TI 2026-013/IV
Tinbergen Institute Discussion Paper

Use of Shared and Private Information in Long-Term Care Risk Perceptions

Lisa Voois[1]
Teresa Bago d-Uva[2]
Owen O'Donnell[3]

1 Erasmus University Rotterdam, Tinbergen Institute

2 Erasmus University Rotterdam, Tinbergen Institute

3 Erasmus University Rotterdam, Tinbergen Institute

Use of Shared and Private Information in Long-Term Care Risk Perceptions

Lisa Voois[1], Teresa Bago d'Uva[2], Owen O'Donnell[3]

[1] Erasmus School of Health Policy and Management, Erasmus University Rotterdam, voois@eshpm.eur.nl

[2] Erasmus School of Economics, Erasmus University Rotterdam, bagoduva@ese.eur.nl

[3] Erasmus School of Economics & Erasmus School of Health Policy and Management, Erasmus University Rotterdam & Tinbergen Institute, ORCID 0000-0002-6289-1924, odonnell@ese.eur.nl

Abstract

Misperception of long-term care (LTC) risk may distort insurance and saving decisions. Comparing older Americans' subjective probabilities of nursing home entry with realized outcomes, we find LTC risk perceptions are inaccurate, partly due to inappropriate weighting of risk factors insurers can observe. Risk perceptions capture only 37% of the potential discriminatory power of this shared information. Private information offsets only one third of the resulting inaccuracy. LTC insurance take-up is positively associated with perceived risk even after adjusting for confounders and reverse causality. These findings are consistent with selection out of insurance partly due to underutilization of shared information.

Keywords: subjective probability, information friction, mental gap, cognitive bias, behavioral insurance, long-term care insurance

JEL: D82, D83, D84, I13, J14

Acknowledgements: The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. We thank participants at various conferences and seminars for comments.

1. Introduction

Take-up of private long-term care insurance is surprisingly low given that formal-care spending is a major financial risk in old age and public insurance is limited in most countries. Misperception of long-term care (LTC) risk could explain this puzzle (Boyer et al., 2019; De Donder & Leroux, 2013; Pauly, 1990; Pestieau & Ponthière, 2012). Perceptions of this risk have been shown to increase with risk factors and to predict LTC use (Akamigbo & Wolinsky, 2006; Finkelstein & McGarry, 2006; Holden et al., 1997; Lindrooth et al., 2000; Taylor et al., 2005) but their accuracy has not been measured. Comparison with objective risks predicted from risk factors (Boyer et al., 2019) can gauge use of information on observable LTC predictors but does not allow for private information causing appropriate deviation of the perceived from the objective risk. We measure the accuracy of older Americans' LTC risk perceptions by comparing them with event outcomes (moving to a nursing home or not), and we decompose that accuracy to reveal the power of perceptions to discriminate between outcomes, as well as the extent to which accuracy is improved through utilization of private information and worsened by underutilization of information that insurance applicants must share with insurers.

Formation of accurate LTC risk perceptions involves acquisition of extensive health information, recognition of risk factors and exertion of cognitive effort to process information. Information frictions and cognitive limitations or biases (Handel & Kolstad, 2015; Handel et al., 2019; Handel & Schwartzstein, 2018; Spinnewijn, 2017) can cause incorrect (even zero) weighting of relevant risk factors (Schwartzstein, 2014) and diversion of attention to salient but irrelevant factors (Bordalo et al., 2012; Gennaioli & Shleifer, 2010). Each error, in isolation, would distort the subjective expectation of uninsured LTC costs away from the actuarially fair insurance price calculated conditional on risk factor information that an applicant would be obliged to share with an insurer (Baillon et al., 2022). Underutilization of this shared

information may lead to rejection of insurance offers that would be accepted if risk perceptions were accurate. In that case, distortion of risk perceptions may partly offset the influence of private information, constraining the scope for adverse selection (Handel, 2013).

We use data from the Health and Retirement Study (HRS) to assess the accuracy of LTC risk perceptions by comparing reported subjective probabilities of moving to a nursing home within five years with the individual-specific outcomes of that event. We decompose the inaccuracy (mean squared prediction error) of the subjective probabilities into bias, noise and discriminatory power of the subjective probabilities, and outcome variability (Yates, 1982). We further decompose the discriminatory power – the difference in risk perceptions of those who do and do not move to a nursing home – into the predictability of nursing home entry, the underutilization of shared information due to inappropriate weighting of risk factors and the use of private information (Bago d'Uva & O'Donnell, 2022). This is done by modeling both the subjective probabilities and the outcomes as functions of risk factors that LTC insurance applicants and providers can both observe.

On average, older Americans overestimate their chances of moving to a nursing home by almost five percentage points. This bias is a relatively small contributor to the inaccuracy of the risk perceptions. Outcome variability, which increases prediction difficulty, and noise in the subjective probabilities contribute most. This is partly offset by their ability to discriminate, around 37% of which comes from private information. While there is previous evidence that these probabilities contain private information (Finkelstein & McGarry, 2006; Hendren, 2013), this is the first study to quantify the importance of that information to the accuracy of LTC risk perceptions. The remaining 63% of the discriminatory power is from use of shared information, although only 37% of the potential of this information to discriminate between those who move to a nursing home and those who do not is realized. The unrealized potential indicates that weights implicitly placed on risk factors in the formation of subjective probabilities deviate

substantially from the error-minimizing weights that an insurer could estimate by regressing the outcome on the same risk factors. Age is the most underweighted risk factor, followed by health indicators and prior LTC use. While there is overestimation of LTC risk on average, many report a subjective probability below their objective risk predicted from observable risk factors. They would perceive insurance as overpriced if offered at a premium that is actuarially fair given those risk factors. This is consistent with high price being the most common reason given for not purchasing LTC insurance (Brown et al., 2012).

We examine heterogeneity by wealth, since misperception of LTC risk is expected to be more consequential for wealthier individuals who are less protected through means-tested public insurance (Medicaid) (Braun et al., 2019; Brown & Finkelstein, 2011), and by education and cognition because each may affect exposure to information frictions and proneness to cognitive limitations/biases. The least wealthy, educated and cognitively able have the least accurate LTC risk perceptions. Differences by wealth and education are fully explained by differences in cognitive ability. The least cognitively able report the noisiest subjective probabilities that contain the least private information and make the least use of the available shared information. The bottom quartile cognition group fails to use 71% of the potential discriminatory power of shared information, in contrast with 5% for the top quartile.

Evidence of inaccurate LTC risk perceptions potentially helps explain the LTC insurance puzzle only if insurance decisions are based on the reported risks. Consistent with previous evidence (Brown et al., 2012; Finkelstein & McGarry, 2006; Zhou-Richter et al., 2010), we find that holding private LTC insurance is positively associated with the reported subjective probability of moving to a nursing home. Given the potential for this correlation to arise from confounders, such as past behavior that separately influences both risk perceptions and insurance, we confirm robustness to controlling for an extensive battery of controls and calculating Oster (2019) bounds. We also show that LTC insurance is associated with the lagged subjective probability,

suggesting that the contemporaneous correlation is not due to insurance raising the perceived likelihood of moving to a nursing home. Finally, the association remains after instrumenting the subjective probability with heteroskedasticity-based instruments (Lewbel, 2012). While we do not claim evidence that inaccurate reported risk perceptions cause mistaken insurance choices, the robust positive correlation is at least consistent with this interpretation.

Previous evidence that reported subjective probabilities of moving to a nursing home correlate with risk factors and predict the outcome (Akamigbo & Wolinsky, 2006; Finkelstein & McGarry, 2006; Holden et al., 1997; Lindrooth et al., 2000; Taylor et al., 2005) does not indicate the extent to which risk factors are weighted correctly and risk perceptions are accurate. We address these limitations by using the mean squared prediction error of subjective probabilities vis-à-vis realized outcomes to measure and decompose the accuracy of LTC risk perceptions.

Evidence from the US (Finkelstein & McGarry, 2006) and Canada (Boyer et al., 2019) indicates that (pessimism) bias in LTC risk perceptions is quite small, but there is much variation, suggesting uncertainty about future LTC needs, with potential consequences for insurance and saving behavior (Ameriks et al., 2020; De Donder & Leroux, 2013). We confirm and extend these findings by showing that subjective probabilities of moving to a nursing home reflect individuals' risk profiles to some extent, although large mistakes are made, with substantial underweighting of the importance of risk factors that are observable to insurers. We also show there is considerable uncertainty due to limited potential to predict nursing home admission, even if the shared information were used optimally.

Previous studies infer the existence of private information on LTC risks from observing that subjective probabilities predict nursing home admission even when conditioning on risk factors observed by insurers (Finkelstein & McGarry, 2006; Hendren, 2013; Lambregts & Schut,

2024). There is also evidence that suggests adverse selection on this private information may be partially offset by advantageous selection (de Meza & Webb, 2001) of low risks on risk preferences (Finkelstein & McGarry, 2006) and numeracy (Lambregts & Schut, 2024), and constrained by rejection of the insurance applications of high risks (Braun et al., 2019; Hendren, 2013).[1] Using HRS data on LTC and other risk perceptions, Solomon (2026) shows that high risks tend to underestimate their risk (compared with risk-factor-predicted risks) and low risks tend to overestimate theirs, which, in theory, may exacerbate welfare loss from adverse selection – because insurers find it more difficult to separate high risks from low risks – but could possibly make a pooling equilibrium feasible. Motivated by concerns about adverse selection, all these studies focus on the detection of private information. We show that this affects the accuracy of LTC risk perceptions less than the underutilization of shared information, which suggests that the latter may be the more important source of selection in the LTC insurance market.

Consistent with observational evidence that LTC insurance take-up is associated with LTC risk perceptions (Brown et al., 2012; Finkelstein & McGarry, 2006; Zhou-Richter et al., 2010), Boyer et al. (2020) find, in an experiment, that stated take-up of LTC insurance increases with perceived risk but predict that eliminating risk misperceptions would raise take-up only marginally because the mean perception error is close to zero. This assumes that under- and overestimation of the risk have equal but opposite impacts on insurance demand. The authors measure perception error as the deviation of perceived risk from risk-factor-predicted risk, which does not allow for private information nor its separation from the underutilization of

[1] Hendren (2013) shows that the additional power of subjective probabilities to predict nursing home admission comes from high risks whose LTC insurance applications would be rejected. Similarly, Braun et al. (2019) show that high risks hold more private information and are more likely to be denied insurance.

shared information in determining risk perception accuracy. We overcome these limitations by using data on the realized event outcome – moving to a nursing home or not.

In the context of research examining health-related insurance allowing for information frictions and cognitive limitations/biases (Abaluck & Gruber, 2011, 2016; Baicker et al., 2015; Bhargava et al., 2017; Handel, 2013; Handel & Kolstad, 2015; Handel et al., 2019; Handel & Schwartzstein, 2018; Ho et al., 2017; Ketcham et al., 2015; Solomon, 2026), we make three main contributions to evidence on LTC risk perceptions, which we show are positively associated with LTC insurance. First, even though subjective probabilities of moving to a nursing home predict that outcome, they are inaccurate, and this is mostly because they are noisy and the outcome is difficult to predict. Second, inaccuracy also stems from underutilization of information that is shared with insurers on application and private information only partially offsets this source of inaccuracy. Third, the least cognitively able hold the least accurate LTC risk perceptions because their subjective probabilities are noisier, contain less private information and make less use of shared information.

Section 2 describes the data. Section 3 explains how we measure risk perception accuracy and decompose that measure before presenting regression methods to estimate partial associations between LTC insurance and risk perceptions. Results are presented in section 4 and the final section discusses their implications and limitations.

2. Data

We use data from the US Health and Retirement Study (HRS), a biennial longitudinal survey of older (50+) Americans (Health and Retirement Study, 2021).[2]

[2] The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

Risk perceptions. We measure risk perceptions with reported subjective probabilities of moving to a nursing home. Respondents who are not living in a nursing home, are at least 65 years old, and who answer three prior expectations questions are asked: "What is the percent chance that you will move to a nursing home in the next five years?".[3] Answers can take any value from 0 ("Absolutely no chance") to 100 ("Absolutely certain"). We rescale them to the 0-1 range. Nonresponse is 3.7%.[4] We use these data from wave 11 (2012) of the HRS because this is the most recent sample for which we can determine whether each respondent moved to a nursing home within a period of five years that does not include the COVID-19 pandemic. This sample includes individuals born in the period 1924-1947.

Outcome. Respondents are asked whether they currently reside in a nursing home, whether they had an overnight stay in a nursing home since the previous wave and, if so, the number of nights of each stay. When a respondent is asked to report the probability of moving to a nursing home, it is likely that they do not consider the possibility of a short stay for rehabilitation after medical treatment. Hence, for consistency, we define the outcome as a nursing home stay of at least 21 consecutive nights. The threshold is chosen because Medicare fully reimburses rehabilitative stays of up to 20 nights in skilled nursing facilities (Medicare, n.d.).[5] We assess robustness to alternative thresholds. For deceased HRS respondents, we include nursing home stays of a) any duration that end with death, and b) $\geq$ 21 nights before death while not in a nursing home.

[3] Respondents are told: "Nursing homes are institutions primarily for people who need constant nursing supervision or are incapable of living independently. Nursing supervision must be provided on a continuous basis for the institution to qualify as a nursing home. Please don't include stays in adult foster care facilities or other short-term stays in a hospital". Prior to this question, there are three questions on expectations about home values and inheritance. Those who give a "don't know" response or refuse to answer these questions are not asked the nursing home question.

[4] This is nonresponse conditional on being asked the question. 6,297 respondents aged 65+ for whom we can establish whether they moved to a nursing home within five years are asked three filter questions on expectations. Among these, 1.3% do not respond to these questions and so are not asked to report their probability of moving to a nursing home within five years.

[5] Medicare partially reimburses rehabilitative nursing home stays of 21-100 nights. It does not cover custodial care.

Family members of the deceased provide the required information. Nursing home stays reported in waves 12 and 13 are within the 5-year period from wave 11 referred to in the subjective probability question. For stays reported in wave 14, we use the date of nursing home entry reported in that wave along with the wave 11 interview date to determine whether the entry is within the 5-year period.

LTC insurance. We use an indicator of reporting having any private LTC insurance that would, in most cases, cover nursing home care.[6]

Risk factors. We regress the outcome, the subjective probability and LTC insurance take-up on LTC risk factors that can be observed by an insurer on processing an application. Following Finkelstein & McGarry (2006), we include indicators of age and sex, limitations in activities of daily living (ADLs) and instrumental activities of daily living (IADLs), body mass index (BMI), cognitive impairment, depression, incontinence, use of prescription medicines, use of mobility and breathing aids, previous LTC use, alcohol use and smoking, diagnosed and medicated diseases/conditions, marital status and spouse's age, and income and wealth (see Appendix A, Table A1). We use several domains of cognitive functioning obtained through validated tests (Ofstedal et al., 2005; Fisher et al., 2017) as combined in the HRS total cognition score (0-35) that aggregates measures of episodic memory and intact mental status and is increasing in cognitive functioning (Table A1 for details). We use an indicator of cognitive impairment corresponding to a score less than or equal to 8 (Mehta et al., 2003).[7]

[6] We use any LTC insurance as the baseline specification of our outcome, since individuals with high perceived nursing home risk but strong aversion to institutional care may purchase in-home-only policies. A follow-up question in the HRS asks respondents what their LTC insurance covers. Only 2.7% of LTC insurance holders report coverage for in-home care only. We nevertheless test robustness to using an alternative indicator of LTC insurance that covers nursing home care (see Appendix Table A9).

[7] It is not uncommon for insurers of long-term care services to administer cognition tests to potential insurees (Dupont, 2024).

Stratifiers. We examine heterogeneity in the accuracy of risk perceptions by wealth, education and cognition. We use quartile groups of total net household wealth, excluding primary housing, social security and employer-sponsored pension wealth because these are excluded in the Medicaid assets test to determine eligibility for LTC services in most states (American Council on Aging, 2021). We distinguish between four levels of education: high-school dropout/General Educational Development (GED), high-school graduate, some college and college graduate. We use quartile groups of cognitive functioning measured by the total cognition score.

Sample. Our main sample includes respondents aged 65-88 in 2012 who a) in wave 11, report their subjective probability of moving to a nursing home within five years, b) can be traced through full, proxy, or exit interviews in subsequent waves to establish if they did move to a nursing home within five years, and c) have full item response for all the risk factors used to predict the outcome.[8] The sample is smaller for the analysis of LTC insurance take-up due to nonresponse on additional covariates used to control for preferences. For the analysis of the association between insurance and the one-wave-lagged subjective probability of moving to a nursing home, we used a sample aged 40-64.

Descriptives. Figure 1 shows the distribution of subjective probabilities of moving to a nursing home. Around 40% report a zero probability. About 12% report a fifty-fifty chance, which could be an expression of not knowing the probability (epistemic uncertainty) rather than a belief that it is precisely 0.5 (Fischhoff & Bruine de Bruin, 1999; Bruine de Bruin & Carman, 2012). We check robustness to dropping respondents who report a 0.5 probability. The mean subjective probability (0.165) overestimates the sample base rate (0.117) – the proportion of

[8] Appendix A, Table A2 gives the number of respondents dropped at each stage to reach the analysis sample. Analysis of heterogeneity by education excludes one respondent missing on education.

the analysis sample that move to a nursing home within five years – by almost 5 percentage points (pp).[9]

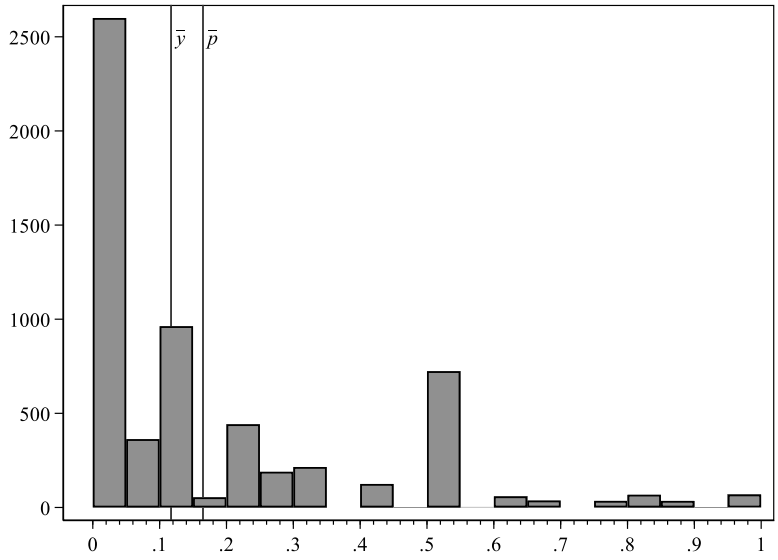

Figure 1. Distribution of subjective probabilities of moving to a nursing home within 5 years

Notes: Bin size is 0.05. y-axis shows frequencies. Vertical lines show the proportion who move to a nursing home within 5 years ($\bar{y}$ = 0.117) and the mean reported subjective probability of moving to a nursing home within 5 years ($\bar{p}$ = 0.165). n = 5,987.

3. Methods

3.1 Risk perception inaccuracy

We measure the average inaccuracy of the reported risk perceptions with their sample mean squared prediction error:

$$MSPE = \frac{1}{n}\sum(p_i - y_i)^2 \in [0,1], \quad (1)$$

where p_i is individual i's reported subjective probability of moving to a nursing home within five years, y_i = 1 if that event occurs and the nursing home stay lasts at least 21 consecutive nights or ends in death, y_i = 0 otherwise, and n is the sample size.

[9] The base rate increases to 0.122 when including those who do not answer the subjective probability question, i.e., they are more likely to enter a nursing home.

The MSPE increases with the variance of the outcome: $Var(y) = \bar{y}(1-\bar{y})$, where $\bar{y} = {}^{1}/_{n}\sum y_i$. Greater variance makes prediction more difficult. Inaccuracy also increases with (squared) bias of the subjective probabilities: $bias = \bar{p} - \bar{y}$, where $\bar{p} = {}^{1}/_{n}\sum p_i$. On the other hand, inaccuracy decreases with the power of the subjective probabilities to distinguish between individuals who move and do not move to a nursing home. This can be measured by the difference in the outcome-conditional means of the subjective probabilities, the *discrimination slope*: $\Delta p = \bar{p}_1 - \bar{p}_0$, where $\bar{p}_k = {}^{1}/_{n_k}\sum 1(y_i = k)p_i$, $n_k = \sum 1(y_i = k)$, $k \in \{0,1\}$. For binary outcomes, this measure partly determines the outcome-prediction covariance: $Cov(p,y) = \Delta p Var(y)$ (Yates, 1982). Finally, inaccuracy increases with the variance of the subjective probabilities, $Var(p)$. Part of this variance is not explained by the outcome: $noise = Var(p) - \Delta p^2 Var(y)$. This can result from predictions that are influenced by factors irrelevant to the risk of moving to a nursing home. It can also be due to measurement error deriving from inability to report probabilities that reflect true beliefs or limited understanding of the probability question. The remainder of the variance of the subjective probabilities is the part explained by the outcome: $signal = \Delta p^2 Var(y)$.

We decompose the MSPE as follows (Yates, 1982):

$$MSPE = Var(y) + bias^2 - 2\Delta p Var(y) + signal + noise \qquad (2)$$

This quantifies the contributions of outcome prediction difficulty (variance) and the subjective probabilities' bias and noise to (reported) risk perception inaccuracy and the offsetting contribution of the subjective probabilities' discriminatory power.

3.2 Use of shared and private information

To assess the extent to which available information is used to form more accurate risk perceptions, we regress the subjective probabilities and the outcome on risk factors for moving to a nursing home ($\boldsymbol{X}$) that insurance applicants would be required to share with insurers:

$$p_i = \sum_{j=1}^{J} \beta_j^p X_{ji} + \varepsilon_i \tag{3}$$

$$y_i = \sum_{j=1}^{J} \beta_j^y X_{ji} + \upsilon_i\,, \tag{4}$$

where β_j^p is the partial association of the subjective probabilities of moving to a nursing home with the j^{th} risk factor, and so the implicit (average) weight individuals give to it when forming their risk perceptions, β_j^y is the partial association of the outcome with the respective risk factor, and ε_i and υ_i are random errors. Regressions (3) and (4) are estimated by ordinary least squares (OLS) and so their estimated coefficients give the weights that best (linearly) predict the subjective probabilities and the outcome, respectively, from the jointly observed risk factors.[10]

We use the regressions to decompose the discrimination slope of the subjective probabilities, Δp, into three parts (Bago d'Uva & O'Donnell, 2022):

$$\Delta p = \Delta\hat{p} + \Delta\hat{\varepsilon} = \Delta\hat{y} - (\Delta\hat{y} - \Delta\hat{p}) + \Delta\hat{\varepsilon}, \tag{5}$$

where $\Delta z = \bar{z}_1 - \bar{z}_0,\ \bar{z}_k = {}^{1}/_{n_k} \sum 1(y_i = k) z_i,\ k \in \{0,1\},\ z_i \in \{p_i, \hat{p}_i, \hat{\varepsilon}_i, \hat{y}_i\}$, with $\hat{p}_i$and $\hat{y}_i$ fitted values from estimation of (3) and (4), respectively, and $\hat{\varepsilon}_i$ residuals from estimation of (3).

The term $\Delta\hat{y}$ measures the extent to which moving to a nursing home can be (linearly) predicted from the jointly observed risk factors. It shows the potential to discriminate between event

[10] We do not exploit the panel nature of the HRS since it would not be possible to separate the use of shared and private information in a fixed effects specification.

outcomes using shared information. The term $\Delta\hat{p}$ measures the discrimination achieved through subjective processing of that information. It shows the extent to which the subjective weights placed on the jointly observed risk factors give predictions that discriminate those who move to a nursing home from those who do not. The term $\Delta\hat{y} - \Delta\hat{p}$ measures the deviation of the subjective weights from the objective weights. It captures the loss of discriminatory power of subjective probabilities due to suboptimal use of shared information. In linear models, this term can be further decomposed to reveal information extraction from each risk factor or from a set of risk factors.[11] Finally, $\Delta\hat{\varepsilon}$ is the accuracy-increasing discriminatory power that derives from private information (not observed by insurers) used in forming the subjective probabilities that is associated with the outcome but not associated with the jointly observed risk factors. Use of this private information can partly offset underuse of shared information.

In our main analysis, we estimate (3) using the wave 11 (2012) sample and the subjective probabilities and risk factors reported, or measured, in the same wave, and we estimate (4) using risk factors observed in wave 8 (2006) for a comparable sample and nursing home stays over the five years subsequent to that wave.[12] This is motivated by the fact that wave 11 respondents could not have been aware of how the risk factors measured in that wave would eventually relate to future nursing home admission. We presume that the best source of information for their subjective weights is the observation of characteristics of people who moved, and did not move, to a nursing home over the previous five years. The estimated relationships between the risk factors of the wave 8 sample and movements of this sample into nursing homes over the subsequent five years then constitute the shared information that could possibly have been

[11] $\Delta\hat{y} - \Delta\hat{p} = \sum_{j=1}^{J}(\hat{\beta}_j^y - \hat{\beta}_j^p)\Delta X_j$, where $\Delta X_j = \bar{X}_{j1} - \bar{X}_{j0}$, $\bar{X}_{jk} = {}^{1}/_{n_k}\sum 1(y_i = k)X_{ji}$, $k \in \{0,1\}$. Any interactions must be treated as a set of observed risk factors (Bago d'Uva & O'Donnell, 2022). Our main analysis does not include interactions but we test robustness to introducing them.

[12] The wave 8 and wave 11 samples are constructed in the same way. Each includes respondents who are 65 and over, who answered the subjective probability question about moving to a nursing home within five years, as well as all the questions used to construct the risk factors, and for whom we can observe the outcome – that is whether they move to a nursing home within five years. See Table A1 for means of the risk factors for both samples.

known by wave 11 respondents when forming their subjective probabilities, as well as by insurers when pricing contracts offered to them. We are then comparing the risk factor weights used to form subjective probabilities with the objective weights that could have been known at the time. We check robustness to estimating (4) with the wave 11 risk factors and nursing home admissions over the five years after that wave.

We also check robustness to using random forest regression and (fractional) logit models, rather than OLS regressions (3) and (4), to predict the subjective probabilities and the outcome from the risk factors.[13] We calculate bootstrap standard errors for the MSPE and each of its components in the decompositions (2) and (5). We use 100 replications to directly bootstrap the standard errors and, for the main estimates, confirm that 1000 replications yield practically the same standard errors.

3.3 Risk perceptions and insurance

To assess whether the demand for LTC insurance is associated with risk perceptions, we use wave 11 data to estimate

$$LTCI_i = \alpha + \gamma p_i + \boldsymbol{\psi X}_i + \boldsymbol{\xi W}_i + u_i, \tag{6}$$

where $LTCI_i = 1$ if the individual reports having any private LTC insurance and $\boldsymbol{X}$ is the set of nursing home risk factors, also used in (3) and (4), that are observable to insurers and may be used to price contracts and to reject high-risk applicants (Hendren, 2013). The vector $\boldsymbol{W}$ contains additional controls to allow for preference shifters and nonlinearities through, respectively, covariates and two-way and three-way interactions between sex and age groups and each of three risk factors: the number of ADLs, the number of IADLs and the total cognition score (Finkelstein & McGarry, 2006). Inclusion of the total cognition score – in addition to the

[13] See Appendix B for details of the random forest regression.

cognitive impairment indicator included in $\boldsymbol{X}$ – gives better control for any direct effect of cognition on the insurance decision in addition to an indirect effect through price. Preference shifters include education level, number of children, and seatbelt use and gender-specific preventive health activities to proxy risk preferences (ibid.) (see Appendix Table A3).

Even with an extensive set of controls, we do not interpret an OLS estimate of γ in (6) as an estimate of the causal effect of risk perception on insurance take-up since the OLS coefficient potentially reflects correlated unobservables, measurement error in subjective probabilities and reverse causality. We conduct three analyses to assess whether these potential sources of endogeneity do affect the partial association captured by the OLS coefficient. First, assuming that selection on unobservables is not stronger than that on the larger array of observable covariates, we calculate Oster (2019) bounds on a (omitted variable) bias-adjusted estimate. Second, we estimate γ in (6) with heteroskedasticity-based instruments for the subjective probability, p_i (Lewbel, 2012). We assume that the identifying source of heteroskedasticity comes from the total cognition score and its interaction with sex, which provides reasonably strong instruments.[14] Given the scope for reasonable doubt about the validity (and transparency) of identification from higher-order moments, we do not claim that this instrumental variable (IV) approach definitively estimates the causal effect of risk perceptions on LTC insurance. Third, to address the concern that a positive OLS estimate of γ in eq. (6) may reflect LTC insurance raising the perceived and reported probability of moving to a nursing home, we

[14] We used heteroskedasticity-based instruments because there are no plausibly exogenous and strong external instruments. The gender composition of children proved to be a weak instrument, and it may violate the exclusion restriction through, for example, bequest motives that affect insurance take-up. We estimate $p_i = \eta + \boldsymbol{\varphi X}_i + \boldsymbol{\zeta W}_i + \nu_i$, and then construct the heteroskedasticity-based instruments, $(\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})\hat{\nu}_i$, where $\hat{\nu}_i = p_i - \hat{p}_i$, $\boldsymbol{Z}_i \subset \{\boldsymbol{X}_i \cup \boldsymbol{K}_i\}$ includes the total cognition score and its interaction with sex, and $\overline{\boldsymbol{Z}} = {}^{1}/_{n} \sum \boldsymbol{Z}_i$. These variables induce heteroskedasticity, $cov(\boldsymbol{Z_i}, {\nu_i}^2) \neq 0$, that strongly predicts p_i, satisfying the first criterion for heteroskedasticity-based instruments (Lewbel, 2012). Such heteroskedasticity likely reflects both variation in beliefs about moving to a nursing home and variation in the error with which those beliefs are reported. Conditional on an extensive battery of controls, we suggest that it is plausible to assume that cognition and its interaction with sex are not correlated with the product of the unobservables that determine the subjective probability and LTC insurance, i.e., $cov(\boldsymbol{Z_i}, u_i \nu_i) = 0$, which is the second (untestable) assumption required for instrument validity (ibid.).

regress LTC insurance on the subjective probability of moving to a nursing home reported in the *previous* wave. Given the persistence of LTC insurance status above the age of 65, we estimate this regression with a sample aged 40-64, in which there is greater between-wave variation in insurance.[15] For respondents aged 40–64, we use the subjective probability of *ever* moving to a nursing home because the 5-year probability is not elicited below age 65. This lifetime probability is asked only in the first wave in which a respondent participates, and only if they are younger than 65. Hence, for this analysis, we do not use subjective probabilities reported in wave 11 but in the first wave of each respondent. In one specification, we control for LTC cover in the previous wave. In another, we restrict the sample to those who do not have LTC insurance when they report their subjective probability of ever moving to a nursing home. In each case, we estimate the association between that reported risk perception and the acquisition of LTC insurance.

We use these three analyses to assess robustness of the sign and significance of the OLS estimate of γ in (6) to potential sources of endogeneity. Our aim is to gauge whether the data are consistent, at least, with risk perceptions influencing the decision to purchase LTC insurance; we do not claim to obtain an unbiased estimate of the magnitude of such an effect.

4. Results

4.1 Risk perception inaccuracy

We obtain a MSPE of the subjective probabilities of moving to a nursing home equal to 0.14. This value is significantly (p-value < 0.01) below a benchmark of 0.25 that would be obtained if everyone were to report a 50-50 chance. It is the value that would be obtained if, for example,

[15] Median (mean) age at first reported LTC insurance cover is 63 (64). These estimates are likely upward biased because some report having LTC insurance in all waves and could have purchased it before we first observe them. Of those aged 65+, only 5% report switching insurance status in the next two years, versus 9% for those aged 40-64.

all those who moved to a nursing home were to report a probability of 0.63 and all of those who did not were to report a probability of 0.37.[16] Another common benchmark is the outcome variance, which is the MSPE that would arise if all were to report the sample base rate, such that the subjective probabilities would display no bias nor noise but neither would they discriminate. Here, the MSPE is significantly larger than the outcome variance ($Var(y) = 0.10$), indicating less accuracy than would be achieved with universal reporting of the base rate and implying that the discriminatory power of the subjective probabilities is not sufficient to offset inaccuracy arising from bias and noise.

Panel A of Table 1 shows the results of decomposing the MSPE measure of risk perception inaccuracy using eq. (2). The variance in nursing home admission is the largest contributor, accounting for 74% of the inaccuracy in subjective predictions of this outcome, followed by noise in the subjective probabilities, which accounts for 36%. The latter could imply that a great deal of attention is paid to irrelevant factors when forming beliefs about the likelihood of moving to a nursing home and/or there is difficulty in expressing those beliefs in reported probabilities. There is very little contribution from the square of the bias – the difference of almost 5 percentage points (pp) between the mean subjective probability and the sample base rate. The covariance of the subjective probabilities with the outcome reduces inaccuracy by only about 11.5% of what it would have been if the subjective probabilities had no discriminatory power.[17]

[16] To be precise, an absolute prediction error of 0.3728 (= |0.6272 – 1| = 0.3728 – 0) for all respondents would give the estimated MSPE = 0.3728^2 = 0.139.
[17] 0.018/(0.139+0.018)×100 = 11.465.

Table 1. Decomposition of risk perception inaccuracy and discrimination

		Estimate	(SE)
A. MSPE	$\mathbf{\frac{1}{n}\sum(p_i - y_i)^2}$	**0.139**	**(0.004)**
Decomposition, eq.(2)			
outcome variance	$Var(y)$	0.103	(0.003)
bias^2	$(\bar{p} - \bar{y})^2$	0.002	(0.000)
Covariance	$-2(\Delta p)Var(y)$	-0.018	(0.002)
Signal	$(\Delta p)^2 Var(y)$	0.001	(0.000)
Noise	$Var(p) - (\Delta p)^2 Var(y)$	0.050	(0.001)
B. Discrimination slope	$\mathbf{\Delta p = \bar{p}_1 - \bar{p}_0}$	**0.086**	**(0.011)**
Decomposition, eq.(5)			
outcome predictability	$\Delta\hat{y}$	0.147	(0.009)
inappropriate weighting	$-(\Delta\hat{y} - \Delta\hat{p})$	-0.093	(0.009)
	$[100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}]$	$[63.3\%]$	
private information	$\Delta\hat{\varepsilon}$	0.032	(0.009)
Mean y	$\bar{y}$	0.117	
Mean p	$\bar{p}$	0.165	
Sample size	n	5,987	

Notes: Panel A gives eq. (2) decomposition of MSPE of subjective probabilities of moving to nursing home within 5 years. Panel B gives eq. (5) decomposition of the discrimination slope of the subjective probabilities. For any variable or prediction z, its discrimination slope is $\Delta z = \bar{z}_1 - \bar{z}_0$, $\bar{z}_k = {}^1/_{n_k} \sum 1(y_i = k) z_i,\ k \in \{0,1\}$. See equations and text for other notation. Bootstrap standard errors (100 replications) in parentheses. See Table A4 for OLS estimates of regressions (3) and (4) used in $\Delta\hat{y}$ and $\Delta\hat{p}$. Sample includes HRS wave 11 respondents aged 65-88 in 2012 with full item response on subjective probabilities and risk factors, and for whom it is possible to determine if they moved to a nursing home within 5 years.

Panel B shows the eq. (5) decomposition of the discrimination slope, Δp – the 8.6 pp difference between the mean subjective probabilities of those who do and do not move to a nursing home. About 37% (3.2/8.6 × 100) of this discriminatory power comes from private information, with the remainder gleaned from shared information contained in jointly observed risk factors. If this information were used to predict objective probabilities of moving to a nursing home using OLS weights from the outcome regression, there would be a 14.7 pp difference between the mean predictions for those who do and do not move to a nursing home ($\Delta\hat{y} = 0.147$). Around 63% of this potential discriminatory power of shared information remains unused due to inappropriate weighting.

Table 2 shows contributions of sets of jointly observed risk factors to decompositions of a) the potential discrimination slope if those risk factors were weighted optimally using OLS estimates from the outcome regression, $\Delta\hat{y}$, b) the discrimination slope achieved with subjective weights estimated from the subjective probabilities regression, $\Delta\hat{p}$, and c) the shortfall of b) from a) due to inappropriate weighting of the risk factors, $\Delta\hat{y} - \Delta\hat{p}$. Applying the optimal weights to differences in age and sex between those who move to a nursing home and those who do not gives a between-group difference of 6.7 pp in the probability of moving to a nursing home. Applying the weights implicit in the subjective probabilities to the same differences in age and sex, we would predict that those who move to a nursing home would have only a 2.5 pp higher probability of doing so. The shortfall of 4.2 pp indicates a lack of appreciation of the extent to which nursing home risk is associated with age and sex. This makes the largest contribution of any set of risk factors to the shortfall of the achieved from the potential discrimination slope. Almost all this shortfall comes from underestimation of the extent to which the risk increases with age, particularly above the age of 85 (Table A4).

There is also underweighting of the risks associated with diagnosed and medicated conditions, mobility and breathing aids, prior LTC use, and ADLs/IADLSs. Those who are cognitively impaired do not even adjust their corresponding risk perceptions in the correct direction. Conditional on the other risk factors, they report lower subjective probabilities of moving to a nursing home despite cognitive impairment being associated with a higher likelihood of that event.

Table 2. Contributions of risk factors to potential and achieved discrimination slopes

	Potential $\Delta\hat{y}$	Achieved $\Delta\hat{p}$	Shortfall $\Delta\hat{y} - \Delta\hat{p}$
Total	**0.147 (0.009)**	**0.054 (0.005)**	**0.093 (0.009)**
Contributions			
Age & sex	0.067 (0.006)	0.025 (0.004)	0.042 (0.007)
ADLs & IADLs	0.016 (0.005)	0.007 (0.003)	0.008 (0.006)
Miscellaneous health	0.002 (0.002)	0.005 (0.002)	-0.003 (0.003)
Mobility & breathing aids	0.018 (0.007)	0.007 (0.003)	0.011 (0.007)
Alcohol & smoking	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)
Diagnosed & medicated conditions	0.019 (0.003)	0.003 (0.003)	0.016 (0.004)
Prior LTC use	0.014 (0.004)	0.004 (0.002)	0.010 (0.004)
Cognitively impaired	0.005 (0.002)	-0.002 (0.001)	0.007 (0.003)
Sociodemographics	0.006 (0.002)	0.003 (0.002)	0.002 (0.003)
n	5,987	5,987	5,987

Notes: For any variable or prediction z, its discrimination slope is $\Delta z = \bar{z}_1 - \bar{z}_0$, $\bar{z}_k = {}^1/_{n_k} \sum 1(y_i = k) z_i,\ k \in \{0,1\}$. The top row gives two of the three components of the eq.(5) decomposition of the discrimination slope of the subjective probabilities using OLS estimates of eqns. (3) and (4). The middle cell of this row gives the difference between these two components – the discrimination slope of the fitted subjective probabilities. Other rows give the contributions of sets of risk factors to the measures in the top row. The left-hand column gives, in each row for the set of risk factors Ω, $\sum_{j\in\Omega} \hat{\beta}_j^y \Delta X_j$. The middle column gives $\sum_{j\in\Omega} \hat{\beta}_j^p \Delta X_j$. The right-hand column gives $\sum_{j\in\Omega} (\hat{\beta}_j^y - \hat{\beta}_j^p) \Delta X_j$. Bootstrap standard errors (100 replications) in parentheses. See Table A1 for the risk factors included in each set. See Table A4 for the OLS estimates $\hat{\beta}_j^y$ and $\hat{\beta}_j^p$ for all j. Sample includes HRS wave 11 respondents aged 65-88 in 2012 with full item response on subjective probabilities and risk factors, and for whom it is possible to determine if they moved to a nursing home within 5 years.

Robustness. Measured risk perception inaccuracy (MSPE of the subjective probabilities), increases when the outcome is defined as any nursing home stay and it decreases when the minimum length of stay (not ending in death) is set to 100 nights, rather than 21 nights used in the main analysis (Appendix B, Table B1). These changes are almost entirely attributable to a shorter minimum length of stay driving the mean outcome towards 0.5 and so increasing the variance, which makes prediction more difficult. Apart from these changes in the outcome variance, the main findings from the MSPE decomposition remain. That is, the subjective probabilities are noisy but also have discriminatory power that comes less from private information and more from use of shared information on jointly observed risk factors, despite at least half of the potential discriminatory power of this shared information remaining unused due to incorrect weighting of risk factors.

Excluding respondents who give a focal response of 0.5 to the subjective probability question reduces the MSPE (Table B1), indicating that these responses are less accurate.[18] With this exclusion, the fraction of the discriminatory power of the subjective probabilities that comes from use of shared information ($\Delta\hat{p}/\Delta p$) falls from 63% to 51% and the underutilization of this information due to inappropriate weighting of risk factors, $(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$, increases from 63% to 72%. These changes suggest that not all those giving a 0.5 response are expressing epistemic uncertainty. The main patterns observed in the decomposition for the full sample are, nevertheless, robust to exclusion of focal responses (Table B1).

Decomposition of the discrimination slope of the subjective probabilities into outcome predictability, inappropriate weighting of risk factors and private information is robust to using the same sample to estimate both (3) and (4), alternative specifications of the covariates included in those regressions and estimation with random forest regression and with (fractional) logit rather than OLS (Appendix B, Table B2).

4.2 Heterogeneity

Table 3 shows evidence of heterogeneity in the inaccuracy of risk perceptions by wealth, education and cognitive functioning. It is obtained by regressing the squared prediction error of each respondent's subjective probability of moving to a nursing home, $(p_i - y_i)^2$, on those characteristics – separately and jointly – plus controls for age, sex and marital status.[19] Coefficients correspond to deviations in the (conditional) MSPE from the reference category of each characteristic. Columns (1)-(3) show that higher risk perception inaccuracy is associated with lower wealth, education and cognitive functioning. That wealthier individuals perceive the risk of moving to a nursing home more accurately is somewhat reassuring for this sub-

[18] This is because a reported probability of 0.5 gives a squared prediction error of 0.25 for a binary (0-1) outcome, which is larger than the MSPE of the sample that does not give this focal response.
[19] See Appendix Table A5 for estimation results without controls.

population since Medicaid protection against LTC risk is withdrawn as wealth rises, and so the risk perceptions of wealthier individuals are potentially more consequential for private LTC insurance demand in this part of the population.[20]

Table 3. Heterogeneity in risk perception inaccuracy (MSPE)

	(1)		(2)		(3)		(4)	
Wealth (ref. Richest quartile)								
Poorest quartile	0.023	(0.010)					-0.003	(0.011)
2nd Poorest quartile	0.015	(0.009)					0.001	(0.010)
2nd Richest quartile	0.002	(0.009)					-0.004	(0.009)
Education (ref. College graduate)								
High school dropout or GED			0.032	(0.010)			-0.007	(0.012)
High school graduate			0.017	(0.008)			-0.003	(0.009)
Some college			0.022	(0.009)			0.009	(0.009)
Cognitive functioning (ref. Top quartile)								
Bottom quartile					0.083	(0.009)	0.087	(0.011)
2nd Bottom quartile					0.056	(0.009)	0.057	(0.009)
2nd Top quartile					0.026	(0.007)	0.026	(0.007)
n	5,987		5,986		5,987		5,986	

Notes: Columns (1)-(3) show estimates from separate OLS regressions of the squared prediction error of the subjective probability of moving to a nursing home within 5 years ($(p_i - y_i)^2$) on indicators of each of household wealth quartile group, educational attainment, total cognition score quartile group, respectively, plus controls for sex, 5-year age groups (up to ≥ 85 years), and marital status (married/partnered). Column (4) shows estimates from a regression in which wealth, education, and cognitive functioning are all included. Robust standard errors in parentheses. The MSPE of the reference groups are 0.122, 0.113, and 0.076, for wealth, education, and cognitive functioning, respectively. Sample sizes differ due to one item non-response for education.

Regressing the squared prediction errors on wealth, education and cognitive functioning simultaneously (column 4), reveals that the MSPE differences by wealth and education are fully explained by the lower cognitive functioning of the less wealthy and lower education groups. There remains a clear gradient in the accuracy of risk perceptions by cognition: a MSPE difference of 8.7 points between the bottom and top quartile groups is substantial compared with an overall MSPE of 14 points.

[20] Brown & Finkelstein (2011) estimate that Medicaid, which covers 60% of LTC costs overall, crowds out private LTC insurance for a majority of the wealth distribution, e.g. by transferring assets to children (Bassett, 2007), while Braun et al. (2019) find almost complete crowd-out for the poor.

Table 4 shows the decompositions of the MSPE and the discrimination slope of the subjective probabilities – eq. (2) and eq. (5), respectively – for each quartile group of cognitive functioning. Panel A shows that the greater inaccuracy of the lower cognition groups is because they are exposed to greater outcome variance, which makes their prediction task more difficult, and their subjective probabilities are noisier. The latter may reflect a greater tendency of the less cognitively able to pay attention to irrelevant factors when forming an expectation about moving to a nursing home. It could also be due to low cognitive functioning impeding ability to express beliefs about that expectation in a probability format.

The top row of Panel B in Table 4 shows that the subjective probabilities of the top two cognition quartiles discriminate best between those who move to a nursing home and those who do not. This is despite the lower predictability of the outcome from the jointly observed risk factors in the top cognition group compared with the bottom. This greater predictability of the outcome for the bottom group has the potential to contribute to higher discrimination power (and so accuracy) of their subjective probabilities. However, this potential is not realized because they weight the risk factors less appropriately - the lowest quartile leaves 71% of the potential discriminatory power of the risk factors unused, while the top quartile extracts much more of the shared information from the risk factors and leaves unused only 5% of their discrimination potential.

Higher cognitive functioning is not only associated with better use of shared information but also with having more private information. In the top cognition quartile, there is a difference of 8.2 pp in the mean subjective probability model residuals between those who move to nursing home and those who do not ($\Delta\hat{\varepsilon} = 0.082$). In the second bottom and bottom cognition quartiles, the respective differences are only -0.4 and 3.3 pp, respectively. This indicates that, after controlling for the information extracted from the jointly observed risk factors, the lower

cognition groups have less additional information to call on, they are less able to use what they have, or they have less and use less of it.

Table 4. Decomposition of risk perception inaccuracy and discrimination by cognition

		Quartile group of total cognition score			
		Bottom	2nd Bottom	2nd Top	Top
A. MSPE	$\frac{1}{n}\sum(p_i - y_i)^2$	**0.201**	**0.152**	**0.116**	**0.076**
		(0.008)	**(0.007)**	**(0.006)**	**(0.005)**
Decomposition, eq.(2)					
outcome variance	$Var(y)$	0.156	0.105	0.084	0.049
		(0.006)	(0.006)	(0.007)	(0.005)
bias^2	$(\bar{p} - \bar{y})^2$	< 0.001	0.002	0.006	0.008
		(0.000)	(0.001)	(0.002)	(0.001)
covariance	$-2(\Delta p)Var(y)$	-0.024	-0.008	-0.018	-0.014
		(0.005)	(0.004)	(0.005)	(0.004)
signal	$(\Delta p)^2 Var(y)$	0.001	< 0.001	0.001	0.001
		(0.000)	(0.000)	(0.001)	(0.001)
noise	$Var(p) - (\Delta p)^2 Var(y)$	0.068	0.052	0.044	0.032
		(0.003)	(0.002)	(0.002)	(0.002)
B. Discrimination slope	$\Delta \boldsymbol{p} = \bar{\boldsymbol{p}}_1 - \bar{\boldsymbol{p}}_0$	**0.078**	**0.036**	**0.110**	**0.137**
		(0.017)	**(0.020)**	**(0.027)**	**(0.036)**
Decomposition, eq.(5)					
outcome predictability	$\Delta\hat{y}$	0.155	0.120	0.127	0.058
		(0.019)	(0.018)	(0.024)	(0.027)
inappropriate weighting	$-(\Delta\hat{y} - \Delta\hat{p})$	-0.111	-0.080	-0.060	-0.003
		(0.019)	(0.019)	(0.024)	(0.028)
	$[100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}]$	[*71.2%*]	[*66.6%*]	[*47.1%*]	[*5.2%*]
private information	$\Delta\hat{\varepsilon}$	0.033	-0.004	0.043	0.082
		(0.013)	(0.017)	(0.022)	(0.025)
Mean y	$\bar{y}$	0.193	0.119	0.092	0.052
Mean p	$\bar{p}$	0.178	0.168	0.171	0.139
Sample size	n	1,671	1,345	1,591	1,380

Notes: Contents of table, samples, and methods are as Table 1 except here the sample is stratified by quartile of the total cognition score (0-35). Scores for bottom, 2nd bottom, 2nd top, and top groups are ≤ 19, 20-22, 23-25, and >25, respectively. The group sizes are unequal due to the discrete distribution of the score and its density which is concentrated. Regressions (3) and (4) are estimated separately for each quartile group. Because we stratify by cognition, we do not include the indicator of cognitive impairment in the regressions. See Table A6 for the inappropriate weighting of sets of risk factors by these groups.

4.3 Risk perceptions and insurance

Table 5, columns (1)-(3) give OLS estimates of the (partial) association of private LTC insurance with the subjective probability of moving to a nursing home within five years. The unconditional estimate in column (1) indicates that an increase in the subjective probability from 0 to 1 is associated with a 10.9 pp increase in the likelihood of being insured. This is a 71% increase on the proportion with LTC insurance (0.154). The association hardly changes when controlling for risk factors potentially observed by applicants and insurers ($\boldsymbol{X}$) (column (2)). The continued positive and significant association is consistent with selection into insurance partly based on private information used to form risk perceptions. In column (3), the partial association remains stable in magnitude and statistical significance after also controlling for preference shifters and interactions to allow for nonlinearities, i.e. $\boldsymbol{W}$ in eq. (6). This robustness is consistent with the partial association not being fully attributable to confounders.

Appendix Table A7 reports heterogeneity in the association shown in column (3) of Table 5 by cognition, wealth and education. Across all cognition groups, there is a significant correlation between LTC insurance and the subjective probability of moving to a nursing home within five years. The strength of the association increases with cognition, but the estimate remains significant and meaningful even for the lowest cognition quartile. The association is small and insignificant for the poorest wealth quartile, which may reflect crowd-out of private insurance by Medicaid. The second poorest wealth quartile shows a moderate association, while the two richest quartiles display stronger and more significant correlations. The pattern by education is less clear: the two lowest education groups have weaker correlations than the two highest groups, but there is no monotonic gradient.

Table 5. Association of LTC insurance with risk perceptions

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	Bias-adjusted	IV
Subjective probability of moving to	0.109	0.108	0.101	0.098	0.089
nursing home within 5 years, *p*	(0.022)	(0.021)	(0.021)		(0.062)
Controls					
Risk factors, ***X***	No	Yes	Yes	Yes	Yes
Others, ***W***	No	No	Yes	Yes	Yes
Effective *F*-statistic					68.31
[critical value]					[9.99]
Hansen J test, p-value					0.880
R^2	0.005	0.079	0.092		
Mean outcome	0.154	0.154	0.154	0.154	0.154
n	5,705	5,705	5,705	5,705	5,705

Notes: In all columns, the dependent variable is an indicator of having private LTC insurance, all data are from wave 11 (2012) and the sample (aged 65+) has full item response to the full set of controls. Columns (1)-(3) show OLS estimates of the coefficient on the subjective probability of moving to a nursing home within 5 years in a linear probability model of that indicator – γ in eq. (6). Risk factors (***X***) are listed in Table A1. Other controls (***W***) are the variables listed in Table A3, plus interactions of sex and age groups with number of ADLs/IADLs and total cognition score. The bias-adjusted estimate in column (4) is the Oster (2019) bound calculated using a maximum $R^2 = 1.3 \times R^2$ in column (3) and assuming proportional selection of observables and unobservables. Table A7 gives additional bound estimates. Column (5) shows the Lewbel (2012) IV estimate using instruments constructed from heteroskedasticity related to the total cognition score and its interaction with sex. Effective *F*-statistic is the Montiel Olea & Pflueger (2013) weak IV test. The critical value is for a 5% significance level and allowing for bias up to 10% of the worst-case bias. The Hansen J test is for the joint null hypothesis that the instruments are valid. Robust standard errors are shown in parentheses.

We calculate Oster bounds for our estimate in the full sample (Table 5, column (3)). Assuming that selection on unobservables is of the same magnitude as selection on observables and the maximum R^2 from a regression that included the unobservables would be 1.3 times the R^2 achieved with all the observed controls (Oster 2019), we get a bound on a bias-adjusted coefficient (column (4)) that is only marginally smaller than the estimate obtained with controls for all observables. Even if the maximum R^2 with controls for unobservables, as well as observables, were to be three times the achieved R^2, the biased-adjusted coefficient would still be about 78% of the unadjusted coefficient (Appendix Table A8). Selection on unobservables would have to be more than 17 times stronger than selection on observables to eliminate the partial association of insurance with risk perceptions (Table A8).

The heteroskedasticity-based IV estimate of the partial association of LTC insurance with the subjective probability of moving to a nursing home (column (5)) is only slightly smaller than the OLS and bias-corrected estimates but it is not statistically significant.[21] Table A9 shows robustness of Table 5 estimates to replacing any LTC insurance with an indicator of reporting having LTC insurance that covers nursing home care, i.e. removing from the indicator a few who report purchasing in-home-only policies.

Table 6 shows estimates of the association between having LTC insurance and the subjective probability of *ever* moving to a nursing home reported by respondents aged 40-64 (at HRS entry, *t*), controlling for year fixed effects. Column (1) reports the contemporaneous association, while the subsequent columns use LTC insurance reported in the next wave (*t*+1). The estimate of column (1) is substantially smaller than those in Table 5, likely reflecting differences in the samples and subjective probability measures used. Column (2), which uses LTC insurance reported in the subsequent wave, shows a larger association. Given that future LTC insurance cannot influence earlier reported subjective probabilities, this positive and significant association would seem to suggest that reverse causality does not drive the estimate in column (1). However, insurance status can also be persistent even in this younger sample, so in column (3) we control for LTC insurance at the time when the subjective probability was measured (*t*) and in column (4) we restrict the sample to those without LTC insurance at *t*. This reduces the association, but the estimates remain positive and statistically significant, which suggests that, while reverse causality may partly contribute to the estimated positive

[21] An effective *F*-statistic greater than the critical value (Montiel Olea & Pflueger, 2013) indicates that heteroskedasticity from the total cognition score and its interaction with sex provides sufficiently strong instruments and the Hansen J overidentification test suggests that these instruments are valid. Using only the total cognition score to construct a heteroskedasticity-based instrument gives a very similar estimate (Appendix Table A10). The estimate is also robust to adding controls that proxy for numeracy, which could be associated with both cognition and LTC insurance though, e.g., financial literacy (Table A10).

contemporaneous associations between LTC insurance and risk perceptions, it does not account for all of it. [22]

Table 6. Association of current and future LTC insurance with risk perception

	LTC insurance at *t*	LTC insurance at *t*+1		
	(1)	(2)	(3)	(4)
Subjective probability of *ever* moving	0.030	0.041	0.029	0.014
to a nursing home (at *t*, HRS entry)	(0.008)	(0.008)	(0.007)	(0.006)
Controls				
Year fixed effects	Yes	Yes	Yes	Yes
LTC insurance at *t*	No	No	Yes	No
Sample restriction				
No LTC insurance at *t*	No	No	No	Yes
R^2	0.003	0.003	0.153	0.002
Mean outcome	0.076	0.075	0.075	0.045
n	15,181	15,181	15,181	14,033

Notes: In all columns, the dependent variable is an indicator of having private LTC insurance and estimates are from linear probability models of that indicator. Column (1) gives the contemporaneous OLS estimate of the coefficient on the subjective probability of *ever* moving to a nursing home. Columns (2)-(4) give OLS estimates of the association between the subjective probability of *ever* moving to a nursing home (reported at *t*) and LTC insurance in the next wave (*t*+1). The sample is aged 40-64 on entry to the HRS, which is the only wave in which each respondent reports that subjective probability. Accordingly, we use data from that wave and the subsequent wave for each respondent and pool across respondents. All models control for year indicators. Column (2) additionally controls for the indicator of LTC insurance in the wave in which the subjective probability is reported. Column (3) restricts the sample to respondents without LTC insurance in the wave in which the subjective probability is reported. Robust standard errors are shown in parentheses.

Overall, we find a significant positive association between holding private LTC insurance and the subjective probability of moving to a nursing home that persists after taking various steps to account for omitted variables, measurement error and reverse causality, with the exception that significance is lost when using heteroskedasticity-based IV. This is consistent – no more than that – with inaccuracy in LTC risk perceptions being consequential for insurance decisions.

[22] Although we do not control for risk factors and covariates in Table 6, Table 5 estimates show that doing so has little impact on the estimated association.

5. Discussion

Misperception of long-term care (LTC) risk could distort insurance and saving decisions, with important consequences for well-being in old age. We find that, while the average older American overestimates their risk of moving to a nursing home, many who end up in a nursing home perceive the risk as low ex ante. LTC risk perceptions are highly inaccurate. In part, this is due to underutilization of information that could be gleaned from risk factors that LTC insurance applicants are obliged to reveal to insurers. Subjective probabilities capture only 37% of this shared information's potential to discriminate between those who do and those who do not move to a nursing home. While we do not estimate the extent to which insurers use this information, we expect their experience and statistical knowledge to give them an advantage over applicants in prediction from jointly observed risk factors. Individuals who are insufficiently sensitive to shared information may underestimate their risk and so decline insurance offered at a price that is actuarially fair (Baillon et al., 2022). Consistent with this scenario, we find a positive association between LTC insurance take-up and the subjective probability of moving to a nursing home that is robust to extensive controls, adjustment for a plausible magnitude of omitted variables bias, use of heteroskedasticity-based instruments (in which case, statistical significance is lost) and using lags to deal with reverse causality.

Inappropriate weighting of jointly observed risk factors could stem from unawareness of the relevance of this shared information for LTC risk, from ambiguity, uncertainty or imperfect perception about how these factors translate into risk, or from an inability to process this information into a single, reported subjective probability. We find that age is the most underestimated risk factor, particularly for the least cognitively able. Since most older people know their ages and the strong correlation of age with nursing home admission is evident from casual observation and LTC insurance premiums that rise steeply with age (American Association for Long-Term Care Insurance, 2024), underutilization of this shared information

appears to be mainly due to inability to process it. Older individuals may be aware that old age increases LTC risk, but may not believe it applies to them personally or underappreciate the extent to which old age translates into increased LTC risk. We find underappreciation across all age categories, but the upward revision of the subjective probability of moving to a nursing home increasingly fails to keep pace with the rising objective probability as individuals grow older. This suggests that there is underappreciation of the rate at which LTC risk rises in old age which could contribute to low insurance take-up in middle-age.

Our finding that the subjective probability of moving to a nursing home predicts that outcome even after conditioning on a large battery of observable risk factors confirms earlier evidence of private information on LTC risks (Finkelstein & McGarry, 2006; Hendren, 2013). We go beyond detection of private information to quantification of its contribution to the accuracy of risk perceptions. This reveals that use of private information offsets only about one third of the inaccuracy that arises from the underuse of shared information. Insurers can be disadvantaged in the information available to them and yet be better informed because of their advantages in the processing and utilization of that information. It is likely that some insurance applicants can use private information on their personal risks to detect and select contracts that are priced below their expected LTC costs, although Solomon (2026) shows that private information may be misperceived and exacerbate welfare loss. Our estimates suggest that there are many others who, even when in possession of private information, cannot accurately determine whether the price is above or below their true expected cost because they underuse information they are obliged to share with a prospective insurer.

Given imperfections in the US LTC insurance market (Ameriks et al., 2018), regulation to limit the scope for selection arising from asymmetric utilization of shared information need not be welfare improving (Handel, 2013; Solomon, 2026). The experience of removing information frictions in the health insurance market suggests that welfare consequences depend on

microfoundations of a particular market (Handel et al., 2019). While the design of effective information interventions to improve the accuracy of LTC risk perceptions is a challenge worth pursuing, it is unlikely that success would eliminate underinsurance of LTC risks. It would solve only one piece of a complicated puzzle that also involves high administration costs (Braun et al., 2019), low-quality products (Ameriks et al., 2018), financial illiteracy (Brown & Finkelstein, 2009), crowd-out by public insurance (Braun et al., 2019; Brown & Finkelstein, 2011; Lambregts & Schut, 2024) and imperfectly perceived private information (Solomon, 2026).

We find that LTC risk perceptions are much less accurate at lower levels of cognitive functioning. Given the strong correlation between cognition and both wealth and education, cognition-related differences in accuracy may contribute to socioeconomic differences in LTC insurance take-up (Finkelstein & McGarry, 2006; Lambregts & Schut, 2020) and arouse or intensify concerns about inequality in well-being in old age that results from suboptimal insurance and saving decisions (Handel et al., 2024). The less cognitively able face a more difficult prediction task because their higher risk increases the variance of the outcome that is the prediction target. The cognition gradient in accuracy is not merely mechanistic, however. The lower cognition groups hold risk perceptions that are noisier. This is consistent with their limited cognitive functioning posing greater difficulties to report a probability (Handel & Schwartzstein, 2018). Their subjective probabilities also contain less private information and are less effective in discriminating between those who move to a nursing home and those who do not. The lower discriminatory power is mainly due to much lower utilization of shared information. The bottom quartile cognition group makes use of less than 30% of the potential discriminatory power of nursing home risk factors, compared with 95% achieved by the top quartile group. Those with lower cognitive functioning underestimate how the various risk factors jointly influence LTC risk, perhaps because they underappreciate the importance of

certain risk factors or because they are limited in their capacity to process and combine the information on the various risk factors and how these relate to LTC risk.

As with all analyses of reported subjective probabilities, we cannot be sure that the data correspond to true beliefs. Measured inaccuracies could partly reflect reporting error arising from difficulty expressing beliefs in probability formats (Gigerenzer & Hoffrage, 1995) that may manifest through extreme rounding and use of focal responses like 0.5 (Giustinelli et al., 2022a; Kleinjans & van Soest, 2014; Manski & Molinari, 2010) as well as heterogeneous use of the probability scale (Bassett & Lumsdaine, 2001). Modeling of reporting behavior tends to find a modest effect on the distribution of subjective probabilities (Kleinjans & van Soest, 2014) and on their estimated associations with outcomes and risk factors (Bassett & Lumsdaine, 2001; Kleinjans & van Soest, 2014). Our main findings are robust to dropping respondents who report a probability of 0.5.

Observed patterns in subjective probability data that are often ascribed to reporting error may actually signal precision of underlying beliefs that is relevant for behavior (Delavande et al., 2025; Drerup et al., 2017; Enke & Graeber, 2023). About half of HRS respondents are estimated to hold imprecise LTC beliefs (Giustinelli et al., 2022b). Imprecision that reflects ambiguous beliefs can directly predict behavior (Delavande et al., 2025)[23], while imprecision reflecting lack of attention to belief formation or cognitive uncertainty when people face complex decisions and rely on heuristics or advice can attenuate the association between behavior and reported (or even objective) probabilities (Drerup et al., 2017; Enke & Graeber, 2023).

The larger underweighting of risk factors and more attenuated relationship between subjective probabilities and LTC insurance behavior among the less cognitively able that we observe are consistent with individuals facing multiple sources of uncertainty that reduce the precision of

[23] Ambiguity can push reported probabilities toward 0.5, leading to excess flatness over the range of objective probabilities (Hudomiet et al., 2023).

their underlying beliefs, as well as their reported probabilities, and may subdue their demand for LTC insurance. Alternatively, these patterns could be explained by low cognition individuals holding (and reporting) precise but inaccurate beliefs, while being constrained in their ability to act upon those beliefs. They may be rejected more often by LTC insurers or crowd-out of their demand by Medicaid may be stronger. To discriminate between these explanations, future research could try to disentangle the causes of the underutilization of shared information. Asking respondents how uncertain they are about their reported probabilities – and trying to ascertain whether that uncertainty stems from lack of information or processing constraints – would be a promising start (Enke & Graeber, 2023; Giustinelli et al., 2022b).

In conclusion, we show that older Americans have inaccurate perceptions of LTC risk in large part because they underutilize information on risk factors they would have to reveal to prospective insurers and the resulting inaccuracy is only partially offset by private information. This could tilt the balance of asymmetric information in the LTC insurance market in favor of insurers who, presumably, are better placed than consumers to predict risks from jointly observed risk factors and to deal with uncertainty surrounding those risks. Our empirical analysis demonstrates the quantitative importance of the underutilization of shared information, suggesting that it would be worthwhile to conduct theoretical analysis of the consequences for the operation of the insurance market.

References

Abaluck, J., & Gruber, J. (2011). Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program. *American Economic Review*, *101*(4), 1180–1210.

Abaluck, J., & Gruber, J. (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review, 106*(3), 2145–2184.

Akamigbo, A. B., & Wolinsky, F. D. (2006). Reported expectations for nursing home placement among older adults and their role as risk factors for nursing home admissions. *The Gerontologist*, *46*(4), 464–473.

American Association for Long-Term Care Insurance. (2024). *2024 Long-Term Care Insurance Facts – Prices – Data – Statistics – 2024 Reports*. https://www.aaltci.org/long-term-care-insurance/learning-center/ltcfacts-2024.php

American Council on Aging. (2021, December 14). *Spending Down Assets to Become Medicaid Eligible for Nursing Home / Long Term Care.* https://www.medicaidplanningassistance.org/medicaid-spend-down/

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., & Tonetti, C. (2018). The long-term care insurance puzzle: modeling and measurement. *Working paper*. https://ebp-projects.isr.umich.edu/VRI/papers/VRI-LTC-I.pdf

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., & Tonetti, C. (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy*, *128*(6), 2375–2451.

Bago d'Uva, T., & O'Donnell, O. (2022). Explaining probability judgement inaccuracy: a lens model extended decomposition of the Brier score. *Decision*, *9*(1), 74–90.

Baicker, K., Mullainathan, S., & Schwartzstein, J. (2015). Behavioral hazard in health insurance. *The Quarterly Journal of Economics*, *130*(4), 1623–1667.

Baillon, A., Kraft, A. D., O'Donnell, O., & van Wilgenburg, K. (2022). A behavioral decomposition of willingness to pay for health insurance. *Journal of Risk and Uncertainty, 64*(1), 43–87.

Bassett, W. F., & Lumsdaine, R. L. (2001). Probability limits: Are subjective assessments adequately accurate? *Journal of Human Resources*, 327–363.

Bassett, W. F. (2007). Medicaid's nursing home coverage and asset transfers. *Public Finance Review, 35*(3), 414–439.

Bhargava, S., Loewenstein, G., & Sydnor, J. R. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics, 132*(3), 1319–1372.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, *127*(3), 1243–1285.

Boyer, M., De Donder, P., Fluet, C., Leroux, M. L., & Michaud, P. C. (2019). Long-term care risk misperceptions. *The Geneva Papers on Risk and Insurance–Issues and Practice*, *44*(2), 183–215.

Boyer, M. M., De Donder, P., Fluet, C., Leroux, M. L., & Michaud, P. C. (2020). Long-term care insurance: information frictions and selection. *American Economic Journal: Economic Policy*, *12*(3), 134–169.

Braun, R. A., Kopecky, K. A., & Koreshkova, T. (2019). Old, frail, and uninsured: accounting for features of the US long-term care insurance market. *Econometrica*, *87*(3), 981–1019.

Brown, J. R., & Finkelstein, A. (2009). The private market for long-term care insurance in the United States: A review of the evidence. *Journal of Risk and Insurance*, *76*(1), 5–29.

Brown, J. R., & Finkelstein, A. (2011). Insuring long-term care in the United States. *Journal of Economic Perspectives*, *25*(4), 119–142.

Brown, J. R., Goda, G. S., & McGarry, K. (2012). Long-term care insurance demand limited by beliefs about needs, concerns about insurers, and care available from family. *Health Affairs*, *31*(6), 1294–1302.

Bruine de Bruin, W., & Carman, K. G. (2012). Measuring risk perceptions: What does the excessive use of 50% mean? *Medical Decision Making*, *32*(2), 232–236.

De Donder, P., & Leroux, M. L. (2013). Behavioral biases and long-term care insurance: A political economy approach. *The BE Journal of Economic Analysis & Policy*, *14*(2), 551–575.

Delavande, A., Del Bono, E., & Holford, A. (2025). Imprecise health beliefs and health behavior. *Journal of Health Economics*, *102*, 103003.

de Meza, D., & Webb, D. C. (2001). Advantageous Selection in Insurance Markets. *The RAND Journal of Economics*, *32*(2), 249–262.

Drerup, T., Enke, B., & von Gaudecker, H. M. (2017). The precision of subjective data and the explanatory power of economic models. *Journal of Econometrics*, *200*(2), 378–389.

Dupont, R. (2024, February 28). *Memory Testing: How to Qualify for Long-Term Care Insurance.* https://www.aplaceformom.com/caregiver-resources/articles/memory-test-for-long-term-care-insurance

Enke, B., & Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, *138*(4), 2021–2067.

Finkelstein, A., & McGarry, K. (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review*, *96*(4), 938–958.

Fischhoff, B., & Bruine de Bruin, W. (1999). Fifty–fifty = 50%? *Journal of Behavioral Decision Making*, *12*(2), 149–163.

Fisher, G. G., Hassan, H., Faul, J. D., Rodgers, W. L., & Weir, D. R. (2017). Health and Retirement Study: Imputation of cognitive functioning measures: 1992–2014 (Final release version): Data description. Ann Arbor, MI: University of Michigan, Survey Research Center.

Gennaioli, N., & Shleifer, A. (2010). What Comes to Mind? *The Quarterly Journal of Economics, 125*(4), 1399–1433.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Giustinelli, P., Manski, C. F., & Molinari, F. (2022a). Tail and center rounding of probabilistic expectations in the Health and Retirement Study. *Journal of Econometrics*, *231*(1), 265–281.

Giustinelli, P., Manski, C. F., & Molinari, F. (2022b). Precise or imprecise probabilities? Evidence from survey response related to late-onset dementia. *Journal of the European Economic Association*, *20*(1), 187–221.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, *103*(7), 2643–2682.

Handel, B. R., & Kolstad, J. T. (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, *105*(8), 2449–2500.

Handel, B. R., Kolstad, J. T., & Spinnewijn, J. (2019). Information frictions and adverse selection: Policy interventions in health insurance markets. *Review of Economics and Statistics*, *101*(2), 326–340.

Handel, B., Kolstad, J., Minten, T., & Spinnewijn, J. (2024). The socio-economic distribution of choice quality: Evidence from health insurance in the Netherlands. *American Economic Review: Insights, 6*(3), 395–412.

Handel, B., & Schwartzstein, J. (2018). Frictions or mental gaps: what's behind the information we (don't) use and when do we care? *Journal of Economic Perspectives*, *32*(1), 155–178.

Health and Retirement Study (2021). *RAND HRS Longitudinal File 2020 public use dataset*. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Hendren, N. (2013). Private information and insurance rejections. *Econometrica*, *81*(5), 1713–1762.

Ho, K., Hogan, J., & Scott Morton, F. (2017). The impact of consumer inattention on insurer pricing in the Medicare Part D program. *RAND Journal of Economics*, *48*(4), 877–905.

Holden, K., McBride, T., & Perozek, M. (1997). Expectations of nursing home use in the Health and Retirement Study: The role of gender, health, and family characteristics. *The*

Journals of Gerontology Series B: Psychological Sciences and Social Sciences, *52*(5), S240–S251.

Hudomiet, P., Hurd, M. D., & Rohwedder, S. (2023). Mortality and health expectations. In R. Bachmann, G. Topa, & W. van der Klaauw (Eds.), *Handbook of Economic Expectations* (pp. 225–259). Academic Press.

Ketcham, J. D., Lucarelli, C., & Powers, C. A. (2015). Paying attention or paying too much in Medicare Part D. *American Economic Review*, *105*(1), 204–233.

Kleinjans, K. J., & Soest, A. V. (2014). Rounding, focal point answers and nonresponse to subjective probability questions. *Journal of Applied Econometrics*, *29*(4), 567–585.

Lambregts, T. R., & Schut, F. T. (2020). Displaced, disliked and misunderstood: A systematic review of the reasons for low uptake of long-term care insurance and life annuities. *The Journal of the Economics of Aging*, *17*, 100236.

Lambregts, T. R., & Schut, F. T. (2024). Who can see it coming? Demand-side selection in long-term care insurance related to decision-making abilities. *Journal of Risk and Insurance*, *91*(3), 697–719.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67–80.

Lindrooth, R. C., Hoerger, T. J., & Norton, E. C. (2000). Expectations among the elderly about nursing home entry. *Health Services Research*, *35*(5 Pt 2), 1181–1202.

Manski, C. F., & Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics*, *28*(2), 219–231.

Medicare. (n.d.). *What Part A covers*. https://www.medicare.gov/providers-services/original-medicare/part-a

Mehta, K. M., Yaffe, K., Langa, K. M., Sands, L., Whooley, M. A., & Covinsky, K. E. (2003). Additive effects of cognitive function and depressive symptoms on mortality in elderly community-living adults. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *58*(5), M461–M467.

Ofstedal, M. B., Fisher, G. G., & Herzog, A. R. (2005). Documentation of cognitive functioning measures in the Health and Retirement Study. Ann Arbor, MI: University of Michigan, Survey Research Center.

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358–369.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.

Pauly, M. V. (1990). The rational nonpurchase of long-term-care insurance. *Journal of Political Economy*, *98*(1), 153–168.

Pestieau, P., & Ponthière, G. (2012). Long term care insurance puzzle. In J. Costa-Font & C. Courbage (Eds.), *Financing long-term care in Europe: Institutions, markets and models* (pp. 41–52). London: Palgrave Macmillan.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, *12*(6), 1423–1452.

Solomon, A. (2026). Imperfect Private Information in Insurance Markets. *The Review of Economics and Statistics*, *108*(2), 485–503.

Spinnewijn, J. (2017). Heterogeneity, demand for insurance, and adverse selection. *American Economic Journal: Economic Policy*, *9*(1), 308–343.

Taylor, D. H., Jr., Osterman, J., Will Acuff, S., & Østbye, T. (2005). Do seniors understand their risk of moving to a nursing home? *Health Services Research*, *40*(3), 811–828.

Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, *30*(1), 132–156.

Appendices

Appendix A. Additional Tables

Table A1. Means of risk factors in samples used to predict subjective probability of moving to a nursing home within 5 years (model (3)) and that outcome (model (4))

		Sample	
Variable	**Definition**	**Model (3)**	**Model (4)**
Sex & age			
Male	1 if male	0.415	0.410
Age	Years	74.8	73.6
Activities of daily living (ADLs)			
Bathing	1 if have any difficulty with activity, 0 otherwise	0.058	0.052
Eating		0.026	0.023
Dressing		0.091	0.087
Toileting		0.047	0.053
Walking		0.058	0.059
Number ADLs	Count of number of ADLs have any difficulty with	0.324	0.317
Instrumental Activities of Daily Living (IADLs)			
Grocery shopping	1 if have any difficulty with activity, 0 otherwise	0.083	0.084
Medication manage		0.024	0.022
Number IADLs	Count of number of IADLs have any difficulty with	0.200	0.181
Miscellaneous health			
Underweight	1 if body mass index < 18, 0 otherwise	0.017	0.016
Obese	1 if body mass index ≥ 30, 0 otherwise	0.309	0.273
Depressed	1 if CES-D8 > 3, 0 otherwise.	0.181	0.201
Incontinence	1 if lost any amount of urine beyond control in last 12 months, 0 otherwise	0.296	0.239
Prescription drugs	1 if reports regular use of prescription drugs, 0 otherwise	0.912	0.888
Mobility & breathing aids			
Wheelchair	1 if use, 0 otherwise	0.024	0.021
Walker		0.077	0.053
Oxygen		0.030	0.023
Cane		0.112	0.091
Crutches		0.002	0.001
Alcohol & smoking			
Drinking problem	1 if report having ≥ 3 alcoholic drinks per day, 0 otherwise	0.050	0.054
Currently smokes	1 if report currently smokes tobacco, 0 otherwise	0.082	0.098
Prior LTC use			
Nursing home care	1 if used in the previous two years, 0 otherwise	0.034	0.026
Home care		0.099	0.078
Diagnosed & medicated conditions			
Arthritis	1 if ever been told by a doctor that have condition, 0 otherwise	0.690	0.645
Cancer		0.210	0.180
Diabetes		0.244	0.199
Chronic lung disease		0.158	0.147
Psychiatric problems		0.149	0.115
Heart condition		0.358	0.332
Stroke		0.101	0.086
High blood pressure		0.681	0.605
Insulin	1 if used insulin for diabetes, 0 otherwise	0.068	0.050
Kidney failure	1 if ever told by a doctor that have kidney failure due to diabetes, 0 otherwise	0.060	0.043
Heart medication	1 if currently taking medication for heart condition, 0 otherwise	0.270	0.241
Heart attack	1 if ever told by a doctor that have had heart attack, 0 otherwise	0.131	0.107
Heart failure	1 if ever told by a doctor that have congestive heart failure, 0 otherwise	0.086	0.067
Hip fracture	1 if report has ever broken hip, 0 otherwise	0.028	0.029
Injuries due to a fall	1 if report injury seriously enough to need medical treatment, 0 otherwise	0.292	0.230
Cognitively impaired	1 if total cognition score (0-35) ≤ 8, 0 otherwise.	0.012	0.011
Sociodemographics			
Married	1 if reported being married or living with partner, 0 otherwise	0.619	0.609
Age spouse	Years	72.5	71.3
Wealth	Total net household wealth, excluding housing, social security and pension wealth. Quartile groups.		
Income	Respondent and spouse earnings, pensions and annuities, SSI and Social Security Disability, Social Security retirement, unemployment and workers compensation, other government transfers, household capital income, and other income. Quartile groups.		

n	5,987	6,849

Notes: Analysis sample for model (3) includes HRS wave 11 respondents aged 65-88 in 2012 with full item response on subjective probabilities and risk factors, and for whom it is possible to determine if they moved to a nursing home within 5 years. For model (4) we use a corresponding sample of wave 8 respondents because wave 11 respondents could not have known how their outcomes would be associated with risk factors measured in that wave. The estimated relationships between the risk factors of the wave 8 sample and movements of this sample into nursing homes over the subsequent 5 years is assumed to constitute the shared information that could possibly have been known by wave 11 respondents when forming their subjective probabilities, as well as by insurers when pricing contracts offered to them. In models, age is entered as indicators for 5-year age groups up to ≥ 85 years. CES-D8 is the Center for Epidemiologic Studies Depression (CESD) scale. Total cognition score is increasing in cognitive functioning. It is the sum of the word recall and mental status summary scores. The word recall summary score (0-20) is the sum of the immediate and delayed word recall scores. The word list contains 10 words. The mental status summary score (0-15) is the sum of scores on serial sevens test, backwards counting from 20, and object, date, and President/Vice-President naming tasks. See HRS codebook 2012 and RAND for detailed definitions of all variables.

Table A2. Sample selection

	Number of respondents
Aged 65-88 and not in nursing home in wave 11	**10,284**
Proxy interview	*-602*
Not asked subjective probability of moving to nursing home within 5 years	*-183*
Non-response to subjective probability of moving to nursing home within 5 years	*-453*
Reported subjective probability of moving to nursing home within 5 years	**9,046**
Cannot determine if moved to nursing home within 5 years	*-1,850*
Observe if moved to nursing home within 5 years	**7,186**
Missing on risk factors	*-1,199*
Full item response on all risk factors	**5,987**

Notes: Respondents are not asked to report their subjective probability of moving to a nursing home within 5 years if they do not give numerical responses to three prior questions about expectations of house values and giving/receiving an inheritance.

Table A3. Variables used in regression (6) to estimate LTC insurance take-up that are not used in regressions (3) and (4)

Variable	Definition	Mean (SD)
Private LTC insurance	1 if have private long-term care insurance, 0 otherwise	0.154
Cognition score	Total cognition score (0-35), increasing in cognitive functioning (see Table A1 notes).	21.9 (4.86)
Education	Highest level of education based on reported years of education and degrees/diplomas.	
Less than high school	Including GED.	0.224
High school graduate		0.320
Some college		0.227
College graduate	At least bachelor's degree.	0.228
Seatbelt use	1 if report always wear seatbelt, 0 otherwise	0.876
Preventive health activities	Proportion of gender-specific health activities that respondents partake in. These include a flu shot, a blood test for cholesterol, monthly self-checks for breast lumps, a mammogram, a pap smear and a check for prostate cancer.	0.739
Number of children	Alive and in-contact children of the respondent and their spouse	3.40 (2.17)
n		5,705

Notes: The sample size is smaller than the main analysis sample of n = 5,987 used to measure and decompose LTC risk perception inaccuracy due to item non-response on the variables listed in this table. See also RAND codebook for details of these variables.

Table A4. OLS estimates of regressions for subjective probability of moving to a nursing home within 5 years (p) and an indicator of that outcome (y)

	p: regression (3)		y: regression (4)	
	Estimate	SE	Estimate	SE
Sex & age				
Male	-0.003	(0.007)	0.001	(0.009)
Age (ref. ≥ 65 & < 70)				
Aged ≥ 70 & ≤ 74	0.015	(0.007)	0.026	(0.007)
Aged ≥ 75 & ≤ 79	0.028	(0.009)	0.043	(0.010)
Aged ≥ 80 & ≤ 84	0.050	(0.012)	0.115	(0.015)
Aged ≥ 85	0.093	(0.015)	0.245	(0.022)
Activities of daily living (ADLs)				
Bathing	0.058	(0.028)	0.075	(0.039)
Eating	0.052	(0.035)	-0.003	(0.046)
Dressing	0.025	(0.027)	0.050	(0.033)
Toileting	0.061	(0.029)	0.094	(0.038)
Walking	0.047	(0.030)	0.080	(0.037)
Number of ADLs	-0.033	(0.020)	-0.039	(0.025)
Instrumental Activities of Daily Living (IADLs)				
Grocery shopping	-0.006	(0.024)	-0.038	(0.036)
Medication manage	0.069	(0.037)	-0.091	(0.044)
Number of IADLs	0.005	(0.015)	0.048	(0.022)
Miscellaneous health				
Underweight	-0.035	(0.022)	0.078	(0.039)
Obese	-0.004	(0.007)	-0.013	(0.008)
Depressed	0.022	(0.009)	0.015	(0.011)
Incontinence	0.019	(0.007)	-0.010	(0.009)
Prescription drugs	0.009	(0.010)	-0.001	(0.010)
Mobility & breathing aids				
Wheelchair	0.013	(0.028)	0.003	(0.044)
Walker	0.004	(0.017)	0.084	(0.030)
Oxygen	0.036	(0.022)	0.010	(0.034)
Cane	0.026	(0.013)	0.016	(0.020)
Crutches	-0.071	(0.076)	-0.107	(0.110)
Alcohol & smoking				
Drinking problem	-0.029	(0.011)	-0.012	(0.013)
Currently smokes	0.003	(0.011)	0.022	(0.012)
Prior LTC use				
Used nursing home care	0.040	(0.021)	0.113	(0.035)
Used home care	0.001	(0.012)	0.027	(0.019)
Diagnosed & medicated conditions				
Arthritis	0.012	(0.006)	0.012	(0.007)
Cancer	0.005	(0.007)	0.022	(0.010)
Diabetes	0.012	(0.008)	0.021	(0.011)
Chronic lung disease	-0.007	(0.006)	0.008	(0.010)
Psychiatric problems	0.018	(0.009)	0.012	(0.014)
Heart condition (any)	-0.009	(0.006)	0.005	(0.010)
Stroke	0.004	(0.011)	0.046	(0.017)
High blood pressure	0.018	(0.006)	0.013	(0.008)
Used insulin for diabetes	-0.007	(0.015)	0.068	(0.025)
Kidney failure due to diabetes	0.014	(0.017)	-0.012	(0.025)
Mediation for heart condition	0.004	(0.010)	-0.007	(0.015)
Heart attack	0.006	(0.012)	0.030	(0.017)
Congestive heart failure	0.015	(0.014)	-0.006	(0.020)
Hip fracture	0.008	(0.021)	-0.037	(0.029)
Injuries due to a fall	-0.011	(0.007)	0.027	(0.010)
Cognitively impaired	-0.063	(0.038)	0.162	(0.061)
Sociodemographics				
Married	-0.117	(0.041)	-0.095	(0.048)
Age spouse	0.001	(0.001)	0.001	(0.001)
Wealth quartile group (ref. Poorest)				
2nd Poorest	0.026	(0.010)	-0.013	(0.012)
2nd Richest	0.035	(0.009)	-0.015	(0.011)
Richest	0.013	(0.009)	-0.011	(0.011)
Income quartile group (ref. Poorest)				
2nd Poorest	0.021	(0.011)	0.009	(0.013)
2nd Richest	0.021	(0.010)	-0.001	(0.012)
Richest	0.017	(0.010)	0.003	(0.012)
Constant	0.068	(0.013)	0.014	(0.013)
R^2	0.060		0.152	
Mean dependent variable	0.165		0.106	

n	5,987	6,849

Notes. Regression (3) estimated using HRS wave 11 respondents aged 65-88 in 2012 with full item response on subjective probabilities and risk factors, and for whom it is possible to determine if they moved to a nursing home within 5 years. Regression (4) is estimated with a corresponding sample observed in wave 8 (2006). The dependent variable in (4) is an indicator of having moved to a nursing home for at least 21 consecutive nights or until death within 5 years of wave 8 interview. The covariates for this model are reported/measured in wave 8. See notes to Table A1 for explanation of why we use different samples for the two regressions. Robust standard errors in parentheses.

Table A5. Heterogeneity in risk perception inaccuracy (MSPE) without controls for sex, age, and marital status

	(1)		(2)		(3)	
Wealth (ref. Richest quartile)						
Poorest quartile	0.032	(0.010)				
2nd Poorest quartile	0.023	(0.009)				
2nd Richest quartile	0.010	(0.009)				
Constant	0.122	(0.006)				
Education (ref. College graduate)						
High school dropout or GED			0.044	(0.010)		
High school graduate			0.029	(0.009)		
Some college			0.028	(0.010)		
Constant			0.113	(0.006)		
Cognitive functioning (ref. Top quartile)						
Bottom quartile					0.125	(0.009)
2nd Bottom quartile					0.076	(0.009)
2nd Top quartile					0.040	(0.007)
Constant					0.076	(0.005)
n	5,987		5,986		5,987	

Notes: Columns (1)-(3) show estimates from separate OLS regressions of the individual squared prediction error of the subjective probability of moving to a nursing home within 5 years $((p_i - y_i)^2)$ on indicators of each of household wealth quartile group, educational attainment, total cognition score quartile group, respectively. Differently from Table 3, the regressions used here do not control for sex, age and marital status. Robust standard errors in parentheses.

Table A6. Inappropriate weighting of risk factors by cognition

	Quartile group of total cognition score			
	Bottom	2nd Bottom	2nd Top	Top
Total inappropriate weighting, $\Delta\hat{y} - \Delta\hat{p}$	**0.111**	**0.080**	**0.060**	**0.003**
Contributions				
Age & sex	0.072	0.033	0.020	-0.005
	(0.015)	(0.011)	(0.011)	(0.015)
ADLs	0.018	-0.008	-0.007	-0.008
	(0.009)	(0.010)	(0.013)	(0.012)
IADLs	-0.003	0.011	0.004	-0.006
	(0.007)	(0.010)	(0.011)	(0.007)
Miscellaneous health	-0.003	0.003	0.004	-0.009
	(0.006)	(0.007)	(0.007)	(0.006)
Mobility & breathing aids	0.006	0.012	0.019	0.027
	(0.010)	(0.014)	(0.017)	(0.021)
Alcohol & smoking	-0.001	-0.001	0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)
Diagnosed & medicated conditions	0.016	0.010	0.018	-0.002
	(0.009)	(0.010)	(0.011)	(0.015)
Prior LTC use	0.009	0.019	-0.002	0.001
	(0.007)	(0.012)	(0.009)	(0.007)
Sociodemographics	-0.004	0.001	0.003	0.005
	(0.008)	(0.007)	(0.005)	(0.008)
n	1,671	1,345	1,591	1,380

Notes: Top row gives $\Delta\hat{y} - \Delta\hat{p}$ for each cognition group. See notes to Table 4 for definitions of groups and Table 2 for notation and samples. Other rows give $\sum_{j\in\Omega}(\hat{\beta}_j^y - \hat{\beta}_j^p)\Delta X_j$. Table A1 describes the risk factors included in each set. Because we stratify by cognition, we do not include the indicator of cognitive impairment in these regressions. Bootstrap standard errors (100 replications) in parentheses.

Table A7. Heterogeneity in association of LTC insurance with risk perceptions

	Cognition			
	Bottom quartile	2nd Bottom quartile	2nd Top quartile	Top quartile
Subjective probability of moving to nursing home within 5 years, *p*	0.075	0.109	0.123	0.131
	(0.027)	(0.047)	(0.048)	(0.065)
Mean outcome	0.078	0.150	0.183	0.215
n	1,572	1,277	1,523	1,333
	Education			
	High school dropout or GED	High school graduate	Some college	College graduate
Subjective probability of moving to nursing home within 5 years, *p*	0.036	0.116	0.166	0.074
	(0.030)	(0.035)	(0.050)	(0.069)
Mean outcome	0.066	0.134	0.148	0.272
n	1,280	1,828	1,294	1,303
	Wealth			
	Poorest quartile	2nd Poorest quartile	2nd Richest quartile	Richest quartile
Subjective probability of moving to nursing home within 5 years, *p*	0.017	0.066	0.183	0.155
	(0.026)	(0.038)	(0.050)	(0.062)
Mean outcome	0.056	0.108	0.188	0.261
n	1,427	1,411	1,429	1,438

Notes: The dependent variable is an indicator of having private LTC insurance, all data are from wave 11 (2012) and the sample (aged 65+) has full item response to the full set of controls. The table shows OLS estimates of the coefficient on the subjective probability of moving to a nursing home within 5 years in linear probability models of that indicator, γ in regression (6), stratified by wealth, education and cognition. Table 5, column (3), shows this estimate for the full sample. Controls include risk factors (***X***), as listed in Table A1, and other controls (***W***), as listed in Table A3, plus interactions of sex and age groups with number of ADLs/IADLs and total cognition score.

Table A8. Oster (2019) bounds for partial association of LTC insurance with subjective probability of moving to nursing home within 5 years

	(1) R^2_{max} = 1.3 R^2	(2) R^2_{max} = 2 R^2	(3) R^2_{max} = 3 R^2
Bias-adjusted $\hat{\gamma}$ when $\delta = 1$	0.098	0.091	0.079
$\tilde{\delta}$ required for $\hat{\gamma} = 0$	17.46	5.71	2.91

Notes: R^2 is the (proportionate) explained variance of an indicator of LTC insurance obtained from an OLS regression on the subjective probability of moving to nursing home within 5 years, risk factors and other observed controls, i.e. eq.(6). See Table 5 column (3). R^2_{max} is the assumed maximum R^2 that could be achieved with additional control for unobservables potentially correlated with the subjective probability. δ is the degree of selection into insurance on unobservables relative to that on observables. In the top row, selection on unobservables is assumed proportional to that on observables and the bias-adjusted partial association is calculated for different values of R^2_{max}. The second row shows the relative degree of selection on unobservables that would be required to drive the partial association to zero at different values of R^2_{max}. We follow Oster (2019) in using R^2_{max} = 1.3 R^2 to calculate the main bias-adjusted coefficient and showing estimates with 2 R^2 and 3 R^2 as more conservative alternative scenarios.

Table A9. Association of nursing-home-inclusive LTC insurance with risk perceptions

	(1)	(2)	(3)	(4)	(5)
		OLS		Bias-adjusted	IV
Subjective probability of moving to	0.108	0.109	0.104	0.102	0.085
nursing home within 5 years, *p*	(0.021)	(0.020)	(0.020)		(0.061)
Controls					
Risk factors, ***X***	No	Yes	Yes	Yes	Yes
Others, ***W***	No	No	Yes	Yes	Yes
Effective *F*-statistic					67.66
[critical value]					[9.98]
Hansen J test, p-value					0.4873
R^2	0.005	0.082	0.094		
Mean outcome	0.145	0.145	0.145	0.145	0.145
n	5,675	5,675	5,675	5,675	5,675

Notes: In all columns, the dependent variable is an indicator of having private nursing-home-inclusive LTC insurance, all data are from wave 11 (2012) and the sample (aged 65+) has full item response to the full set of controls. Columns (1)-(3) show OLS estimates of the coefficient on the subjective probability of moving to a nursing home within 5 years in a linear probability model of that indicator – γ in regression (6). Risk factors (***X***) are listed in Table A1, Other controls (***W***) are the variables listed in Table A3, plus interactions of sex and age groups with number of ADLs/IADLs and total cognition score. The bias-adjusted estimate in column (4) is the Oster (2019) bound calculated using a maximum $R^2 = 1.3\ R^2$ in column (3) and assuming proportional selection of observables and unobservables. Table A7 gives additional bound estimates. Column (5) shows the Lewbel (2012) IV estimate using instruments constructed from heteroskedasticity related to the total cognition score and its interaction with sex. Effective *F*-statistic is the Montiel Olea & Pflueger (2013) weak IV test. The critical value is for a 5% significance level and allowing for bias up to 10% of the worst-case bias. The Hansen J test is for the joint null hypothesis that the instruments are valid. Robust standard errors are shown in parentheses.

Table A10. Heteroskedasticity-based IV estimates of effect of subjective probability of moving to nursing home within 5 years on probability of having LTC insurance

	(1)	(2)	(3)
Subjective probability of moving to	0.085	0.106	0.107
nursing home within 5 years, *p*	(0.069)	(0.076)	(0.076)
Controls			
Risk factors (***X***) and preference shifters (***W***)	Yes	Yes	Yes
Numeracy score	No	No	Yes
Effective *F*-statistic	147.23	53.88	53.80
[critical value]	[23.11]	[9.24]	[9.23]
Hansen J test p-value		0.314	0.318
n	5,705	4,405	4,405

Notes: This table shows Lewbel (2012) IV estimates of the effect of the subjective probability of moving to a nursing home within 5 years on the probability of having LTC insurance using heteroskedasticity-based instruments for the subjective probability and controlling for risk factors and preference shifters. In our main estimate shown in Table 5, column (5), we construct instruments from heteroskedasticity related to cognition and its interaction with sex. Column (1) of this table gives an estimate using only heteroskedasticity related to the total cognitive functioning score, not interacted with sex, to construct an instrument. Column (2) uses the same instruments and covariates used to produce the main estimate but uses a sub-sample of respondents who answered three questions on percentages, division and compound interest in 2010 that are intended to provide a proxy measure of numeracy. Column (3) uses this sample (and the same instruments used in (2)) and controls for a numeracy score equal to the number of correct answers to the three questions. 15% answer all three questions incorrectly and 8.4% answer all three correctly. The effective *F*-statistic is the Montiel Olea & Pflueger (2013) test for weak instruments, with the critical value given for a 5% significance level and allowing bias up to 10% of the worst-case bias. The Hansen J test is for the joint null hypothesis that the instruments are valid. This is not relevant in column (1) since this model is just identified. Robust standard errors are shown in parentheses.

Appendix B. Robustness of decompositions of risk perception inaccuracy and discriminatory power

Decomposition of risk perception inaccuracy and discrimination

Table B1. Robustness of decomposition of risk perception inaccuracy and discrimination to definition of outcome and exclusion of focal point subjective probabilities (p)

		Baseline	Outcome (y): nursing home		Drop if
			≥ 1 night	≥ 100 nights	p = 0.5
		(1)	(2)	(3)	(4)
A. MSPE	$\mathbf{\frac{1}{n}\sum(p_i - y_i)^2}$	**0.139**	**0.165**	**0.118**	**0.123**
		(0.004)	**(0.004)**	**(0.003)**	**(0.004)**
Decomposition, eq. (2)					
outcome variance	$Var(y)$	0.103	0.133	0.074	0.098
		(0.003)	(0.003)	(0.003)	(0.003)
bias^2	$(\bar{p} - \bar{y})^2$	0.002	< 0.000	0.007	0.000
		(0.000)	(0.000)	(0.001)	(0.000)
covariance	$-2(\Delta p)Var(y)$	-0.018	-0.019	-0.014	-0.015
		(0.002)	(0.003)	(0.002)	(0.002)
signal	$(\Delta p)^2 Var(y)$	0.001	0.001	0.001	0.001
		(0.000)	(0.000)	(0.000)	(0.000)
noise	$Var(p) - (\Delta p)^2 Var(y)$	0.050	0.050	0.050	0.040
		(0.001)	(0.001)	(0.001)	(0.002)
B. Discrimination slope	$\mathbf{\Delta p = \bar{p}_1 - \bar{p}_0}$	**0.086**	**0.071**	**0.093**	**0.076**
		(0.011)	**(0.010)**	**(0.013)**	**(0.011)**
Decomposition, eq.(5)					
outcome predictability	$\Delta\hat{y}$	0.147	0.158	0.127	0.141
		(0.009)	(0.008)	(0.010)	(0.009)
inappropriate weighting	$-(\Delta\hat{y} - \Delta\hat{p})$	-0.093	-0.108	-0.068	-0.102
		(0.009)	(0.008)	(0.010)	(0.009)
	$100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$	63.3%	68.2%	53.2%	72.2%
private information	$\Delta\hat{\varepsilon}$	0.032	0.021	0.033	0.037
		(0.009)	(0.008)	(0.011)	(0.009)
Sample size	n	5,987	5,987	5,987	5,263

Notes: Column (1) reproduces estimates given in Table 1. Columns (2) and (3) use different lengths of stay – ≥ 1 night and ≥ 100 nights, respectively – to define the outcome (move to a nursing home). Baseline uses ≥ 21 nights. In the sample, outcome means using definitions of a stay of ≥ 1 night, ≥ 21 nights, and ≥ 100 nights are 0.158, 0.117, and 0.081, respectively. Column (4) shows estimates after dropping from the sample those reporting a subjective probability of moving to a nursing within 5 years equal to 0.5.

Decomposition of risk perception discrimination slope

Based on regressions (3) and (4) for the subjective probability of moving to a nursing home within five years (p) and that outcome (y), respectively, panel B of Table 1 gives results from using eq. (5) to decompose the discrimination slope of the subjective probabilities, $\Delta p = \bar{p}_1 - \bar{p}_0$, into outcome predictability, $\Delta\hat{y}$, inappropriate weighting of risk factors, $\Delta\hat{y} - \Delta\hat{p}$, and private information, $\Delta\hat{\varepsilon}$. Table B2 demonstrates robustness of this decomposition to use of different samples and specifications of regressions (3) and (4), and to using random forest regression, rather than OLS, to predict the subjective probabilities and the outcome.

Alternative sample. The main results in Table 1 – reproduced in column (1) of Table B2 – use estimates of (4) obtained by regressing an indicator of moving to a nursing home within five years of wave 8 on risk factors observed in that wave. Column (2) of Table B2 shows estimates obtained when we use the same sample to estimate both (3) and (4). In that case, the outcome used in (4) is an indicator of moving to a nursing home within five years after wave 11 and the risk factors are those reported and measured in that wave. In this case, the contribution of outcome predictability to the discrimination slope increases. This is expected since predictions of the outcome are now made within sample, not out of sample as is the case with the approach taken for the main estimates. However, the increase is marginal (from 0.147 to 0.155). Consequently, the fraction of the potential discriminatory power of the observed risk factors that is unrealized because of inappropriate weighting in formation of the subjective probabilities ($(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$) rises by less than 2 pp.

Alternative specifications. To obtain the main estimates, we do not include interactions between risk factors in equations (3) and (4). This makes the detailed decomposition presented in Table 2 feasible. Table B2 shows robustness to relaxing this restriction; column (3) shows estimates of the (non-detailed) decomposition of the discrimination slope obtained with interactions

between sex and age groups and each of the number of ADLs/IADLs and an indicator of cognitive impairment in (3) and (4), as in Finkelstein & McGarry (2006). This only slightly increases outcome predictability and has even smaller impacts on the magnitudes of inappropriate weighting and private information. To be consistent with the eligibility criteria for Medicaid coverage of nursing home expenses, we exclude housing wealth from the wealth measure. Table B2 also shows robustness to this choice; including housing wealth, as in Finkelstein and McGarry (2006), has a negligible impact on each component of the decomposition (column (4)). Finally, we assess robustness to using an indicator of cognitive impairment in equations (3) and (4), as Finkelstein & McGarry (2006). Since less than 2 percent of the sample is cognitively impaired by this measure, we replace it with an indicator of being below the first (bottom) quartile of the total cognition score. Column (5) shows that outcome predictability increases slightly and there is a negligible impact on the other results.

Alternative estimators. The linear specification of equations (3) and (4) facilitate the detailed decomposition given in Table 2. OLS estimation of the parameters of equation (4) gives a set of risk factor weights that minimize the outcome MSPE. These provide an appropriate benchmark against which to evaluate the weights implicit in the subjective probabilities that we aim to recover through OLS estimation of (3). Notwithstanding these advantages of linear models estimated by OLS, machine learning methods can allow for extensive nonlinearity that would be expected to give better predictions of the outcome from the risk factors, and so increase the outcome predictability component ($\Delta\hat{y}$) of the discrimination slope decomposition. Machine learning may also be better at modeling the subjective probabilities, with consequences for the inappropriate weighting and private information components of the discrimination slope decomposition, eq. (5).

For these reasons, we assess robustness of the decomposition to using random forest regression to predict the reported subjective probability of moving to a nursing within five years and the

realization of that outcome from the risk factors. Since our sample is relatively small compared with many random forest applications, we use 80% of it to train each model and 20% for testing (rather than a 50-50 split). We use the mean squared prediction error as the splitting criterion at each internal node, and set the minimum node size to 10 to limit overfitting.

Comparing column (1) and (6) of Table B2 reveals the surprising result that outcome predictions from the random forest regression performs slightly worse than those from OLS in discriminating between those who move to a nursing home and those who do not. The discrimination slope of those random forest (RF) predictions is smaller than that of the OLS ones: ($\Delta\hat{y}^{RF} < \Delta\hat{y}^{OLS}$). The reason is that the outcome predictions use estimates from models that are fitted to data on wave 8 risk factors and nursing home admissions over the subsequent five years. However, $\Delta\hat{y}$ $\left(= {}^{1}/_{n_1} \sum 1(y_i = 1)\hat{y}_i - {}^{1}/_{n_0} \sum 1(y_i = 0)\hat{y}_i\right)$ measures the extent to which these predictions discriminate between those who move a nursing home and those who do not within five years of wave 11. Random forest regression gives more accurate predictions than OLS when applied within the sample *period* used for estimation. But it performs worse than OLS when the estimates obtained using wave 8 data (+ 5 years) are used to make predictions from wave 11 data. Despite the precautions taken to reduce the risk of overfitting, it appears that the random forest regression estimates remain more prone to this.

The private information term ($\hat{\varepsilon}$) obtained using the random forest regression of the subjective probabilities (column 6) is slightly larger than the respective term obtained with OLS (column 1). This implies that the random forest estimates give predictions of the subjective probabilities that discriminate between those who move to a nursing home and those who do not to a lesser extent than is achieved with predictions obtained from the OLS estimates ($\Delta\hat{p}^{RF} < \Delta\hat{p}^{OLS}$). This second surprising result can also be explained. The random forest regression does predict the subjective probabilities more accurately from the risk factors: ${}^{1}/_{n} \sum (p_i - \hat{p}_i^{RF})^2 =$

$0.036 < {}^{1}/_{n}\sum(p_i - \hat{p}_i^{OLS})^2 = 0.048$. However, the predictions of the subjective probabilities obtained from the random forest estimates do not discriminate as well as the OLS predictions between the values of the outcome (*y*). The random forest estimator is better at modeling the mistakes made in forming subjective probabilities – capturing also variation in those probabilities that is not correlated with nursing home admission.

Finally, column (7) gives estimates obtained with the same sample and specification of covariates used in column (1) for the main estimates but using a fractional logit model for the subjective probabilities (bounded in the interval [0, 1]) and a logit model for the (binary) outcome, instead of using the linear specifications of equations (3) and (4). Use of these nonlinear estimators makes no notable difference to the decomposition results, showing that the linear specifications are not too restrictive.

Table B2. Robustness of decomposition of discrimination slope of subjective probabilities

	Main estimates	Model outcome with					
		wave 11-14 data	wave 8-11 data, as in baseline				
			With interactions	Include housing wealth	Alternative cognition indicator	Random forest	(Fractional) Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome predictability	0.147	0.155	0.151	0.147	0.150	0.130	0.156
$\Delta\hat{y}$	(0.009)	(0.011)	(0.010)	(0.009)	(0.009)	(0.002)	(0.011)
Inappropriate weighting	-0.093	-0.101	-0.096	-0.092	-0.095	-0.079	-0.101
$-(\Delta\hat{y} - \Delta\hat{p})$	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.002)	(0.010)
$100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$	*63.3%*	*65.1%*	*63.3%*	*62.9%*	*63.7%*	*60.7%*	*65.2%*
Private information	0.032	0.032	0.031	0.032	0.032	0.035	0.032
$\Delta\hat{\varepsilon}$	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.001)	(0.009)
n	5,987	5,987	5,987	5,987	5,987	5,987	5,987

Notes: All columns give the eq. (5) decomposition of $\Delta p = \bar{p}_1 - \bar{p}_0$: the difference between the means of the subjective probabilities of those who do and do not move to a nursing home within 5 years. In each case, Δp = 0.086 (SE = 0.011). Column (1) reproduces the main estimates from Table 1. In column (2), $\Delta\hat{y}$ is obtained from eq. (4) estimated by regressing an indicator of moving to a nursing home within 5 years of wave 11 on risk factors observed in that wave. In columns (1) and (3)-(6), $\Delta\hat{y}$ is obtained from eq. (4) estimated by regressing an indicator of moving to a nursing home within 5 years of wave 8 on risk factors observed in that wave. In column (3), we add interactions of sex and age groups with number of ADLs/IADLs and cognitive impairment to the baseline specification of models (3) and (4). In column (4), we form wealth quartile groups from total household wealth including housing wealth. In column (5), we replace the cognitive impairment indicator with an indicator of being below the lowest quartile of total cognition score. In column (6), we use random forest regression, instead of OLS, to predict the subjective probabilities and the outcome from the risk factors. In column (7), we use fractional logit to model the subjective probabilities and logit to model the outcome. Bootstrap standard errors (100 replications) in parentheses.