Long-Term Care Risk Perceptions, Information, and Insurance

Lisa Voois Teresa Bago d'Uva Owen O'Donnell

Erasmus University Rotterdam & Tinbergen Institute

February, 2024

Abstract

We measure the accuracy of older Americans' long-term care (LTC) risk perceptions by comparing subjective probabilities of moving to a nursing home with outcomes of that event. We estimate the contributions to accuracy of two categories of information: private and shared with insurers. We find inaccuracy that is partly due to inappropriate weighting of the risk factors that insurers can observe. Only 37% of the potential discriminatory power of this shared information is realized. Private information offsets only around one third of the resulting inaccuracy. We also find that lower cognition is associated with risk perceptions that are less accurate, utilize less shared information, and contain less private information. Perceived risk is positively associated with LTC insurance, and this persists after adjusting for extensive controls, using lagged perceived risk to avoid reverse causality, and instrumenting individuals' perceived risk with their number of children. These findings point to the potential for *behavioral selection* out of insurance due to underutilization of shared information that may partly offset adverse selection.

Keywords: long-term care, risk perception, subjective probability, information friction, behavioral 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 (LTCI) is surprisingly low given that formal care spending is a major financial risk in old age and that public insurance is limited in most countries. Misperception of long-term care (LTC) risk could explain this puzzle. We measure the accuracy of older Americans' perceptions of this risk and estimate the extent to which that accuracy is improved through use of private information and worsened by underutilization of information that is shared with insurers.

Formation of accurate LTC risk perceptions requires ability to acquire extensive health information, recognize risk factors, and cognitively process all this information. There is scope for relevant risk factors being weighted incorrectly, or ignored entirely, and for diversion of attention to salient but irrelevant circumstances and characteristics. Even without possession of private information, the subjective expectation of uninsured LTC costs may not be consistent with 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 people to decline insurance offers that they would take if their risk perceptions were accurate. This *behavioral selection* may partly offset the influence of private information and so constrain its scope to generate adverse selection. This balance of information depends not only on consumers' possession of private information but also on comparative ability on the two sides of the market to process and utilize shared information.

We use data from the Health and Retirement Study (HRS) to measure the accuracy of LTC risk perceptions by comparing subjective probabilities of moving to a nursing home within five years with the actual outcomes of that event. The subjective probabilities and the outcomes are both modelled as functions of risk factors that insurers can also observe. Model estimates are used to decompose the inaccuracy of the subjective probabilities – their mean squared prediction error – into outcome variability that is not predictable from the risk factors, bias, noise, and the offsetting discriminatory power of the subjective probabilities (Bago d'Uva & O'Donnell, 2022). Discriminatory power is given by the difference in mean subjective probability between those who enter a nursing home within five years and those who do not, the discrimination slope (Yates, 1982). We measure the extent to which this increases with use of private and shared information and decreases with underuse of information due to inappropriate weighting of the jointly observed risk factors.

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. Nevertheless, many underestimate the risk. Unpredictable outcome variability and noise in the subjective probabilities contribute most to their inaccuracy. This is partly offset by the discriminatory power of the subjective probabilities, of which around 37% comes from private information. There is previous evidence that these probabilities contain private information (Finkelstein & McGarry, 2006; Hendren, 2013) but this is the first study to quantify the importance of that to the formation of accurate LTC risk perceptions.

The remainder of the discriminatory power of the subjective probabilities comes from use of shared information. There is far from full utilization of this information. Only 37% of its potential discriminatory power is realized. The rest is unused; 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 diagnosed and medicated health conditions, reliance on mobility and breathing aids, prior LTC use, and limitations in (instrumental) activities of daily living.

We examine heterogeneity by wealth because means-tested public insurance (Medicaid), which covers more than 60% of LTC costs, imposes a substantial implicit tax on private LTCI that varies with wealth. Brown & Finkelstein (2011) estimate that Medicaid crowds out private LTCI for a majority of the wealth distribution, while Braun et al. (2019) find almost complete crowd-out for the poor. Misperceptions of LTC risks are therefore expected to be more consequential for wealthier individuals, who are less protected through public insurance. We also examine heterogeneity by education and cognition because each may affect ability to process information and financial literacy (Lusardi & Mitchell, 2007) and have been found to correlate with accuracy of subjective survival probabilities (Bago d'Uva et al., 2020).

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 least use of the available shared information. The bottom quartile cognition group fails to use 71% of the potential discriminatory power of the shared information, in contrast with 5% for the top quartile. Given the strong correlation between cognition and both wealth and education, these cognition-related differences in the accuracy of LTC risk perceptions may explain socioeconomic differences in LTCI (Finkelstein & McGarry, 2006; Lambregts & Schut, 2020a). They are also consistent with a socioeconomic gradient in choice quality (Handel et al., 2022).

Inaccurate risk perceptions, measured by subjective probabilities, will only lead to behavioral selection to the extent that they influence insurance decisions. We find a positive association between the subjective probability of moving to a nursing home and holding private LTCI. It is however difficult to assess whether people act on their subjective probabilities. They may take past, correlated, behavior into account when reporting those beliefs (de Paula et al., 2014).[1] We therefore confirm robustness of that association to adding an extensive battery of controls, and calculating Oster (2019) bound estimates, as well as using the lagged subjective probability to avoid reverse causality – the insured may perceive a higher likelihood of using a nursing home given their coverage. We also instrument the subjective probability with the number of children of the respondent and their spouse. As the main providers of informal care, the number of children would be expected to lower the expectations about future need for formal care, and so the perceived risk of moving to a nursing home. It is also plausible that this does not influence demand for LTCI through other channels. The instrumental variable estimate is also positive and significant, consistent with a higher perceived LTC risk raising the likelihood of purchasing LTCI. This suggests that inaccurate reported risk perceptions may indeed imply mistaken insurance choices. We do not have incontrovertible evidence for this interpretation. Our various estimates are however all consistent with it.

Previous research demonstrates that 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). However, correlation does not imply that risk factors are weighted correctly, nor does predictive power equate to accuracy of risk perceptions. Subjective probabilities can correlate highly with the realized risk without being close, on average, to that risk. Optimal individual decisions require perceived risks that correspond to objective risks. We address these limitations of correlation studies by measuring the accuracy of LTC risk perceptions, namely, using the mean squared error of subjective probabilities of moving to a nursing home vis-à-vis the realized outcomes.

Evidence from the US (Finkelstein & McGarry, 2006) and Canada (Boyer et al., 2019) indicates that, while (upward/pessimistic) bias in perception of nursing home risk is quite

[1] Behavior has been shown to respond to experimentally manipulated subjective probabilities (Delavande & Kohler, 2016; Delavande et al., 2022). Hurwitz & Mitchell (2022) show that the provision of information on the probability of survival to old age increases regret about not having purchased LTCI.

small, there is much variation. This suggests much 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, although subjective probabilities of moving to a nursing home reflect, to some extent, individuals' risk profiles, woefully large mistakes are made, with severe 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 when the shared information is used optimally.

Some previous studies have also inferred the existence of private information on LTC risks from evidence that subjective probabilities predict nursing home admission even when conditioning on risk factors observed by insurers (Finkelstein & McGarry, 2006; Hendren, 2013; Lambregts & Schut, 2020b). These studies suggest that adverse selection on this private information may be offset by advantageous selection (de Meza & Webb, 2001) of low risks on risk preferences (Finkelstein & McGarry, 2006) and numeracy (Lambregts & Schut, 2020b) and constrained by rejection of the insurance applications of high risks (Braun et al., 2019; Hendren, 2013).[2] We highlight another mechanism that can weaken the link between private information and adverse selection: differential utilization of shared information. We suggest that this can lead to behavioral selection on any discrepancy between the price that is actuarially fair, given the information available to both sides of the market, and the price the consumer perceives to be fair, given their inferior ability to process that information. Underweighting jointly observed risk factors when forming subjective expectations of LTC costs is consistent with people most frequently citing the high price of LTCI as their reason for not purchasing it (Brown et al. 2012). Such underutilization of shared information reduces

[2] Advantageous selection would explain why those purchasing LTCI are not more likely to move to a nursing home (Finkelstein & McGarry, 2006). Hendren (2013) shows that the additional power of subjective probabilities to predict nursing home admission comes from high risks whose LTCI applications would be rejected. Similarly, Braun et al. (2019) show that high risks (and the poor) hold more private information and are more likely to be denied insurance.

the advantage consumers have from any private information. It may even tilt the balance of asymmetric information in favor of insurers who, presumably, are better placed than consumers to predict risks from observed risk factors.

There is previous evidence that LTC risk perceptions are associated with holding LTCI (Brown et al., 2012; Finkelstein & McGarry, 2006; Zhou-Richter et al., 2010). Boyer et al. (2020) confirm this finding in a stated-choice experiment and predict that eliminating risk misperceptions would only slightly increase LTCI take-up because the mean error in risk perceptions is close to zero. This assumes that under- and overestimation of the risk have equal but opposite impacts on the demand for insurance. These authors measure perception error as the deviation of the perceived risk from the risk predicted using an external model containing objective weights on various risk factors. This does not capture private information and so also does not allow to separate its role from that of underutilization of shared information in the determination of risk perception accuracy. We overcome these limitations by using data on the realized risk – moving to a nursing home.

We offer four main findings that add to evidence on LTC risk perceptions and their implications for LTCI that, more generally, feed into knowledge about information frictions and mental gaps in health-related insurance (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). First, we show that, even though subjective probabilities of moving to a nursing home have some power to predict that outcome, they are inaccurate - this is mostly because the outcome is difficult to predict and the subjective probabilities are noisy. Second, we show that the inaccuracy also stems from underutilization of information on risk factors that are shared with insurers and that private information only partially offsets this. This may deviate willingness to pay away from the fair price of insurance, causing behavioral selection that offsets adverse selection. Third, the least cognitively able hold the least accurate LTC risk perceptions because their subjective probabilities are noisier and contain less private, as well as less shared, information. Given the strong correlation between cognition and socioeconomic status, this evidence may arouse or intensify distributional concerns about inequality in well-being in old age that results from suboptimal insurance and saving decisions. Fourth, we show that LTCI is positively associated with LTC risk perceptions and that this is robust to addressing endogeneity with a number of different strategies. This suggests that concern about misperception of a major financial risk in old age distorting insurance choices may well be justified.

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). 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 did move to a nursing home within five years spanning a period that does not include the COVID-19 pandemic. This sample includes individuals born in the period 1924-1959 and their spouses (of any age).

[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. Out of 6297 respondents aged 65+ for whom we can establish whether they moved to a nursing home within five years, and who are asked the three filter questions on expectations, 1.3% do not respond to these questions and so are not asked to report their probability of moving to a nursing home within 5 years.

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. A short stay in a nursing home for rehabilitation after medical treatment may not be contemplated when a respondent is asked to report the probability of moving to a nursing home. Medicare fully reimburses rehabilitative stays of up to 20 nights in skilled nursing facilities, and it partially reimburses such stays of 21-100 nights (American Council on Aging, 2021). To improve consistency with the event referred to in the subjective probability question, we define the outcome as a nursing home stay of at least 21 consecutive nights. We assess robustness to defining the outcome as a stay of a) any duration, and b) more than 100 nights. For deceased HRS respondents, we include nursing home stays of a) any duration that end with death, and b) ≥ 21 nights before death while not in a nursing home. 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.

We model the outcome, and its subjective probability, with LTC risk factors that can be observed by insurers. Following Finkelstein and McGarry (2006), we include indicators of age and sex, limitations in activities of daily living (ADLs), 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 examine heterogeneity in the accuracy of risk perceptions measured by the subjective probabilities by wealth, education, and cognition. We use quartile groups of total net household wealth, excluding housing, social security, and pension wealth, as in the Medicaid assets test to determine eligibility for long-term care services.[5] We distinguish between four levels of education: high-school dropout/General Educational Development (GED), high-school graduate, some college, and college graduate. We make use of HRS data on several domains of cognitive functioning obtained through validated tests (Ofstedal et al., 2005; Fisher et al., 2017). We use the HRS total cognition score (0-35), which aggregates measures of episodic memory and intact mental status and is increasing in cognitive functioning (see Table A1 for more details on the score). We consider as risk factor an indicator of cognitive impairment, corresponding to cognition score equal to or lower than 8 (Mehta et al., 2003). In the heterogeneity analyses, we use four quartile groups of cognitive functioning.

Our 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.[6]

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 rather than a belief that it is precisely 0.5 – epistemic uncertainty (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 objective

[5] See https://www.verywellhealth.com/your-assets-magi-and-medicaid-eligibility-4144975 and https://www.verywellhealth.com/irrevocable-trust-medicaid-4173386.
[6] Appendix A, Table A2 gives the number of respondents dropped at each stage to reach the analysis sample.

probability of moving to a nursing home within five years – by almost 5 percentage points (pp).[7]

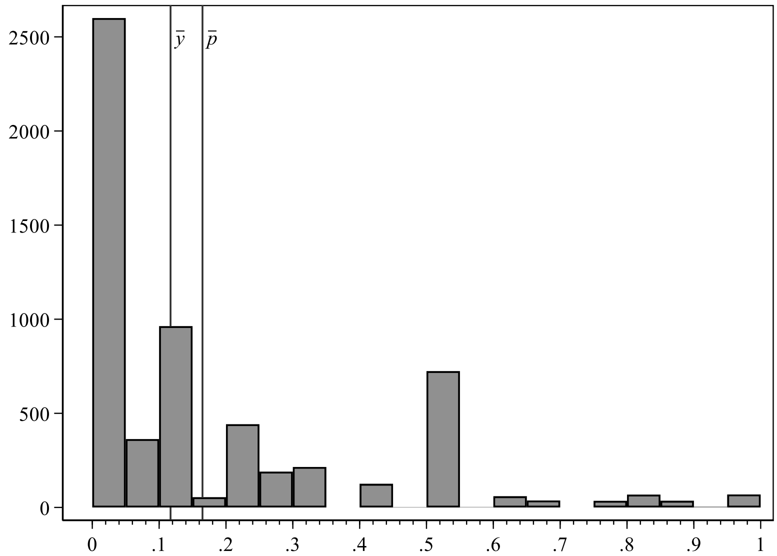

Figure 1. Distribution of subjective probabilities of moving to a nursing home within 5 years.

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: prediction difficulty, discriminatory power and noise

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

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

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.

The MSE 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

[7] The objective probability 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.

(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 increasing discriminatory power of the subjective probabilities, i.e., the extent to which they are associated with the outcome. This can be measured by the difference in their outcome-conditional means, 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, as ours, this discrimination slope relates to outcome-prediction covariance in the following way: $Cov(p, y) = \Delta p Var(y)$. Finally, inaccuracy increases with the variance of the subjective probabilities, $Var(p)$. Part of this variance is not explained by the outcome and is termed noise: $noise = Var(p) - \Delta p^2 Var(y)$. 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 captures signal, i.e., the extent to which it is explained by the outcome: $signal = \Delta p^2 Var(y)$. In sum, inaccuracy of subjective probabilities (as measured by the MSE), increases with the prediction difficulty (captured by outcome variance), with bias and noise in subjective probabilities, and decreases with their discriminatory power. These four determinants of inaccuracy are captured in this decomposition (Yates, 1982):

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

3.2 Use of available – shared and private – information in forming risk perceptions and their discriminatory power

To assess the extent to which available information is used to form accurate risk perceptions, we model the subjective probabilities and the outcome each as functions of nursing home admission risk factors ($\boldsymbol{X}$) 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 expectations with the j^{th} risk factor, and so the implicit (average) weight individuals give to it when forming their subjective expectations of moving to a nursing home; β_j^Y is the partial association of that outcome with the respective risk factor; and ε_i and υ_i are random errors. Models (3) and (4) are estimated by OLS and so their estimated coefficients give the weights that best predict the outcome and the subjective probability, respectively, from the jointly observed risk factors. $\hat{p}_i$ and $\hat{y}_i$ are the respective fitted values, while residuals $\hat{\varepsilon}_i$ capture weight given to other risk factors that are unobserved by insurers and are uncorrelated with the observed ones (Bago d'Uva & O'Donnell, 2022).

We decompose the discrimination slope of the subjective probabilities, Δp, into two parts. The first one reflects the utilization of the shared information - the jointly observed risk factors $\boldsymbol{X}$ (contained in $\hat{p}_i$). The other part represents prediction accuracy deriving from use of private information, i.e., of other risk factors that are unobserved by insurers and that are unrelated to the jointly observed risk factors (contained in $\hat{\varepsilon}_i$). Importantly, the extent to which the subjective probabilities capture information on the risk of moving to a nursing home depends also on the relationship between that outcome and the jointly observed risk factors (captured by $\hat{y}_i$). These three components of the discriminatory power of the subjective probabilities can be separated as follows:

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

where $\Delta z = \bar{z}_1 - \bar{z}_0$, $\bar{z}_k = {}^1\!/_{n_k} \sum 1(y_i = k) z_i,\ k \in \{0,1\}$, $z_i \in \{p_i, \hat{p}_i, \hat{\varepsilon}_i, \hat{y}_i\}$, and $\hat{p}_i$, $\hat{y}_i$ and $\hat{\varepsilon}_i$ are as defined above (Bago d'Uva & O'Donnell, 2022).

The term $\Delta\hat{y}$ therefore measures the extent to which moving to a nursing home can be (linearly) predicted from the jointly observed risk factors. This predictability increases the discriminatory power of the subjective probabilities, and so their accuracy (eq.(2)). The term $\Delta\hat{p}$ captures the realization of that predictability into the subjective probabilities, i.e., the subjective weights placed on the jointly observed risk factors. For example, if they predict the outcome but not the subjective probability, then that predictability is not realized and so also does not contribute to increased accuracy. The term $\Delta\hat{y} - \Delta\hat{p}$ then measures the deviation of the subjective weights from the objective weights. This term captures the loss of discriminatory power due to suboptimal use of shared information. In linear models, it can be further decomposed to reveal information extraction from each risk factor or from a set of risk factors.[8] Finally, $\Delta\hat{\varepsilon}$ is the discriminatory power that derives from private information used in forming the subjective probabilities that is not associated with the jointly observed risk factors – this contributes to increased prediction accuracy. Use of such private information can partly offset underuse of shared information.

In our main analysis, we estimate model (3) using the wave 11 (2012) sample and the subjective probabilities and risk factors reported, or measured, in the same wave, and (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.[9] This is motivated by the fact that wave 11

[8] $\Delta\hat{y} - \Delta\hat{p} = \sum_{j=1}^{J} (\hat{\beta}_j^y - \hat{\beta}_j^p) \Delta X_j$, 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.

[9] 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.

respondents could not have been aware of how the risk factors measured in that wave would eventually relate to future nursing home admission. We assume 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 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 and the objective weights that could have been known at the time. We nevertheless 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, rather than linear models (3) and (4), to predict the subjective probabilities and the outcome from the risk factors.[10] We calculate bootstrap standard errors for the MSE and each of its components in eqns. (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 LTC risk perceptions appear to influence the demand for insurance, which would give cause for concern about inaccurate perceptions possibly resulting in suboptimal insurance, we regress LTCI enrollment on the subjective probability of moving to a nursing home. Using wave 11 data, we estimate

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

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

where $LTCI_i = 1$ if the individual has private LTCI and $\boldsymbol{X}_i$ is the set of nursing home risk factors used in the MSE decomposition. These should affect the price of LTCI, and possibly its availability given that insurers often reject high-risk applicants (Hendren, 2013). Among them is a binary indicator of cognitive impairment as it is potentially observable and so usable in pricing by a prospective insurer.[11] The vector $\boldsymbol{Z}$ contains additional control variables, namely, the total cognitive functioning score, preference shifters and interactions between sex-specific age groups and the number of ADLs, the number of IADLs, and the total cognition score. The total cognition score gives better control than solely the indicator of cognitive impairment for any direct effect of cognitive ability on the insurance decision in addition to an indirect effect through price. Preference shifters include indicators of education levels and seatbelt use, and gender-specific preventive health activities as proxies for risk preferences (see Appendix Table A3 for descriptive statistics of $\boldsymbol{Z}$ control variables).

Even with an extensive set of controls, we do not claim that an OLS estimate of γ in (6) can be given a causal interpretation. There is potential for correlated unobservables, measurement error in the risk perceptions, and reverse causality – having LTCI cover would be expected to raise the perceived likelihood of moving to a nursing home. We use three strategies to assess the extent to which we can rule out that that estimate is driven solely by these potential sources of endogeneity.

To assess the potential for confounding by unobservables, we compare OLS estimates of γ as more observable controls are added to the model and use Oster (2019) bounds to obtain a bias-adjusted estimate assuming that selection on unobservables is equal to that on observables. To assess the potential for bias through reverse causality, we estimate a simple version of (6) in which the wave 11 value of $LTCI_i$ is replaced with the value of that indicator

[11] It is not uncommon for insurers of long-term care services to administer cognition tests for potential insurees, see e.g., https://www.aplaceformom.com/caregiver-resources/articles/memory-test-for-long-term-care-insurance.

in the next wave (12, two years later). Insurance cover held in 2014 cannot possibly affect the subjective probability reported in 2012. However, given the persistence of insurance status above the age of 65, there is still scope for a positive association in this revised specification to partly, or fully, result from the insured reporting a higher likelihood of moving to a nursing home. Therefore, we supplement this analysis with another that regresses $LTCI_i$ on the lagged value of the subjective probability of *ever* moving to a nursing home that is reported (once) by respondents aged 40-64 years, who have more changes in LTCI status. In this sample, we test whether the acquisition of LTCI is associated with the subjective lifetime probability reported in the previous wave.

Finally, we instrument p_i in (6) with the respondent and spouse's number of children who are alive and reported to be in contact with the respondent/spouse $(Children_i)$. The first stage equation is:

$$p_i = \eta + \theta Children_i + \boldsymbol{\varphi X}_i + \boldsymbol{\zeta Z}_i + \nu_i \,. \qquad (7)$$

Having more children – the main providers of informal care (Van Houtven & Norton, 2004; Charles & Sevak, 2005) – would be expected to lower the perceived risk of needing formal care. Conditional on our extensive battery of controls, it is plausible to assume that the number of children only influences the demand for LTCI via the perceived risk of needing to move to a nursing home.

There are nevertheless conceivable circumstances in which the exclusion restriction would be violated. For a given perceived risk of moving to a nursing home, older people with more children may be more likely to insure in order to protect wealth they intend to bequeath. As with all instruments that are not randomly assigned, doubt about the validity of this instrument cannot be fully eliminated. We use this IV estimator, along with the other strategies, as means of checking the robustness of the sign and significance of the OLS estimate of γ in (6) to correcting, as best as possible, for potential endogeneity bias. We do not claim we obtain consistent estimates of the magnitude of the causal effect of LTC risk perceptions on the demand for LTCI. We therefore remain cautious in interpreting the estimates obtained as we cannot fully rule out the presence of endogeneity that is not tackled by the approaches above, in which case there could still be a positive estimate of γ even in the absence of a true causal effect. Our aim with these analyses is rather to document whether the data are consistent with risk perceptions influencing the decision to purchase LTCI. Such evidence would support legitimate concern about behavioral consequences of inaccurate risk perceptions.

4. Results

4.1 Risk perception inaccuracy

We obtain a MSE of the subjective probabilities of moving to a nursing home equal to 0.14. This is the same value that would be obtained if, for example, 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.[12] This value is significantly ($p < 0.01$) below a benchmark of 0.25, which would be obtained if everyone were to report a 50-50 chance of moving to a nursing home. It is significantly greater (less accurate) than the MSE of 0.10 that would arise if all were to report the sample base rate (in which case $MSE = Var(y)$). This means that any discriminatory power in the subjective probabilities is more than offset by their variance, which also includes noise.

Panel A of Table 1 shows the decomposition of the MSE using eq.(2). The variance in nursing home admission (0.10) is the largest contributor to inaccuracy in predictions of this outcome, followed by noise in the subjective probabilities (0.05), which accounts for more than 30% of

[12] To be precise, an absolute prediction error of 0.3728 (= |0.6272 – 1| = 0.3728 – 0) for all respondents would give the estimated MSE = 0.3728^2 = 0.139.

the MSE. This implies 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 that those beliefs cannot be expressed accurately in a probability. The square of the bias – the difference of almost 5 pp between the mean subjective probability and the sample base rate – contributes very little to inaccuracy. The covariance of the subjective probabilities with the outcome reduces the MSE (inaccuracy) by only about 11.5% of what it would have been if the subjective probabilities had no discriminatory power.[13]

Table 1. Decomposition of risk perception inaccuracy and discrimination

		Estimate	SE
A. MSE	$\frac{1}{n}\sum(\boldsymbol{p_i} - \boldsymbol{y_i})^2$	**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	$\Delta \boldsymbol{p} = \bar{\boldsymbol{p}}_{\mathbf{1}} - \bar{\boldsymbol{p}}_{\mathbf{0}}$	**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 MSE of subjective probabilities of moving to nursing home within 5 years. o*utcome variance* is $Var(y)$. *covariance* is shorthand for $2Cov(p, y)$. *noise* is $Var(p) - signal$. 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 models (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.

[13] (0.018/(0.139+0.018))×100 = 11.465.

Panel B shows the eq.(5) decomposition of the discrimination slope Δp – the difference between the mean subjective probabilities of those who do and do not move to a nursing home – into: the predictability of the outcome from the jointly observed risk factors ($\Delta\hat{y}$); the shortfall in the utilization of this shared information due to inappropriate weighting of those risk factors ($\Delta\hat{y} - \Delta\hat{p}$); and private information that is not (linearly) correlated with the jointly observed risk factors ($\Delta\hat{\varepsilon}$). Those who move to a nursing home report, on average, a probability that is 8.6 percentage points higher than the mean probability reported by those who do not move to a nursing home (Δp = 0.086). Less than two-thirds (63%) of this discriminatory power is gleaned from shared information ($\Delta\hat{p}$ = 0.054, SE = 0.005), with the rest deriving from use of private information ($\Delta\hat{\varepsilon} = 0.032$). There is far from full utilization of that shared information ($\Delta\hat{y} - \Delta\hat{p} = 0.093$). If people were to predict risks using OLS weights on the jointly observed risk factors, then there would have been a 14.7 pp difference between the mean subjective probabilities of those who do and do not move to a nursing home ($\Delta\hat{y} = 0.147$). Around 63% of this potential discriminatory power remains unused due to inappropriate weighting ($100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$).

Table 2 decomposes the terms of eq.(5) further into contributions of specific sets of risk factors to: a) the discrimination slope that potentially could be achieved using estimates of the optimal weights ($\Delta\hat{y}$), i.e., the predictability of the outcome from those risk factors; b) the discrimination slope that is actually achieved with estimated weights implicit in formation of the subjective probabilities ($\Delta\hat{p}$); and c) the shortfall of b) from a) due to inappropriate weighting of the jointly observed risk factors, i.e., due to not fully realizing that predictability ($\Delta\hat{y} - \Delta\hat{p}$). 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. This means there is 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 of this shortfall comes from underestimation of the age-related risk, particularly above the age of 85 (Table A4).

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.

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.

Decomposition of the discrimination slope of the subjective probabilities into outcome predictability, inappropriate weighting of risk factors, and private information is robust to changes to the estimation sample and model specifications, and to using random forest regression, rather than OLS, to predict the subjective probabilities and the outcome (Appendix B, Table B1).

Risk perception inaccuracy (MSE) 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 B2). 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 MSE decomposition continue to hold. The subjective probabilities are noisy but also have discriminatory power that comes from the use of shared information in the jointly observed risk factors more than from private information. However, as for the main analysis, we observe far from full utilization of the shared information – at least half of its discriminatory power remains unused due to incorrect weighting of the risk factors.

Excluding from the sample respondents who give a focal response of 0.5 to the subjective probability question, which may indicate that they simply not know the risk, reduces the MSE (Table B2). This is because the part of the sample that gives a 0.5 probability has a MSE of 0.25, as our outcome is binary, which is larger than the MSE of the remaining sample. Noise falls and squared bias becomes smaller than 0.001. On the other hand, the fraction of the discriminatory power that comes from use of shared information also falls ($\Delta\hat{p}/\Delta p$, from 63% to 51%) and inappropriate weighting rises ($(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$, from 63% to 72%), which suggests that not all those giving a 0.5 response may be expressing epistemic uncertainty. The main patterns observed in the decomposition for the full sample are nevertheless also robust to this change.

4.2 Heterogeneity in risk perception inaccuracy

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 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.[14] Coefficients correspond to shifts in the (conditional) MSE from that of the respective reference category. 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-population given that it has the least protection against the risk through Medicaid. Risk perceptions in this part of the population are potentially more consequential for private LTCI demand.

Regressing the squared errors of the subjective probabilities on wealth, education, and cognitive functioning simultaneously (column 4), reveals that the MSE 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 MSE difference of 8.7 points between the bottom and top quartile groups is substantial compared with an overall MSE of 14 points.

[14] See Appendix Table A5 for estimation results without controls.

Table 3. Heterogeneity in risk perception inaccuracy (MSE)

	(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 error of the subjective probability of moving to a nursing home within five 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 MSE of the reference groups are 0.122, 0.113, and 0.076, for wealth, education, and cognitive functioning, respectively.

Table 4 shows the decompositions of the MSE and the discrimination slope of the subjective probabilities – eqns. (2) and (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 tendency of the less cognitively able to pay more 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.

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

		Quartile group of total cognition score			
		Bottom	2nd Bottom	2nd Top	Top
A. MSE	$\boldsymbol{\frac{1}{n}\sum(p_i - y_i)^2}$	**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	$(\bar{p} - \bar{y})^2$	< 0.000	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	< 0.000	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	$\boldsymbol{\Delta p = \bar{p}_1 - \bar{p}_0}$	**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. Models (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.

The top row of Panel B 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 weigh 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 information from the risk factors and leaves unused only 5% of their discrimination potential. This explains the higher discrimination of subjective probabilities for the highest cognition groups, in spite of their lower predictability from the jointly observed risk factors.

Higher cognitive functioning is not only associated with better use of shared information contained in the jointly observed risk factors but also with greater use of 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 either have less additional information to call on to form expectations, or they are less able to use it.

4.3 Risk perceptions and insurance

Panel A of Table 5 gives OLS estimates of the (partial) association of holding private LTCI 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 10.7 pp increase in the likelihood of having LTCI. This is a 69% increase on the proportion with LTCI (0.154). The association diminishes only slightly when the jointly observed risk factors are added as controls. The continued positive and significant association is consistent with selection into insurance partly on the basis of private information that is used in formation of the subjective probabilities. The partial association remains stable in magnitude and statistical significance after controlling further for preferences for LTCI. This robustness is consistent with the partial association not being fully attributable to correlated unobservables. Following Oster (2019) in assuming that selection on unobservables is of the same magnitude as selection on observables and the maximum R-squared from a regression that included the unobservables would be 1.3 times the R-squared achieved with all the observed controls, we get a bias-adjusted estimated coefficient of 0.097, which is only marginally less than the estimate with all controls (see Panel A of Appendix Table A7, which also shows in Panel B that effects of unobservable confounders must be large to eliminate the relation between LTCI and the subjective probability of moving to a nursing home, which seems unlikely, given the large set of controls we already include).

While it appears from the results above that the estimated coefficient of the subjective probability is reasonably robust to correcting for omitted variable bias, it could still be biased by reverse causality. In panel B, we regress an indicator of holding LTCI in 2014 on the subjective probability of moving to a nursing home reported in 2012. The estimate from the bivariate regression in column (1) is the same as the respective contemporaneous association estimate in panel A. However, those holding LTCI in 2014 may also have been covered in 2012, which may in turn have influenced their perceptions of the likelihood that they would move to a nursing home. The estimated coefficient falls in size and becomes statistically insignificant when we either control for the lagged dependent variable (column 2) or restrict the sample to those without LTCI in 2012 (column 3).

Table 5. (Partial) Association of LTCI with subjective probability of moving to nursing home

	Sample	(1)	(2)	(3)
A. LTCI in year *t* (mean=0.154)	Aged 65+ in *t*=2012			
Sub. prob. nursing home ≤ *t*+5 years		0.107	0.103	0.100
		(0.021)	(0.021)	(0.021)
Control for risk factors			Yes	Yes
Other controls				Yes
R^2		0.005	0.074	0.091
n		5,814	5,814	5,814
B. LTCI in year *t*+2 (mean=0.160)	Aged 65+ in *t*=2012			
Sub. prob. nursing home ≤ *t*+5 years		0.107	0.013	-0.009
		(0.023)	(0.013)	(0.012)
Control for LTCI at *t*			Yes	
Restrict to without LTCI at *t*				Yes
R^2		0.004	0.665	0.000
n		5,473	5,473	4,610
C. LTCI in year *t*+2 (mean=0.075)	Aged 40-64 in *t*=1996-2016			
Sub. prob. nursing home ever after *t*		0.041	0.029	0.014
		(0.008)	(0.007)	(0.006)
Year fixed effects		Yes	Yes	Yes
Control for LTCI at *t*			Yes	
Restrict to without LTCI at *t*				Yes
R^2		0.003	0.153	0.002
n		15,181	15,181	14,033

Notes: All panels show OLS estimates of the coefficient on the subjective probability of moving to a nursing home in a linear probability model of having private LTCI. In panel A, both the dependent variable and the subjective probability are measured in wave 11 (2012) and the latter is the probability of moving to a nursing home within 5 years of that wave. The sample has full item response to the full set of controls. *Control for risk factors* refers to the inclusion of all the risk factors in Table A1, those contained in $\boldsymbol{X}$ in eq.(6). *Other controls* are those contained in $\boldsymbol{Z}$ in eq.(6), which include those in Table A3 (except for the number of children) and interactions of sex and age groups with number of ADLs/IADLs and the cognition score. In panel B, the dependent variable is having private LTCI in wave 12 (2014), while the subjective probability remains that reported in wave 11. In panel C, we use the respondent's subjective probability of ever moving to a nursing home in their lifetime. This is reported in only one wave. The dependent variable is having private LTCI in the subsequent wave. In this panel, the sample includes those aged 40-64, while the samples used in panels A and B include those aged 65+. In panels B and C, column (2) controls for LTCI cover in the wave prior to that used to measure the dependent variable and column (3) restricts the samples to those without LTCI in the previous wave. No control for risk factors and other controls in panels B and C. Robust standard errors are shown in parentheses.

The disadvantage of the strategies used to obtain the estimates in columns (2) and (3) of Panel B is that they leave little variation in LTCI to be potentially explained by the subjective probabilities. This is because first enrolment in LTCI tends to occur before the age of 65 and insurance status does not change much after that.[15] Panel C tackles this by using a sample aged 40-64 to estimate the (partial) association between the subjective probability of ever moving to a nursing home and holding LTCI in the wave after this probability is reported. The bivariate association is substantially smaller than the estimates that are potentially biased by reverse causality and are obtained from older samples (Panel A). Controlling for the lagged dependent variable or restricting the sample to those without LTCI in the previous wave reduces the magnitude of the estimate further. However, unlike for the older sample for which this strategy is less informative, after taking all of these steps to reduce the potential for reverse causality in the younger sample, LTCI remains positively and significantly associated with the subjective probability of moving to a nursing home (columns (2) and (3) of Panel C).

Table 6 gives IV estimates of the effect of the subjective probability of moving to a nursing home within five years on the likelihood of holding LTCI. The first stage and reduced form estimates show that the instrument – the respondent and their spouse's number of in-contact children – significantly reduces the reported subjective probability of moving to a nursing home and the objective probability of having private LTCI. These estimates are consistent with people with more children perceiving a lower risk of needing to move to a nursing home and so having a lower demand for insurance. The effective first stage *F*-statistic (22.96) is very slightly below the critical value (23.11) at 5% significance with bias exceeding 10% of the "worst case" bias (Montiel Olea & Pfleuger, 2013). This indicates that the null of a weak instrument is not rejected using a robust test. For this reason, and because *t*-ratio based

[15] Median age at which we observe the first occurrence of LTCI is 63, mean is 64. These estimates are likely upward biased, since some individuals report to have LTCI through all HRS waves, which means they could have purchased it before we observe them. Of those aged 65 and over, only 5% report to switch insurance status in the next two years, versus 10% for those aged 40-64.

inference can be underpowered to detect a null effect even with a large F-statistic (Keane & Neal, 2023; Lee et al., 2022), we use weak-instrument inference.[16]

Table 6. Effect of subjective probability of moving to nursing home on LTCI

	OLS (1)	IV (2)	First stage (3)	Reduced Form (4)
Sub. prob. nursing home within 5 years	0.103 (0.021)	0.680 [0.106, 1.509]		
Number of children			-0.007 (0.001)	-0.005 (0.002)
Controls	Yes	Yes	Yes	Yes
F-statistic			22.96	5.27
n	5,705	5,705	5,705	5,705

Notes: Dependent variable in columns (1), (2) & (4) is holding private LTCI. Subjective probability of moving to a nursing home within five years is instrumented with the number of alive and in-contact children of the respondent and their spouse. All data are from wave 11 (2012). Controls are those contained in ***X*** in eq.(6), the risk factors in Table A1, and variables contained in ***Z*** in eq.(6), which include those in Table A3 (except for the number of children) and interactions of sex and age groups with number of ADLs/IADLs and the cognition score. Sample restricted to observations with full item response on LTCI, subjective probability, number of children, and controls. Robust standard errors in parentheses. In brackets is 95% confidence interval calculated using weak-instrument robust inference. F-statistics are Montiel Orea & Pfleuger (2013) effective first-stage F-stat and Anderson & Rubin (1949) weak-instrument robust test for the reduced form.

Subject to validity of the exclusion restriction on the instrument in eq.(6) and given that the instrument is significant in the first stage, its significance in the reduced form (p=0.022) implies rejection (at the same level of significance) of the null that the subjective probability has no effect on LTCI ($H_0: \gamma = 0$) (Keane & Neal, 2023). This Anderson-Rubin (1949) test has the correct size irrespective of the strength of the instrument (Keane & Neale, 2023). The weak-instrument robust 95% confidence interval for the IV estimate of γ is wide, but does not contain zero. The IV point estimate is substantially larger than the OLS estimate, which is inconsistent with reverse causality or omitted variables upwardly biasing the latter. We argue that the IV interval estimate gives reasonable grounds to believe that risk perceptions,

[16] On the other hand, Angrist & Kolesár (2023) find that *t*-ratio based inference in the case of just-identified IV models is usually reliable, as endogeneity is typically not sufficiently severe to result in misleading *t*-tests. Using standard robust inference, our IV estimate is significant at the 5% level: $\hat{\gamma}_{IV}/_{robust\ SE}$= 0.680/0.322 = 2.11 > 1.96.

measured by subjective probabilities of moving to a nursing home, do influence the decision to insure. We would have less confidence in a claim that the IV point estimate gives a reasonable estimate of the magnitude of that effect. But estimation of this magnitude is not our objective. As discussed above, our aim with these various strategies is to weigh the evidence that risk perceptions influence insurance behavior. The evidence shown supports this and so points to risk perceptions being consequential.

5. Discussion

Misperception of LTC risk can distort saving and insurance decisions with important consequences for well-being in old age. We find that older Americans, on average, tend to overestimate their risks of moving to a nursing home, and their risk perceptions are inaccurate. Many make large mistakes. In part, this is because they underutilize information in risk factors that they are obliged to share with insurers on application for LTCI. Subjective probabilities encapsulate only 37% of the potential that these risk factors have to discriminate between those who do and those who do not move to a nursing home. We do not present evidence on the extent to which insurers use this shared information. It seems nevertheless safe to assume that the experience and statistical knowledge of their underwriters allow them to do much better than insurance applicants. Those with a risk perception that is insufficiently sensitive to shared information may underestimate their risk and so decline insurance offered at a price that is actuarially fair for that risk (Baillon et al. 2022). This potential for *behavioral selection* will materialize if reported risk perceptions influence insurance decisions. Consistent with this scenario, we find an association between LTCI and subjective probabilities of moving to a nursing home that is robust to extensive controls, to using lags to deal with reverse causality, as well as instrumenting subjective probabilities with number of children that, as the main providers of informal care, reduce the perceived risk of needing nursing home care.

Inappropriate weighting of the jointly observed risk factors could stem from unawareness of the relevance of this shared information for LTC risk or from an inability to process this information into a subjective probability. We find that age is the most underestimated risk factor, particularly for the least cognitively able. Since a majority of older people continue to know their ages and the strong correlation of age with nursing home admission is evident from casual observation, it appears that a substantial part of the underutilization of shared information is due to inability to process that information. The upside of the discovery that inaccurate LTC risk perceptions are partly due to underestimation of age-related risk is that this source of inaccuracy may be less consequential for insurance decisions. Most LTCI is purchased before people reach the old ages at which the upward revision of the subjective probability of moving to a nursing home fails to keep pace with the rising objective probability. However, if there is underappreciation of the rate at which LTC risk will rise in old age, then this error could still contribute to low take-up of LTCI in middle-age.

Our finding that subjective probabilities of moving to a nursing home predict that outcome even after conditioning on a large battery of risk factors confirms earlier evidence of private information on LTC risks (Finkelstein & McGarry, 2006; Hendren, 2013). We go beyond the detection of private information by also quantifying 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, effectively, be better informed because of their advantage in the processing and utilization of that information. No doubt, some insurance applicants can use private information on their personal risks to detect and select contracts that are priced below their expected LTC costs. But our estimates suggest that there are likely 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 share with the insurer.

Given imperfections in the US LTCI market (Ameriks et al. 2018), regulation to limit the scope for behavioral selection arising from asymmetric utilization of shared information need not be welfare improving (Handel, 2013). The experience of removing information frictions in the health insurance market suggests that welfare consequences depend on microfoundations in a particular market (Handel et al., 2019). While the challenge of designing effective information interventions that make LTC risk perceptions more accurate is worth pursuing, success will unlikely 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), and crowd-out by public insurance (Braun et al., 2019; Brown & Finkelstein, 2011; Lambregts & Schut, 2020b).

We find that LTC risk perceptions are much less accurate at lower levels of cognitive functioning. The less cognitively able face a more difficult prediction task because their higher risk increases the variance of the prediction target. The cognition gradient in accuracy is however not merely mechanistic. 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. This suggests that there may be limited scope to improve risk perception accuracy through informing people of risk factors.

Many lack the cognitive ability to process this information. On the other hand, low cognitive functioning is strongly correlated with old age. There seems therefore to be greater potential to improve risk perceptions that would be more consequential for LTCI decisions in middle-age and early old-age, when cognition is less of a constraint and when those decisions are mainly taken.

As with all analyses of data on reported subjective probabilities, we cannot ensure that they correspond to true beliefs. Measured inaccuracies could reflect reporting error arising from the difficulty of expressing beliefs in probability formats that many people experience (Gigerenzer & Hoffrage, 1995). Indeed, we find that the least cognitively able report the noisiest probabilities. Measurement error may manifest through extreme rounding of reported probabilities and use of focal responses, such as 0.5. Modelling of this reporting behavior tends to suggest that it only modestly biases probabilistic beliefs (Basset & Lumsdaine, 2001; Giustinelli et al., 2022; Kleinjans & van Soest, 2014; Manski & Molinari, 2010) and their measured associations with observed variables (Kleinjans & van Soest, 2014). Our main findings are robust to dropping respondents who report a probability of 0.5.

We show that older Americans have inaccurate perceptions of LTC risks partly because they underutilize information on risk factors that they would be obliged to share with their desired insurers, and that the resulting inaccuracy is only partially offset by private information. This potentially has consequences for behavioral selection and the operation of the LTCI market. Our empirical analyses reveal that the underutilization of shared information is quantitatively important and suggests that (theoretical) analyses of such consequences would be worthwhile.

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), 2145–84.

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.

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.

American Council on Aging. (2021, December 14). *Spending Down Assets to Become Medicaid Eligible for Nursing Home / Long Term Care.* Retrieved January 10, 2022, from https://www.medicaidplanningassistance.org/medicaid-spend-down/

Anderson, T.W. & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 46-63.

Angrist, J., & Kolesár, M. (2023). One instrument to rule them all: The bias and coverage of just-id iv. *Journal of Econometrics*.

Bago d'Uva, T., O'Donnell, O., & van Doorslaer, E. (2020). Who can predict their own demise? Heterogeneity in the accuracy and value of longevity expectations. *The Journal of the Economics of Ageing*, *17*, 100135.

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, 43-87.

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

Bhargava, S., Loewenstein, G. & Sydnor, J.R. (2017). Choose to lose: Health plan choices from a menu with dominated option. *Quarterly Journal of Economics* 132(3), 1319-72.

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), 134-69.

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), 119-42.

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.

Charles, K.K. & Sevak, P. (2005). Can family caregiving substitute for nursing home care? *Journal of Health Economics* 24(6), 1174–90.

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. & Kohler, H-P. (2016). HIV/AIDS-related expectations and risky sexual behaviour in Malawi. *The Review of Economic Studies*, 83(1), 118–164.

Delavande, A., Del Bono, E., Holford, A., Sen, S., Lesic, V. (2022). Expectations about the productivity of effort and academic outcomes: Evidence from a randomized information intervention. *Working paper*: https://drive.google.com/file/d/1CXYu67DR9opFmoQJTdZDmliYKotS9b3Q/view

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

de Paula, Á., Gil, S., & Todd, P. E. (2014). How beliefs about HIV status affect risky behaviors: evidence from Malawi. *Journal of Applied Econometrics*, 29, 944–964.

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.

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

Giustinelli, P., Manski, C. F., & Molinari, F. (2022). Tail and center rounding of probabilistic expectations in the health and retirement study. *Journal of Econometrics*, *231*(1), 265-281.

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. (2022). The social determinants of choice quality: Evidence from health insurance in the Netherlands. *Working paper*: https://static1.squarespace.com/static/5ee3119aa4c9ed2dd490b6ff/t/614383f10085c57a49ff5d6c/1631814644287/HKMS_draft_web_Sep2021.pdf

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*, 2017, 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.

Hurwitz, A., & Mitchell, O. S. (2022). *Financial Regret at Older Ages and Longevity Awareness* (No. w30696). National Bureau of Economic Research.

Keane, M., & Neal, T. (2023). Instrument strength in IV estimation and inference: A guide to theory and practice. *Journal of Econometrics*, *235*(2), 1625-1653.

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. (2020a). 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 Ageing*, *17*, 100236.

Lambregts, T. R., & Schut, F. T. (2020b). Who can see it coming? Demand-side selection in long-term care insurance related to decision-making abilities. *Working paper:* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4052261

Lee, D. S., McCrary, J., Moreira, M. J., & Porter, J. (2022). Valid t-ratio Inference for IV. *American Economic Review*, *112*(10), 3260-90.

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.

Lusardi, A. & Mitchell, O. (2007). Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education. *Business Economics*, 42, 35-44. 10.2145/20070104

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

Medicare. (n.d.). *Medicare Part A coverage – nursing home care.* Retrieved January 10, 2022, from https://www.medicare.gov/what-medicare-covers/what-part-a-covers/medicare-part-a-coverage-nursing-home-care

Medicaid. (n.d.). *Nursing Facilities.* Retrieved January 10, 2022, from https://www.medicaid.gov/medicaid/long-term-services-supports/institutional-long-term-care/nursing-facilities/index.html

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.

Montiel Olea, J.L. & Pflueger, C.E. (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics* 31, 358-369.

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.

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

Taylor Jr, D. H., 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.

Van Houtven CH, Norton EC. (2004). Informal care and health care use of older adults. *Journal of Health Economics* 23(6), 1159–80.

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

Zhou-Richter, T., Browne, M. J., & Gründl, H. (2010). Don't they care? or, are they just unaware? risk perception and the demand for long-term care insurance. *Journal of Risk and Insurance*, *77*(4), 715-747.

Appendices

Appendix A. Additional Tables and Figures

Table A1. Nursing home risk factors and means for samples used to estimate models (3) and (4)

Variable	Definition	Mean Sample Model (3)	Mean Sample 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 your control during 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 (any)		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. Score, which is increasing in cognitive functioning, sums 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.	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: In models, age is entered as indicators for 5-year age groups up to ≥ 85 years. 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. Sample for model (4) is corresponding sample including wave 8 respondents. CES-D8 is the Center for Epidemiologic Studies Depression (CESD) scale. See HRS codebook 2012: https://hrs.isr.umich.edu/sites/default/files/meta/2012/core/codebook/h12_00.html
and RAND: https://hrsdata.isr.umich.edu/sites/default/files/documentation/other/1680723673/randhrs1992_2020v1.pdf 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*
Item response on all risk factors	5,987
Item response also on education	5,986

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 to estimate models (6) and (7) that are not used to estimate models (3) and (4)

Variable	**Definition**	**Mean (SD)**
Private LTCI	1 if have private long-term care insurance, 0 otherwise	0.154
Cognition score	Total cognition score (0-35), increasing in cognitive functioning. Derived from word recall and mental status summary scores (see also Table A1).	21.9 (4.88)
Education		
Less than high school	1 if highest level of education is below high school (no high school diploma) or a GED, 0 otherwise	0.226
High school graduate	1 if highest level of education is a high school diploma or if have 12 years of education and no college degree, 0 otherwise	0.321
Some college	1 if have a school diploma or GED and have more than 12 years of education or another degree below bachelor's, 0 otherwise	0.227
College graduate	1 if bachelor's degree or higher, 0 otherwise	0.226
Seatbelt use	1 if 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.	0.738
Number of children	Alive and in-contact children of the respondent and their spouse	3.40 (2.17)

Notes: n = 5,814, as in model (6), except for the number of children, where n = 5,705, as in model (7). See also RAND codebook: https://hrsdata.isr.umich.edu/sites/default/files/documentation/other/1680723673/randhrs1992_2020v1.pdf.

Table A4. OLS estimates of models for subjective probability of moving to a nursing home within 5 years (p) and indicator of actually moving to a nursing home within 5 years (y)

	Model (3) of p		Model (4) of y	
	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-squared	0.060		0.152	
Mean dependent variable	0.165		0.106	

n	5,987	6,849

Notes. Model (3) estimated using HRS wave 11 data and sample that includes 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. Model (4) is estimated with a corresponding sample observed in wave 8 (2006). The dependent variable in this model is an indicator of having moved to a nursing home for at least 21 consecutive nights or until death within five years of wave 8 interview. The covariates for this model are reported/measured in wave 8. Robust standard errors in parentheses.

Table A5. Heterogeneity in risk perception inaccuracy (MSE) 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 error of the subjective probability of moving to a nursing home within five years $((p_i - y_i)^2)$ on indicators of each of household wealth quartile group, educational attainment, total cognition score quartile group, respectively. 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 $\Delta\hat{y} - \Delta\hat{p}$	**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 $\sum_{j\in\Omega}(\hat{\beta}_j^y - \hat{\beta}_j^p)\Delta X_j$. See Table A1 for the risk factors included in each set. Because we stratify by cognition, we don't include our cognitively impaired dummy in our regressions. Bootstrap standard errors (100 replications) in parentheses.

Table A7. Oster bounds for subjective probability of moving to nursing home within 5 years, $\hat{\beta}$, and $\tilde{\delta}$

	(1) $R_{max} = 1.3\ R^2$	(2) $R_{max} = 2\ R^2$	(3) $R_{max} = 3\ R^2$
Panel A			
Bias adjusted $\hat{\beta}$ when $\delta = 1$	0.097	0.090	0.079
Panel B			
$\tilde{\delta}$ for $\hat{\beta} = 0$	18.48	6.03	3.08

Notes: R_{max} is the maximum R-squared, the R^2 value that corresponds to an OLS model of LTCI on the subjective probability of moving to nursing home within 5 years with full controls included, see eq.(6) and Table 5, panel A, column 3. δ is the relative degree of selection of observables and unobservables, which is assumed proportional in Panel A. In Panel B the delta which corresponds to $\hat{\beta} = 0$ is reported for different R_{max} values. The R_{max} value 1.3 R^2 is chosen according to Oster (2019), with 2 R^2 and 3 R^2 representing more conservative values.

Appendix B. Robustness analysis

Table 1, panel B 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 B1 demonstrates robustness to changes in the samples and model specifications used to estimate models (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 use estimates of model (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. If, instead, we use the nursing home indicator and risk factors for the wave 11 sample, then outcome predictability increases, as it should since predictions are then 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) and so the fraction of the risk factors' potential discriminatory power that is unrealized because of their inappropriate weighting in formation of the subjective probabilities ($(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$) rises by less than 2 pp (Table B1, column (2)).

Alternative specifications. To obtain the main estimates, we do not include interactions between risk factors in models (3) and (4). This makes the detailed decomposition presented in Table 2 feasible. Allowing interactions between sex and age groups and each of the number of ADLs/IADLs and an indicator of cognitive impairment, as in Finkelstein & McGarry (2006), only slightly increases outcome predictability and has even smaller impacts on the magnitudes of inappropriate weighting and private information (Table B1, column (3)). To be consistent with the eligibility criteria for Medicaid coverage of nursing home expenses, we exclude housing wealth from the measure of household wealth. Including housing wealth, as in Finkelstein and McGarry (2006), has a negligible impact on each component of the decomposition (Table B1, column (4)). Following Finkelstein & McGarry (2006), we use an indicator of cognitive impairment in models (3) and (4). 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 quartile of the total cognition score, which is increasing in cognitive ability. Outcome predictability increases slightly and there is a negligible impact on the other results (Table B1, column (5)).

Alternative estimator. The linear models (3) and (4) facilitate the detailed decomposition given in Table 2. OLS estimation of model (4) parameters gives a set of risk factor weights that minimize the MSE of the outcome predictions. These provide an appropriate benchmark against which to evaluate the weights implicit in the subjective probabilities. Notwithstanding these advantages of linear models estimated by OLS, using machine learning methods to allow for extensive nonlinearity 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 modelling the subjective probabilities, with consequences for the inappropriate weighting and private information components of the discrimination slope decomposition (eq.(5)).

For these reasons, we check robustness of the decomposition to using random forest regression to predict from the risk factors the reported subjective probability of moving to a nursing within five years and the realization of that outcome. Since our sample is relatively small compared with many random forest applications, we use 80% of the sample for training each model and 20% for testing, rather than the 50-50 split used with larger samples. We use the mean squared error as the splitting criterion at each internal node, and set the minimum node size to 10 to limit overfitting.

Comparing columns (6) and (1) of Table B1 reveals the surprising result that random forest regression actually performs slightly worse than OLS in discriminating between those who move to a nursing home and those who do not. The discrimination slope of the random forest (RF) outcome predictions is smaller than that of the OLS predictions: ($\Delta\hat{y}^{RF} < \Delta\hat{y}^{OLS}$). 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 are 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 ($\Delta\hat{p}^{RF} < \Delta\hat{p}^{OLS}$). While this may also seem a surprising result, it can also be explained. The random forest regression does predict the subjective probabilities more accurately from the risk factors: ${}^{1}\!/_{n} \sum (p_i - \hat{p}_i^{RF})^2 = 0.036 < {}^{1}\!/_{n} \sum \left(p_i - \hat{p}_i^{OLS}\right)^2 = 0.048$. 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. The random forest is better at modelling the mistakes made in forming subjective probabilities – variation in those probabilities than is not correlated with nursing home admission.

Table B1. Robustness of decomposition of discrimination slope of subjective probabilities to sample, model specification, and estimator

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

Notes: All columns give the eq.(5) decomposition of $\Delta\boldsymbol{p} = \overline{\boldsymbol{p}}_{\boldsymbol{1}} - \overline{\boldsymbol{p}}_{\boldsymbol{0}}$. In each case, $\Delta\boldsymbol{p}$ = 0.086 (SE = 0.011). Column (1) gives the baseline estimates from Table 1. In column (2), $\Delta\hat{y}$ is obtained from model (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 model (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. Bootstrap standard errors (100 replications) in parentheses.

Table B2. Robustness of decomposition of risk perception inaccuracy and discrimination to definition of outcome and exclusion of focal point subjective probabilities

		Baseline	Outcome: nursing home for		Drop if sub. prob. = 0.5
			≥ 1 night	≥ 100 nights	
		(1)	(2)	(3)	(4)
A. MSE	$\boldsymbol{\frac{1}{n}\sum(p_i - y_i)^2}$	**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	$\boldsymbol{\Delta p = \bar{p}_1 - \bar{p}_0}$	**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: Table contents as Table 1 in paper. Notes to that table apply. Column (1) gives the baseline estimates given in that table. Columns (2) and (3) vary the length of stay – ≥ 1 night and ≥ 100 nights, respectively – used to define the outcome (move to a nursing home). Baseline using ≥ 21 nights. In the sample, outcome mean 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 five years equal to 0.5.