TI 2015-052/V
Tinbergen Institute Discussion Paper

Who can predict their own Demise? Accuracy of Longevity Expectations by Education and Cognition

Teresa Bago d'Uva[1]
Esen Erdogan-Ciftci[2]
Owen O'Donnell[1,3]
Eddy van Doorslaer[1]

[1] Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands;
[2] Novartis, Turkey;
[3] University of Macedonia, Greece.

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

WHO CAN PREDICT THEIR OWN DEMISE? ACCURACY OF LONGEVITY EXPECTATIONS BY EDUCATION AND COGNITION

TERESA BAGO D' UVA[a,b], ESEN ERDOGAN-CIFTCI[c],
OWEN O'DONNELL[a,b,d] and EDDY VAN DOORSLAER[a,b]

May 2015

Abstract

Biased longevity expectations will lead to suboptimal decisions regarding saving, retirement, annuitization and health, with consequences for wellbeing in old age. Systematic differences in the accuracy of longevity expectations may partly explain heterogeneity in economic behaviour by education and cognitive functioning. Analysis of eight waves of the US Health and Retirement Study reveals that individuals with lower levels of education and cognitive functioning report survival probabilities that are less accurate in predicting their in-sample mortality. There is little evidence that the gradients in the veracity of expectations are due to the less educated and cognitively able responding less to changes in objective mortality risks. However, high school dropouts and the least cognitively able report survival probabilities that are less stable and display greater unexplained variability. These disadvantaged groups appear to be less confident in their longevity beliefs, which is justified given that their expectations are less accurate.

JEL Classification: D83, D84, I12, J14

Keywords: Expectations, Mortality, Health, Cognition, Education

---

Acknowledgements: We thank Rob Alessie, Susann Rohwedder, Paul Lau, Martin Salm and Sílvia Sousa, as well as seminar and conference participants at Lancaster, Manchester, Netspar (Paris), SSPH+ (Grindelwald), ASHE (Los Angeles) and the University of Southern California for helpful comments. Early stages of this research were supported by Netspar under theme *Health, income, work and care across the life cycle II*. The research has also been supported by the National Institute of Aging (R01AG037398-01). Owen O'Donnell acknowledges support from the project: *Thalis Panteion Investigating Crucial Interdisciplinary Linkages in Ageing Societies*, which is part of the *Education and Lifelong Learning* programme of the Greek Ministry of Education, Religion, Culture and Sport (MIC 380266). Teresa Bago d'Uva was funded by a Veni grant of the Dutch Organisation for Scientific Research (NWO).

a Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands.
b Tinbergen Institute, the Netherlands
c Novartis, Turkey
d University of Macedonia, Greece

1. INTRODUCTION

Decisions about saving, retirement, pension annuitization and investment in human capital are commonly assumed to depend on predicted longevity. Biased expectations will lead to sub-optimal choices and reduced wellbeing. Inadequate retirement savings as a consequence of excessive pessimism regarding survival chances might result in an impoverished old age, while undue optimism in evaluating the mortality risk associated with illness may cause non-adherence to medication, inadequate adjustment of health behaviour and further deterioration in health. If ability to acquire health knowledge, perceive risks and process information differ by education and cognitive functioning, then the accuracy of longevity expectations would be anticipated to vary with these characteristics. This may help explain the accumulating evidence of heterogeneity by education and cognition in behaviours that would optimally arise from planning household finances over a lifetime horizon (Fang et al. 2008, Banks et al. 2010, Christelis et al. 2010, Smith et al. 2010, Behrman et al. 2012, Agarwal and Mazumder 2013).

We use eight waves of the US Health and Retirement Study (HRS) of older Americans to examine whether ability to predict longevity varies with education and cognitive functioning. We find that even high school dropouts and the least cognitively able report survival probabilities that predict within sample mortality. However, these groups are less accurate than their more educated and cognitively able contemporaries. This is relevant to positive analysis of economic and health behaviour and to normative evaluation of inequality in the resulting outcomes. It also has implications for the design of policy related to financial advice and health promotion.

To cast light on the source of variation in the accuracy of longevity expectations, we examine whether there are differences by education and cognitive functioning in the extent to which new information arising from the onset of illness is incorporated into subjective survival probabilities. This extends analyses based on the first two waves of the HRS that examined how survival expectations are revised on average (Hurd and McGarry 2002) and by the smoking status of the respondent (Smith et al. 2001). Exploiting many more waves of HRS to identify health changes that could be used to revise expectations, we do not find that the response to the onset of health conditions is systematically and consistently lower among the least educated and cognitively able.

There are clear differences in the stability of expectations over time. There is less persistence in the survival probabilities reported by individuals in the lowest education and cognition groups. Interpreted within a Bayesian learning framework, this would suggest that these individuals have less confidence in their prior estimate of the probability of survival. The same groups report survival probabilities that display greater unexplained variation. High school dropouts and the least cognitively able appear to find it more difficult to estimate their longevity: their estimates fluctuate more over time, are explained less by objective risk factors and are weaker predictors of actual mortality.

Education and cognition are highly but not perfectly correlated. When we allow heterogeneity in both dimensions simultaneously, the difference in the accuracy of expectations by education is stronger. This would suggest that low education is associated with a knowledge deficit, and possibly also a deprivation of experience, that hampers the evaluation of survival chances. Using data on numeracy available only in later waves of the survey, we establish that differences in the veracity of expectations by numeracy are considerably stronger than those by education and cognition. Individuals with poor numerical skills either formulate the least accurate longevity beliefs or have the greatest difficulty in expressing their beliefs in a probability.

Poor numeracy and cognitive skills are known to impede understanding of probabilities (Bruine de Bruin et al. 2000, Peters 2008). Of particular relevance to our findings are studies showing that low numeracy is associated with lack of comprehension of medical advice concerning treatments associated with different survival probabilities (Schwartz et al. 1997, Reyna and Brainerd 2007, Zikmund-Fisher et al. 2007). The less numerate and cognitively able may therefore be expected to have difficulty utilizing information on mortality risks and formulating a longevity belief in terms of a survival probability.

This limited ability to think probabilistically about longevity, and possibly other prospects, is a potential explanation of the strong correlation between cognitive functioning and decision taking with respect to household finances (Banks and Oldfield 2007, Banks et al. 2010, Smith et al. 2010, Christelis et al. 2010, Behrman et al. 2012, Agarwal and Mazumder 2013), health insurance (Fang et al. 2008) and health (Gottfredson and Deary 2004, Cutler and Lleras-Muney 2010). Individuals with limited education and/or low cognitive ability may be less likely to save for retirement because they have difficulty making a probabilistic assessment of the horizon over which they should plan. The same constraint offers a potential explanation of experimental evidence demonstrating that individuals with low education, limited numerical ability and poor financial literacy skills have greater difficulty valuing an annuity (Brown et al. 2013). If one finds it difficult to evaluate the probability of living to 75, then the value of insurance against longevity risk will be even more difficult to establish. More generally, incapacity to form accurate longevity expectations may be one factor contributing to impaired decision-making ability and resulting in poor quality decisions (Choi et al. 2014).

Variation in the accuracy of subjective survival probabilities is relevant not only to evaluation of observed differences household finances. It may also help explain the well-established education gradient in health, which is increasingly attributed, at least in part, to a knowledge and cognition deficit that impedes the ability of the less educated to process information on risky health behaviours (Kenkel 1991, Cutler and Lleras-Muney 2010). Lange (2011) finds that the less educated are less likely both to incorporate objective risks of cancers into subjective assessments of those risks and to take preventive measures to reduce the risks. While we do not find that the survival chances reported by the least educated are least responsive to a condition such as cancer, we do show that this group is least successful in predicting longevity. This may distort decisions taken with respect to preventive care and lifestyle choices with consequences for health.[1]

Other studies have established that individuals are able, on average, to report subjective survival probabilities that predict within sample mortality in the HRS (Smith et al. 2001, Hurd and McGarry 2002, Siegel et al. 2003, Hurd 2009) and other surveys (Van Doorn and Kasl 1998, Kutlu-Koc and Kalwij 2013). We extend the evidence by documenting heterogeneity in the

[1] Without examining the education gradient, Hurd (2009) finds that smokers and heavy drinkers overestimate their survival chances relative to actual survival rates within the HRS sample. On the other hand, Smith, Taylor et al. (2001), also using the HRS, find that smokers are more responsive to smoking–related mortality risks. Using Dutch data. Kutlu Koc and Kalwij (2013) find that subjective longevity expectations are less responsive to smoking, alcohol, overweight and obesity than is actual longevity.

accuracy of longevity expectations by education and cognition. In doing so, we move beyond the question of whether a representative agent possesses and can process information on his longevity prospects that are instrumental to solving lifecycle planning problems. Rather, we address a question that has a distributional motivation: Which individuals are more likely to hold inaccurate expectations and, consequently, make suboptimal decisions?

Elder (2013) challenges the claim that reported subjective survival probabilities provide useful information, demonstrating that life table survival probabilities do a better job of predicting the in-sample mortality of HRS respondents. Indeed, the life table probability should fail to predict death by a certain age, on average, only to the extent that the life table does not incorporate cohort effects. The more pertinent issue is whether, given the life table probability, the subjective probability provides additional information that explains inter individual variation in longevity. Elder remains sceptical after finding that the perceived deviation of survival chances from the life table average accounts for only 35% of the cross individual variation in reported survival probabilities in the HRS. This suggests that there is a great deal of noise in reported survival chances. We examine whether the signal to noise ratio is lowest for the least educated and cognitively able.

Delavande and Rohwedder (2011) are considerably more optimistic regarding the information content of subjective survival probabilities, arguing that the measure provides a meaningful proxy for longevity that allows the socioeconomic gradient in mortality to be estimated without observing actual mortality. They show that the gradient by wealth tercile in the reported probability of survival to the age of 75 is remarkably similar to the gradient in actual survival to 75 among the HRS panel respondents who potentially could have reached that age. On the other hand, the gradient in reported probabilities by education (and income) is markedly shallower than that in actual survival (see Delavande and Rohwedder 2011, online Appendix). This suggests that the accuracy of subjective survival probabilities differs by education.[2]

Rather than comparing the education (and cognition) gradients in subjective and actual longevity, we assess heterogeneity in the accuracy of expectations directly by testing whether subjective survival probabilities reported by individuals with lower levels of education (cognition) are less predictive of actual survival than are the probabilities reported by those with higher levels of education (cognition). We confirm that the survival expectations reported by the less cognitively able and educated are indeed less accurate and demonstrate that this is only partly driven by the greater tendency of these groups to give focal responses of 0, 50 and 100 percent survival chances.

The paper is organized as follows. Section 2 presents the HRS data on subjective survival probability, education, cognition and mortality. The extent to which subjective survival probabilities predict mortality and how this varies with education, cognition and numeracy is examined in Section 3. In Section 4, we examine the extent to which survival chances are revised in response to new information on objective mortality risks and test whether this response differs by education and cognitive functioning. The final section concludes.

[2] Using Dutch data from a smaller, younger sample in which fewer deaths are observed, Kutlu-Koc and Kalwij (2013) find no education gradient in subjective survival probabilities, but neither do they obtain a significant gradient in actual mortality.

2. DATA AND DESCRIPTIVES

We use data mostly from the original cohort of the US Health and Retirement Study (HRS), which is a nationally representative sample of individuals born between 1931 and 1941, who were aged 50-62 years in 1992 and have been interviewed every two years since. We include spouses of the original HRS cohort and of other HRS cohorts only if they were born between 1931 and 1941.[3] We analyse a sample of individuals who were all born within this decade.

We took all variables from the RAND HRS data files (St. Clair et al. 2009). We use data from wave three (1996), from which point consistent measures of cognitive functioning are available, until wave ten (2010), at which point the cohort was aged 68-80. Some individuals belonging to the 1931-41 birth cohort enter our sample after wave 3 either because of missing covariates in that wave or because they are new spouses or spouses of cohorts added to the survey in later waves. The analyses are based on 32,437 individual×wave observations for which there is complete information on subjective survival probabilities, vital status, education, cognition and covariates.

2.1 SURVIVAL EXPECTATIONS

The HRS asks the respondent to report the chance that s/he will live to a specified target age on a scale of 0-100, with 0 indicating 'absolutely no chance' and 100 corresponding to 'absolutely certain'. The target age depends on the respondent's current age. All those aged 65 and younger are asked to report the percentage chance of living to at least the age of 75. In the first four waves, those aged 75 or less (including those <66) were asked to evaluate their chances of living to at least the age of 85. From the fifth wave, instead of the target age being fixed at 85 for those older than 65, it varied with the respondent's age. The target ages used in the analysis by wave and age are identified in (online) Appendix A Table A1.

We deal with differences in the target age by standardising the reported probability on the implied life table probability of survival to the specified target given the respondent's age and gender.[4] Specifically, we take the difference between the reported probability and the life table probability. A positive value does not necessarily imply that the respondent is overly optimistic. Given health, health behaviour and family history, his or her survival chances may very well be objectively above the actuarial average based only on age and gender. The reason for the transformation is principally to overcome the differences in the questions and so maximise the number of observations and waves from which information can be used. In addition, taking the deviation from the life table probability eliminates the difference in longevity expectations that is

---

[3] Direct members of the AHEAD (born<1924), CODA (1924-30), War Babies (WB) (1942-47) and Early Baby Boomers (EBB) (1948-53) cohorts are not included since they were not born between 1931 and 1941. Our estimation sample consists of 32,437 individual×wave observations from the HRS original cohort (core members & spouses), 154 from the AHEAD/HRS overlap, 85 spouses from AHEAD, 24 spouses from CODA, 130 spouses from WB and 2 spouses from EBB.

[4] Life table probabilities are produced by the National Center for Health Statistics (http://www.cdc.gov/nchs/products/life_tables.htm#life). These probabilities are available within the RAND HRS data files. The probability of survival to age T is calculated as the number of a (sex specific) birth cohort surviving to that age divided by the number of the same cohort surviving to the respondent's age at the time of the survey. Life tables for the year of the survey are used, except that that 2006 life table is used for wave 9 since the 2008 data were not yet available.

implicit when the same probability of survival to a given target age is reported by individuals who differ in age and sex.

The mean reported survival probability in our estimation sample is five percentage points less than the corresponding life table probability, although the median is closer (Table 1). The average degree of pessimism is even greater than that indicated by this deficit since the life table tends to underestimate the longevity of current cohorts. There is a great deal of variation in the difference between the reported and life table probabilities: the standard deviation (29.5) is about five times the mean. Variability is to be expected since the life table gives an average, around which mortality risks obviously do vary.

As has previously been recognised (Hamermesh 1985, Hurd and McGarry 1995, Gan et al. 2005, Perozek 2008, Elder 2013), women tend to be more pessimistic than men with respect to survival chances at all ages (Figure 1). Despite this, we find no evidence that the accuracy of subjective survival probabilities in predicting mortality differs significantly by gender and allowing for such a difference does not affect the estimated heterogeneity by education and cognition. We therefore present estimates from the more parsimonious specification that imposes homogeneity in the accuracy of expectations by gender.

The age pattern in the deviation of the subjective from the life table probability is similar across genders (Figure 1). Between the ages of 54 and 64, both males and females become more pessimistic. This may reflect failure to revise the probability of survival to the target age of 75 sufficiently upward as the respondent nears that age (Elder 2013). After the age of 65, the difference between the subjective and objective probability narrows and becomes positive beyond the age of 70 for males and 75 for females. This tendency for individuals to become optimistic in their survival expectations as they enter old age is well-established (Hamermesh 1985, Hurd et al. 1998, Bissonnette et al. 2012, Elder 2013, Ludwig and Zimper 2013, Wu et al. 2014). In these data, it appears to be mainly driven by spikes at the ages of 70 and 75, and to a lesser extent at 66, that coincide with changes in the target age. This suggests a tendency to insufficiently adjust the reported probability downward when the target is raised and the objective likelihood of reaching it falls. Between changes in the target age (67-69, 71-74, >75), the deviation of the subjective from the life table probability tends to fall.

About 4% of observations (across all waves) do not answer the survival probability question (Table 1).[5] This may reflect difficulty in completing the abstract task of formulating and reporting a probability. Focal answers of 0, 50 and 100 percent may be given by individuals who have trouble performing the exercise. Around one-quarter of respondents report a value of 50 percent. Such bunching has been attributed to an extreme form of rounding (Gan et al. 2005, Manski and Molinari 2010, Kleinjans and Van Soest 2014), epistemic uncertainty (50%='I haven't a clue') (Fischhoff and Bruine De Bruin 1999, Bruine de Bruin and Carman 2012) and extreme ambiguity (50%='I am very unsure') (Daniel Hill, et al. 2004, Hudomiet and Willis 2013). About 5% of the sample give an answer of 0 percent to the survival expectations question, while three times that fraction answer 100 percent.

[5] In modelling mortality we exclude these observations. Including them reveals that non-response is correlated with a higher risk of death in the subsequent period.

Table 1 & Figure 1 here

2.2 Education and Cognitive Functioning

Education groups are formed from years of schooling and highest qualification obtained, as follows: i) high school drop outs and General Educational Development (GED) (0-11 years of schooling and no qualification other than GED)[6]; ii) high school graduate (12 years of schooling and high school diploma); some college (more than 12 years of schooling and high school diploma or GED); and, iv) college and above (bachelor degree or higher).

There is a pronounced education gradient in longevity expectations that is in the same direction as actual differences in mortality risks.[7] On average, high school dropouts report a survival probability about 13 percentage points below the actuarial average, while college graduates report a mean probability one point above the life table average (Table 1).

The least educated appear to have the greatest difficulty in providing subjective survival probabilities. Across all waves, around 9% of high school dropouts do not answer the question, compared with less than 2% of college graduates. When they do answer, the lesser educated are significantly more likely to give a focal response of 0, 50 or 100 percent. Kleinjans and Van Soest (2014) find, using the HRS, that modelling non-response and focal responses makes little difference to the estimated education gradient in longevity expectations.[8] Nonetheless, we examine whether estimates of education-related differences in the accuracy of longevity expectations are robust to dropping observations giving focal responses.

In addition to higher non-response and focal response rates, the less educated have been found to be more prone to give answers to subjective probability questions that violate monotonicity and adding up (Dominitz and Manski 1997, Dominitz and Manski 2006, Delavande and Rohwedder 2008, Van Santen et al. 2012).[9] On the whole, it seems likely that there is more noise in the subjective probabilities reported by the least educated.

The HRS contains measures of cognitive functioning in several domains based on well validated measured tests (Ofstedal et al. 2005, Fisher et al. 2012). We restrict attention to the measures that are directed at respondents of all ages and have been fielded in a consistent format since wave 3.[10] These measures relate to *episodic memory* and *intact mental status*. The former enables recollection of experiences and specific events from the past (Tulving 1972). It may be required

---

6 The GED is an achievement test intended to identify cognitive aptitude equivalent to that of a high school graduate. GED is commonly taken by immigrants, war veterans, the home schooled and high school dropouts. We categorise those with GED together with high school dropouts since the purpose of the classification is to demarcate levels of educational experience (Heckman et al. 2014) Cognition is identified separately using the tests in the HRS.

7 Population vital statistics indicate that in 2007 the mortality rate of high school dropouts aged 55-64 was 2.4 times higher than that of individuals with at least some college education in the same age range (Xu et al. 2010) (Table 26). The education gradient in mortality rates in our sample is given in Table 4.

8 However, taking an approach that imposes fewer identifying assumptions, Manski and Molinari (2010) find that after allowing for non-response and rounding even an age difference in subjective survival probabilities is not discernible.

9 On the other hand, a measure of coherence in subjective expectations of stock market returns was found to be unrelated to education in a US sample (Gouret and Hollard 2011).

10 At least for first interviews. Some of the measures are, from 1998 onwards, available only for respondents older than 65 or those interviewed for the first time.

for reasoning (Smith et al. 2010). In the HRS it is assessed through two word recall tasks. The interviewer reads a list of ten common nouns (e.g., book, child, hotel) and the respondent is asked immediately to recall as many as possible in any order.[11] After five minutes or so, during which other questions are answered, the respondent is again requested to recall the words. The total recall score is the sum of the number of words recalled in these immediate and delayed recall tests (ranging from 0 to 20).

Mental status refers to the intactness of the neuro-cognitive system essential to communication and learning (Smith et al. 2010). It is measured through parts of the Telephone Interview for Cognitive Status (TICS) (Brandt et al. 1988). One dimension of mental status is *executive functioning*, which refers to the cognitive processes that facilitate the use of past experience in current action (National Center for Learning Disabilities,. 2012). These processes are used in planning, organization and management and are likely to be of considerable importance in relation to decisions taken concerning pensions, saving, health insurance, retirement and health behaviour. The cognitive processes relevant to the tasks we use from the HRS are working memory, attention and problem solving. Working memory refers to the ability to store and process information simultaneously (Baddeley and Hitch 1974). It, and the other two processes referred to, is assessed by asking respondents to subtract 7 from 100 and continue subtracting 7 from the answer for a total of five subtractions. The test score is the number of correct subtractions (0-5). This serial 7's test is part of the Mini-Mental State Examination (Folstein et al. 1975), in which it assesses the attention and calculation dimensions of mental status. A second test, which additionally assesses processing speed, asks the respondents to count backwards as quickly as possible, starting from 20 for 10 continuous numbers. Respondents are allowed two trials for this exercise. Scores are recorded as 0 if incorrect or "don't know/unable to do" on both tries, 1 if incorrect on the first try but correct on the second try, and 2 if correct on the first try. Three additional elements of mental status are assessed by orientation in time, ability to name objects and to recall the names of the President and Vice-President.[12] The total mental status score sums those of the serial 7's, backwards counting, date/object/president naming tests and ranges from 0 to 15.

Following Ofstedal, Fisher et al. (2005), we aggregate the measures of episodic memory and mental status into a total cognition score, which ranges from 0 to 35.

Cognitive functioning, particularly episodic memory, decreases with age (Anderson and Craik 2000). Our aim is not to identify the extent to which expectations become biased as individuals age and cognitive functioning declines. Rather, we wish to assess whether individuals who enter middle age with relatively low cognitive ability hold less accurate expectations of their length of life. If this is the case, it may lead to suboptimal decisions that result in lower wellbeing at older ages. Consistent with this objective, we test whether there is heterogeneity in the accuracy and the updating of survival expectations by cognition measured at the first time an individual enters

---

11 To avoid respondents acquiring familiarity with the words, they are asked to recall from a different list in each successive wave.

12 Orientation in time is tested by asking the date and the day of the week. Object naming asks questions such as: "What do you usually use to cut paper?". The score of each test is the sum the number of correct answers and ranges from 0 to 4 (date), and 0 to 2 (object and President, Vice-President). These questions are asked of all respondents in wave 3 but only of new entrants and those aged 65 and above from wave 4 onwards.

our sample.[13] We standardize these first-entry cognition scores for age and sex, and examine heterogeneity across quartile groups of these standardized scores.[14]

Given the positive correlation between cognitive functioning and educational attainment, the former is also positively correlated with subjective survival chances (Table 1, middle panel), although not as strongly as is education.[15] Reporting the probability of survival to a specific age is a task that involves cognitive processes, including abstraction, concentration, retrieval and processing of information from health experiences, comprehension of the concept of probability, and calculation. One would expect ability to accomplish this task to be related to cognitive functioning. This appears to be evident in the data. The percentage of the lowest cognition group that does not report a survival probability is more than three times that of the highest functioning group.

Previous analyses of the HRS reveal that respondents who score poorly on cognition tests are less likely to respond to a variety of subjective probability questions and those who do are more likely to give focal responses of 0, 50 or 100 percent (Hurd and McGarry 1995, Kleinjans and Van Soest 2014, Johannes Binswanger and Martin Salm. 2014).[16] This is evident in Table 1, although the relationship between cognition and the likelihood of reporting a probability of 50 percent is neither monotonic nor as strong as that between education and this response.

Numeracy is measured in every second wave from wave 6 (2002).[17] A numeracy score is given by the sum of correct answers to three questions involving calculation of a percentage, division and compound interest.[18] The score is age-sex standardized as for cognition and quartile groups formed using the first measure taken for each individual.

As for education and cognition, low numeracy is strongly associated with lower subjective survival probability, higher non-response and a greater tendency to report survival probabilities of 0 and 100 percent (Table 1, bottom panel).[19]

13 For most individuals, this is wave 3. For some spouses and those with missing values at wave 3, it is later waves. While we examine heterogeneity in the accuracy of survival expectations with respect to the initial measure of cognition, the models estimated allow mortality risk to vary with time varying cognitive functioning.

14 The standardised score is the residual from a linear regression of the score on a full set of sex-specific age-year dummies. We form gender specific quartile groups.

15 The correlation coefficient between years of education and the total cognition score is 0.426 (p-value<0.001). Mean age-sex standardised cognition scores are: high school dropouts – 28.9, high school graduates – 46.5, some college- 53.9, college graduates – 63.1. Examination of the mean of the difference between the reported and life table survival probability across levels of cognition within an education category, and vice versa, reveals that longevity expectations vary in both dimensions, although the education gradient is stronger and more consistent (see Appendix Table A2).

16 Kleinjans and Van Soest (2014) find that low cognitive functioning, measured by word recall score, is not significantly correlated with a higher probability of reporting a 50% probability, although it is associated with a greater tendency to round responses.

17 From wave 6, numeracy is also measured in the wave in which a new respondent enters the panel.

18 The questions are: 1) If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease? 2) If 5 people all have the winning numbers in the lottery and the prize is two million dollars, how much will each of them get? 3) Let's say you have $200 in a savings account. The account earns 10 percent interest per year. How much would you have in the account at the end of two years? Respondents who answer 1) and 2) incorrectly are not asked 3). Answers of 'Don't know' and refusals are considered wrong.

19 The correlation coefficient between years of education and the numeracy score is 0.4522 (p-value<0.001). Education differences in longevity expectations are stronger than those by numeracy. The mean deviation of the

2.3 Mortality

The death of a HRS cohort member is reported by a relative contacted by an interviewer. A sample participant is presumed to be alive if s/he cannot be contacted but there is information obtained from a relative or another source that the person is alive. If no information is available, possibly because the respondent has dropped out of the HRS when alive and refused to participate further in the survey, then vital status is classified as unknown. We trace the vital status of each respondent from wave 3 (or the wave in which s/he first appears in our sample) until wave 10, which corresponds to a period of 14 years (1996-2010). Of the 6816 individuals who appear in our sample at some time, 1347 (19.8%) die by wave 10 and the vital status is unknown for 256 (3.8%).

In Table 2 we present the 14 year mortality rates for all respondents in the estimation sample who are observed in wave 3 (N=5210).[20] Over the period, around 21% of the respondents died. There is a clear gradient in mortality by both education and cognitive functioning. The mortality rate among high school dropouts is around twice that of college graduates. The rate for those in the bottom quartile group of cognition is slightly less than three quarters greater than that of those in the top quartile group.

Mortality is also strongly correlated with longevity expectations. Of the 312 individuals who report a zero probability of survival to the age of 75 in 1996, 142 (45%) do not survive to 2010. The mortality rate falls steadily as the reported survival probability rises until reaching the top of the distribution. The mortality rate among the one fifth of respondents who report a 100 percent chance of surviving to 75 is around 20%, which is substantially greater than the rate for those reporting a probability between 51 and 99 percent. While a reported probability of zero does appear to contain information on mortality risk, the accuracy of a focal response of 100 is particularly low.[21]

Table 2 here

The relationship of longevity expectations to mortality risk by education and cognition is shown in Figures 2a and 2b respectively. Those who reported a zero chance of survival experienced an elevated mortality rate at all levels of education. The fall in the death rate as the reported probability of survival turns positive is steepest for those with a college degree. This is consistent with the hypothesis that (low) longevity expectations of the most highly educated are most accurate. However, one should be careful of reading too much into the variation at the bottom of the distribution since only 20 individuals (2%) in the highest education group report a zero probability of survival to 75 (see online Appendix A, Table A4). In contrast, 14% of high school dropouts report a zero survival chance. A distinct kink in the death rate at a reported survival

---

reported from the life table survival probability differs significantly across education levels within each quartile group of numeracy. On the other hand, within two of the four education groups (*high school dropouts* and *some college*), there is no significant difference in expectations by numeracy (see Appendix Table A2)

20 In this table, we restrict attention to those observed in wave 3 in order to compute the mortality rate over a fixed period (14 years) and because all respondents reported survival chances to a target age of 75 in that wave.

21 The mortality rate among individuals who do not give a response to the survival probability question in wave 3 is 28.9% compared with 23.5% among those who do respond. The mortality rate among those missing on covariates is also slightly higher. Mortality in our estimation sample is therefore slightly lower than it is in the full HRS cohort aged 54-66 in wave 3. However, the gradients by education, cognition and longevity expectations are affected little by this selection. See Appendix, Table A3.

chance of 50 percent is evident only for college graduates, while the rate turns upwards at a response of 100 percent for all groups. On the whole, it is not immediately evident that longevity expectations reported by the more highly educated are more accurate. Likewise, Figure 2b does not support the hypothesis that the more cognitively able report more accurate expectations.

This simple analysis may not be indicative of heterogeneity in the accuracy of longevity expectations. The mortality rate between 1996 and 2010 of high school dropouts is twice that of college graduates (Table 2). This gradient is not fully reflected in the expectations, which suggests that the relationship between the actual and perceived mortality risks does vary with education.[22] Besides the lack of control for covariates, Figure 2 utilises survival probabilities reported in wave 3 only, which may precede the onset of health conditions. To more fully exploit the information potentially in the data, we estimate the mortality hazard as a function of longevity expectations.

Figure 2 here

3. ACCURACY OF LONGEVITY EXPECTATIONS

We assess the accuracy of longevity expectations by examining the extent to which the deviation of the reported from the life table probability of survival to a target age predicts death within the period of observation. Our main focus is on whether this predictive power varies with educational attainment and cognitive functioning.

3.1 MODELLING LONGEVITY

We observe a sample of individuals who were alive and aged between 54 and 65 at the time wave 3 of the HRS was conducted. We observe whether each individual is dead by each subsequent wave (4-10). Using these data, we model the probability of death at each age by specifying a discrete time hazard. This can be implemented by pooling the data over all waves from the third until the tenth, or until the wave in which the individual is recorded as being dead, and estimating a binary model of vital status at each wave (Jenkins 1995).

Let τ_i indicate the age of individual *i* when she is first observed. Each individual is observed until the wave in which her death is recorded or wave 10 if she survives.[23] Let $t = \tau_i + s_i$ indicate age at which a spell is completed, such that s_i represents the years from first observation until either death or the end of the observation period. The duration elapsed until the age at death is modelled via the probability of a spell ending at each t given survival to that age, which is the discrete time hazard rate (Jenkins 1995),

$$h_{it} = P\left(T_i = t \middle| T_i \geq t; X_{it}\right), \quad (1)$$

[22] The mean reported probability of death by 75 of high school dropouts in wave 3 is around 60% greater than the probability reported by college graduates. This is considerably less than the 100% difference in the actual death rates by 2010, although it is recognised that death by 2010 is not the same event as death by the age of 75. We are defining the reported probability of death by 75 as one minus the reported probability of survival to 75.

[23] A few individuals drop out of the sample without their vital status being established. Such individuals are used in the estimation until the wave in which they drop out, after which they effectively become right-censored observations.

where T_i is a discrete random variable representing the age at which a spell ends.[24] These hazards map into the probabilities of varying lengths of survival.

If one ignores potential unobservable heterogeneity in the probability of survival to the age at first observation, then the log-likelihood for the discrete time duration model is equivalent to that for a binary indicator of vital status at each wave (Jenkins 1995).[25] Let y_{it} be an indicator of the occurrence of death at age *t*. For those who are still alive in wave 10, $y_{it}=0$ at all observed ages. For those who die during the observation period, $y_{it}=0$ at all ages except that in the wave in which death is recorded, when $y_{it}=1$. Such individuals obviously do not appear in subsequent waves. The log-likelihood of observing the vital status history of the sample can be written as

$$Log\,L=\sum_{i=1}^{n}\sum_{t=\tau_i}^{\tau_i+s_i} y_{it}\log h_{it}+\left(1-y_{it}\right)\log\left(1-h_{it}\right). \qquad (2)$$

We specify the hazard rate to be complementary log-log, which is the discrete-time counterpart of the hazard for an underlying continuous-time proportional hazards model (Prentice and Gloeckler 1978, Jenkins 1995). We estimate hazards of the form,

$$h_{it}=1-\exp\left[\exp\left(f\left(\tilde{P}_{it-2},\mathbf{Z_{it\text{-}2}}\right)+\theta_g\left(t\right)\right)\right] \qquad (3)$$

where $f(\ )$ is a function of $\tilde{P}_{it-2}$ - the difference between the reported and life table probability of survival to a target age measured at the previous wave, which is two years previous to the point at which vital status is established – and covariates, $\mathbf{Z_{it\text{-}2}}$. The function $\theta_g(t)$ represents the baseline hazard of death at each age for a given gender (g), which we specify as piecewise-constant by including a dummy for each gender-age combination giving a semi-parametric specification of the discrete-time duration model (Jones et al. 2012).

We estimate a variety of models distinguished by the specification of $f(\ )$. To begin with, this function permits no interaction between longevity expectations and either educational attainment or cognitive functioning. Estimates from this specification will confirm whether subjective survival probabilities predict the length of life. We then allow for an explicit interaction between longevity expectations and one or both of education and cognition. We test whether the extent to which expectations predict the age at death varies with each of these factors. Using the data from wave 6 onwards, we examine whether the predictive power of expectations varies with numeracy.

---

24 We do not use the precise age at death but rather whether an individual dies between waves that are separated by around two years. We assume that the underlying survival time is continuous and that deaths occur at a uniform rate within each two year period (Jones et al. 2012) (pp.182-218).

25 This is not the case if unobserved heterogeneity in survival to the time of inclusion in the stock sample is taken into account since then terms corresponding to the probability of being alive at each wave preceding the observation period do not cancel out.

3.2 EMPIRICAL SPECIFICATION

We first examine estimates with no control for covariates other than the gender-age dummies.[26] The purpose is to test whether respondents are able to use all available information, including that correlated with observable characteristics, to formulate and report survival probabilities that predict longevity. Subsequently, we present estimates from two specifications that cumulatively add sets of covariates. We first add socioeconomics characteristics: income, wealth, employment status, race and marital status, in addition to education (see Appendix A, Table A5 for variable descriptions). Objective mortality risks do vary with wealth and race, for example. To the extent that individuals are knowledgeable of this variation and incorporate it into their expectations, the predictive power of expectations should decline when control is made for these covariates. But without adding these controls there is potentially bias due to heterogeneity in use of the reporting scale – for given beliefs about survival prospects – by covariates that are also correlated with mortality risks.

Indicators of health and health behaviour are left out of the first set of covariates in order to test whether individuals can utilize the information contained in these measures to predict their longevity, and whether the accuracy of predictions based on observable health information varies with education and cognition. We then examine the accuracy of longevity expectations conditional on indicators of health, smoking and parental longevity and test whether, conditional on observable mortality risk factors, longevity expectations are formed on the basis of additional information available to the individual, but not observed in the survey data, that is predictive of mortality. If, conditional on these additional controls, reported survival probability is still able to predict mortality, then this would indicate that individuals hold and utilize private information about their survival chances over and above that recorded in a detailed survey like the HRS. Again, we examine whether possession and utilization of such private information varies with education and cognition.

We include indicators of the previous existence and recent onset of eight diseases or risk factors: lung disease, cancer, diabetes, heart disease, stroke, psychiatric illness, high blood pressure and arthritis. Additionally, we include measures of current cognitive functioning[27] and control for psychological distress (depression and anxiety) through a shortened version of the CES-D scale[28] (Steffick 2000).We measure functional status by three Activities of Daily Living (ADL) – Instrumental ADL indices following Wallace and Herzog (1995) using only the ADL-IADL questions asked consistently across waves. We also include indicators of mobility and large muscle problems, as well as fine motor skills. Finally, we control for whether the individual was hospitalized in the previous two years, BMI and for whether the person is a current smoker or a former smoker. We condition not only on current but also on lagged measures of psychological

[26] We have estimated models that indude an interaction between longevity expectations and gender. The respective term is never significant and its inclusion does not alter the gradient in the predictive accuracy of longevity expectations by education or cognition. We therefore present estimates from models that no not allow this explicit heterogeneity by gender.

[27] Current cognition is captured by measures that are available in every wave: the total recall score and the sum of the scores of the serial 7's and backwards counting tests. Both are age-sex standardized. In the models that explicitly allow for heterogeneity in the accuracy of longevity expectations by initial cognitive functioning, the quartile groups of the total cognition score are induded in addition to the current cognition measures.

[28] Center for Epidemiologic Studies Depression Scale (Radloff 1977).

distress, functional limitations, hospitalization and BMI. The complete list and their definitions are given in Appendix A, Table A5.

All these health measures are objective in the sense that the respondent is asked to report a diagnosed condition, experienced limitation or hospital admission. Respondents are also asked to subjectively assess their health as *excellent*, *very good*, *good*, *fair* or *poor*. Answering this question may involve the processing of information on the diagnoses, symptoms, limitations and treatments reported in the previously mentioned measures, along with other complaints and feelings considered relevant to the individual's overall health experience. The respondent may rate not only his or her current functioning and wellbeing but also weigh a prognosis of future health, possibly including survival chances. A response of *poor* may reflect pessimism about survival chances given the prognosis of some current illness. In this case, this ordinal self-assessed health (SAH) question would be at least a partial substitute for the rather more difficult to answer survival probability question (Benítez-Silva and Ni 2008). Indeed, there is abundant evidence that SAH is predictive of mortality even conditional on physiological measures of health (Idler and Benyamini 1997). We purposefully do not control for SAH since the aim is to take account of observable longevity-relevant information but not to control for the processing of that information in some variable other than subjective survival probability.

Through genetics, shared environments and inherited health behaviour, parental longevity provides information on life expectancy. We include indicators of whether the respondent's mother and father are deceased, the age/age-at-death of each parent and an interaction between the indicator of vital status and age-at-death.

3.3 Accuracy of expectations on average

Table 3 shows the coefficient on the longevity expectations variable – the difference between the subjective and life table probability of survival to a target age – in mortality hazard models that do not allow for heterogeneity by education or cognition.[29] Taking the exponent of the significant coefficient of -0.0168 in the model without covariates gives an estimated hazard ratio of 0.9833 with respect to a percentage point increase in the deviation of the subjective probability above the life table probability. That is, the hazard rate of death is 1.67% lower for each unit increase in the longevity expectations variable. Since the magnitude of the coefficient is reasonably small, it approximates this relative change in the hazard rate, which provides a convenient interpretation of all the coefficients in the table.[30]

As anticipated, controlling for socioeconomic covariates that are correlated with mortality risks reduces the magnitude of the longevity expectations coefficient. But it remains highly significant indicating that people utilize other available information on mortality risks – most obviously that obtained from observing their own health – to predict their longevity with some accuracy. The estimate from this model implies that the hazard ratio for a standard deviation increase in the longevity expectation measure is 0.6833, which is very close to that for a standard deviation increase in wealth (0.6572). The relationship between actual and expected longevity is as strong as that between mortality and wealth.

29 The full estimates for the model including all covariates are given in Appendix Table A6.

30 Using $\exp(b) - b \approx 1$ for small b.

When control is made for health indicators, smoking and parental longevity in addition to socioeconomic characteristics (right-hand column), the coefficient falls substantially but remains strongly significant. A percentage point rise in the deviation of the subjective from the life table probability is associated with an approximately 0.5% decrease in the hazard rate, which is 27% of the estimate obtained with no controls (and 36% of the estimate with only socioeconomic controls). Most of the predictive power of the subjective survival probability derives from the incorporation of information on mortality risks that are associated with measured health indicators, smoking and parental longevity. Once these factors are controlled for, individuals apparently possess further information that is useful in predicting their longevity.

Table 3 here

3.4 Accuracy of expectations by education and cognition

In Table 4, we present estimates from models that allow the association of longevity expectations with the mortality hazard to differ across categories of education and cognitive functioning. The table gives the coefficient on the longevity expectation variable for each education/cognition category, i.e. the expectations coefficient for the reference category plus the coefficient for the category-specific interaction with expectations. This approximates the relative effect on the hazard rate of a marginal change in longevity expectations for each group.

Consider the estimates by education reported in the left-hand panel. The point estimate of the expectations effect is increasing with education in all three specifications, which suggests that the better educated are indeed more accurate in forecasting their longevity, although the null of homogeneity is rejected (at 10% significance) only when there is no control for health, smoking and parental longevity (*mortality risks*). Controlling only for age and sex, longevity expectations are a significant predictor of mortality for all education groups. While there is a clear gradient, it should not be overlooked that even high school dropouts are capable of reporting survival probabilities that do, at least to some extent, predict longevity. A unit increase in the subjective survival probability reported by a high school dropout in excess of the life table probability is associated with a decrease in the mortality hazard rate of 1.2%. For college graduates, the equivalent estimate is 2.1% - around three-quarters greater. Individuals with at least some college education are significantly better at predicting their longevity than high school dropouts.

Controlling for socioeconomic characteristics reduces the coefficients but leaves expectations as a significant predictor of mortality for all education groups and does not diminish the education gradient in the accuracy of expectations. Once control is made for the full set of covariates, the expectations reported by high school dropouts no longer significantly predict their longevity, while the expectations variable remains significant for the other education groups. Even after conditioning on a large battery of health indicators, smoking behaviour and parental longevity, individuals who have at least graduated from high school are able to incorporate information in their reported expectations that predicts their longevity. The exploitation of this additional, unmeasured information, and not only differential incorporation of information contained in the indicators of socioeconomic characteristics, health, smoking and parental longevity, appears to give the better educated an advantage in predicting their own demise, although the differences across groups are no longer significant with the complete set of covariates.

In the middle panel, the accuracy of expectations is allowed to differ by cognition, but not education. Controlling only for age and gender, longevity expectations are significant predictors of mortality at all levels of cognitive functioning. Even individuals with a cognition score below the first quartile report survival probabilities that predict longevity. The association between expectations and realisations rises with cognition but not monotonically. The accuracy of expectations for the third quartile group is not significantly different from that of the first quartile group. The significant difference between the top and bottom cognition groups is roughly similar to that between the top and bottom education groups.

Controlling for socioeconomic factors leaves the expectations variable as a significant predictor of mortality for all cognition groups. The gradient in the accuracy of expectations is affected little by adding these controls, although only the top group is significantly different from the bottom. Conditioning on mortality risk factors reduces the coefficient for all groups but leaves expectations as a significant predictor of longevity for three of the four groups. Individuals with the lowest level of cognitive functioning, unlike high school dropouts, appear to form survival expectations on the basis of information on mortality risks that is not captured by the rich battery of health indicators included. While the point estimates suggests that expectations remain more accurate for the highest compared with the lowest cognition group, there is no significant difference. All told, differences in the accuracy of longevity expectations by cognitive functioning are not quite as strong as those by education.

The right-hand column of Table 4 shows estimates from models in which the predictive power of subjective probabilities is permitted to vary with both education and cognition. At each level of education (cognition), we show the effect of the longevity expectations variable for the lowest cognition (education) group. Controlling only for age and sex, survival probabilities reported by all education-cognition groups predict mortality. Within the first quartile of cognitive functioning, the expectations coefficient rises with education but only the difference between college graduates and high school dropouts is significant (at 10%). Among college graduates with the lowest cognitive functioning, a unit increase in expectations is associated with an approximately 1.8% decrease in the hazard rate, compared with a fall of 1.1% for high school dropouts with the same (broad) level of functioning. Among high school dropouts, the estimates suggest a rise in the accuracy of expectations with cognitive functioning that is neither monotonic nor significant. Although it is not apparent in the table, the estimates imply that a percentage point increase (above the life table probability) in the survival probability reported by a college graduate in the top cognition group is associated with a 2.3% decrease in the hazard rate, compared with a 1.1% decrease associated with a marginal change in the survival chance reported by a high school dropout in the lowest cognition group.

Controlling for socioeconomic factors has little impact. After controlling for health, smoking and parental longevity, the residual information in the subjective probabilities reported by high school dropouts in the bottom cognition group no longer significantly predicts longevity. Holding cognition constant, the point estimates suggest a strengthening relationship with rising levels of education, but the difference is only marginally significant (10%) for high school graduates and those with some college education. Differences by cognition are weaker, although the point estimate for the top cognition group is almost three times the magnitude of that for low cognition high school dropouts. The lack of significance of the education and cognition differences in this specification could very well arise from a lack of power given the strong overlap between the two characteristics.

Overall, the analysis indicates that even the least educated individuals with low cognitive ability can formulate and report survival expectations that are accurate in the sense of predicting their own longevity. But higher educated and more cognitively able individuals are significantly and substantially better at predicting their survival chances. For all education and cognition groups, predictive power falls substantially when control is made for measurements of health, smoking behaviour and parental longevity, suggesting that the information on mortality risks associated with these indicators is incorporated into expectations. Remaining differences in the accuracy of expectations by education (in particular) and cognition, although they are not always significant, suggest that the better educated and more cognitively able are not only better in utilizing observed longevity-relevant information, but they hold more additional information on mortality risks that is not observed in the covariates.

Table 4 here

3.5 Role of focal responses

As observed in section 2.2, individuals with lower educational attainment and cognitive functioning are more likely to report survival chances of 0, 50 and 100 percent. If these focal responses reflect difficulty in comprehending the task of reporting probabilities, are due to a high degree of rounding or reflect a good deal of ambiguity over longevity, then they should contain less information on mortality risks. Their higher prevalence among the responses of the less educated and cognitively able could then potentially explain why these groups are less successful in predicting their longevity. To assess the validity of this interpretation of the results, we examine whether they are robust to dropping observations giving a focal response.

The bottom panel of Table 3 shows the impact on the longevity expectations coefficient of dropping focal responses from models that do not permit heterogeneity by education and cognition. Dropping responses of a 50:50 chance of survival has little or no impact on the estimate. This is consistent with Figure 2, which shows little sign of a kink in the relationship between the mortality rate and the subjective survival probability at a response of 50 (see also Delavande and Rohwedder 2008). On average, a reported survival probability of 50 percent does not deviate from the true survival chance any more (or less) than do other responses. Dropping these responses also has little or no impact on the education and cognition gradients in the predictive power of reported survival chances (Table 5).[31]

As would be anticipated from Figure 2, dropping responses of 0 reduces the extent to which expectations predict actual longevity, on average, while dropping those reporting a 100 percent chance of reaching the target age raises the predictive power of expectations (Table 3).[32] On average, a response of 'no chance of survival' is indeed informative of a higher probability of

[31] We show the robustness to dropping focal responses on the estimates of expectations accuracy by education and cognition only for the model that controls for socioeconomic covariates. The changes in the estimates obtained from the other specifications follow similar patterns. The omitted results are available on request.

[32] Irrespective of the specification of covariates, the 95% confidence interval estimate obtained without each of these two focal responses does not include the respective point estimate obtained from the sample including these responses, although the respective confidence intervals do overlap.

death but so too is a response of '100% chance of survival' (Delavande and Rohwedder 2008, Hurd 2009).

Dropping responses of 100 raises the predictive power of the reported survival probability for all groups but has little impact on the education and cognition differences (Table 5). If anything, they widen. Leaving out observations reporting a survival chance of 0 reduces the magnitudes of the point estimates of the differences by education and cognition and renders them insignificant.[33] At least part of the reason the survival chances reported by individuals with low education and cognition are less accurate in predicting longevity is that the difference in mortality risk associated with a zero and a positive reported probability is smaller for these individuals than it is for high education and cognition groups. A zero probability reported by the highly educated and cognitively able is more likely to indicate a greatly inflated mortality risk that is correctly perceived. Those with low education or cognitive functioning are more likely to be overly pessimistic when they report a zero chance of survival.[34]

Table 5 here

3.6 ACCURACY OF EXPECTATIONS BY NUMERACY

The most numerate individuals are likely to be most comfortable with the concept of probability and most successful in expressing a longevity belief in a survival probability. In Table 6 we present estimates of the longevity coefficients for education and numeracy quartile groups in mortality hazard models estimated using data from wave 6 onwards and including age-sex and socioeconomic controls only.[35] With heterogeneity only by education, expectations predict longevity significantly for all groups and the point estimates indicate a gradient in the same direction as that obtained using data from wave 3 onwards. However, the education gradient in the accuracy of expectations is not significant. This is likely attributable to the substantially smaller sample size and the much fewer number of deaths observed over an eight, as opposed to fourteen, year period. Despite this reduced power, there are significant differences in the accuracy of expectations by numeracy skills. The coefficient for the top numeracy quartile group is more than twice the magnitude of that of the bottom quartile group. This significant difference is maintained when the accuracy of expectations is allowed to vary by both numeracy and education. The expectations coefficient of the most numerate high school dropouts is double the magnitude of the coefficient of the least numerate dropouts.

Numeracy skills appear to be particularly pertinent to the formation and reporting of accurate survival probabilities. This is not necessarily inconsistent with the early observation that, if anything, the accuracy of expectations varies more with education than it does with overall cognitive functioning. Numeracy is not cognition. It is a more specific skill that can be taught.

---

33 Without controlling for covariates, the education differences remain significant (10%) when responses of zero are dropped.

34 If observations reporting any one of the three focal responses are dropped, then the education gradient in the accuracy of expectations remains: the coefficient for college graduates is twice the magnitude and significantly different from that of high school dropouts. These results are available on request.

35 The expectations coefficients are larger with control only for age-sex and smaller in magnitude with the full set of controls. Controlling only for age and sex, the effects for the top two numeracy groups are significantly different from that for the bottom group. With the full set of controls, there are no significant differences by numeracy. These results are available on request.

Table 6 here

4. Updating of longevity expectations

We have established that the less educated, cognitively able and numerate hold expectations that are less accurate in predicting longevity. This suggests that these groups have less knowledge of health signals relevant to mortality risks or are less able to process information on such risks. It could also be that the less educated and cognitively able are overly responsive to events that are only marginally relevant to longevity. Or they may simply be less able to express beliefs in a number, such that there is more noise in their subjective probabilities. A further possibility is that the disadvantaged groups are more vulnerable to mortality risks that are more difficult to predict. While the HRS, indeed any available data, does not allow one to discriminate between these explanations, some insight into why there are education and cognition gradients in the predictive accuracy of longevity expectations can be gained by exploring heterogeneity in the updating of subjective survival probabilities in response to events that carry varying degrees of information on mortality risks.

4.1 Modelling expectations

We suppose that an individual's subjective survival probability relative to the life table probability at a given point in time is a function of that reported in the previous period with an adjustment in response to any new information on mortality risk that has become available. We estimate the following model,

$$\tilde{P}_{it} = \alpha + \beta \tilde{P}_{it-1} + \Delta\mathbf{Health}_{\mathbf{it}}\boldsymbol{\delta} + \Delta\mathbf{Relative}_{\mathbf{it}}\boldsymbol{\theta} + f\left(Age_{it}, Gender_i\right) + \varepsilon_{it} \quad (4)$$

where t is used here to index the survey wave, $\Delta\mathbf{Health}_{\mathbf{it}}$ is a vector of indicators of the onset of disease and health risk factors, and $\Delta\mathbf{Relative}_{\mathbf{it}}$ contains indicators of parental longevity and spousal death. While the dependent variable is already standardized on age and gender by taking the deviation from the life table probability, we further control for age and gender to allow for the patterns observed in Figure 1, including the spikes at ages corresponding to changes in the target age.

The model is estimated on the same sample used for the mortality analysis. In addition to pooling across groups, we look for evidence of heterogeneity in the updating of expectations by estimating a separate model for each education/cognition group.

One might interpret (4) as deriving from an assumption that expectations evolve according to a Bayesian learning process (Viscusi and O'Connor 1984, Viscusi 1991, Smith et al. 2001) (see online Appendix B). Since the observed differences in accuracy by education and cognition may reflect different degrees to which belief formation indeed adheres to Bayesian updating, we do not limit our interpretation to this model. If one were to assume that all respondents behave as if they are Bayesian, then the coefficient on the lagged dependent variable would represent the precision of belief in the prior probability of survival relative to the precision of the estimated mortality risk formed on the basis of all new information. Further, division of the coefficient on each health change by one minus the coefficient on the lagged dependent variable would provide an estimate of the extent to which that change shifts the *risk equivalent* - the probability assessed on the basis of all information obtained since the prior was formed. Parameter heterogeneity across education and cognition groups would then be interpreted as reflecting differences in the degree of confidence placed in the prior probability and in the assessed mortality risk associated with a health condition, as well as the precision of the belief in this risk assessment.

We estimate (4) by ordinary least squares, which is inconsistent in the presence of time invariant unobservable determinants of expectations that would necessarily be correlated with the lagged dependent variable. If we use fixed effects and omit the lagged dependent variable, then the same conclusion is reached with respect to the main hypothesis of interest – there are little or no consistent differences by education (or cognition) in the response of expectations to changes in health that represent objective mortality risks (see online Appendix Table A8).[36]

4.2 EMPIRICAL SPECIFICATION

We do not include in Δ**Health** all the health indicators that were used in the previous section to predict mortality. Doing so would make it difficult to interpret the coefficient on any one indicator. For example, this would identify the extent to which longevity expectations are revised in response to the experience of a stroke when there is no change in mobility and activities of daily living that may be impeded as a consequence of a stroke. We focus instead on the response to the onset of the eight diseases or risk factors (lung disease, cancer, diabetes, heart disease, stroke, psychiatric illness, high blood pressure and arthritis) without conditioning on the measures of functional and mobility limitations, and of psychological distress. The coefficient on a health condition should then capture the reassessment of survival chances arising from learning of the diagnosis and experiencing the physical and mental consequences of the condition. We do include an indicator of whether the person was admitted to hospital in the past two years (and not in two years before the previous interview) since this may cause one to re-evaluate life expectancy and admission could be for treatment of a condition other than the eight we identify.

We aim to establish whether subjective survival chances are revised most in response to conditions that carry the greatest mortality risk and whether the more educated and cognitively able are most responsive to such conditions. In order to establish the objective risks, we estimate the mortality hazard as a function of ever having been diagnosed with each of the eight health conditions and having been hospitalized in the past two years, plus age-gender dummies to capture the baseline hazard. This reveals that lung disease and cancer pose the greatest threat to life (see online Appendix A, Table A7). The hazard ratio for each of these diseases, as well as a hospital admission, is more than 2.3, indicating an increase in the probability of death between waves of more than 130%. Diabetes is the next most life-threatening condition, with a hazard ratio of 1.9, which reflects the high prevalence of comorbidities experienced by diabetics. Stroke, psychiatric illness and heart disease follow in the *death list*, each with a hazard ratio of around 1.4, which corresponds to less than half of the relative increase in mortality associated with diabetes.

---

[36] We do not apply system GMM (Blundell and Bond 1998), which in principle can be used to estimate the lagged dependent variable model while allowing for time invariant unobservables. The recent literature has drawn attention to the often poor finite sample performance of this estimator and has argued that it is highly reliant on the assumption that the correlated unobservables are indeed time invariant (Roodman 2009, Bun and Sarafidis forthcoming). This assumption is questionable in the present context. Unobservable differences in ability to form longevity expectations need not be fixed over time. Rather, more able individuals may learn more from accumulating experience and information such that the error in their expectations diminishes.

Neither arthritis nor hypertension (conditional on the other conditions) is associated with reduced longevity.[37]

On the basis of these findings, we categorise the conditions into the following groups ranked by decreasing mortality risk: 1) lung disease and cancer; 2) hospital admission; 3) diabetes; 4) stroke and heart disease; 5) psychiatric illness; and, 6) hypertension and arthritis. Since psychiatric illness is fundamentally different from physical illness and may be anticipated to have a different impact on longevity expectations, we keep this condition separate from stroke and heart disease despite the objective mortality risks being similar.

Heterogeneity in the extent to which expectations are revised in response to health conditions would be justified if the objective risks varied by education and cognition. We tested this but found only a few significant differences by education/cognition in the hazard ratios of specific conditions.[38]

We include in **∆Relative** indicators of whether the respondent's father and mother was alive in previous wave. If a parent was alive at the time the previous expectation was formed, then there is an opportunity to learn more about parental health and longevity that could potentially be used to revise expectations. The new information could arise from a parental death, but it could also be obtained from witnessing a parent continuing to live when death was anticipated. We have determined that the revision of expectations does not differ significantly between those who experience a parental bereavement and those whose parent continues to live. We include an indicator of whether the respondent has been widowed since the previous wave. While spousal longevity need not provide any information on one's own health and longevity, except to the extent that health environments and habits are shared, the death of a partner may be expected to prompt contemplation of survival prospects. In addition, there may be a direct impact of bereavement on health and longevity (Van den Berg et al. 2011).

Demographic effects are captured by a gender dummy and a quadratic in age. This was chosen by testing down from more general specifications that allowed for higher order polynomials and gender specific age effects. Specifying the age effect as quadratic, rather than a set of year dummy variables, allows us to include indicators of the target age change, which depends only on age and wave. These indicators pick up the spikes observed in Figure 1.

---

37 The ranking of the conditions is not entirely robust to controlling for measures of functional and mobility limitations and psychological distress. This is to be expected. For example, conditioning on mobility problems has little impact on the cancer coefficient but reduces that of lung disease, leaving the former as the most life-threatening disease. Presumably this reflects the greater mobility restrictions arising from respiratory problems. Controlling for the CESD depression score reduces the coefficient on psychiatric illness by more than half and leaves it insignificant. This confirms that depression is the most life-threatening form of psychiatric illness and is indicative of the interpretation problems that would arise if we were to condition on the health measures, in addition to the conditions, in the model of expectations.

38 The hazard ratio for heart disease is a significant 1.55 for high school dropouts and an insignificant 1.06 for college graduates. The difference between the two groups is significant (p=0.034), and the null of equality across all four education groups is rejected at the 10% level (p=0.088). Stroke has no significant impact on mortality for the highest and lowest education groups, but significantly raises the risk for the two middle groups. The null of homogeneity across the four groups is rejected for this condition (p=0.032). There are no other significant differences by education. With respect to cognition, the only significant difference is that hypertension is associated with an increased mortality risk for the lowest quartile group but not for the others. The null of equality in the risk across the four groups is rejected (p=0.074). The detailed results are available on request.

4.3 UPDATING OF EXPECTATIONS ON AVERAGE

The left-hand column of Table 7 gives estimates of model (4). Reporting of the probability of survival displays persistence with the coefficient on the lagged dependent variable only marginally greater than previous estimates obtained using only the first three waves of the HRS (Smith et al. 2001, Smith et al. 2001).[39]

The degree to which the survival probability is revised downward on occurrence of each group of health conditions is broadly, but not entirely, consistent with their rankings by objective mortality risks. The strongest reaction is to the onset of the most life-threatening conditions – lung disease and cancer. And there is no significant revision of longevity expectations with the onset of hypertension or arthritis, which do not carry any mortality risk. While the subjective and objective risk rankings are consistent at the top and bottom, they differ in between. Stroke/heart disease and psychiatric illness prompt greater downward revisions of survival expectations than does diabetes, although the latter carries a greater objective risk. It seems plausible that depression, which is highly correlated with the reporting of any kind of psychiatric illness, causes a downward revision of longevity expectations to an extent that is excessive relative to the response to more life-threatening physical illnesses. The strong response to stroke may be due to the salience of the event, the effects of which can leave the sufferer with a permanent reminder of the fragility of her health. The weaker response to diabetes may be because there is a lack of appreciation of diabetes being a precursor to a number of comorbidities than can be the eventual cause of death.

Having a mother or a father alive at the time the survival probability was previously reported is associated with an upward revision of survival expectations. As mentioned above, this effect does not differ significantly depending on whether or not the parent has died since the last wave. It suggests that individuals whose parents were alive are more likely to receive some information that causes them to revise their longevity expectations upward. Consistent with this, it could be that those with surviving parents experience a slower deterioration in health but do not fully appreciate the information on their advantageous health trajectory that is provided by their parents' long life. Only as time passes is it observed that health is improving relative to that of peers and survival expectations are revised upward. Of course, there is much supposition in this explanation.

The age effects reflect the pattern observed in Figure 1. As the cohort ages the subjective survival probability falls further below the life table probability before the gaps begins to closer and then becomes positive mainly through the spikes at the changes in the target age. Females become increasingly more pessimistic than males over time.

If one interprets the estimates as corresponding to parameters of a Bayesian learning model, then the coefficient on the lagged dependent variable implies that the precision associated with the prior probability accounts for more than half (55%) of the precision attached to all beliefs on which expectations are formed. Equivalently, the precision of the risk estimate based on all new

[39] Strictly, our estimate is not directly comparable to the earlier ones because the dependent variables differ. We take the difference of the subjective from the life table probability, while the earlier estimates are based on the reported probability of survival to 75. If individuals do not sufficiently revise their probability upward as they approach the target age, then there should be less persistent in our measure than in the raw subjective probability.

information is less than the precision attached to prior.[40] Since the prior is established from experience over many years, while the new information is obtained in the last two years, it stands to reason that there is more confidence in the accuracy of the estimate of the prior.

According to this Bayesian learning interpretation, the onset of lung disease or cancer is estimated to be assessed as equivalent to a 15 percentage point reduction in the probability of survival to the target age.[41] The downward revision in the reported probability is less than half of this (6.8) because of the uncertainty attached to the subjective estimate of the longevity consequences of these diseases. The *risk equivalents* of the other health conditions are simple transformations of the respective coefficients and necessarily rank in the same order.

Table 7 here

4.4 UPDATING OF EXPECTATIONS BY EDUCATION & COGNITION

Education group specific estimates of model (4) are given in the right-hand panel of Table 7. There are relatively few significant differences across groups in the coefficients of specific variables. One clear difference is the coefficient on the lagged dependent variable, which increases monotonically and significantly in moving from the lowest to the highest education group. The reporting of survival probabilities is more stable over time among the more highly educated.

If one were to assume Bayesian updating, then this would suggest that more highly educated individuals place greater weight on their prior in the process of updating their expectations. This does not necessarily imply that the better educated are less responsive to new information. It is only possible to identify the precision of the prior relative to the precision of the estimate derived from new information. It could be that the better educated hold more precise estimates both of the prior and the risk equivalent obtained from new information. Further, all experiences that influence the revision of expectations are not necessarily objectively relevant to longevity. It could be that the less educated place undue weight on irrelevant events.

In assessing the incorporation of new information into expectations one needs to pay attention not only to the coefficients but also to the proportion of the variance that is explained. There is clear heterogeneity by education in this respect. The explained variability of subjective probabilities rises from 29.8% among high school dropouts to 36.5% among college graduates. There is more variation in reported longevity expectations that is random, or at least not systematically related to observed covariates, among the least educated. Presumably this contributes to the survival expectations of high school dropouts being less predictive of their longevity than is the case for the higher education groups.

---

[40] The estimates imply that the precision of the new information is 83% of the precision of the prior $\left(0.83 = \frac{1}{0.5464} - 1\right)$. See online Appendix B for the correspondence between the lagged dependent variable coefficient and the precision parameters.

[41] This *risk equivalent* is given by $\hat{\delta}_j \big/ \left(1-\hat{\beta}\right)$. See Appendix B.

There are little or no systematic or significant differences across education groups in the revision of expectations in response to the onset of health conditions.[42] The downward revision of the survival probability in response to psychiatric illness is strongest amongst the least educated and weakest among the most educated. High school dropouts reduce their longevity expectations more in response to psychiatric illness than they do after the onset of cancer or lung disease. This may indicate overreaction.[43]

There are differences in the effects of demographics. The rate at which female high school dropouts become more pessimistic over time compared to their male counterparts is significantly greater than the rate of increase in the relative pessimism of women with higher levels of education. It would appear that low educated women appreciate least their longevity advantage over men as they age. Socioeconomic inequality in longevity is less pronounced among females than it is among males (Mustard and Etches 2003)[44] and consequently women have a greater longevity advantage at low levels of education. This does not appear to be understood.

Heterogeneity in the updating of survival expectations by cognition and numeracy is similar to that by education (see Appendix A, Tables A9 and A10). Expectations are more stable for higher cognition and numeracy groups. In fact, this gradient is steeper by cognition and numeracy than it is by education. A larger proportion of the variability in expectations is explained for higher cognition and numeracy groups, with these gradients again more pronounced than that observed by education.

Unlike for education, there are significant differences by cognition in the response to the most life-threatening conditions. The downward revision of the survival probability on the occurrence of lung disease or cancer made by the lowest quartile cognition group is no more than half of that of the other groups and is significantly less than the revision made by two of the three other groups. But there are no clear or significant differences in the response to other conditions.

5. CONCLUSION

Our analysis confirms that older US citizens can, at least to some extent, predict their longevity. This is reassuring for those utilizing economic models that presume agents hold longevity expectations that are based on actual mortality risks. It is more disconcerting from the perspective of markets in life insurance and pension annuities, which are vulnerable to adverse selection if customers hold and exercise private information on their chances of survival to old age. The fact that subjective survival probabilities predict longevity even after conditioning on a battery of health measures, smoking behaviour and parental longevity, suggests that the average

[42] If the model is run with each health condition entered separately, then the point estimate of the response to cancer is smallest for high school dropouts but it does not rise monotonically across the groups and any differences are not significant.

[43] The education difference in the response to psychiatric illness is also apparent in fixed effects estimates (see Appendix A, Table A8 – the p-value for the test of no difference between the bottom and top education group is 0.107). Also consistent in the fixed effects estimates is that more of the within individual variability in expectations is explained for the highest education group, although the gradient in the R-squared is not so clear across all groups as it is in the lagged dependent variable model.

[44] The evidence indicates that absolute differences in longevity by socioeconomic status, which are relevant here, are smaller for females than males. Relative differences are not (Mustard and Etches 2003).

individual holds and processes information that is useful in predicting mortality and which is unlikely to be observable to providers of the aforementioned financial products.

Our focus has not been on the average accuracy of longevity expectations but on heterogeneity by education and cognition. The hypothesis was that the least educated and cognitively able are less likely to hold accurate expectations and this may result in mistaken decisions with respect to retirement, pensions, saving and health behaviour. While there is support in the data for the first part of this hypothesis, the gradients revealed in the accuracy of expectations should not obscure the fact that even the least educated and cognitively able are capable of predicting their longevity with some degree of accuracy. Even these individuals can, to an extent, be represented as agents forming expectations on the basis of relevant information. But the differences in the accuracy of expectations are marked. A marginal increase in the subjective survival probability reported by a high school dropout is around half as strongly correlated with a reduced actual mortality risk as is the same report made by a college graduate. After conditioning on health measures, smoking and parental longevity, the least educated can process no further information that is useful in predicting longevity. The higher education groups do successfully utilize further information. If there is an advantage to be reaped from private information and adverse selection, it is more likely to be enjoyed by the better educated.

Differences in the accuracy of expectations by cognition are slightly weaker than those by education. With controls for objective mortality risk, there are no significant differences in the predictive power of expectations by cognition and the lowest cognition group, unlike high school dropouts, is still able to predict longevity. But differences by numeracy are stronger than those by education. An affinity with numbers appears to be particularly important to the formation and/or reporting of accurate survival probabilities.

There is heterogeneity by education (and cognition) in the updating of longevity expectations. Higher education groups display greater consistency over time in the reporting of subjective probabilities. In a Bayesian framework, this would imply that they have more confidence in their prior and place relatively greater weight on it. We find little evidence that the degree to which survival chances are revised downward with the onset of a health condition varies by education, although high school dropouts are more (possibly over) responsive to psychiatric illness. Evidence that the higher education (and cognition) groups formulate expectations more systematically is provided by the fact that a substantially larger proportion of the variability in subjective survival probabilities can be explained for these groups.

Given that the least educated and cognitively able hold less accurate longevity expectations, one would expect their economic and health investment decisions to deviate further from those that are optimal given the objective circumstances and risks they are exposed to. We have not tested whether there are behavioural consequences of mistaken expectations. Binswanger and Salm (2014) find that the stock holdings of less numerate HRS respondents are more weakly associated with reported probabilities of stock market returns. Rather than sub-optimal decisions being taken on the basis of biased expectations, this suggests that decisions of the least numerate are not taken on the basis of probabilities at all. Either way, our findings support a case for targeting advice regarding complex financial decisions that involve planning over a lifetime horizon at the least educated, cognitively able and numerate. In particular, these individuals need help in formulating more accurate assessments of how long they can expect to live.

REFERENCES

Agarwal, S. and B. Mazumder (2013) 'Cognitive Abilities and Household Financial Decision Making', *American Economic Journal: Applied Economics* 5(1): 193-207.

Anderson, N.D. and F.I.M. Craik (2000) 'Memory in the Aging Brain', in E. Tulving and F.I.M. Craik (eds) *The Oxford Handbook of Memory,* pp. 411-425. Oxford: Oxford University Press.

Baddeley, A.D. and G.J.L. Hitch (1974) 'Working Memory', in G.A. Bower (ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory,* (vol. 8 edn). pp. 47-89. New York: Academic Press.

Banks, J. and Z. Oldfield (2007) 'Understanding Pensions: Cognitive Function, Numeracy and Retirement Saving ', *Fiscal Studies* 28: 143-170.

Banks, J., C. O'Dea and Z. Oldfield (2010) 'Cognitive Function, Numeracy and Retirement Saving Trajectories', *The Economic Journal* 120(548): F381-F410.

Behrman, J., O.S. Mitchell, C. Soo and D. Bravo (2012) 'How Financial Literacy Affects Household Wealth Accumulation', *American Economic Review* 102(3): 300-304.

Benítez-Silva, H. and H. Ni (2008) 'Health Status and Health Dynamics in an Empirical Model of Expected Longevity', *Journal of Health Economics* 27(3): 564-584.

Binswanger, J. and Salm, M., Anonymous 2014. Does Everyone use Probabilities? the Role of Cognitive Skills. Tilburg: Department of Economics, Tiburg University.

Bissonnette, L., M.D. Hurd and P.C. Michaud (2012) 'Individual Survival Curves Combining Subjective and Actual Mortality Risks', in L. Bissonnette (ed.) *PhD Thesis: Essays on Subjective Expectations and Stated Preferences,* pp. 67-94. Tilburg, NL: Tilburg University.

Blundell, R. and S. Bond (1998) 'Initial Conditions and Moment Restrictions in Dynamic Panel Data Models', *Journal of Econometrics* 87(1): 115-143.

Brandt, J., M. Spencer and M.F. Folstein (1988) 'The Telephone Interview for Cognitive Status', *Neuropsychiatry, Neuropsychology and Behavioral Neurology* 1: 111-117.

Brown, J.R., A. Kapteyn, E.F.P. Luttmer and O.S. Mitchell (2013) 'Cognitive Constraints on Valuing Annuities. NBER Working Paper', No. 19168. Cambridge, MA: NBER.

Bruine de Bruin, W., B. Fischhoff, S.G. Millstein and B.L. Halpern- Felsher (2000) 'Verbal and Numerical Expressions of Probability: `It's a Fifty Chance'', *Organizational Behavior and Human Decision Processes* 81(1): 115-131.

Bruine de Bruin, W. and K.G. Carman (2012) 'Measuring Risk Perceptions: What does the Excessive use of 50% Mean?', *Medical Decision Making* 32(2): 232-236.

Bun, M.J.G. and V. Sarafidis (forthcoming) 'Dynamic Panel Data Models', in B. Baltagi (ed.) *Oxford Handbook on Panel Data,* Oxford: Oxford University Press.

Choi, S., S. Kariv, W. Müller and D. Silverman (2014) 'Who is (More) Rational?', *American Economic Review* 104(6): 1518-1550.

Christelis, D., T. Jappelli and M. Padula (2010) 'Cognitive Abilities and Portfolio Choice', *European Economic Review* 54(1): 18-38.

Cutler, D.M. and A. Lleras-Muney (2010) 'Understanding Differences in Health Behaviors by Education', *Journal of health economics* 29(1): 1-28.

Delavande, A. and S. Rohwedder (2011) 'Differential Survival in Europe and the United States: Estimates Based on Subjective Probabilities of Survival', *Demography* 48(4): 1377-1400.

Delavande, A. and S. Rohwedder (2008) 'Eliciting Subjective Probabilities in Internet Surveys', *Public Opinion Quarterly* 72: 866-891.

Dominitz, J. and C. Manski (1997) 'Using Expectations Data to Study Subjective Income Expectations', *Journal of the American Statistical Association* 92: 855-867.

Dominitz, J. and C.F. Manski (2006) 'Measuring Pension-Benefit Expectations Probabilistically', *Labor* 20: 201-236.

Elder, T. (2013) 'The Predictive Validity of Subjective Mortality Expectations: Evidence from the Health and Retirement Study', *Demography* 50(2): 569-589.

Fang, H., M.P. Keane and D. Silverman (2008) 'Sources of Advantageous Selection: Evidence from the Medigap Insurance Market', *Journal of Political Economy* 226(2): 303-350.

Fischhoff, B. and W. Bruine De Bruin (1999) 'Fifty?Fifty=50%?', *Journal of Behavioral Decision Making* 12(2): 149-163.

Fisher, G.G., H. Hassan, W.L. Rodgers and D.R. Weir (2012) 'Health and Retirement Study Imputation of Cognitive Functioning Measures: 1992 – 2010 Early Release.'. Ann Arbor, MI: Survey Research Centre, University of Michigan,.

Folstein, M.F., S.E. Folstein and P.R. McHugh (1975) 'Mini-Mental State: A Practical Method for Grading the Cognitive State of Patients for the Clinician', *Journal of Psychiatric Research* 12: 189-198.

Gan, L., M.D. Hurd and D. McFadden (2005) 'Individual Subjective Survival Curves', in D.A. Wise (ed.) *Aspects of Economics of Aging*, pp. 377-412. Cambridge MA: NBER.

Gottfredson, L.S. and I.J. Deary (2004) 'Intelligence Predicts Health and Longevity, but Why?', *Current Directions in Psychological Science* 13(1): 1-4.

Gouret, F. and G. Hollard (2011) 'When Kahneman Meets Manski: Using Dual Systems of Reasoning to Interpret Subjective Expectations of Equity Returns', *Journal of Applied Econometrics* 26(3): 371-392.

Hamermesh, D.S. (1985) 'Expectations, Life Expectancy and Economic Behavior', *Quarterly Journal of Economics* 100(2): 398-408.

Heckman, J.J., J.E. Humphries and T. Kautz (eds) (2014) *The Myth of Achievement Tests: The GED and the Role of Character in American Life.* Chicago: The University of Chicago Press.

Hill, D., Perry, M. and Willis, R.J., Anonymous 2004. Estimating Knightian Uncertainty from Survival Probability Questions on the HRS. Ann Arbor: .

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

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

Hurd, M.D., D. McFadden and L. Gan (1998) 'Subjective Survival Curves and Life Cycle Behavior', in D. Wise (ed.) *Inquiries in the Economics of Aging,* Chicago: University of Chicago Press.

Hurd, M.D. and K. McGarry (2002) 'The Predictive Validity of Subjective Probabilities of Survival', *Economic Journal* 112(442): 966-985.

Hurd, M.D. and K. McGarry (1995) 'Evaluating Subjective Probabilities of Survival in the Health and Retirement Study', *Journal of Human Resources* XXX(supplement): S268-S292.

Idler, E. and Y. Benyamini (1997) 'Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies', *Journal of Health and Social Behavior* 38(1): 21-37.

Jenkins, S.P. (1995) 'Easy Estimation Methods for Discrete-Time Duration Models', *Oxford Bulletin of Economics and Statistics* 57(1): 129-138.

Jones, A.M., N. Rice, T. Bago d'Uva and S. Balia (2012) *Applied Health Economics.* (Second edn) London: Routledge.

Kenkel, D., S. (1991) 'Health Behavior, Health Knowledge, and Schooling', *Journal of Political Economy* 99(2): 287-305.

Kleinjans, K.J. and A. Van Soest (2014) 'ROUNDING, FOCAL POINT ANSWERS AND NONRESPONSE TO SUBJECTIVE PROBABILITY QUESTIONS', *Journal of Applied Econometrics* 29(4): 567-585.

Kutlu-Koc, V. and A. Kalwij (2013) 'Individuals' Survival Expectations and Actual Mortality', No. 05/2013-013. Tilburg: Netspar.

Lange, F. (2011) 'The Role of Education in Complex Health Decisions: Evidence from Cancer Screening.', *Journal of Health Economics* 30(1): 43-54.

Ludwig, A. and A. Zimper (2013) 'A Parsimonuous Model of Subjective Life Expectancy', *Theory and Decision* 75: 519-541.

Manski, C.F. and F. Molinari (2010) 'Rounding Probabilistic Expectations in Surveys', *Journal of Business & Economic Statistics* 28(2): 219-231.

Mustard, C.A. and J. Etches (2003) 'Gender Differences in Socioeconomic Inequality in Mortality', *Journal of epidemiology and community health* 57(12): 974-980.

National Center for Learning Disabilities, (Last updated 2012) 'What is Executive Function?'. Accessed September 13 2012 <http://www.ncld.org/types-learning-disabilities/executive-function-disorders/what-is-executive-function>.

Ofstedal, M.B., G.G. Fisher and A.R. Herzog (2005) 'Documentation of Cognitive Functioning Measures in the Health and Retirement Study', No. HRS Documentation Report DR-006. Ann Arbor, Michigan: Survey Research Center, University of Michigan.

Perozek, M. (2008) 'Using Subjective Expectations to Forecast Longevity: Do Survey Respondents Know Something we Don'T Know?', *Demography* 45(1): 95-113.

Peters, E. (2008) 'Numeracy and the Perception and Communication of Risk', *Annals of the New York Academy of Science* 1128: 1-7.

Prentice, R. and L. Gloeckler (1978) 'Regression Analysis of Grouped Survival Data with Applications to Breast Cancer Data', *Biometrics* 34: 57-67.

Radloff, L.S. (1977) 'The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.', *Applied Psychological Measurement* 1(3): 385-401.

Reyna, V.F. and C.J. Brainerd (2007) 'The Importance of Mathematics in Health and Human Judgment: Numeracy, Risk Communication, and Medical Decision Making', *Learning and Individual Differences* 17(2): 147-159.

Roodman, D. (2009) 'A Note on the Theme of Too Many Instruments', *Oxford Bulletin of Economics and Statistics* 71: 135-158.

Schwartz, L.M., S. Woloshin, W.C. Black and H.G. Welch (1997) 'The Role of Numeracy in Understanding the Benefit of Screening Mammography', *Annals of Internal Medicine* 123(11): 966-972.

Siegel, M., E.H. Bradley and S.V. Kasl (2003) 'Self-Rated Life Expectancy as a Predictor of Mortality: Evidence from the HRS and AHEAD Surveys', *Gerontology* 49: 265-71.

Smith, J.P., J.J. McArdle and R. Willis (2010) 'Financial Decision Making and Cognition in a Family Context', *The Economic Journal* 120(November): F363-F380.

Smith, V.K., D.H. Taylor, F.A. Sloan, F.R. Johnson and W.H. Desvousges (2001) 'Do Smokers Respond to Health Shocks', *Review of Economics and Statistics* 83(4): 675-687.

Smith, V.K., D. Taylor and F. Sloan (2001) 'Longevity Expectations and Death: Can People Predict their Own Demise?', *American Economic Review* 91(4): 1126-1134.

St. Clair, P., D. Blake, D. Bugliari, S. Chien, O. Hayden, M. Hurd, S. Ilchuk, F.Y. Kung, A. Miu, C. Panis, P. Pantoja, A. Rastegar, S. Rohwedder, E. Roth, J. Carroll and J. Zissimopoulos (2009) '*RAND HRS Data Documentation, Version I,"* '. Santa Monica, CA.: RAND Labor and Population Program, RAND Center for the Study of Aging.

Steffick, D.E. (2000) 'Documentation of Affective Functioning Measures in the Health and Retirement Study'. Ann Arbor, MI: HRS/AHEAD Documentation Report DR-05. Survey Research Center, University of Michigan.

Tulving, E. (1972) 'Episodic and Semantic Memory', in E. Tulving and W. Donaldson (eds) *Organization of Memory,* pp. 381-403. New York: Academic Press.

Van den Berg, G.J., M. Lindeboom and F. Portrait (2011) 'Conjugal Bereavement Effects on Health and Mortality at Advanced Ages', *Journal of health economics* 30(4): 774-794.

Van Doorn, C. and S.V. Kasl (1998) 'Can Parental Longevity and Self-Rated Life Expectancy Predict Mortality among Older Persons? Results from an Australian Cohort', *Journal of Gerontology: Social Sciences* 53B(1): S28-S34.

Van Santen, P., R. Alessie and A. Kalwij (2012) 'Probabilistic Survey Questions and Incorrect Answers: Retirement Income Replacement Rates', *Journal of Economic Behavior & Organization* 82(1): 267-280.

Viscusi, W.K. (1991) 'Age Variations in Risk Perceptions and Smoking Decisions', *The review of economics and statistics* 73(4): 577-588.

Viscusi, W.K. and C.J. O'Connor (1984) 'Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?', *The American Economic Review* 74(5): 942-956.

Wallace, R.B. and A.R. Herzog (1995) 'Overview of the Health Measures in the Health and Retirement Study', *Journal of Human Resources* 30(5): 84-107.

Wu, S., R. Stevens and S. Thorp (2014) 'Die Young Or Live Long: Modelling Subjective Survival Probabilities'. Sydney: School of Risk and Actuarial Studies, University of New South Wales.

Xu, J., K.D. Kochanek, S.L. Murphy and B. Tejada-Vera (2010) 'Deaths: Final Data for 2007', *National Vital Statistics Report* 58(19): 1-133.

Zikmund-Fisher, B., D.M. Smith, P.A. Ubel and A. Fagerlin (2007) ' Validation of the Subjective Numeracy Scale: Effects of Low Numeracy on Comprehension of Risk Communications and Utility Elicitations', *Medical Decision Making* 27: 663-671.

FIGURES

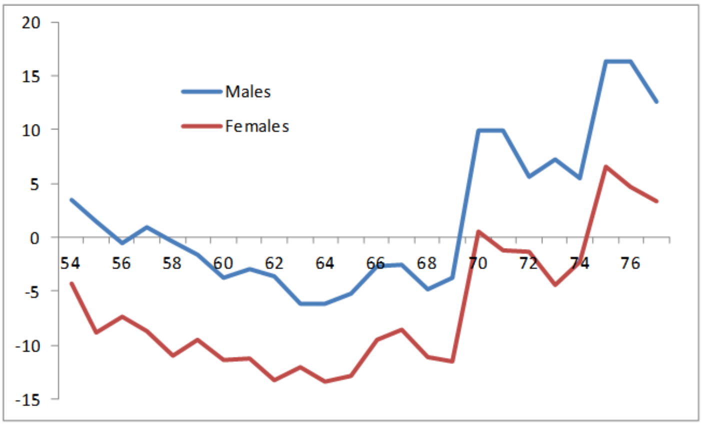

Figure 1: Mean difference between reported and life table survival probability by gender and age
Notes: as Table 1

a) by education **b) by cognitive functioning**

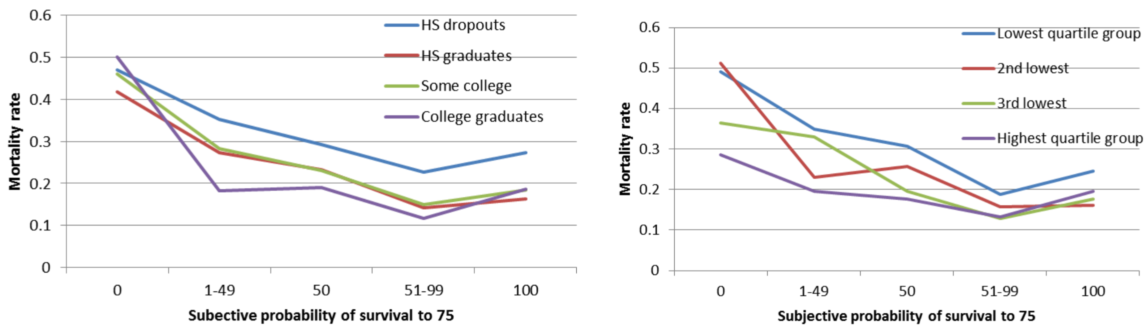

Figure 2: Proportion of individuals who die between 1996 and 2010 by subjective probability of survival to 75

TABLES

Table 1: Difference between reported and life table survival probability by education, cognition and numeracy

	Mean	Median	% reporting probability of			Non-response (%)	N
			0	50	100		
Total	-5.38	-2.20	5.14	25.80	15.75	4.09	32437
Education							
High school dropout & GED	-13.28	-13.45	11.28	26.77	17.84	9.04	6966
High school graduate	-6.24	-4.72	4.47	28.30	16.23	3.03	11175
Some College	-2.77	0.63	3.42	24.93	17.20	2.65	7132
College	1.04	4.69	1.91	21.83	11.46	1.79	7164
Equality across groups: p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	
Cognitive functioning							
Lowest quartile group	-9.20	-6.72	8.38	25.26	18.93	7.79	8223
2nd lowest	-7.01	-4.91	5.69	27.36	14.58	3.57	8209
2nd highest	-4.31	-1.25	3.31	26.87	13.98	2.61	7967
Highest quartile group	-0.87	2.83	3.07	23.70	15.39	2.34	8038
Equality across groups: p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	
Numeracy							
Lowest quartile group	-7.09	-6.48	8.02	25.42	15.88	7.88	4477
2nd lowest	-3.85	-2.48	5.37	26.87	12.77	3.59	4324
2nd highest	-1.52	-0.78	4.35	28.63	10.90	2.72	4118
Highest quartile group	-1.57	2.15	4.01	26.06	9.37	2.39	4044
Equality across groups: p-value	<0.001	<0.001	<0.001	0.006	<0.001	<0.001	

Notes: Reported probability is the stated percentage chance of surviving to a specified target age. Life table probability is that percentage implied by the Vital Statistics life tables given current age and gender. Data from HRS waves 3-9 (total, education & cognition) and waves 6-9 (numeracy). Sample size (N) and all statistics are with respect to the estimation sample except for percentage non-response. Cognition/numeracy categories are quartile groups of age-sex standardized total cognition/numeracy score. 'Equality across groups' gives p-value for null of equal means/medians/proportions. For mean, it is a Wald test that allows for unequal variances across groups. For the median, it is implemented by the joint significance of the group indicators in a least absolute deviation regression in which those indicators are the only regressors. For proportions, it is the Pearson chi-square test.

Table 2: Percentage of individuals who die between 1996 and 2010 by education, cognition and subjective probability of survival to 75

	Percent	N
Education		
High school dropout & GED	30.57	1184
High school graduate	19.98	1802
Some College	19.91	1130
College	15.45	1094
Test equal proportions, p-value	<0.001	
Cognitive functioning		
Lowest quartile group	28.52	1413
2nd lowest	21.34	1326
2nd highest	18.26	1243
Highest quartile group	16.53	1228
Test equal proportions, p-value	<0.001	
Subjective survival probability to 75		
0	45.51	312
1-49	28.71	498
50	23.96	1327
51-99	14.86	1918
100	19.74	1155
Test equal proportions, p-value	<0.001	
Total	21.42	5210

Notes: Sample includes respondents used in estimation of mortality models who are observed alive in wave 3 (1996). See Appendix Table A3 for mortality rates for sample that includes observations with missing values of subjective survival probability and model covariates. Cognition groups defined as in notes to Table 1. Subjective survival probability is the value reported in wave 3. Test of equal proportions is the Pearson Chi-square test.

Table 3: Longevity expectations coefficient in mortality hazard models

		Controlling for:	
	Age-sex	Age-sex + socioeconomics	Age-sex + socioeconomics + mortality risks
Full sample	-0.0168***	-0.0129***	-0.0046***
	[0.0012]	[0.0012]	[0.0012]
Log likelihood	-3944.918	-3808.477	-3368.997
N	32437	32437	32437
Drop if report survival probability:			
0	-0.0131***	-0.0102***	-0.0035**
	[0.0015]	[0.0015]	[0.0015]
N	30770	30770	30770
50	-0.0166***	-0.0130***	-0.0048***
	[0.0012]	[0.0013]	[0.0013]
N	24068	24068	24068
100	-0.0204***	-0.0159***	-0.0063***
	[0.0014]	[0.0015]	[0.0015]
N	27332	27332	27332

Notes: Coefficients on difference between the subjective and life table probability of survival to a target age (*longevity expectation*) from complementary log-log hazard models of mortality. *Mortality risks* in third column refer to indicators of health, smoking and parental longevity. Full estimates from the most general specification are given in Appendix Table A6. Standard errors adjusted for clustering at individual level in parentheses. *, ** and *** indicate significance at 10%, 5% and 1% respectively. Bottom panel shows estimates obtained from sub-samples that exclude observations reporting subjective survival probability of 0, 50 and 100 percent.

Table 4: Longevity expectations coefficients in mortality hazard models with heterogeneity by education and cognition

	Education, with controls for:			Cognition, with controls for:			Education & Cognition, with controls for:		
	Age-sex	Age-sex + socio-economics	Age-sex + socio-economics+ mortality risks	Age-sex	Age-sex + socio-economics	Age-sex + socio-economics + mortality risks	Age-sex	Age-sex + socio-economics	Age-sex + socio-economics + mortality risks
Education									
High school dropout	-0.0121***	-0.0094***	-0.0023				-0.0107***	-0.0083***	-0.0020
	[0.0019]	[0.0019]	[0.0019]				[0.0021]	[0.0020]	[0.0020]
High school graduate	-0.0160***	-0.0131***	-0.0050**				-0.0134***	-0.0109***	-0.0043*
	[0.0022]	[0.0021]	[0.0021]				[0.0027]	[0.0026]	[0.0025]
Some College	**-0.0188*** **	**-0.0156*** **	-0.0063***				-0.0159***	-0.0130***	-0.0053*
	[0.0026]	**[0.0025]**	[0.0024]				[0.0032]	[0.0031]	[0.0030]
College graduate	**-0.0210*** **	**-0.0176*** **	-0.0072**				**-0.0178*** **	**-0.0148*** **	-0.0062
	[0.0033]	**[0.0032]**	[0.0030]				**[0.0041]**	**[0.0039]**	[0.0038]
Cognitive functioning									
Lowest quartile group				-0.0127***	-0.0099***	-0.0032*	-0.0107***	-0.0083***	-0.0020
				[0.0018]	[0.0017]	[0.0018]	[0.0021]	[0.0020]	[0.0020]
2nd lowest group				-0.0178***	-0.0143***	-0.0049**	-0.0142***	-0.0114***	-0.0027
				[0.0024]	[0.0023]	[0.0023]	[0.0030]	[0.0029]	[0.0029]
2nd highest group				-0.0164***	-0.0128***	-0.0039	-0.0123***	-0.0093***	-0.0014
				[0.0027]	[0.0027]	[0.0026]	[0.0034]	[0.0033]	[0.0033]
Highest quartile group				**-0.0203*** **	**-0.0175*** **	-0.0082***	-0.0157***	-0.0135***	-0.0055
				[0.0029]	**[0.0028]**	[0.0027]	[0.0037]	[0.0036]	[0.0034]
Homog. $\chi^2(k-1)$ (p-value)	7.64 (0.0540)	7.16 (0.0670)	2.98 (0.3943)	6.28 (0.0987)	6.27 (0.0990)	2.76 (0.4304)	10.12 (0.1198)	9.93 (0.1278)	4.40 (0.6226)
Log likelihood	-3931.5119	-3804.487	-3367.3963	-3926.5436	-3802.6306	-3367.3553	-3920.2381	-3800.5753	-3366.4387
N	32437	32437	32437	32437	32437	32437	32437	32437	32437

Notes: As Table 3, with the following additions. This table gives estimates from models that include interactions between the longevity expectation variable and education and/or cognition categories. The coefficient on the longevity expectation variable for each education/cognition category is presented (i.e. coefficient for reference category plus coefficient on category-specific interaction). These approximate the relative effect on the hazard rate of a marginal change in longevity expectations for each category. For the model allowing heterogeneity by education and cognition (right-hand panel), the effects by education are for the lowest quartile group of cognition and the effects by cognition are for high school dropouts. Standard errors in parentheses are estimated by the delta method with adjustment for clustering at the individual level. Bold indicates that the coefficient differs at least at 10% from that for the lowest education/cognition group. Homog. indicates a Wald test of the null that the variation in the hazard with longevity expectations does not differ by education/cognition. k is the number of education and/or cognition groups.

Table 5: Longevity expectations coefficients in mortality hazard models – robustness to dropping focal responses

	Dropping subjective survival probability responses of:					
	0		50		100	
	Education	Cognition	Education	Cognition	Education	Cognition
Lowest group	-0.0078***	-0.0086***	-0.0092***	-0.0100***	-0.0118***	-0.0119***
	[0.0023]	[0.0021]	[0.0019]	[0.0017]	[0.0023]	[0.0021]
2nd lowest group	-0.0085***	-0.0097***	-0.0132***	-0.0144***	-0.0157***	-0.0161***
	[0.0025]	[0.0027]	[0.0020]	[0.0024]	[0.0025]	[0.0027]
2nd highest group	-0.0131***	-0.0115***	**-0.0155*** **	-0.0136***	**-0.0181*** **	-0.0161***
	[0.0029]	[0.0032]	[0.0026]	[0.0026]	[0.0029]	[0.0030]
Highest group	-0.0147***	-0.0128***	**-0.0190*** **	**-0.0171*** **	**-0.0214*** **	**-0.0225*** **
	[0.0036]	[0.0033]	[0.0032]	[0.0028]	[0.0033]	[0.0030]
Homog. $\chi^2(k-1)$ (p-value)	4.11(0.2502)	1.38(0.7095)	8.70(0.0336)	5.71(0.1268)	6.87(0.0760)	8.91(0.0304)
Log likelihood	-3296.6838	-3296.4852	-2831.0412	-2831.3772	-3330.7778	-3326.9095
N	30770	30770	24068	24068	27332	27332

Notes: as Table 4. Estimates correspond to those from models including age-sex and socioeconomic controls reported in Table 4 with observations reporting subjective survival probabilities of 0, 50 and 100 percent dropped from the sample. Lowest group is high school dropouts for education and lowest quartile group for cognition. Highest group is college graduates for education and highest quartile group for cognition. Intermediate groups as in Table 4. Homogeneity test as in Table 4. Bold indicates that the group coefficient is significantly different from that of the lowest group at 10% level or less.

Table 6: Longevity expectations coefficients in mortality hazard models with heterogeneity by education and numeracy

	Education	Numeracy	Education & Numeracy
Education			
High school dropout	-0.0122***		-0.0099***
	[0.0025]		[0.0028]
High school graduate	-0.0125***		-0.0082
	[0.0027]		[0.0031]
Some College	-0.0161***		-0.0110
	[0.0033]		[0.0040]
College graduate	-0.0178***		-0.0115
	[0.0036]		[0.0046]
Numeracy			
Lowest quartile group		-0.0096***	-0.0099***
		[0.0023]	[0.0028]
2nd lowest group		-0.0144***	-0.0099***
		[0.0032]	[0.0038]
2nd highest group		-0.0155***	-0.0146***
		[0.0030]	[0.0040]
Highest quartile group		**-0.0206*** **	**-0.0153*** **
		[0.0035]	[0.0045]
Homog. $\chi^2(k-1)$ (p-value)	2.38 (0.4980)	7.87 (0.0488)	8.76 (0.1875)
Log likelihood	-2340.65	-2335.537	-2335.138
N	16963	16963	16963

Notes: As Table 4. Estimates obtained from models estimated using data from wave 6 onwards including age-sex and socioeconomic controls only. Numeracy categories are quartile groups formed from age-sex standardised numeracy score. Bold indicates significant difference from lowest group at 5% level or less.

Table 7: Updating of longevity expectations – full sample and by education

	Full sample	High school dropouts	High school graduates	Some college	College graduates
Lagged longevity expectation	0.5463***	0.4957***	**0.5268*** **	**0.5689*** **	**0.5788*** **
	[0.0054]	[0.0113]	[0.0091]	[0.0112]	[0.0115]
Onset of health condition					
Lung disease/Cancer	-6.7950***	-6.2421***	-7.2578***	-6.6255***	-6.3499***
	[0.8068]	[1.8196]	[1.3545]	[1.6486]	[1.6703]
Diabetes	-2.0164**	-1.7168	-2.2653	-1.402	-0.3005
	[0.9147]	[2.0117]	[1.6158]	[1.6970]	[1.6782]
Psychiatric illness	-6.7744***	-9.4413***	-5.4153**	-6.4518***	**-2.5837**
	[1.2754]	[2.6060]	[2.2020]	[2.2931]	[2.8032]
Stroke/Heart disease	-3.5715***	-2.1705	-5.1990***	-3.4582**	-2.8791**
	[0.7626]	[1.8607]	[1.2781]	[1.5708]	[1.4152]
Hypertension/Arthritis	0.3562	0.6423	-0.7564	1.7896*	0.2424
	[0.4816]	[1.2301]	[0.8636]	[0.9158]	[0.8280]
New hospitalization	-1.8658***	-1.6205	-1.7517**	-1.1172	-2.5242***
	[0.4210]	[1.0131]	[0.7443]	[0.7944]	[0.7662]
Mother alive at last wave	2.8253***	3.3160***	2.1832***	2.4056***	2.7292***
	[0.3136]	[0.8588]	[0.5499]	[0.6034]	[0.5223]
Father alive at last wave	3.0075***	2.7183*	2.9248***	3.2064***	1.6713**
	[0.4701]	[1.5462]	[0.8289]	[0.8272]	[0.7802]
Widowed since last wave	-0.9443	0.0975	-1.0854	0.4621	-1.6033
	[0.9117]	[1.9548]	[1.3914]	[2.1031]	[1.7996]
Female	-3.4906***	-5.1497***	**-3.4970*** **	**-2.4513*** **	**-1.7765*** **
	[0.2714]	[0.7129]	[0.4724]	[0.5427]	[0.4731]
Age - 54	-0.6637***	-0.9026***	-0.6748***	-0.5799**	-0.5521***
	[0.1175]	[0.3141]	[0.2015]	[0.2292]	[0.2014]
(Age -54) squared	0.0439***	0.0577***	0.0446***	0.0368***	0.0385***
	[0.0054]	[0.0140]	[0.0091]	[0.0106]	[0.0094]
Target age changed from 75 to 80	2.4269***	3.1540***	1.6492**	2.4694***	2.6973***
	[0.4469]	[1.1274]	[0.7691]	[0.9007]	[0.7914]
Target age changed from 80 to 85	10.5858***	13.1162***	**9.0061*** **	10.5469***	10.4555***
	[0.5378]	[1.3138]	[0.8976]	[1.1486]	[0.9934]
Target age changed from 85 to 90	8.7081***	11.1887***	8.6473***	6.4543***	8.1964***
	[1.1192]	[2.7281]	[1.8533]	[2.4772]	[2.0083]
Intercept	-1.0092	-4.0361**	-0.7352	**-0.5466**	**0.4012**
	[0.6247]	[1.6977]	[1.0867]	[1.2074]	[1.0541]
R-squared	0.3366	0.2983	0.3125	0.3518	0.3648
N	32437	6966	11175	7132	7164

Notes: Least squares estimates of model (4) for full sample and for each education group. Robust standard errors in parentheses. Bold indicates coefficient is significantly different from respective coefficient estimated for high school dropouts at least at 10% level.

ONLINE APPENDICES

Appendix A: Additional Tables

Table A1: Target ages used in subjective probability of survival question

	Respondent's current age:				
Wave	**< 65**	**65**	**66-69**	**70-74**	**75-79**
3 (cohort aged 54-65)	75	75	-	-	-
4 (cohort aged 56-67)	75	75	-[a]	-	-
5 (cohort aged 58-69)	75[b]	80[c]	80	-	-
6 (cohort aged 60-71)	75[b]	80[c]	80	85	-
7 (cohort aged 62-73)	75[b]	80[c]	80	85	-
8 (cohort aged 64-75)	75[b]	80[c]	80	85	90
9 (cohort aged 66-77)	-	-	80	85	90

Notes: a. In wave 4, HRS documentation indicates that the question was to be skipped if the respondent was older than 75. But from the data it appears that the question was skipped if older than 65. This means that there is no information on survival expectations for individuals aged 66 and 67 in our sample in wave 4. b. Also asked to report the chance of living to age 80 (85 in wave 8). c. Also asked to report the chance of living to age 75. We opted to use the age 80 target because there are many missing observations for target age of 75 in wave 8, and also to maintain greater consistency across waves.

Table A2: Mean difference between reported and life table survival probability by education, cognition and numeracy

Education	Quartile group of total cognition score				Test of equal means: p-value
	Lowest quartile	2nd lowest	2nd highest	Highest quartile	
High school dropout and GED	-13.42	-14.21	-12.64	-11.41	0.234
	(N=3550)	(N=1619)	(N=1112)	(N=685)	
High school graduate	-8.54	-6.57	-4.80	-4.74	<0.001
	(N=2814)	(N=3233)	(N=2793)	(N=2335)	
Some college	-3.79	-6.49	-2.57	0.99	<0.001
	(N=1286)	(N=1881)	(N=1855)	(N=2110)	
College graduate	1.52	-0.74	-0.98	3.38	<0.001
	(N=573)	(N=1476)	(N=2207)	(N=2908)	
Test of equal means: p-value	<0.001	<0.001	<0.001	<0.001	<0.001
	Quartile group of numeracy score				
	Lowest quartile	2nd lowest	2nd highest	Highest quartile	
High school dropout and	-10.10	-11.90	-8.82	-13.10	0.1683
	(N=1840)	(N=850)	(N=594)	(N=316)	
High school graduate	-7.61	-3.12	-3.38	-4.27	<0.001
	(N=1563)	(N=1686)	(N=1384)	(N=1202)	
Some college	-0.66	-2.57	-0.38	-2.07	0.2737
	(N=708)	(N=954)	(N=1003)	(N=1041)	
College graduate	-2.13	1.40	3.56	3.41	0.001
Test of equal means: p-value	(N=366)	(N=834)	(N=1137)	(N=1485)	
	<0.001	<0.001	<0.001	<0.001	<0.001

Notes: as Table 1 in main text.

Table A3: Table 2 for sample including observations missing on subjective survival probability and covariates

	As Table 2 + missing on covariates		As Table 2 + missing on covariates & subjective survival probability	
Education	Percent	N	Percent	N
High school dropout & GED	32.78	2004	32.68	2240
High school graduate	21.02	2374	21.04	2472
Some College	21.05	1425	21.2	1462
College	16.39	1287	16.74	1314
Test equal proportions, p-value	<0.001	7090	<0.001	7488
Cognitive functioning				
Lowest quartile group	33.93	1677	34.09	1889
2nd lowest	24.13	1832	24.09	1914
2nd highest	18.66	1801	18.59	1861
Highest quartile group	17.98	1780	18.15	1824
Test equal proportions, p-value	<0.001	7090	<0.001	7488
Subjective survival probability to 75				
0	48.28	495		
1-49	31.28	729		
50	25.15	1793		
51-99	16.51	2435		
100	21.18	1638		
Test equal proportions, p-value	<0.001	7090		
Total	23.51	7090	23.80	7488

Notes: As Table 2 in text.

Table A4: Table corresponding to Figure 2 - Percentage of individuals who die between 1996 and 2010 by subjective probability of survival to 75, education and cognition

	Subjective probability of survival to 75					
	0	1-49	50	51-99	100	Test equality
Education						p-value
High school drop out	46.95	35.22	29.28	22.71	27.34	<0.001
N	164	159	321	273	267	
High school graduate	41.76	27.22	23.19	14.26	16.36	<0.001
N	91	169	526	582	434	
Some college	45.95	28.28	23.05	15.01	18.38	<0.001
N	37	99	269	453	272	
College graduate	50.00	18.31	18.96	11.80	18.68	<0.001
N	20	71	211	610	182	
Cognitive functioning						
Lowest quartile group	49.02	34.86	30.71	18.78	24.48	<0.001
	153	175	368	378	339	
2nd lowest	51.25	23.08	25.76	15.78	16.07	<0.001
	80	117	361	488	280	
2nd highest	36.36	33.03	19.58	12.80	17.62	<0.001
	44	109	337	492	261	
Highest quartile group	28.57	19.59	17.62	13.21	19.64	<0.001
	35	97	261	560	275	

Notes: As Table 2 and Figure 2.

Table A5: Definitions of socioeconomic, health, health behaviour and parental longevity covariates used in models of mortality

Variable	Description
Education	In categories: high school drop outs and General Educational Development (GED) (0-11 years of schooling and no qualification other than GED); high school graduate (12 years of schooling and high school diploma); some college (more than 12 years of schooling and high school diploma or GED); college and above (bachelor degree or higher).
Education parents	Number of education years of mother and of father.
Income	Logarithm of total household income.
Wealth	Logarithm of total wealth (excluding secondary residence) and indicator of positive wealth (0,1).
Employment	Indicator of whether currently working for pay (0,1).
Race	Categories White/Caucasian, Black/African American and other (reference category).
Marital status	Indicator of whether married/partnered (1) or widowed, divorced or never married (0).
Lung disease	Respondent has ever been told by a doctor that he/she has the condition. Mortality models include an indicator of whether condition reported in the previous wave (0,1) and another of whether did not report condition in last wave but did in present wave (*X since last wave* (0,1)). Expectations updating model includes the latter only in lagged dependent variable specification and fixed effects model includes indicator of currently having the condition.
Cancer	
Diabetes	
Stroke	
Psychiatric illness	
Heart disease	
Hypertension	
Arthritis	
Cognition – recall	Age-sex standardized total score on immediate and delayed word recall tests in current wave
Cognition – count	Age-sex standardized total score on the serial 7's and backwards counting tests.
CES-D score	Score (0-8) is the sum of five "negative" indicators minus two "positive" indicators. Negative indicators correspond to reports that all or most of the time experience: depression, feeling that everything is an effort, restless sleep, loneliness, sadness, and feeling that could not get going. Positive indicators are whether the respondent felt happy and enjoyed life all or most of the time. We include the current and the lagged score.
Mobility index	Current and lagged score (0-5) that is the sum of indicators of some difficulty with: walking several blocks, walking one block, walking across the room, climbing several flights of stairs and climbing one flight of stairs. We include the current and the lagged score.
Large muscle index	Current and lagged score (0-4) that is the sum of indicators of some difficulty with: sitting for two hours, getting up from a chair, stooping or kneeling or crouching, and pushing or pulling a large object.
ADL index	Current and lagged score (0-5) that sums indicators of difficulty with the activities of daily living: bathing, eating, dressing, and getting in or out of bed. Difficulty walking across a room excluded because it is included in the mobility index.
Fine motor skills indicator	Current and lagged indicator (0,1) of whether individual has some difficulty picking up a dime. Difficulty with eating and dressing excluded because these are included in the ADL index.
iADL index	Current and lagged score (0-3) that is sum of indicators of some difficult with the instrumental activities of daily living: taking medication, using a telephone and handling money.
Hospitalization	Indicator (0,1) of whether the respondent reports any overnight hospital stay since the last interview, or the last 2 years for new interviewees. We also include the lag of this indicator.
BMI	Current and lagged body mass index - weight in kilograms divided by the square of height in meters.
Smoker	Whether current smoker, former smoker or never smoked (reference).
Cognition – count	Age-sex standardized total score on the serial 7's and backwards counting tests.
Parental longevity	Separately for father and mother: age of the parent (currently or at death, if already deceased); indicator of whether parent is deceased; and interaction between age (-at-death) and vital status.

Table A6: Mortality hazard coefficients from model with all controls and without heterogeneity in longevity coefficient by education and cognition (Table 3, right-hand column)

Longevity expectation	-0.0046***	[0.0012]			
White	0.0303	[0.1996]	Heart disease at last wave	0.1661**	[0.0842]
Black	0.0432	[0.2116]	Heart disease since last wave	0.1732	[0.1551]
Married or partnered	-0.2136***	[0.0820]	Stroke at last wave	-0.0244	[0.1264]
High school graduate	0.1666*	[0.0934]	Stroke since last wave	-0.0016	[0.2299]
Some college	0.2974***	[0.1079]	Psychiatric illness at last wave	0.0209	[0.1055]
College graduate	0.3207**	[0.1267]	Psychiatric illness since last wave	0.1134	[0.2206]
Education years mother	-0.0073	[0.0131]	Arthritis at last wave	-0.1491*	[0.0810]
Education years father	-0.0014	[0.0117]	Arthritis since last wave	-0.0522	[0.1773]
Currently working for pay	-0.4930***	[0.0948]	Lagged mobility index	0.0864**	[0.0377]
Ln(Income)	0.0159	[0.0301]	Current mobility index	0.2859***	[0.0356]
Ln(Wealth)	-0.0427**	[0.0200]	Lagged large muscle index	-0.1029**	[0.0409]]
Has positive wealth	0.2157	[0.2288]	Current large muscle index	-0.0494	[0.0399]
Cognition – recall	-0.1306***	[0.0389]	Lagged ADL index	-0.0098	[0.0609]
Cognition – count	-0.0636*	[0.0354]	Current ADL index	0.069	[0.0552]
Age father	-0.0311	[0.0326]	Lagged fine motor skills	-0.0556	[0.1385]
Age mother	-0.0162	[0.0161]	Current fine motor skills	-0.0657	[0.1337]
Father died	-2.6021	[2.8120]	Lagged iADL index	-0.0227	[0.1049]
Mother died	-1.2449	[1.3989]	Current iADL index	0.0308	[0.0944]
Father died x father's age-at-death	0.0285	[0.0327]	Lagged CES-D score	-0.0764***	[0.0229]
Mother died x mother's age-at-death	0.0145	[0.0162]	Current CES-D score	0.0338	[0.0206]
High blood pressure at last wave	0.0786	[0.0775]	Lagged hospitalization	0.3340***	[0.0771]
High blood pressure since last wave	-0.0031	[0.1651]	Current hospitalization	0.4798***	[0.0806]
Diabetes at last wave	0.5997***	[0.0816]	Lagged BMI	0.0664***	[0.0125]
Diabetes since last wave	0.2019	[0.1870]	Current BMI	-0.1055***	[0.0136]
Cancer at last wave	0.6125***	[0.0916]	Former smoker)	0.3601***	[0.0880]
Cancer since last wave	1.4437***	[0.1183]	Current smoker	0.7490***	[0.1074]
Lung disease at last wave	0.3078***	[0.0933]	Constant	0.4916	[3.1287]
Lung disease since last wave	0.5908***	[0.1627]			
N	32437		Log likelihood	0.4916	

Notes: Table shows coefficients from the complementary log-log hazard models of mortality corresponding to right-hand column of Table 3. Model also includes gender specific age-year dummy variables. Standard errors allowing for clustering at the individual level in parentheses. *, ** and *** indicate significance at 10%, 5% and 1% respectively.

Table A7: Mortality hazard coefficients with respect to health conditions

	Coefficient	Robust SE
Lung disease	0.8840***	[0.0805]
Cancer	0.8546***	[0.0747]
Diabetes	0.6396***	[0.0739]
Stroke	0.3415***	[0.1033]
Psychiatric illness	0.3379***	[0.0866]
Heart disease	0.3062***	[0.0757]
Hypertension	0.0871	[0.0734]
Arthritis	-0.0741	[0.0733]
Hospitalization	0.8446***	[0.0723]
Log-likelihood	-3679.7839	
N	32437	

Notes: Table shows coefficients from a complementary log-log hazard model of mortality. Hazard ratio given by exponentiating coefficient. Each health condition is represented by an indicator of ever having been diagnosed with the condition. Hospitalization is an indicator of having been hospitalized in the past two years. Model also includes a dummy for gender and a complete set of age (year) indicators. Standard errors allowing for clustering at the individual level in parentheses. *** indicates significance at 1%.

Table A8: Fixed effects model of longevity expectations – full sample and by education

	Full sample	**High school dropouts**	**High school graduates**	**Some college**	**College graduates**
Ever had health condition					
Lung disease/Cancer	-5.1413***	-4.9959***	-5.6298***	-4.4667***	-5.0658***
	[0.7644]	[1.9028]	[1.2561]	[1.4953]	[1.5297]
Diabetes	-0.4931	-1.248	-0.0918	-2.6018	2.433
	[0.8269]	[1.9450]	[1.2510]	[1.7455]	[1.5800]
Psychiatric illness	-3.1104***	-5.8389**	-2.3067	-2.9894	-0.3406
	[1.1293]	[2.3865]	[2.0230]	[2.0890]	[2.4610]
Stroke/Heart disease	-2.7541***	-3.7930*	-1.6658	-2.5138*	-3.6262***
	[0.7402]	[1.9544]	[1.2553]	[1.5050]	[1.2515]
Hypertension/Arthritis	0.1465	-0.7376	0.129	0.9788	-0.0845
	[0.5804]	[1.5467]	[1.0702]	[1.1912]	[0.9394]
Hospitalization	-1.3874***	-1.4394	-1.5280**	-1.5211**	-0.9943
	[0.3569]	[0.8777]	[0.6243]	[0.6970]	[0.6481]
Mother alive at current wave	-0.2965	1.5099	-1.7444	-0.0633	-0.0959
	[0.6018]	[1.7047]	[1.0867]	[1.0767]	[0.9881]
Father alive at current wave	-0.0702	-3.0062	1.4211	0.9894	-1.395
	[0.8982]	[3.1465]	[1.5649]	[1.5564]	[1.5060]
Widow/er at current wave	0.1129	-0.5254	-0.0669	1.439	-0.2245
	[0.8051]	[1.7554]	[1.2910]	[1.7007]	[1.5879]
Age - 54	-0.9694***	-0.8644**	-1.0090***	-1.0436***	-0.9497***
	[0.1354]	[0.3631]	[0.2343]	[0.2693]	[0.2378]
(Age - 54) squared	0.0145**	0.0048	0.0194	0.0176	0.0145
	[0.0072]	[0.0190]	[0.0122]	[0.0150]	[0.0129]
Target age is 80	4.1481***	5.5848***	3.7536***	3.8418***	3.6776***
	[0.4830]	[1.2720]	[0.8314]	[0.9729]	[0.8479]
Target age is 85	16.7964***	20.3906***	**14.9010*** **	16.3239***	16.6756***
	[0.9190]	[2.3371]	[1.5326]	[1.9677]	[1.6835]
Target age is 90 or more	26.5212***	33.6395***	**25.0570*** **	**22.0048*** **	25.8518***
	[1.6695]	[4.2597]	[2.7310]	[3.6755]	[3.0250]
Intercept	0.3485	-6.4934***	**-0.3592**	**2.8850* **	**6.0956*** **
	[0.7568]	[2.0718]	[1.3490]	[1.5319]	[1.2400]
R-squared	0.0498	0.0482	0.0447	0.0499	0.0732
N	32437	6966	11175	7132	7164

Notes: Estimates of fixed effects model for full sample and for each education group. Instead of indicators of health/widowhood/target age changes between the previous and the current waves (as in model (4)), this model includes current indicators. In particular, we include indicators of: ever having been diagnosed with each condition; of having been hospitalized in the past two years; of being currently a widow/er, and of each target age (reference is 75). Parental living status is represented by indicators of whether mother/father is currently alive. Age and gender are modelled as in model (4). Robust standard errors in parentheses. Bold indicates coefficient is significantly different from respective coefficient estimated for high school dropouts at least at 10% level.

Table A9: Updating of longevity expectations by cognition

	Lowest quartile	2nd lowest	2nd highest	Highest quartile
Lagged longevity expectation	0.4686***	**0.5658*** **	**0.5831*** **	**0.5812*** **
	[0.0109]	[0.0101]	[0.0102]	[0.0107]
Onset of health condition				
Lung disease/Cancer	-3.8018**	**-7.9288*** **	**-8.6482*** **	-7.3121***
	[1.7812]	[1.5849]	[1.5733]	[1.4781]
Diabetes	-1.9809	-0.7162	-1.733	-3.2172**
	[1.9627]	[1.7783]	[1.6960]	[1.5810]
Psychiatric illness	-6.5160**	-9.5461***	-3.7537	-5.7612**
	[2.6044]	[2.0488]	[2.6882]	[2.7243]
Stroke/Heart disease	-3.5747**	-3.6532**	-3.7931***	-3.1317**
	[1.7686]	[1.4960]	[1.3493]	[1.3663]
Hypertension/Arthritis	0.1537	0.3945	-0.5256	1.2842
	[1.1506]	[0.9361]	[0.8829]	[0.8323]
New hospitalization	-2.4718**	-1.4451*	-1.9134**	-1.3409*
	[0.9684]	[0.8089]	[0.7786]	[0.7385]
Mother alive at last wave	3.3267***	2.4240***	2.7418***	2.4147***
	[0.7389]	[0.6226]	[0.5682]	[0.5595]
Father alive at last wave	4.1210***	3.3170***	1.8760**	2.2680***
	[1.2206]	[0.9347]	[0.8340]	[0.7966]
Widowed since last wave	1.551	-0.9871	**-3.2048***	-1.5648
	[1.7087]	[1.9219]	[1.9277]	[1.7328]
Female	-4.5706***	**-2.9848*** **	**-2.6276*** **	-3.6386***
	[0.6317]	[0.5312]	[0.4956]	[0.4929]
Age – 54	-1.1121***	**-0.4811** **	**-0.4043***	-0.6723***
	[0.2792]	[0.2240]	[0.2140]	[0.2147]
(Age - 54) squared	0.0696***	**0.0352*** **	**0.0301*** **	**0.0404*** **
	[0.0127]	[0.0102]	[0.0100]	[0.0098]
Target age changed from 75 to 80	2.9533***	1.4370	1.6245**	3.7217***
	[1.0273]	[0.8839]	[0.8246]	[0.7957]
Target age changed from 80 to 85	12.5981***	10.4589***	10.4421***	**8.8431*** **
	[1.2575]	[1.0411]	[1.0511]	[0.9568]
Target age changed from 85 to 90	9.6208***	6.1292***	11.1600***	8.0242***
	[2.7155]	[2.2393]	[2.1455]	[1.9439]
Intercept	-1.9348	-2.3565**	-1.4026	**1.6878**
	[1.4940]	[1.1949]	[1.1191]	[1.1338]
R-squared	0.2670	0.3516	0.3759	0.3743
N	8223	8209	7967	8038

Notes: As for Table 7, except that here groups are based on quartiles of total cognition score

Table A10: Updating of longevity expectations by numeracy

	Lowest quartile	**2nd lowest**	**2nd highest**	**Highest quartile**
Lagged longevity expectation	0.4841***	**0.5771*** **	**0.5843*** **	**0.6080*** **
	[0.0144]	[0.0140]	[0.0142]	[0.0142]
Onset of health condition				
Lung disease/Cancer	-4.5317**	-3.4926*	-7.6592***	-6.8073***
	[2.2152]	[1.8809]	[1.9788]	[1.8556]
Diabetes	-1.0091	-4.2286*	-2.3673	-0.9311
	[2.3953]	[2.5054]	[2.1200]	[2.1219]
Psychiatric illness	-4.3694	-4.2132	-10.8610***	-6.0394*
	[3.7043]	[3.2229]	[3.3918]	[3.2051]
Stroke/Heart disease	-2.9018	0.1637	-2.9794	-3.9389**
	[2.0986]	[1.8742]	[1.8582]	[1.7310]
Hypertension/Arthritis	0.4551	0.5789	1.8736	-0.7856
	[1.4711]	[1.2457]	[1.2340]	[1.1322]
New hospitalization	-1.2904	-1.8633*	-1.7234	-1.5502
	[1.2057]	[1.0838]	[1.0492]	[0.9996]
Mother alive at last wave	3.6096***	3.5276***	**0.7197**	4.3028***
	[1.1238]	[0.9478]	[0.9879]	[0.8255]
Father alive at last wave	5.2294**	2.8733*	3.5598**	5.7351***
	[2.4138]	[1.6220]	[1.7794]	[1.4002]
Widowed since last wave	-4.3240*	-1.0822	-1.2721	0.1283
	[2.5657]	[2.1249]	[2.0189]	[2.6325]
Female	-4.8475***	-3.6370***	-3.9664***	**-2.8236*** **
	[0.8504]	[0.7343]	[0.7333]	[0.6954]
Age – 54	-1.3675*	**0.6742**	**0.9307**	0.4036
	[0.8210]	[0.7567]	[0.8337]	[0.7046]
(Age -54) squared	0.0772**	**-0.0073**	**-0.0135**	**0.0081**
	[0.0303]	[0.0276]	[0.0284]	[0.0253]
Target age changed from 75 to 80	3.0330**	2.2605**	1.9859*	3.0527***
	[1.2054]	[1.0629]	[1.1249]	[1.0713]
Target age changed from 80 to 85	13.5035***	11.9616***	**8.4481*** **	**7.4986*** **
	[1.3177]	[1.1004]	[0.9903]	[1.0133]
Target age changed from 85 to 90	12.2427***	8.0156***	11.3444***	6.7202***
	[3.0688]	[2.5683]	[2.0295]	[2.2312]
Intercept	-0.1163	-9.6026*	-9.694	-7.7117
	[5.3618]	[4.9357]	[6.0811]	[4.6932]
R-squared	0.2936	0.3627	0.3757	0.411
N	4477	4324	4118	4044

Notes: As for Table 7, except that here groups are based on quartiles of total numeracy score and estimates obtained using data from wave 6 onwards.

Appendix B: Derivation of empirical specification of expectations updating from Bayesian learning model

If expectations are updated in a Bayesian fashion, then the posterior probability of survival to a target age (P_{it}) is a weighted average of the prior (P_{it-1}) and the risk assessed on the basis of new information that may be gleaned, for example, from the onset of disease or the death of a parent,

$$P_{it} = \frac{\varphi}{\varphi+\xi} P_{it-1} + \frac{\xi}{\varphi+\xi} r_{it} \;, \qquad \varphi > 0, \xi > 0,\; 0 < r_{it} < 100 \qquad \text{(A1)}$$

where r_{it} is the *risk equivalent* of the new information utilized, and φ and ξ represent the precision of beliefs in the prior and the risk assessment respectively (Viscusi and O'Connor 1984, Viscusi 1991). The individual is assumed to process information, such as being diagnosed with cancer, into a probability of survival – a risk level (r_{it}). If the individual were 100% confident in this new information, then the prior would be completely discarded. But there is likely to be uncertainty in the prognosis for a given diagnosis, such that the posterior shifts only marginally from the prior. The greater the *relative* precision of the belief in the prior, the less will be the deviation from it.

Let the risk equivalent weighted by its precision parameter be a linear function of observable information and a random error, such that (A1) becomes,

$$P_{it} = \frac{\varphi}{\varphi+\xi} P_{it-1} + \alpha + \Delta\mathbf{Health_{it}}\boldsymbol{\delta} + \Delta\mathbf{Relative_{it}}\boldsymbol{\theta} + g\left(Age_{it}, Gender_i\right) + u_{it} \,. \qquad \text{(A2)}$$

Note that age and gender are relevant, even though their values at *t* are predictable when expectations are formed at *t-1* because the probability of surviving to a target age rises in a way that depends on age and gender at time passes. Consequently, conditional on the previously reported probability, age and gender always provide relevant information in reassessing the probability of surviving to the target age.

Replacing each subjective probability with its difference from the respective life table probability $\left(P_{it}^{L}\right)$ gives,

$$\tilde{P}_{it} = \frac{\varphi}{\varphi+\xi} \tilde{P}_{it-1} + \alpha + \Delta\mathbf{Health_{it}}\boldsymbol{\delta} + \Delta\mathbf{Relative_{it}}\boldsymbol{\theta} + g\left(Age_{it}, Gender_i\right) - \left(P_{it}^{L} - \frac{\varphi}{\varphi+\xi} P_{it-1}^{L}\right) + u_{it} \qquad \text{(A3)}$$

The term in parentheses is a function of the change in the life table probability of reaching a target age. This varies only with age and gender. Subsuming it into the demographic effects results in (4).