 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 $h_{it} = 1$ and $h_{it} = 0$.

The variable *h* is 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 low health and 1 high health as before. Then, $d_{it} \equiv d_{it}(z_{it}) \in \{0,1\}$ is the person's *realized outcome*, that is, the labor supply *i* selects when the realized health state is z_{it}; whereas, $d_{it}(1 - z_{it}) \in \{0,1\}$ is the *counterfactual outcome,* that is, the labor supply *i* would have selected had the realized health state been $(1 - z_{it})$.

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, i.e., $\Delta_{it+1} = d_{it+1}(L) - d_{it+1}(H)$. Of course, this individual-level treatment effect cannot be observed because only one of the two health states is realized.

Consider the case that 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 realization data 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 persons rather than within person.[6] 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, and $E\left[d_{it}(h)\right] \equiv P\left[d_{it}(h) = 1\right]$ as d is binary (covariates are omitted for simplicity). 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, $E\left[d_{it}(1) \mid z_{it} = 0\right]$ and $E\left[d_{it}(0) \mid z_{it} = 1\right]$. Thus, without additional assumptions on these counterfactual moments, the ATE parameter is not (point) identified.

[6] The increasing availability of large data sets (aka big data) and the related bourgeoning of statistical learning methods (aka machine learning) have stimulated the development of a rapidly growing econometric literature on heterogeneous treatment effects. For example, see Athey and Imbens (2016) and Wager and Athey (2018) for recent theoretical contributions and Davis and Heller (2017) and Knaus, Lechner, and Strittmatter (2018) for applications. While these techniques can control for very rich observed heterogeneity, they still rely on standard identification assumptions to address unobservability of counterfactual moments.

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, realized outcomes in each group can be used as a measure of the counterfactual outcomes 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. Hence, one must grapple with non-random selection of individuals into health states. For example, high-health individuals may have higher (unobserved) preference for work that might persist should these individuals experience a negative health shock, and *vice versa*.

Whenever longitudinal data on health and labor supply realizations are available, a popular strategy has been to identify the effect of interest off health shocks.[7] This approach requires that the health change be uncorrelated with respect to other factors affecting retirement in order to identify a casual effect or the difficult task of finding an instrument for health.

Treatment Effects in Expectations. The approach we advance in this paper circumvents the impossibility of simultaneously observing $d_{it}(1)$ and $d_{it}(0)$ *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.[8] We define the *Subjective ex ante Treatment Effect*

$$\begin{aligned} SeaTE(i,t,\tau) &= E_{i,t-\tau}(\Delta_{it}) \\ &= E_{i,t-\tau}\left[d_{it}(1)\right] - E_{i,t-\tau}\left[d_{it}(0)\right] \\ &= P_{i,t-\tau}\left[d_{it}(1)=1\right] - P_{i,t-\tau}\left[d_{it}(0)=1\right], \end{aligned} \tag{1.4}$$

[7] Health shocks that can be observed in panel data have been exploited in both structural models and more reduced-form analyses; e.g., see Bound, Schoenbaum, Stinebrickner, and Waidmann (1999), Cai (2010), Disney, Emmerson and Wakefield (2006), Maurer, Klein and Vella (2011), McGeary (2009), van der Klaauw and Wolpin (2008), Garcia-Gomez (2011), Blundell, Britton, Costa Dias, and French (2016), and Jones, Rice, and Zantomio (2016).

[8] 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.

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 in a survey τ periods in advance.

These individual-level effects 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. Measurement and interpretation of the individual-level S*ea*TE does not rely on specific assumptions about the nature of respondents' expectations. Yet, as discussed by AHMR, "If individuals form rational expectations over their future outcomes, and in the absence of unanticipated aggregate shocks, this parameter [the AS*ea*TE] coincides with the mean (ex post) effect of treatment on outcome." The individual-level shocks that make *ex ante* and *ex post* different average to zero under rational expectations.[9]

B. Dynamic Programming Interpretation of S*ea*TE

In this section, we relate the components of S*ea*TE to the individuals' decision problem. Doing so allows interpretation of the conditional probabilities in terms of the individuals' optimization problem.

Individual agents are represented by primitives, $u_{it}\left(s_{it},d_{it}\right)$ and $\pi_{it}\left(s_{i,t+1}\,|\,s_{it},d_{it}\right)$. As before, *i* indexes individuals and *t* time periods, with $i=1,\ldots,N$ and $t=0,1,...,T<\infty$. $u_{it}(s_{it},d_{it})$ 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, $\pi_{it}(s_{i,t+1}\,|\,s_{it},d_{it})$ 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.

[9] 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.

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

$$U_{it} = \sum_{j=0}^{T} \beta^j u_{i,t+j}\left(s_{i,t+j}, d_{i,t+j}\right), \tag{1.5}$$

where β is the discount factor. The agent behaves optimally according to the expected-utility maximizing decision rule, $\delta_{it}^*\left(s_{it}\right)$, 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,

$$\delta_{it}^*\left(s_{it}\right) = \underset{d_{it}\in\{0,1\}}{\arg\max}\left\{u_{it}\left(s_{it}, d_{it}\right) + \beta \sum_{s_{i,t+1}} V_{i,t+1}^*\left[s_{i,t+1}, \delta_{i,t+1}^*\left(s_{i,t+1}\right)\right] \cdot \pi_{it}\left(s_{i,t+1} \mid s_{it}, d_{it}\right)\right\}, \tag{1.6}$$

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 $\delta_{it}^*\left(s_{it}\right)$ is a deterministic function of s_{it}, given the primitives.[10]

Dynamic programming fixing health. Our approach elicits respondents' expectations about their labor supply in τ periods, both unconditionally and upon fixing the level of future health. 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.

[10] 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.

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 (1.5) becomes

$$U_{it} = \sum_{j=0}^{T} \beta^j u_{i,t+j} \left[\left(x_{i,t+j}, y_{i,t+j}, \varepsilon_{i,t+j} \right), d_{i,t+j} \right], \tag{1.7}$$

and the related expression for the agent's optimal solution in equation (1.6) becomes

$$\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right) = \arg\max_{d_{it} \in \{0,1\}} \; u_{it}\left[\left(x_{it}, y_{it}, \varepsilon_{it}\right), d_{it}\right] +$$
$$\beta \sum_{\left(x_{i,t+1}, y_{i,t+1}, \varepsilon_{i,t+1}\right)} V_{i,t+1}^{*}\left[\left(x_{i,t+1}, y_{i,t+1}, \varepsilon_{i,t+1}\right), \delta_{i,t+1}^{*}\left(x_{i,t+1}, y_{i,t+1}, \varepsilon_{i,t+1}\right)\right] \cdot \pi_{it}^{xy\varepsilon}\left[\left(x_{i,t+1}, y_{i,t+1}, \varepsilon_{i,t+1}\right) | \left(x_{it}, y_{it}, \varepsilon_{it}\right), d_{it}\right]$$

(1.8)

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).[11] 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, so even if health is endogenous, 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).

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

[11] 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.

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

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

Note that ε here is unknown to both the econometrician *and* the individual at the time of elicitation.[12] 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 (1.8) becomes

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

(1.9)

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

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

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

(1.10)

where the expression for $\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)$ is given in (1.9) and $P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right) = 1 | x_{it}\right]$ in the last line of expression (1.10) is the individual's expected optimal decision in period t, obtained by *integrating* $\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)$ 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.

[12] 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.

Consider the implications (1.10) for a survey response. From the viewpoint of a respondent at time (*t*-1), the optimal choice at time *t*, $\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)$, 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, $\pi_{it}^{x}\left(x_{i,t+1} \mid x_{it}, y_{it}, d_{it}\right)$. 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 state*.

Specifically, we elicit all components of (1.10), as follows:

(i) On the right-hand side of (1.10), the probability of working *given* fixed 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 (1.10), the *unconditional* probability of working, $P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)=1\right]$.

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}, (for example, 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 conditional 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 the S*ea*TE, which is given by the difference in subjective conditional 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,

$$\mathrm{S}ea\mathrm{TE}\left(i, t, 1\right)=P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=1\right]-P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=0\right]. \tag{1.11}$$

As long as *y* may depend on *h*, in equations (1.9) and (1.10) 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 (1.11) 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.[13]

Econometric implementation: ex ante and conditional value functions. To implement the model econometrically it is useful to re-write the conditional choice probabilities in (1.10) in terms of the *ex ante value function* and the *conditional value function*. The DP literature makes specific additivity and orthogonality assumptions that permit estimation.

Following Arcidiacono and Ellickson (2011),[14] the *ex ante* (or *integrated*) *value function* at a generic future time *t*, $\overline{V}_{it}^{*}\left(x_{it}\right)$, 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,

$$\begin{aligned}
\overline{V}_{it}^{*}\left(x_{it}\right) &= \int_{\varepsilon_{it}}\int_{y_{it}} V_{it}^{*}\left[\left(x_{it}, y_{it}, \varepsilon_{it}\right), \delta_{it}^{*}\left(x_{it}, y_{it}, \varepsilon_{it}\right)\right]\cdot \pi_{i,t-1}^{y}\left(y_{it} \mid x_{it}, y_{i,t-1}, d_{i,t-1}\right) dy_{it} \cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it} \\
&= \int_{\varepsilon_{it}}\int_{y_{it}} \left[\left[u_{it}\left(x_{it}, y_{it}, d_{it}\right) + \varepsilon_{it}\left(d_{it}\right)\right] + \beta \sum_{x_{i,t+1}} \overline{V}_{i,t+1}^{*}\left(x_{i,t+1}\right) \pi_{it}^{x}\left(x_{i,t+1} \mid x_{it}, y_{it}, d_{it}\right)\right] \\
&\quad \cdot \pi_{i,t-1}^{y}\left(y_{it} \mid x_{it}, y_{i,t-1}, d_{i,t-1}\right) dy_{it} \cdot \pi^{\varepsilon}\left(\varepsilon_{it}\right) d\varepsilon_{it}.
\end{aligned} \tag{1.12}$$

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. In our setting, the respondent must further carry out the integration with respect to the distribution of *y*. In the simple model of health and labor supply that we specify in 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,

$$v_{it}\left(x_{it}, y_{it}, d_{it}\right) = u_{it}\left(x_{it}, y_{it}, d_{it}\right) + \beta \sum_{x_{i,t+1}} \overline{V}_{i,t+1}^{*}\left(x_{i,t+1}\right) \pi_{it}^{x}\left(x_{i,t+1} \mid x_{it}, y_{it}, d_{it}\right). \tag{1.13}$$

[13] One could decompose the effect of health that operates through wages versus other factors by conditioning on wages and health jointly in the elicitation task. We are pursuing this approach in future surveys.

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

The conditional value function is a key component for forming the conditional choice probabilities that we measure and that use as a basis for estimation of the parameters of the simple structural model that we specify below. Specifically, equation (1.9) can be re-written as

$$\delta_{it}^{*}\left(x_{it},y_{it},\varepsilon_{it}\right)=\underset{d_{it}\in\{1,0\}}{\arg\max}\left[v_{it}\left(x_{it},y_{it},d_{it}\right)+\varepsilon_{it}\left(d_{it}\right)\right], \tag{1.14}$$

and the conditional choice probabilities in equation (1.10) can be re-written as

$$\begin{aligned}
&P_{i,t-1}\left[\delta_{it}^{*}\left(x_{it},y_{it},\varepsilon_{it}\right)=1\mid x_{it}\right]\\
&=\int_{\varepsilon_{it}}\int_{y_{it}}\delta_{it}^{*}\left(x_{it},y_{it},\varepsilon_{it}\right)\cdot\pi_{i,t-1}^{y}\left(y_{it}\mid x_{it},x_{i,t-1},y_{i,t-1},d_{i,t-1}\right)dy_{it}\cdot\pi^{\varepsilon}\left(\varepsilon_{it}\right)d\varepsilon_{it}\\
&=\int_{\varepsilon_{it}}\int_{y_{it}}\underset{d_{it}\in\{1,0\}}{\arg\max}\left[v_{it}\left(x_{it},y_{it},d_{it}\right)+\varepsilon_{it}\left(d_{it}\right)\right]\cdot\pi_{i,t-1}^{y}\left(y_{it}\mid x_{it},x_{i,t-1},y_{i,t-1},d_{i,t-1}\right)dy_{it}\cdot\pi^{\varepsilon}\left(\varepsilon_{it}\right)d\varepsilon_{it}.
\end{aligned} \tag{1.15}$$

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 (1.15) links our data to the primitives of the model.

Interpreting conditional probability responses in DP framework. We now present a simple model to illustrate how our dynamic programming framework can be used to elicit information about individual-specific valuations using the standard assumption that underlie econometric implement of the DP framework. As above, we treat work and health as binary.

First, consider the case where there are no unspecified state variables. As in the discussion of decision tree (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 of individual i of being in state h and making choice d at time t given expectation and optimization from t+1 onward. Then

$$\begin{aligned}
V_{it}\left(h_{it},d_{it}\right)=(1-h_{it})&\left\{d_{it}\left(v_{it}\left(H,W\right)+\varepsilon_{it}\left(W\right)\right)+(1-d_{it})\left(v_{it}\left(H,\sim W\right)+\varepsilon_{it}\left(\sim W\right)\right)\right\}\\
+h_{it}&\left\{d_{it}\left(v_{it}\left(L,W\right)+\varepsilon_{it}\left(W\right)\right)+(1-d_{it})\left(v_{it}\left(L,\sim W\right)+\varepsilon_{it}\left(\sim W\right)\right)\right\},
\end{aligned} \tag{1.16}$$

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.

$$
\begin{aligned}
&\text{When } H, \\
&\quad W \text{ if } 0 \le v_{it}\left(H,W\right) - v_{it}\left(H,\sim W\right) + \varepsilon_{it}\left(W\right) - \varepsilon_{it}\left(\sim W\right) \\
&\sim W \text{ otherwise.} \\
&\text{When } L, \\
&\quad W \text{ if } 0 \le v_{it}\left(L,W\right) - v_{it}\left(L,\sim W\right) + \varepsilon_{it}\left(W\right) - \varepsilon_{it}\left(\sim W\right) \\
&\sim W \text{ otherwise.}
\end{aligned}
\tag{1.17}
$$

Again, as traditional in dynamic programming applications, we define objects that are the *differenced conditional value functions* and of the residual components. Let

$$
\begin{aligned}
\tilde{v}_{it}^{H} &= v_{it}\left(H,W\right) - v_{it}\left(H,\sim W\right) \\
\tilde{v}_{it}^{L} &= v_{it}\left(L,W\right) - v_{it}\left(L,\sim W\right) \\
\tilde{\varepsilon}_{it} &= \varepsilon_{it}\left(W\right) - \varepsilon_{it}\left(\sim W\right),
\end{aligned}
\tag{1.18}
$$

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

$$
\begin{aligned}
\text{When } H,\ \delta_{it}^{*} &= 1 \text{ if } 0 \le \tilde{v}_{it}^{H} + \tilde{\varepsilon}_{it} \\
&= 0 \text{ otherwise.} \\
\text{When } L,\ \delta_{it}^{*} &= 1 \text{ if } 0 \le \tilde{v}_{it}^{L} + \tilde{\varepsilon}_{it} \\
&= 0 \text{ otherwise.}
\end{aligned}
\tag{1.19}
$$

where we are using the 0/1 notation for working to be transparently analogous to discrete choice dynamic programming econometrics.

We now show how conditional probability estimates 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 precisely into the discrete choice problem in (1.19). 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

$$P_{i,t-1}^{H} \equiv P_{i,t-1}\left[\delta_{it}^{*}=1 \mid h_{it}=H\right] \tag{1.20}$$

and

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

yields

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

Then (1.19) given the distributional assumption for $\tilde{\varepsilon}_{it}$ implies

$$\begin{aligned} P_{i,t-1}^{H} &= \Phi\left(\tilde{v}_{it}^{H}\right) \\ P_{i,t-1}^{L} &= \Phi\left(\tilde{v}_{it}^{L}\right). \end{aligned} \tag{1.22}$$

We invert these expressions to yield

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

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.

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 (1.23) 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;[15] (iv) who gave complete and consistent

[15] 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.

responses to the latter battery; and (iv) who reported being in high health in Survey 4.[16] Table A1 in the online appendix summarizes the selection process.

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 summary 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) used to select the sample, they are broadly similar to those from the HRS and the Survey of Consumer Finances (SCF) (Ameriks, Caplin, Lee, Shapiro and Tonetti, 2015).

C. S*ea*TE and Its Components

In the work-health expectations battery, eligible VRI respondents are first asked for the percent chance out of 100 that they will be working in 2 years ahead. 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. These questions were then repeated for the 4 years horizon.[17]

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 2 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. This respondent is asked the following sequence of questions for the 2-year horizon:

[16] 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).
[17] 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 this questions is designed to inform. It has been shown to be reliable in many contexts. Alternatively, one could ask about work under various diagnoses (e.g., high blood pressure, cancer, in ability to lift, cognitive decline). This could not be practical, especially because different conditions would be relevant for different types of work and because there are competing, multiple health risks. Heterogeneity in interpretation of the scale is not necessarily 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 conditional probability of working in low health, the conditional probability of working in high health, and the S*ea*TE. 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 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 as the 4-year horizon.

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

The next two columns show the empirical distributions of the working probabilities conditional on experiencing low health and 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 conditional working probabilities given high health results from the high and relatively undispersed expected 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 conditional 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 years case, the median drops from 68 to 20 and the mean from 59 to 33. We discuss the S*ea*TE below.

In Figure 3, we show box-and-whisker plots of the conditional 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 A1 in the online appendix display analogous box and whisker plots for the unconditional working and health probabilities.

In the left plots of Figure 3, 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 3 reveals that the age-specific mean and median decrease sharply and steadily from the ≤ 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 3 represent graphically and by age group a basic finding in Table 1, that 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 third 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 less work 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 working less in low health. Most of the remaining respondents have zero S*ea*TE, which means 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, 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 expect always to work while the remaining 21 percent are interior. The fractions who expect to never work and always work flips at 4 years as the tendency to retire regardless of 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 health shocks are expected to reduce work, there remains considerable heterogeneity in the size of the effect of health.

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 covariates. Except for age, S*ea*TE is little predicted by covariates. Hence, most of the heterogeneity is unobserved. 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 health has a bigger effect on work for those over age 60. The difference in the effect is non-monotonic in age and only statistical 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 birthday). 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 largely small and statistically insignificant. Since most of the heterogeneity is unobserved, these findings imply that a including 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 in their responses. The respondents are quite good at applying the LTP.

Figure 4 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.[18] 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 heaping at the corners,

[18] 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.

this vast majority of corner respondents are getting the LTP exactly right. This is consistent with more recent evidence on rounding and probability imprecision suggesting that responses of 0 and 50 percent are not more rounded or imprecise than other responses and, if anything, responses of 0 percent are less so (Giustinelli, Manski, and Molinari 2019).

V. Estimates using Dynamic Programming Framework

A. DP: Eliciting Utility

We now proceed with analysis derived from the dynamic programming specification of Section III. The elicited conditional 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 $\tilde{v}^L$ from equation (1.23) for the 2-year and 4-year-ahead conditional probabilities. 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 conditional probabilities.[19] 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 5A shows a scatter plot of the values $\tilde{v}^H$ and $\tilde{v}^L$ for the 2-year horizon. Figure 5B shows an analogous plot for the 4-year horizon. These plots illustrate many features of the value of work conditional on health 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,

[19] Note for the purpose of the analysis in this paper, we are treating both the horizons as different "one-period ahead" expectations. That is, we are not modeling the transition from 2 to 4 year, but instead presenting them as separate (though obviously related) measurements.

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 individual 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

$$\begin{aligned} \tilde{v}_i^H &= \alpha^H + \upsilon_i^H \\ \tilde{v}_i^L &= \alpha^L + \gamma\upsilon_i^H + \upsilon_i^L, \end{aligned} \tag{1.24}$$

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 in the high health state. The heterogeneity in the low health states 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 (1.24). The orthogonal decomposition is a convenient way to summarize the observed heterogeneity.

Table 7 presents estimates of the parameters specified in equation (1.24).[20] 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 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 ahead horizon, 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).

[20] 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.

For interpreting these results, it is important to bear in mind that the estimate is based on a single cross section. The willingness to work declines sharply with age in the age range of the VRI sample, and this decrease is much greater in low health. The estimate of the coefficient γ, controlling the correlation within individual of the willingness to work across health states, decrease slightly to 0.66. Online Appendix Table A3 show the regression with covariates. As expected, there are significant shifts in the value of work with age, but not with covariates.

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 as opposed to data on realized behavior. Consider the estimate of a linear regression model using data on working d_i and health h_i realizations,

$$d_i = b_0 + b_1 h_i + e_i. \tag{1.25}$$

As discussed in Section III, the least squares estimate of b_1 will be an unbiased estimate of the ATE only when health is exogenous. In the terms of our DP model, exogeneity will fail when realized health is correlated across individuals with the values $\tilde{v}_i^H$ or $\tilde{v}_i^L$. Our elicitation approach is designed to render this heterogeneity observable *ex ante.* Specifically, the average S*ea*TE across individuals will be an unbiased estimate the causal effect of health b_1 even if there is heterogeneity that would be unobserved with conventional data on realizations of health and work.

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 (1.19) implies that the realized decision to work is

$$d_i = (1-h_i)\mathbf{I}[\tilde{v}_i^H + \tilde{\varepsilon}_i] + h_i\mathbf{I}[\tilde{v}_i^L + \tilde{\varepsilon}_i], \tag{1.26}$$

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 observed heterogeneity, we use the measured conditional value functions ($\tilde{v}_i^H$ or $\tilde{v}_i^L$) and simulated realizations of health (h_i), and the residual component ($\tilde{\varepsilon}_i$) to calculate simulated decisions according to equation (1.26). 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 implement of the DP formulation, $\tilde{\varepsilon}_i$ is simulated as standard normal.

We consider three cases for correlation of the health transition probability π_i^h :

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.[21]

The first case implies health is exogenous, so the OLS estimate of (1.25) 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 SES) 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 (because 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 values even after conditioning a rich set of covariates. Therefore, conditioning on such covariates in

[21] Specifically, the simulations are based on adjusting the π_i^h by subtracting 0.1 from individuals in the bottom quintile of $\tilde{v}_i^H$, subtracting 0.05 from those in the second quintile of $\tilde{v}_i^H$, leaving the middle quintile unadjusted, and adding 0.075 to the top two quintiles of $\tilde{v}_i^H$. (The top two quintiles are combined because they have a common $\tilde{v}_i^H$ corresponding to individual who gave a 100% change of working when in high health.)

econometric applications using data on realized decisions and states, though helpful, is not likely to eliminate bias from unobserved heterogeneity.

C. DP: Realizations

Taking advantage of the panel structure of the VRI, we now use data on realized labor supply and health of our respondents after two years to validate the 2-year ahead work and health probabilities. In Survey 6, we observe health and labor supply realizations who also took our expectations battery in Survey 4. This set of respondents provides the panel sample for this sub-section.[22]

There are relatively few health transitions from high (excellent, very good, good) to low (fair, poor) in the two years between the Survey 4 and Survey 6. Health does decline within high health. Accordingly, we use the finer, three-way partition of health embodied in the survey design. Table 9 shows mean probabilities or rates (in the case of realizations) by row. Each row conditions on the respondents' health state in Survey 4. The columns correspond to future health state partitioned in the Survey 4 probability questions. The realizations as of Survey 6 correspond to the 2-year ahead horizon in the Survey 4 questions. Recall that these partitions differ by health state at the time of the survey (see Figure 2), so the groupings vary across the rows.

Panel A reports the mean conditional working probabilities by realized health in Survey 6. Panel B reports means for realized work. Comparison of Panel A and Panel B reveals that the health-contingent working probabilities given by the respondents in Survey 4 match up remarkably well with the labor supply realizations in Survey 6, so there is evidence of rational expectations for work. There is a declining work-health gradient both *ex ante* and *ex post*.

Moving to health, a comparison of the health probabilities in Panel C and the health realizations in Panel D reveals some deviations from rational expectations for health notwithstanding the unbiased prediction of the conditional expectations of work. The respondents in excellent health in Survey 4 over-estimate the odds of extreme states (excellent and fair/poor). Respondents in very good and good health in Survey 4 over-estimate the odds of fair/poor. Therefore, while the conditional probabilities that directly enter the calculations of the S*ea*TE and the DP values align very closely with realizations, the pessimism about low health of the respondents does not match

[22] Survey 4 was fielded in late 2015. Survey 6 was fielded in early 2018, roughly two years later. There are 584 respondents in the panel sample. There is no evidence of selective non-response to Survey 6 conditional on age, probability of work, and probability of low health (see Online Appendix Table A4).

the realizations. This failure of rational expectations could arise in principle from a correlated shock, but more likely reflects a systematic bias of being too pessimistic about health transitions.

We now turn to predicting work realizations based on the conditional probability of work. In regressions reported in Table 10, we consider two predictors of working in Survey 6. First, for each respondent, we use the inner product of the work probabilities conditional on health and the probabilities of the corresponding health states (labeled *ex ante* health). Both are measured at Survey 4. Second, for each respondent, we use the work probability conditional on health for health state realized in Survey 6 (labeled realized health). The first and third columns show estimate for this basic specification. The alternate columns report estimates with age interactions.

The results in Columns 1 and 2 are very similar to those in Columns 3 and 4. The conditional working probability measures are strong predictors of work realizations, with a coefficient close to 0.6 and highly statistically significant. In Column 1, the null hypothesis of a unit coefficient of working probability interacted with age is rejected. The age-specific estimates in Column 2 reveal, however, that its estimated coefficient is quite close to one for all age groups between 60 and 69. In particular, the null hypothesis of a unit coefficient cannot be rejected for the 62, 65, and 68-69 years old and is close to the threshold for respondents the other groups, except the 70-71 years old and the over 72. Hence, the conditional work probabilities are very close to being unbiased predictors. Moreover, they also account for a substantial fraction of the cross-sectional variation in work outcomes with an R^2 of about one quarter. Hence, there is good evidence supporting the validity and usefulness of the *ex ante* survey measure in the realizations data.[23] [24]

VI. Replication in the Health and Retirement Study

We replicate the analysis 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

---

[23] In Tables A5 and A6 of the online appendix, we report parallel statistics and estimation results to those shown in Tables 9 and 10, with the conditional working probabilities replaced by the DP values. We also checked to see whether the results change in specifications including the other covariates. The results mirror those presented for the regressions using the conditional probabilities as regressors. The regressions in the text using the conditional probabilities are easier to interpret. Those using the DP values are, however, directly account for unobserved value of work in the form relevant for estimate of models such as (1.19).

[24] Following a similar approach, Gong, Stinebrickner, and Stinebrickner (2019) investigate the relationship between mid- to long-term working and family (marriage and fertility) expectations of Berea College students and their subsequent outcomes.

questions as in the VRI.[25] We analyze a sample of 483 HRS respondents who, in addition to taking the module, met the following criteria: (i) who were 50 or older;[26] (ii) who were in the labor force at the time of the survey; (iii) who gave complete and consistent (or close to consistent) responses to the expectations battery;[27] and (iv) who reported being in high health.[28] Note that there are large differences in the characteristics of the VRI and HRS samples (see Table A2 and Table B2 in the online appendix). The VRI respondents are older, healthier at the same age, more educated, and more affluent. Hence, the results are not meant to be directly comparable, but rather demonstrate the applicability of the approach in different populations.[29]

Tables 11-14 report HRS results parallel to the VRI results in Tables 1-4. In Table 11, 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. (Recall that the respondents to the HRS module are younger than those in the VRI.) They also have higher average probability of entering low health, although the difference is not large, especially at 4 years.

In Table 12, 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 SeaTE at 2 years, 1.6 percent more positive S*ea*TE at 4 years).

In Table 13, even though the proportion of zero SeaTE 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 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 larger.

Among negative S*ea*TE respondents, the distribution of S*ea*TE is remarkably similar in the VRI and HRS samples. See Table 14. Hence, the estimated effect of health on work is quite similar despite the difference in the samples and in responses reflected in Table 13. Online

---

[25] An experimental module is a short battery of questions, taking approximately 3 minutes to complete, that a random subset of HRS respondents are invited to answer after completing the core questionnaire. In this module, respondents were selected only if they were below 65 years old, so the age range is lower than that of the VRI.

[26] The HRS is a representative study of the U.S. population 50 and older. However, the age requirement is only applied to household heads. So a small fraction of HRS respondents, typically female spouses of the household head, may be under 50. We exclude these respondents.

[27] We exclude respondents whose inconsistency (e.g. summing of probabilities to one) exceed 10 percentage points. For the marginally-inconsistent responses, we renormalized the responses.

[28] 296 of the 1082 HRS respondents who took our module reported being in fair or poor health.

[29] We attempted to analyze a subset of the HRS sample with similar characteristics as the VRI along lines of Ameriks, Caplin, Lee, Shapiro, and Tonetti's consideration of a VRI-eligible HRS population. Our samples are too small to provide very meaningful comparisons.

Appendix B reports additional results for the HRS. Note that in HRS responses relative to those in the VRI, the law of total probability does not hold nearly as well (see Figure B3) and there are more inconsistent answers (Table B1).

VII. Conclusion

In this paper, we provide a novel strategy for assessing the causal effect of a treatment on a behavior or outcome at the individual level. We apply our approach to quantify person-specific effects of health on work. 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, its cross-sectional average gives a consistent estimate of the standard average treatment effect (ATE) absent aggregate shocks. We give a formal interpretation to the S*ea*TE and the conditional choice probabilities in the two workhorse frameworks of econometric causality: the potential outcomes framework (POF) and dynamic programming (DP).

We document that the effect of 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. A very few individuals have a positive effect of health on work.

We map the conditional probabilities into a 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 disturbance in a standard discrete choice formulation of the labor supply decision. There is a strong correlation within individuals on the value of 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 is appropriately 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 regardless of health.

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 conditional responses and the law of total probability.

Importantly, we validate the conditional probabilities using panel outcome data. The conditional probabilities interacted with subsequent health outcomes have strong explanatory power for realized work.

Finally, we replicate results using the same question battery in an experimental module in the Health and Retirement Study (HRS). The S*ea*TE in the HRS is quite similar to that in the VRI, though the distribution of underlying conditional probabilities in the HRS differs.

The *ex ante* method gives estimates of potential outcomes. The methodology in this paper could be applied in a wide range of applications beyond health and retirement. More generally, the approach is useful for giving reliable estimates of effects when treatment are difficult to manipulate experimentally or control-for econometrically including the particularly interesting case of policies that have not yet been 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 (1.22) 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 conditional probability of working given health $P_{i,t-1}^{h}$ is the weighted average of the conditional 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 give 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 (1.22).

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

References

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

Ameriks, John and Joseph Briggs and Andrew Caplin and Minjoon Lee and Matthew D. Shapiro and Christopher Tonetti. *Forthcoming*. “Older Americans Would Work Longer If Jobs Were Flexible.” Forthcoming in the *American Economic Journal: Macroeconomics*.

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 and Joseph V. Hotz and Arnaud Maurel and Teresa Romano. 2017. *Ex Ante* Returns and Occupational Choice. Working Paper.

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 and Hector Calvo-Pardo and Chryssi Giannitsarou and Michael Haliassos. 2017. Informative Social Interactions. Working Paper.

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

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.

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. and 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 and Jack Britton and Monica Costa Dias and Eric French. 2016. The Dynamic Effects of Health on the Employment of Older Workers. MRRC Working Paper, 2016-348.

Bordalo, P. and N. Gennaioli and Y. Ma and A. Shleifer. 2017. “Overreaction in Macroeconomic Expectations.” Working Paper.

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

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

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

Carroll, Christopher. 2017. Heterogeneity, Macroeconomics, and Reality. Keynote Lecture given at the Sloan-BoE-OFR Conference on Heterogeneous Agent Macroeconomics, U.S. Department of the Treasury, September 22, 2017.

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.

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.

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

Disney, Richard and 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.

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

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

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

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

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

Fuster, Andreas and Ricardo Perez-Truglia and Mirko Wiederholt and Basit Zafar. 2017. " Expectations with Endogenous Information Acquisition: An Experimental Investigation." Working Paper.

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

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

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 and Charles F. Manski and Francesca Molinari. 2019. Precise or Imprecise Probabilities? Evidence from survey response on Dementia. Working Paper.

Gong, Yifan and Todd Stinebrickner and Ralph Stinebrickner. 2019. Marriage, Children, and Labor Supply: Beliefs and Outcomes. Working Paper.

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

Gustman, Alan L. and 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. and 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." *The Quarterly Journal of Economics*. 100(2): 389-408.

Heckman, James J. 2001. "Esseys. Econometrics and Empirical Economics." *Journal of Econometrics* 100: 3–5.

Heckman, James J. 2005. "The Scientific Model of Causality." *Sociological Methodology* 35(1): 1–97.

Heckman, James J. 2008. "Econometric Causality." *International Statistical Review* 76(1): 1–27.

Heckman, James J. 2010. "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature* 48(2): 356–398.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945-960.

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 and Michael D. Hurd and S. Rohwedder. 2019. The Causal Effects of Economic Incentives, Health and Job Characteristics on Retirement: Estimates Based on Subjective Conditional Probabilities. Presented at the "Health and Retirement: Expectations, Cognition and Behavior" conference in honor of the retirement of Robert Willis, Ann Arbor, May 17, 2019.

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

Jones, Andrew M. and 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 and 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.

Knaus, Michael C. and Michael Lechner and Anthony Strittmatter. 2017. Heterogeneous Employment Effects of Job Search Programs: A machine learning approach, IZA Discussion Paper 10961.

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.

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. 1995. *Identification Problems in the Social Science*. Harvard University 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. 2007. *Identification for Prediction and Decision*. Harvard University Press.

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

Maurer, Jurgen and 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.” *The 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 and Eddy van Doorslaer and Tom van Ourti. 2015. Health and inequality. In Handbook of Income, Volume 2B, eds. AB Atkinson, FJ Bourguignon, Chapter 18, 1419-1533, Elsevier.

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

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

Roy, A.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.

Rubin, Donald. 1976. Inference with Missing Data. *Biometrika* 63(3): 581-592.

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.

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.

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

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. *JASA* 113(523): 1228-1242.

Wiswall, Matthew and Basit Zafar. 2016. “Human Capital Investments and Expectations about Career and Family” NBER Working Paper, 22543.

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

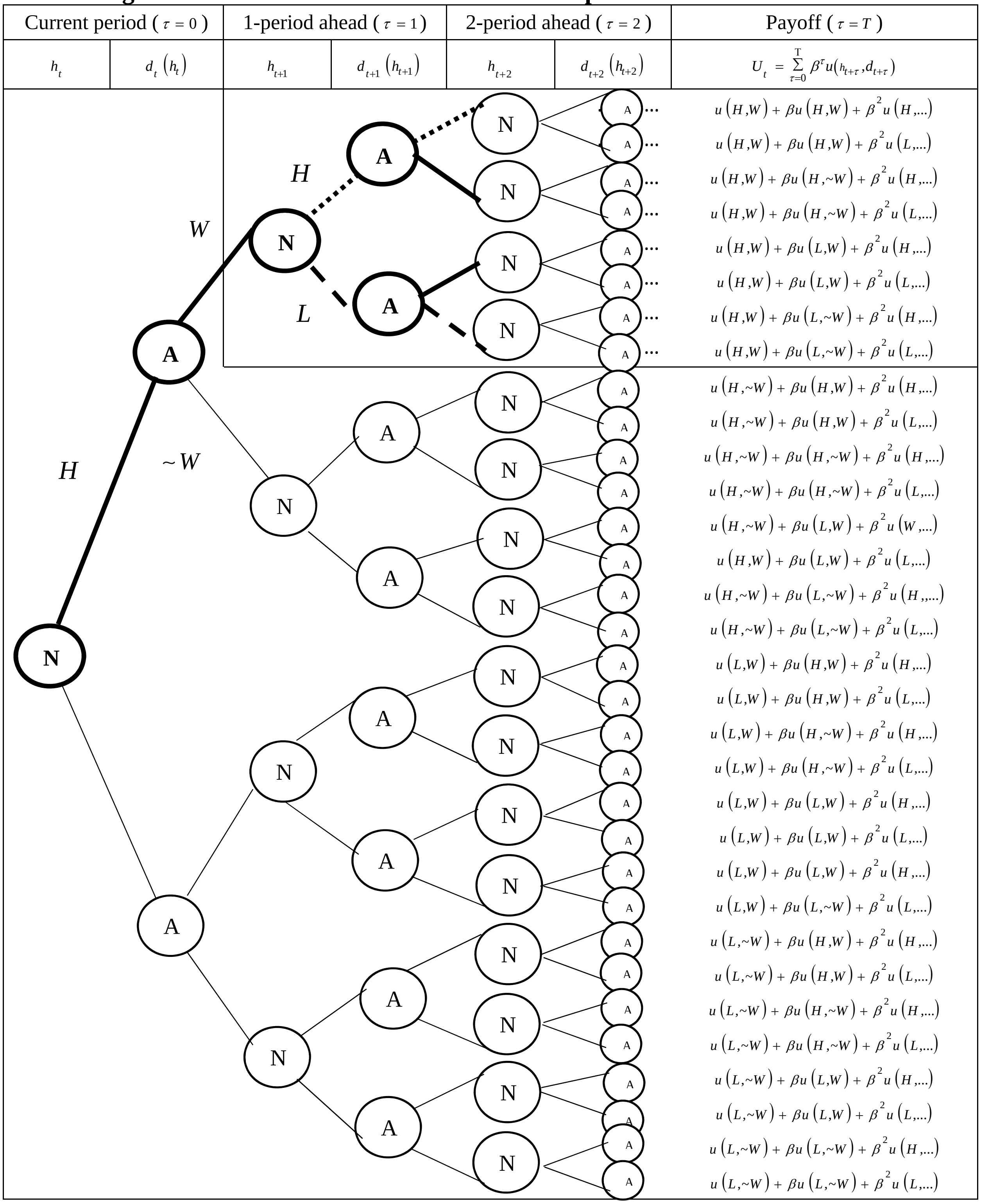
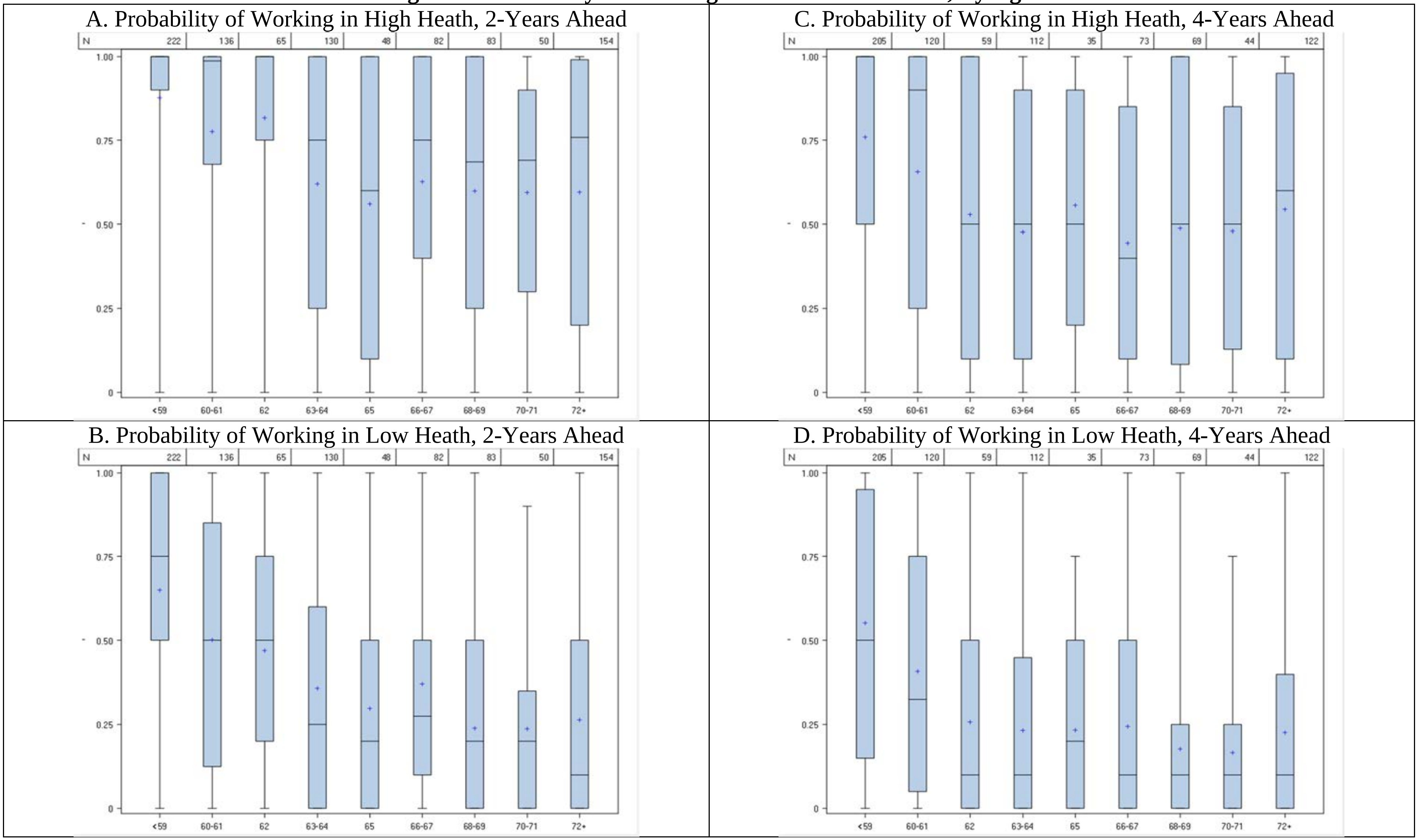


Figure 2. Partition of the Health States for Survey Questions

<table>
<tr><td></td><td colspan="5">Partition of future health state</td></tr>
<tr><td>Current health</td><td>Excellent</td><td>Very Good</td><td>Good</td><td>Fair</td><td>Poor</td></tr>
<tr><td>Excellent</td><td></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>Very Good</td><td colspan="2"></td><td></td><td colspan="2"></td></tr>
<tr><td>Good</td><td colspan="2"></td><td></td><td colspan="2"></td></tr>
<tr><td>Fair</td><td colspan="2"></td><td></td><td colspan="2"></td></tr>
<tr><td>Poor</td><td colspan="2"></td><td colspan="2"></td><td></td></tr>
</table>

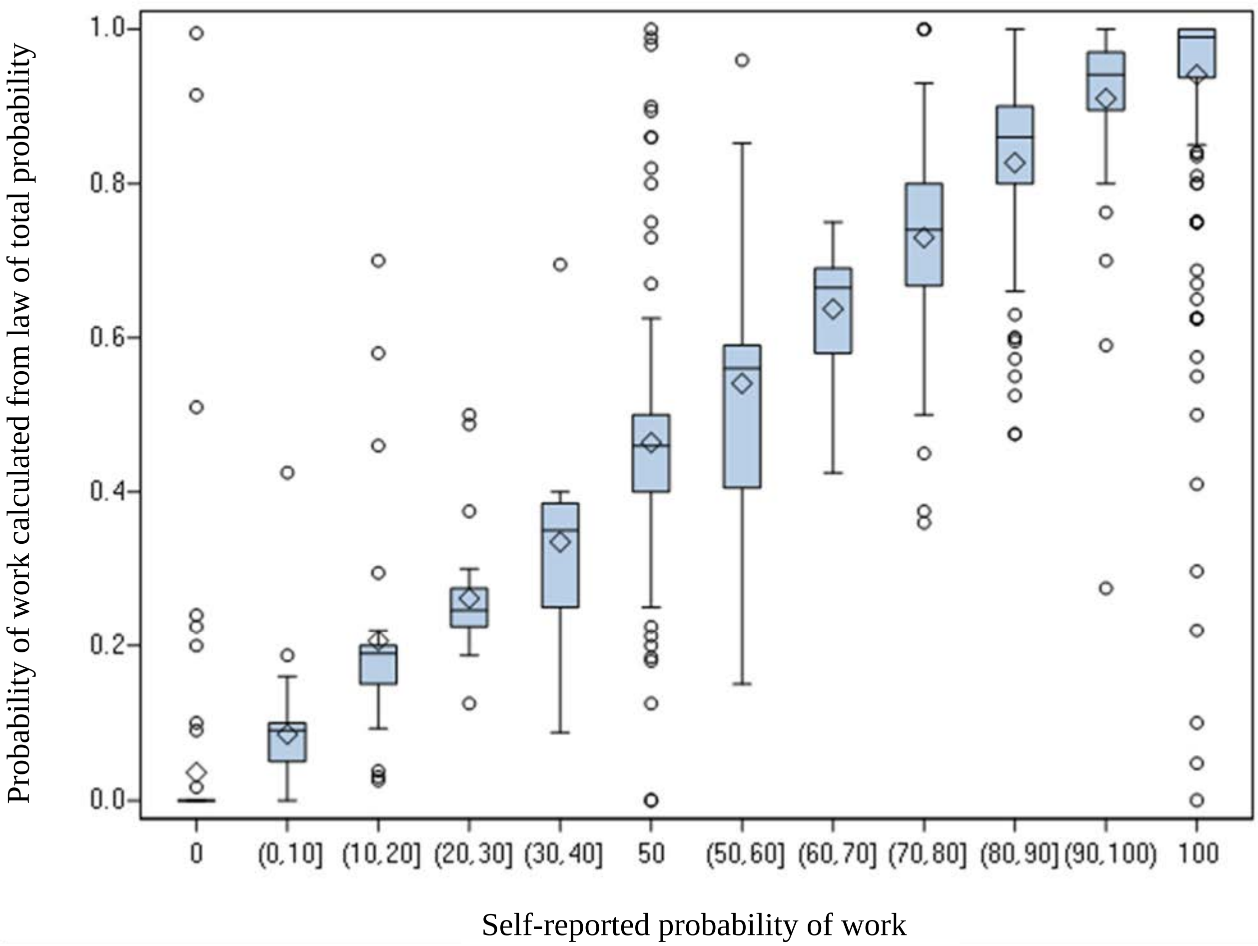
Note: This table shows the partition of the future health states for the expectations questions. The expectations sequence partitions future health states conditional on the respondent's current health. See Appendix for sequence of questions.

Figure 3. Probability of Working Conditional on Health, By Age



Note: Box and whiskers plots of the distribution of probability of working given health 2- and 4-years ahead. The “+” is the mean, the mid-line is the median, and the box shows inter-quartile range. Age as of time of the survey.

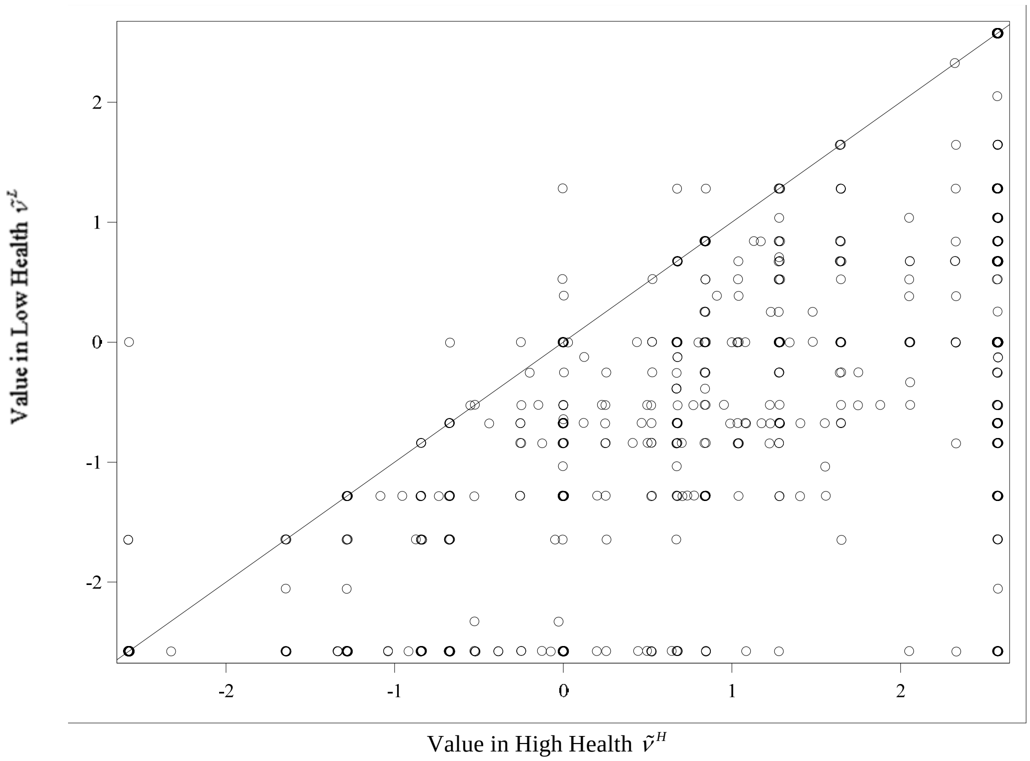
Figure 4. Do Respondents Apply the Law of Total Probability?



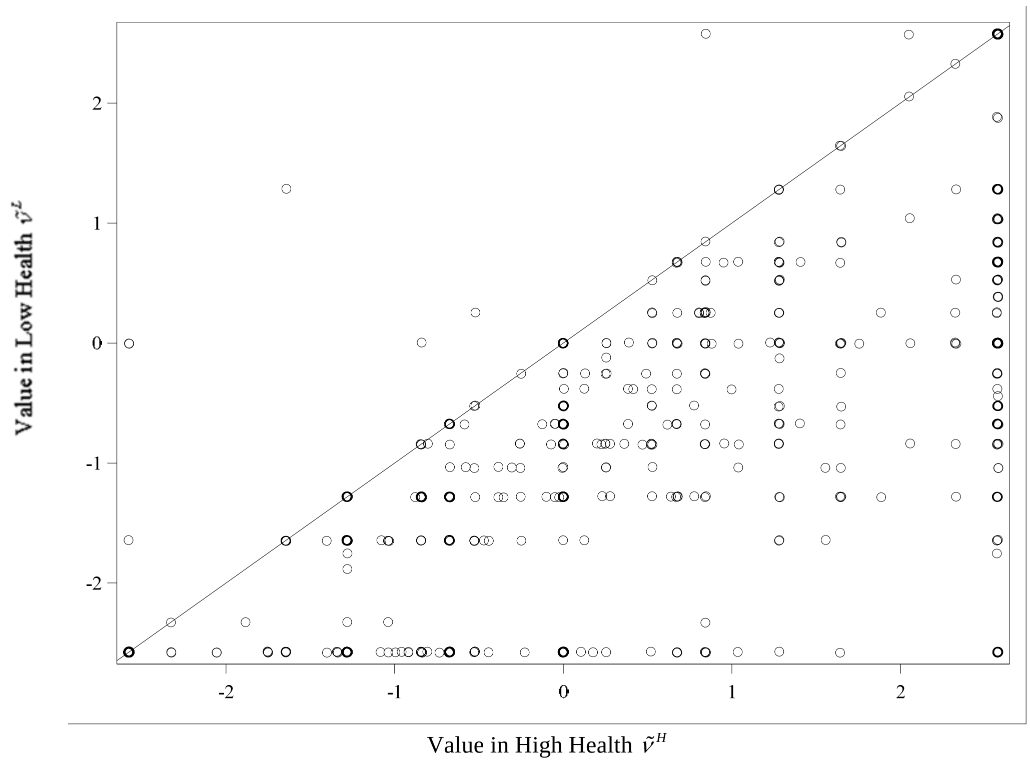
Note: Figure shows the distribution of responses for the unconditional probability of working in 2 years computed using the law of total probability (on the vertical axis) versus the self-reported reported unconditional probability of working in 2 years (on the horizontal axis). The correlation between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

Figure 5. Measured Conditional Value Functions

A. 2-Year Ahead



B. 4-Year Ahead



Note: Figures show the scatter plots of the differenced conditional value functions $\tilde{v}^H$ and $\tilde{v}^L$ for each respondent at 2- and 4-year horizons.

Table 1. Percent Chance of Working, Health, and Working Conditional on Health

	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 sub-sample and 839 for the 4-year sub-sample. Table shows mean, standard deviation, first quartile (Q25), median, and third quartile (Q75) across respondents for each reported probability. Probability of working is calculated from the law of total probability using conditional health and conditional working responses (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), zero S*ea*TE (same chance of working across high and low health), and positive S*ea*TE (greater chance of working in low 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 conditional on the probability of working in high and low health being the same. In both states, never work respondents have zero probability of work, always work have probability one of work, and maybe work have interior probability of work.

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 low 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 (<60 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: *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 6. Conditional Value Functions 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 conditional value functions $\tilde{v}^H$ and $\tilde{v}^L$. See equation (1.23).

Table 7. Quantifying Cross-Sectional Heterogeneity in Conditional Value Functions

	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 conditional value functions $\tilde{v}^H$ and $\tilde{v}^L$ as specified in equation (1.24) of 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 mean; empirical case uses individual-specific elicited health transition probabilities. Highly correlated case has stronger correlation of health transition probability and value of work as described in text. The LHS variable is the simulated realized decision to work (*d*). The RHS variable is simulated realized health state (*h*). *h*=1 is low health and *d*=1 is work.

Table 9. Panel Results: Realizations and Expectations (means)

A. Conditional Working Probability, By *ex ante* Health

<table>
<tr><th></th><th>E</th><th>VG</th><th>G</th><th>F</th><th>P</th></tr>
<tr><td>E</td><td>0.730</td><td colspan="2">0.707</td><td colspan="2">0.390</td></tr>
<tr><td>VG</td><td colspan="2">0.708</td><td>0.691</td><td colspan="2">0.411</td></tr>
<tr><td>G</td><td colspan="2">0.693</td><td>0.675</td><td colspan="2">0.451</td></tr>
</table>

B. Realized Working Status, By Realized Health

<table>
<tr><th></th><th>E</th><th>VG</th><th>G</th><th>F</th><th>P</th></tr>
<tr><td>E</td><td>0.738</td><td colspan="2">0.719</td><td colspan="2">0</td></tr>
<tr><td>VG</td><td colspan="2">0.712</td><td>0.692</td><td colspan="2">--</td></tr>
<tr><td>G</td><td colspan="2">0.679</td><td>0.701</td><td colspan="2">0.667</td></tr>
</table>

C. Unconditional Health Probability, By *ex ante* Health

<table>
<tr><th></th><th>E</th><th>VG</th><th>G</th><th>F</th><th>P</th></tr>
<tr><td>E</td><td>0.792</td><td colspan="2">0.107</td><td colspan="2">0.101</td></tr>
<tr><td>VG</td><td colspan="2">0.728</td><td>0.114</td><td colspan="2">0.158</td></tr>
<tr><td>G</td><td colspan="2">0.322</td><td>0.350</td><td colspan="2">0.328</td></tr>
</table>

D. Realized Health

<table>
<tr><th></th><th>E</th><th>VG</th><th>G</th><th>F</th><th>P</th></tr>
<tr><td>E</td><td>0.689</td><td colspan="2">0.311</td><td colspan="2">0</td></tr>
<tr><td>VG</td><td colspan="2">0.818</td><td>0.175</td><td colspan="2">0.007</td></tr>
<tr><td>G</td><td colspan="2">0.269</td><td>0.644</td><td colspan="2">0.087</td></tr>
</table>

Note: The tables show mean probabilities or rates (in the case of realizations) by row. The rows correspond to heath at the time of Survey 4. The columns correspond to future health state partitioned in the Survey 4 probability questions. The realizations are from Survey 6, which was fielded two years after Survey 4, so the timing matches the *ex ante* probabilities. Recall that these partitions differ by health state at the time of the survey (see Figure 2), so the groupings vary across the rows. The "--" indicates value was suppressed due to low cell count.

Table 10. Panel Results: Predicting Work Using Conditional Expectations

	Ex ante health		Realized health	
	(1)	(2)	(3)	(4)
Constant	0.322	0.447	0.301	0.443
	(0.036)	(0.074)	(0.037)	(0.075)
Work Probability Conditional on Health	0.595		0.590	
	(0.048)		(0.047)	
Age				
≤ 59		-0.188		-0.238
		(0.140)		(0.144)
60-61		-0.247		-0.289
		(0.133)		(0.135)
62		-0.513		-0.528
		(0.178)		(0.181)
63-64		-0.083		-0.127
		(0.115)		(0.118)
65		-0.391		-0.287
		(0.162)		(0.158)
66-67		-0.126		-0.159
		(0128)		(0.130)
68-69		-0.251		-0.256
		(0.130)		(0.131)
70-71		0.283		0.286
		(0.157)		(0.158)
Work Probability Conditional on Health Interacted with Age				
≤ 59		0.695		0.730
		(0.136)		(0.135)
60-61		0.637		0.672
		(0.138)		(0.135)
62		0.896		0.857
		(0.213)		(0.201)
63-64		0.618		0.653
		(0.126)		(0.124)
65		0.828		0.659
		(0.203)		(0.192)
66-67		0.641		0.639
		(0.169)		(0.155)
68-69		0.832		0.784
		(0.158)		(0.149)
70-71		0.107		0.100
		(0.218)		(0.202)
≥ 72		0.438		0.406
		(0.110)		(0.102)
Observations	584	584	584	584
R^2	0.207	0.253	0.216	0.261

Table 11. Percent Change of Working, Health, and Working Conditional on Health: HRS

	Working	Low Health	Working in Low Health	Working in High Health	S*ea*TE
			2-Year Ahead		
Mean	76.2	22.5	54.9	81.5	-26.8
Std. Dev.	29.2	20	33.6	28.7	27.2
Q25	65	5	25	78	-50
Median	88.8	20	50	100	-20
Q75	100	35	80	100	0
			4-Year Ahead		
Mean	67.4	26.7	46.8	73.7	-26.9
Std. Dev.	32.1	20.3	33.7	32.5	27.4
Q25	46	10	15	50	-50
Median	77	20	50	90	-20
Q75	95.5	40	75	100	0

Note: Sample size is 480 for the 2-year sub-sample and 428 for the 4-year sub-sample. Table shows mean, standard deviation, first quartile (Q25), median, and third quartile (Q75) across respondents for each reported probability. Probability of working is calculated from the law of total probability using conditional health and conditional working responses (see text for discussion). S*ea*TE is the different between the probability or working in low versus high health.

Table 12. S*ea*TE: Negative, Zero, or Positive (fraction of responses, percent): HRS

	2-Year Ahead	4-Year Ahead
Negative S*ea*TE	66.67	69.63
Zero S*ea*TE	31.67	27.80
Positive S*ea*TE	1.66	2.57
Observations	480	428

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

Table 13. Unpacking Zero S*ea*TE (fraction of responses, percent): HRS

	2-Year Ahead	4-Year Ahead
Never work	10.53	15.13
Always work	59.21	47.06
Maybe work	30.26	37.82
Observations	152	119

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

Table 14. Unpacking Negative S*ea*TE: HRS

	2-Year Ahead	4-Year Ahead
Mean	-40.7	-39.1
Std. Dev.	22.6	24
Q25	-50	-50
Median	-40	-40
Q75	-20	-20
Observations	320	298

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