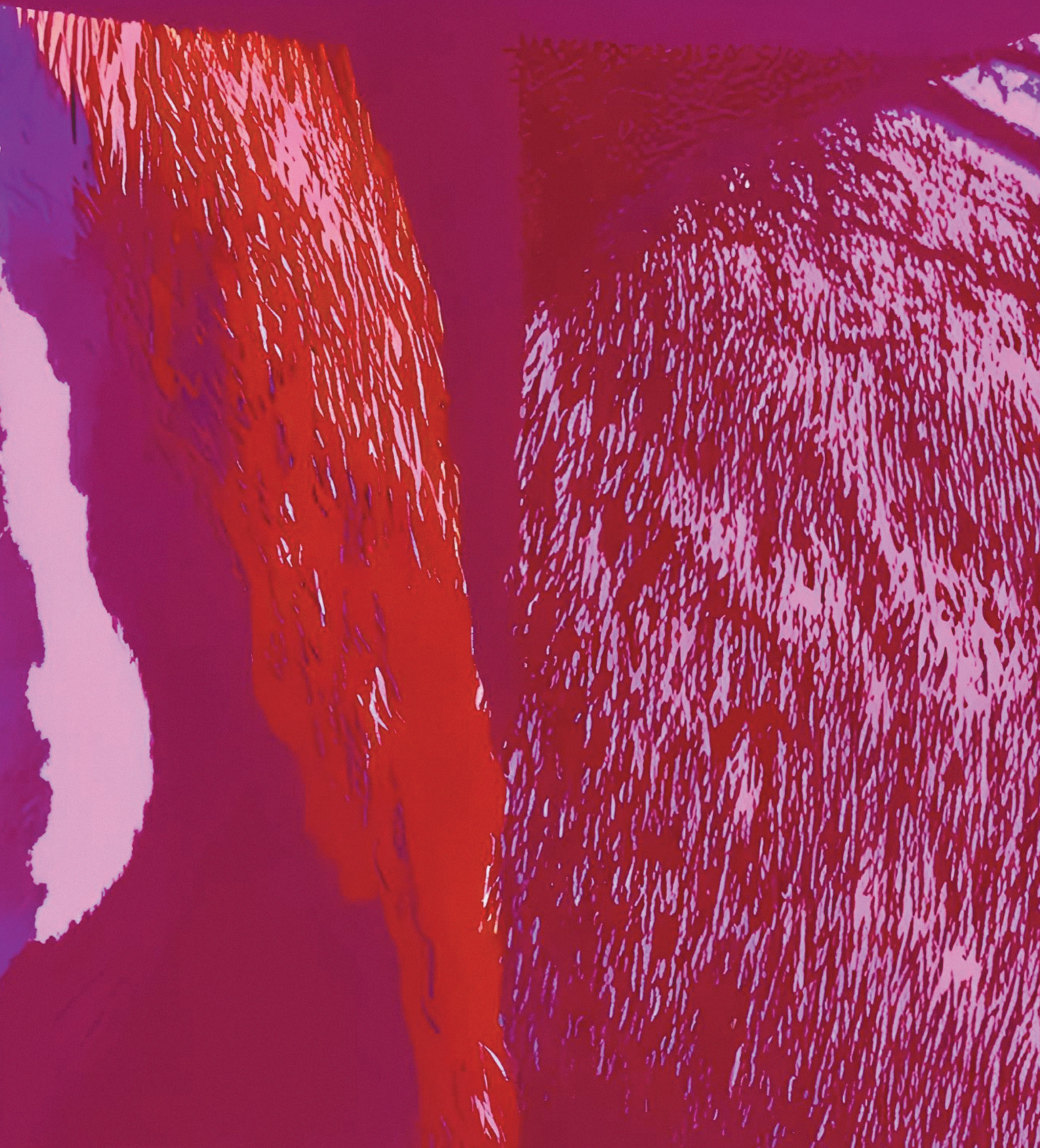
Lisa Voois

Empirical Studies of Health-Related Expectations and Behaviors



Empirical Studies of Health-Related Expectations and Behaviors

Empirical Studies of Health-Related Expectations and Behaviors

Empirische studies over gezondheidsgerelateerde verwachtingen en gedrag

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on
Thursday 3 October 2024 at 15:30 hrs

by

Lisa Voois
born in Dordrecht, the Netherlands.

Erasmus University Rotterdam

Doctoral Committee

Promotor:	Prof.dr. O.A. O'Donnell
Other members:	Prof.dr. J.C. van Ours Prof.dr. R.L. Lumsdaine Prof.dr. M. Salm
Co-promotor:	Dr. T.M. Marreiros Bago d'Uva

Tinbergen Institute, Erasmus School of Economics dissertation series number: 848
ISBN 978-90-5892-705-7

Design: PanArt, www.panart.nl
Print: OBT bv, www.obt.eu
Cover Artwork: AI, Ron & Ola

This publication (cover and interior) is printed on certified FSC® paper PlanoSuperior.

FSC www.fsc.org
MIX
Papier | Ondersteunt verantwoord bosbeheer
FSC® C022586

Mijn oma heeft de lagere school gevolgd
Mijn moeder ging naar de huishoudschool
Ik heb de mogelijkheid gehad om naar de universiteit te gaan
Pap en Mam, bedankt!

Acknowledgments

This thesis would not exist today without the help, support and encouragement of many people.

Teresa, when we first met back in 2016 while I was a teaching assistant for the Methods & Techniques course, I never expected that this could lead to your supervision of my bachelor's and master's theses, and ultimately, to this PhD dissertation. I want to thank you for inspiring and encouraging me to pursue a PhD and for all your help and support in getting me there. I've learned so much from you during the PhD, from developing and progressing with research to navigating the academic world. Thank you for the countless open discussions over the years, whether related to research or not, and I'm very grateful for your continued mentorship.

Owen, your sharp and insightful feedback has been invaluable during my PhD. Thank you for your quick responses and even quicker thinking, for showing me how to structure and write (and rewrite) papers, and for your ability to be practical and provide structure and guidance when I lost sight of the bigger picture. I think you and Teresa make a great supervisory team, and I've enjoyed and learned much from our discussions over the years. I look forward to many more of those, especially as we continue to refine the retirement paper.

To the entire Health Economics group at ESE, both those who were there when I first started and those who joined along the way, thank you. I feel very lucky to have been part of such a friendly and approachable group of great researchers. Your generous feedback on my work has been a tremendous help.

I'd also like to thank my PhD colleagues. Vahid, thank you for being my constant throughout covid, for listening to my (endless) venting and for always discussing research and sharing your valuable insights with me. I'm very grateful that I got to share an office with you, and I hope we can bring some of our research ideas to fruition. Elisa, in het jaar dat jij begon was er maar één PhD vacature in onze groep en ik ben heel blij dat jij hier toen voor bent aangenomen. Bedankt voor je oprechte interesse, je openhartigheid in al onze gesprekken en voor alle gezelligheid op kantoor, tijdens conferenties en daarbuiten. Ik vind het heel fijn dat jij mijn paranimf bent! Joaquim, thank you for being a bright spot in the N-building dungeon. Rita, I really enjoyed sharing an office with you. Thanks for all the fun, laughter and good advice. Dilnoza, thank you for your kindness and all the sincere conversations we've had. Francisca, from following TI courses together to being colleagues at ESHPM. Thank you for your warmth, enthusiasm and genuine interest in all aspects of my life. Merel, bedankt voor jouw vriendschap sinds de bachelor, alle leuke tripjes en alle eettentjes die ik samen met jou heb ontdekt. Ilse, ik vond het erg gezellig om in mijn laatste jaar jouw bureaugenoot te zijn. Bedankt voor alle klets- en klaagsessies. Gianluca, thanks for your warmth and endless positivity. David, Francesco and Andrea, thank you for popping into my office to check in and provide (much needed) distractions from work. Lieke, ik vind het erg leuk dat wij elkaar hebben leren kennen tijdens de PhD, bedankt voor alle gedachtewisselingen over onderzoek en de academische wereld en voor de geweldige tijd in Oslo. To all the other PhDs and colleagues, whether at ESE, ESHPM or those I've met at conferences, which I've not mentioned by name (since there are simply too many of you), thank you. A PhD can be quite lonely, but it still takes a village to finish one. Thank you for all the (silly) conversations, good times, feedback, practical support, and discussions about research and life.

I'd also like to thank my committee members, Jan, Martin, Robin, Carlos and Marieke for your involvement in the ceremony, for carefully reading my dissertation and for all your questions and suggestions, which will help greatly in improving my work.

Marco, bedankt voor het vertrouwen in mij en de vrijheid die ik heb gekregen om mijn PhD af te ronden. Aan al mijn (oud) collega's bij HSI, bedankt voor het warme welkom, de prettige en laagdrempelige sfeer en ik kijk er naar uit om mijn academische carrière bij jullie voort te zetten.

Ik wil ook graag mijn lieve vrienden en familie bedanken. San & Am, jullie hebben mijn hoogte- en dieptepunten de afgelopen jaren van dichtbij meegemaakt. Bedankt voor al jullie steun, vertrouwen, de broodnodige afleidingen, begrip als ik minder aanwezig was dan ik had willen zijn en voor het mij eraan herinneren dat er een zeer mooi en leuk leven bestaat naast werk. Ik waardeer onze hechte vriendschap enorm en ben zo blij dat wij samen delen en meemaken hoe onze levens steeds rijker worden. San, bedankt dat jij mijn paranimf bent. Ik had me dit echt niet anders kunnen voorstellen!

In & Mir, more than 10 years ago we started our university adventure together. Thank you for making that an unforgettable time. It's definitely one of the reasons why I'm still at the university today. Thank you for all our heart-to-hearts, for listening to my endless complaining and for always cheering me on. Now that my 'labor' is done, let's finally plan our next trip.

Aan al mijn andere lieve vrienden en familie, bedankt voor jullie interesse, oprechte aandacht en alle gezellige afleidingen. Ils, bedankt voor je betrokkenheid, alle leuke spelavonden en al ons gegier en gelach. Mam en Pap, bedankt dat jullie altijd in mij hebben geloofd, naar mij hebben geluisterd en mij hebben ondersteund in alles wat op mijn pad kwam. Voor jullie paraatheid, alle adviezen door de jaren heen en jullie bereidheid om mij met alles te helpen. Steeph, jeetje, waar te beginnen. Jij weet als geen ander dat dit PhD avontuur zeker niet altijd even leuk was. Bedankt voor je eindeloze steun, hulp en liefde. Dat je mij weet te aarden als ik het (weer eens) allemaal even niet zag zitten. Voor je geduld en begrip en dat je mij van de beste beslissingen – Ron & Ola – kunt overtuigen. Ik kan niet wachten om te zien wat de toekomst ons brengt!

Lisa Voois,
July 2024

Contents

Chapter 1

Introduction

Most health-related decisions are risky. They have multiple possible outcomes, with some more likely to occur than others. For example, both smokers and non-smokers face a risk of dying prematurely. But the risk is higher for smokers. This thesis is a collection of empirical studies on health-related expectations and behaviors. Two of the studies examine expectations of important later-life events – retirement and nursing home admission. These expectations may influence saving and insurance decisions, with consequences for well-being in old age and at other points in the life cycle. The other two studies investigate health behaviors, including smoking, drinking and sleeping. These behaviors influence future health and may also impact later economic outcomes, such as education and retirement.

Microeconomic analysis of decision-making assumes that individuals hold preferences over outcomes and beliefs about the likelihoods of those outcomes (Manski, 2004; Manski, 2023). It has been common practice to assume rational expectations that are, by and large, objectively correct (Manski, 2004). Subject to the validity of this assumption, preferences can be identified from choices. But do individuals indeed form expectations that encapsulate all the available information? The seminal work of Daniel Kahneman and Amos Tversky (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) suggested that they may not. It demonstrated that individuals typically do not have complete information and do not consistently act rationally on the basis of the information available. Doubts about the validity of rational expectations motivated the direct measurement of expectations through reported subjective probabilities (Manski, 2004)

Subjective probabilities about a range of future life events are now routinely elicited in various surveys (Bruine de Bruine et al., 2023). Expectation questions typically invite respondents to report the likelihood of a certain event as a percent chance. For example: “What is the percent chance that you will live to be 75 or more?”. Initial studies showed that individuals are able and willing to answer these questions. Answers are rounded, sometimes excessively (Lillard & Willis, 2001; Manski, 2004). But reported subjective probabilities about some future life event do vary with known determinants of that event and have predictive power for it (Hurd, 2009; Manski, 2004). For example, in accordance with objective mortality risks, reported subjective survival probabilities are lower for smokers, higher for those with higher educational attainment and they predict actual survival (Hurd & McGarry, 1995; 2002).

Reported expectations about future health can give insight into why individuals engage in certain health behaviors. For example, individuals greatly underestimate the mortality risk of smoking (Bago d'Uva & O'Donnell, 2022). They may then decide to smoke, while, if they were aware of its objective risks, their optimal decision may have been not to smoke. Together, expectations and health behaviors shape outcomes over the life cycle. They influence current and future health, which are important determinants for many economic outcomes, such as education, retirement and well-being in old age.

In four subsequent chapters, this thesis examines older-age expectations and health behaviors. Chapters 2 and 3 investigate the accuracy of subjective probabilities of two major health- and age-related events - nursing home admission and retirement in the United States. Chapters 4 and 5 examine health behaviors. Chapter 4 examines the extent to which gender differences in health behaviors, most notably smoking and drinking, can explain why men die earlier than women in Russia. Chapter 5 examines the consequences of sleep behavior in adolescence for mental health and education outcomes in the United States.

Chapter 2 explores one potential reason for underinsurance of long-term care risk in the US; namely, the underestimation of this risk due to the underweighting of risk factors that insurers observe. Accuracy of older Americans' perceptions of long-term care risk is measured by comparing their subjective probabilities of moving to a nursing home with realizations of that event. Accuracy is decomposed into the use of information on risk factors that is shared with insurers (shared information) and the use of private information that each individual possesses. There is inaccuracy, and it partly derives from the underweighting of risk factors that insurers observe. This is especially pronounced for the least cognitively able. Such underutilization of shared information may lead individuals to decline insurance offers that they would accept if their risk perceptions were accurate.

Chapter 3 examines retirement expectations that, if inaccurate, can have significant adverse impacts on lifetime welfare by impeding appropriate retirement planning. The analysis measures the accuracy of US workers' reported subjective probabilities of continuing to work full-time past the ages of 62 and 65. There is substantial inaccuracy. The subjective probabilities are approximately as accurate as they would be if all respondents viewed their chances of being retired at 62 (or 65) as being determined by a coin toss. This is largely due to uncertainty surrounding work status at ages 62 and 65, but it is also because US workers pay attention to irrelevant factors when forming their expectations. The less educated hold less accurate retirement expectations because they are worse at extracting information from observed objective predictors of retirement, and they also use much less (private) information that is not captured by those predictors. These individuals may therefore be planning inappropriately for retirement and bearing the consequences of it later in life.

Chapter 4 investigates gender differences in health behaviors in Russia and how these contribute to the gender mortality gap, which is very large. By the early 2000s, life expectancy at birth of Russian women was 13 years higher than that of men. Over the next decade, which saw the implementation of tobacco and alcohol control policies, this gap declined by 2 years. This chapter estimates contributions of health behaviors to the gender gap in 5-year mortality and how these changed between 2000-2003 and 2010-2013. In each of those periods, gender differences in health behaviors – particularly smoking and, to a lesser extent, drinking – explain a large share of the gender mortality gap. The joint contribution of smoking and drinking fell substantially over the decade. These results are consistent with men's declining tobacco and alcohol consumption, which may possibly be due to the control policies implemented, explaining most of the narrowing of the gender mortality gap.

Chapter 5 examines adverse consequences of sleep behavior. It focuses on sleep deprivation during adolescence, which is prevalent and may be particularly harmful to health and productivity later in life. Using longitudinal data from the US, I estimate effects of long-term exposure to later sunset time across the US on the sleep duration of adolescents and on the subsequent risk of depression and on educational attainment by young adulthood. Sustained exposure to later sunset times during adolescence reduces sleep duration and increases the probability of being diagnosed with depression. There is no evidence that it affects high school graduation or college entrance, but it does reduce the probability of eventually graduating from college. These findings are consistent with persistent sleep deprivation in adolescence having adverse effects on mental health and education.

Chapters 2 and 3 are co-authored with my supervisors, Teresa Bago d'Uva and Owen O'Donnell. Chapter 4 is co-authored with Teresa Bago d'Uva. For all three co-authored chapters, I had principal responsibility for defining the research question and designing the empirical strategy, and sole responsibility for the data analysis. I wrote the first draft of each of these chapters, after which further revisions were made by all authors. Chapter 5 is a single-authored chapter, which has benefitted greatly from feedback from my supervisors.

Part 1: Expectations

Chapter 2

Long-term care risk perceptions, information, and insurance

Joint work with Teresa Bago d'Uva and Owen O'Donnell

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.

2.1 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, e.g. by transferring assets to children (Bassett, 2007), 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, 2020). They are also consistent with a socioeconomic gradient in choice quality (Handel et al., 2024).

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 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

---

[1] Behavior has been shown to respond to experimentally manipulated subjective probabilities (Delavande & Kohler, 2016; Delavande et al., 2023). Hurwitz & Mitchell (2022) show that the provision of information on the probability of survival to old age increases regret about not having purchased LTCI.

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, 2022). 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, 2022) 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 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

[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.

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.

2.2 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

[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 6,297 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

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 1924-1947.

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 does not cover custodial care, but it fully reimburses rehabilitative stays of up to 20 nights in skilled nursing facilities, and it partially reimburses such stays of 21-100 nights (Medicare, n.d.). 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 (American Council on Aging, 2021). We distinguish between four levels of education: high-school dropout/General Educational Development (GED), high-school graduate, some college, and college graduate. We make use of HRS data on several domains of cognitive functioning obtained through validated tests (Ofstedal et al., 2005;

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.

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.[5]

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 probability of moving to a nursing home within five years – by almost 5 percentage points (pp).[6]

[5] Appendix A, Table A2 gives the number of respondents dropped at each stage to reach the analysis sample.
[6] 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.

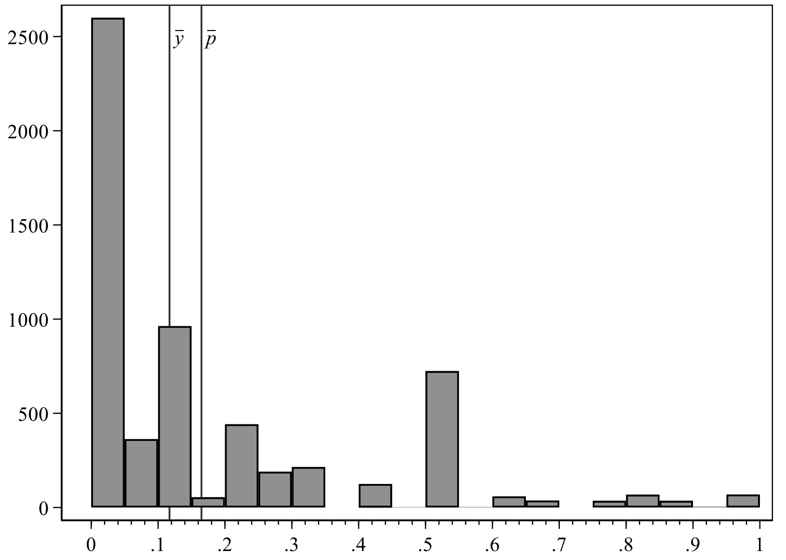


Figure 1. Distribution of subjective probabilities of moving to a nursing home within 5 years

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.

2.3 Methods

2.3.1 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 (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):

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

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.[7] $\hat{p}_i$ 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.

---

[7] We do not exploit the panel nature of the HRS since estimating a fixed effects model does not allow us to separate use of shared and private information.

[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.

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 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.

2.3.3 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

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

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

[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.

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

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 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:

$$p_i = \eta + \theta Children_i + \boldsymbol{\varphi X}_i + \boldsymbol{\zeta Z}_i + \nu_{i.} \tag{7}$$

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

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.

2.4 Results

2.4.1 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.

[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.

Table 1. Decomposition of risk perception inaccuracy and discrimination

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

Notes: Panel A gives eq.(2) decomposition of 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.

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 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]

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).

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

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.

2.4.2 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.

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

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.

Table 3. Heterogeneity in risk perception inaccuracy (MSE)

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

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.

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	$\frac{1}{n}\sum(p_i - y_i)^2$	**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	$\Delta p = \bar{p}_1 - \bar{p}_0$	**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.

2.4.3 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 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]

[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.

[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.

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

	OLS	IV	First stage	Reduced Form
	(1)	(2)	(3)	(4)
Sub. prob. nursing home	0.103	0.680		
within 5 years	(0.021)	[0.106, 1.509]		
Number of children			-0.007	-0.005
			(0.001)	(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 $\boldsymbol{X}$ in eq.(6), the risk factors in Table A1, and variables 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. 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, 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.[17]

[17] The estimates in Panel A of Table 5 and in Table 6 are also robust to controlling for race.

2.5 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, 2022).

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 the scope to improve the risk perceptions of lower-cognition people through informing them of risk factors may be somewhat limited given their restricted 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. Modeling of this reporting behavior tends to suggest that it only modestly biases probabilistic beliefs (Bassett & 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.

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	Highest level of education based on reported years of education and degrees/diplomas.	
Less than high school	Including GED.	0.226
High school graduate		0.321
Some college		0.227
College graduate	At least bachelor's degree.	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	2^{nd} Bottom	2^{nd} 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 modeling the subjective probabilities, with consequences for the inappropriate weighting and private information components of the discrimination slope decomposition (eq.(5)).

For these reasons, we 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: ${}^{1}\!/_{n}\sum(p_i - \hat{p}_i^{RF})^2 = 0.036 < {}^{1}\!/_{n}\sum(p_i - \hat{p}_i^{OLS})^2 =$

0.048. However, the predictions of the subjective probabilities obtained from the random forest estimates do not discriminate as well as the OLS predictions between the values of the outcome. The random forest is better at modeling 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 $\boldsymbol{\Delta p} = \overline{\boldsymbol{p}}_{\boldsymbol{1}} - \overline{\boldsymbol{p}}_{\boldsymbol{0}}$. In each case, $\boldsymbol{\Delta p}$ = 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.
			≥ 1 night	≥ 100 nights	prob. = 0.5
		(1)	(2)	(3)	(4)
A. MSE	$\frac{1}{n}\sum(p_i - y_i)^2$	**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	$\Delta p = \bar{p}_1 - \bar{p}_0$	**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.

Chapter 3

Ready to retire? Accuracy of retirement expectations

Joint work with Teresa Bago d'Uva and Owen O'Donnell

Abstract

Inaccurate retirement expectations potentially lower lifetime welfare. Retiring earlier than anticipated may force low consumption in old age and lead to inability to afford long-term care. Retiring later than expected could result in incomplete exhaustion of savings in old age and unintended bequests. We assess the accuracy of older US workers' retirement expectations by comparing their subjective probabilities of continuing to work full-time past the ages of 62 and 65 with whether they are in fact working full-time once they reach those ages. We measure inaccuracy with the mean squared prediction error (MSE) and decompose it into outcome variance, bias, noise and misuse of information in observed (in the data), known predictors of retirement as well as the offsetting contribution of private information that is not captured by those predictors. There is substantial inaccuracy. The subjective probabilities are approximately as accurate as they would be if all respondents viewed their retirement chances as determined by a coin toss. This is largely due to high variability of work status at ages 62 and 65 but also due to noise in the subjective probabilities, which accounts for more than 40% of the MSE. Use of information in observed predictors and, especially, private information contributes to accuracy and offsets the inaccuracy due to noise. The lower educated hold less accurate retirement expectations because they are worse at extracting information from observed predictors and use much less private information. These individuals may therefore be planning less appropriately for retirement and bearing the consequences of it in retirement.

3.1 Introduction

When to retire is one of the most important economic decisions in life. According to life cycle theory, we decide when to retire by forming expectations of longevity, future health, and returns on savings and investments, comparing the returns to continued work versus retirement at each age, and choosing the retirement age that maximises lifetime utility (Bernheim, 1989). Shocks – ill-health or divorce, for example – will disrupt plans. But the average error will be small if expectations are formed rationally on the basis of all available information (Benitez-Silva & Dwyer, 2005). We assess the credibility of this hypothesis by measuring the accuracy of older US workers' subjective probabilities of retiring at the ages of 62 and 65 and by estimating the extent to which those expectations utilize information contained in measured predictors of retirement that are observed in the data as well as private information that is not captured by those predictors.

The formation of accurate retirement expectations is likely to be consequential for lifetime welfare, even more so in recent decades that have seen a shift in responsibility for securing retirement wealth towards workers in the US and other countries (Lusardi & Mitchell, 2011). Retiring earlier than expected requires stretching accumulated wealth (pension and other) over a longer-than-anticipated period of retirement, which may result in low consumption in old age and inability to afford long-term care. Retiring later than expected could leave consumption lower than it need have been during working life and may result in incomplete exhaustion of savings in old age and unintended bequests. Working for longer than expected may also take a toll on health in old age. Formation of accurate retirement expectations is not only consequential but difficult. There is much information to consider. Relevant predictors of retirement age may be given insufficient weight or ignored entirely. On the other hand, attention may be paid to salient but irrelevant factors. There is scope for psychological biases in the formation of beliefs to result in systematic errors in retirement expectations.

A substantial literature, reviewed by Kézdi and Shapiro (2023), shows that retirement expectations do contain information on actual retirement. Retirement expectations vary with predictors of retirement, such as health and economic status (Dwyer & Hu, 1998), and they predict retirement even conditional on such information (Hurd, 1999). However, they are imperfect predictors of retirement (Kézdi & Shapiro, 2023). They are likely to contain noise (Kézdi & Shapiro, 2023) and to reflect much uncertainty in the outcome (Benitez-Silva & Dwyer, 2005; Caliendo et al., 2023). Our paper is the first to quantify the contributions of information extraction, noise and outcome uncertainty to the inaccuracy of subjective probability measures of retirement expectations. It assesses the extent to which individuals hold expectations that would allow them to plan optimally for their retirement.

We assess the accuracy of retirement expectations by comparing each individual's subjective probability of working full-time after age 62 (and age 65) with their actual work status at that age (the outcome). We measure inaccuracy of the subjective probabilities by their mean squared prediction error, using data from the Health and Retirement Study (HRS). We decompose that error into the variance of the outcome – which captures its uncertainty and so prediction difficulty – and bias and noise in the subjective probabilities, as well as their discriminatory power measured by the difference in the means of the subjective probabilities of those who do and do not work full-time after age 62 (or 65). We further decompose discriminatory power into: a) the objective predictability of retirement from observed predictors, b) the extent to which the subjective probabilities realise that potential, and c) private information contained in those probabilities that is not (linearly) related to the observed predictors (Bago d'Uva & O'Donnell, 2022). This involves modeling both the subjective probabilities and the outcomes as functions of baseline retirement predictors observed in the HRS data – health indicators, financial circumstances, job characteristics and sociodemographics. Comparison of the predictions from the two models gives a measure of the extent to which information contained in the observed predictors is captured by the subjective probabilities. Inappropriate weighting of the predictors results in underuse of the available information. We measure private information by the variation in the subjective probabilities that is explained by the outcome but not by the measured predictors that are observed in the data.

We find that the mean subjective probability of a full-time worker continuing to work full-time at the age of 62 (and 65), 0.54 (0.35), is very close to the proportion that do end up working full-time at that age, 0.54 (0.38). Hence, the subjective probabilities are almost unbiased, which has previously been documented (Kézdi & Shapiro, 2023). There is, nevertheless, a substantial amount of variance in the predictions, most of which is not explained by the outcome. The mean squared error is close to 0.25, which means that the subjective probabilities are as accurate as every participant reporting their retirement probability as they would the probability of getting heads in a coin toss. This inaccuracy is largely due to the high variance of the retirement outcome, given its mean is close to 0.5. The subjective probabilities have considerable discriminatory power. The contribution of this to the accuracy of the predictions is, however, completely offset by noise, which accounts for more than 40% of the MSE. Less than a quarter of the discriminatory power comes from the use of information provided by observed predictors of retirement. The rest is from private information predictive of the outcome, that is not (linearly) related to those predictors but is incorporated in the subjective probabilities. Previous studies have shown that subjective probabilities of retirement age are correlated with objective predictors of retirement and predict retirement conditional on those predictors (see McGarry, 2004; Kézdi

& Shapiro, 2023). This is the first study to isolate and quantify the contributions of the use of observed and private information to the formation of accurate retirement expectations.

Workers do reasonably well in incorporating into their expectations information that is available from predictors of retirement observed in the data. More than three quarters of the potential discriminatory power of this information is realized in the subjective probabilities. The weights attached to the predictors in the formation of subjective probabilities do not differ so much from the objective weights estimated from regression of retirement status on the predictors. In particular, information provided by job characteristics is, on the whole, used appropriately. Health indicators tend to be underweighted. On average, people tend to underestimate the risk of future work-limiting health problems that are predictable from current health status.

Low education and limited cognitive functioning may constrain ability to gather and process information and so lower the accuracy of retirement expectations. In the US, the less educated are much less likely to plan for retirement and they have lower financial literacy (Lusardi & Mitchell, 2011), which likely leaves them ill-prepared for retirement. We find that the least educated and cognitively able hold the least accurate subjective probabilities of continued work beyond the age of 62 (65). Differences in accuracy by cognition are not as strong as those by education. The least educated report the most noisy subjective probabilities with the least discriminatory power, which is due to extracting less information from observed predictors of retirement as well as holding less private information. The subjective probabilities of college graduates contain almost twice as much private information as those of high school dropouts. The latter group use less than half of the potential discriminatory power of observed predictors of retirement, whereas college graduates use more than three quarters. These findings suggest there may be a steep education gradient in the quality of retirement planning that adds to evidence of a gradient in the propensity to plan and in financial literacy (Lusardi & Mitchell, 2011).

Older Americans in higher socioeconomic groups are enjoying longer lives in better health, allowing them to spend more years working but also to have more years in retirement free of disability (Bavafa et al., 2023; Hudomiet et al., 2021a). These gains are not shared with lower socioeconomic groups. On top of these groups having less to look forward to in their retirement, our results add that they are likely to be less well prepared for it because they hold more inaccurate expectations of when they will retire.

Research on retirement expectations measures expectations either with the planned retirement age (Benitez-Silva & Dwyer, 2005; Haider & Stephens, 2007) or the subjective probability of working past a particular age, most notably ages 62 and 65 (Hurd, 2009;

McGarry, 2004). Most studies, like this one, use the latter partly because of measurement and interpretation issues that arise with the former (Kézdi & Shapiro, 2023).[18] Subjective probabilities of continued work have been shown to correlate with observable predictors of retirement and to predict retirement conditional on those predictors (Giustinelli & Shapiro, 2019; Hudomiet et al., 2021b; Hurd, 2009; Kézdi & Shapiro, 2023; McGarry, 2004) and to respond to policies that change incentives for retirement (Ayyagari, 2019; Bottazzi et al., 2006; Lindeboom & Montizaan, 2020; Woodruff, 2020). However, correlation does not capture the magnitude of prediction errors. Nor does it separately quantify the role of noise: subjective probabilities may be (highly) correlated with outcomes and observed predictors and yet still be inaccurate due to bias or incorrect weighting of relevant predictors. Attention may be paid to salient but irrelevant factors. Even if subjective probabilities are close to correct on average, they may fluctuate wildly from person to person. For optimal decision-making, it is important that the subjective probabilities align with objective probabilities.

Using HRS data, Kézdi & Shapiro (2023) show that while the subjective probabilities of working past age 62 (and age 65) of older US full-time workers are essentially unbiased, they rise less steeply with age than the objective probabilities. This attenuation bias, which is smaller at higher educational attainment and cognitive ability, partly comes from reporting error that may reflect difficulty transforming beliefs into probabilities (idem). We extend these findings by quantifying the contributions of noise, which includes reporting error and attention paid to salient but irrelevant factors, and the inappropriate weighting of observed retirement predictors to the inaccuracy of retirement expectations.

Benitez-Silva & Dwyer (2005) do not reject the hypothesis that planned retirement ages reported in the HRS are rational expectations. They do, however, reject the assumption that these expectations display perfect foresight. There is uncertainty regarding retirement timing that individuals cannot perfectly foresee, which makes retirement planning difficult. Caliendo et al. (2023), also using HRS data, estimate that individuals would be prepared to sacrifice up to 3.2% of lifetime consumption to eliminate retirement age risk. We show that risk concerning the timing of retirement derives from the inherently low predictability of retirement from observed predictors. In addition to this, expectations contain a lot of noise.

This paper offers four main findings that add to evidence on retirement expectations and their implications for retirement planning. First, although subjective probabilities of retiring at a particular age do, to some extent, predict retirement status at that age, they are highly inaccurate. This is mainly due to the difficulty of predicting retirement. Noise in the reported

[18] There is a lot of uncertainty around retiring at certain ages, so a single statistic of retirement age tells us relatively little (Kézdi & Shapiro, 2023). It is furthermore unclear what the single statistic that is derived is, it could be the mean age, for example, or the modal age (idem).

subjective probabilities is another important contributor to the inaccuracy. Second, we are the first to quantify the contribution of incorrect weighting of known predictors and we show that this is not an important source of the inaccuracy of retirement expectations. It is the low predictability of retirement that is the greater problem, not the ability of individuals to reach the potential predictability. Third, we show that individuals do use substantial private information – subjective probabilities predict retirement even after controlling for observed predictors. Fourth, the least educated have the most inaccurate retirement expectations. We show that their subjective probabilities contain more noise, have substantially less private information, and capture less of the information that is available in observed predictors.

3.2 Data

We use data from the Health and Retirement Study (HRS), a biennial longitudinal survey of a representative sample of older (50+) Americans (Health and Retirement Study, 2023). Since the first wave in 1992, respondents who are currently working full-time are asked about their probability of doing so after age 62 and, in a follow-on question, after age 65: "What do you think are the chances that you will be working full-time after you reach age 62 (65)?" ($q62$ and $q65$). Answers can take values in the range from 0 ("Absolutely no chance") to 100 ("Absolutely certain"), which we rescale to the 0-1 range. The HRS includes more questions about chances of working for pay in the future but we opt for questions $q62$ and $q65$ as they reflect standard retirement ages and have been fielded consistently which maximises the number of observations for which we can establish whether the retirement event is realized. Respondents who are already 62 or 65 are not asked the respective probability question, nor are respondents who do not answer three prior expectation questions.[19] Those who answer 0 ("Absolutely no chance") to $q62$ are not asked $q65$ and are instead imputed 0 for this question. Nonresponse is low for both questions: less than 1.5% of respondents refused to answer or indicated "don't know". We use subjective probabilities from waves 3-12 (1996-2014) because these waves consistently include predictors of retirement and wave 12 is the latest wave for which we can determine the outcome of the predicted event. Waves 3-12 include respondents born in 1924-1959 and their spouses (any age).

We follow Kézdi & Shapiro (2023) in defining the sample and outcomes. We restrict our analyses to workers who gave a numeric response to the subjective probability question (or were imputed 0 for $q65$, due to responding 0 to $q62$) and who, at that time, worked full-time and were at least 3 and at most 8 years younger than the threshold age (54-59 for $q62$ and

[19] There are some cross wave differences but this is true for but most of the waves. The three prior expectation questions asked are about home values and inheritance. If respondents refuse to answer these probability questions or answer "don't know" they are not asked the subsequent expectations questions.

57-62 for *q*65). We use the labor force status of respondents, reported in each wave, to construct outcomes.[20] To do so, we additionally restrict our analyses to workers for whom we observe the outcomes at the threshold ages 62 and 65 or, for those who were not interviewed at these ages due to the biennial structure of the panel, one year later (at age 63/66 for *q*62/*q*65). This ensures that we are defining a respondent as not working full-time at each threshold age as correctly as possible.[21] The outcome is defined as 1 if a respondent is observed to work full-time at threshold age, t, or at any age up until $t + 5$, and 0 otherwise. The outcome of respondents who die before they reach the threshold also equals 0 (i.e., they are not working full-time at that age). We assess robustness to a) alternative time windows (t, t+3) and (t, t+7), and to b) including only respondents who are still alive by the threshold age.

We model both the subjective probabilities and the outcomes as functions of known predictors of retirement observed in the same wave as the subjective probabilities. We follow McGarry (2004) and include health indicators (self-assessed health, subjective probability of living to age 75, cognitive functioning score, depression CES-D 8 score, existence of self-reported work-limiting health problem, number of functional limitations and number of diagnosed health conditions), financial and job information on earnings, household wealth, individual retirement wealth (private pension wealth), pension (current job, yes/no), union (yes/no), tenure (years at current job), experience (total years worked), employer-sponsored (retiree) health insurance (yes/no), and sociodemographic indicators of age, sex, race, marital status and education (more detail and descriptive statistics can be found in Appendix A, Table A1).

Our samples for *q*62 and *q*65 include observations of respondents aged, respectively, 54-59 and 57-62 in 1996-2014 who a) report their subjective probability of working past the threshold age, b) can be followed over time to determine, as explained above, if they worked full-time past the threshold age or are known to be dead before that age, and c) have no item nonresponse on any of the retirement predictors used.[22] This results in 10,704 observations of 5,746 respondents for *q*62 and 8,819 observations of 4,897 respondents for *q*65.

Figure 1 shows the distribution of subjective probabilities of working full-time past ages 62 and 65. Around 50% of observations are focal responses of 0, 0.5 or 1. About 20% (10%) report to be certain to still be working full-time after age 62 (65), and 15% (30%) report there is no chance of working full-time past that age. Around 15% of respondents, for both

[20] The HRS defines full-time work as working 35 hours or more per week for at least 36 weeks per year.
[21] Suppose that for *q*62 we only observe a respondent to not work full-time, after the threshold age, at age 64. This does not exclude the possibility that this respondent worked full-time at age 63.
[22] Appendix A, Table A2 gives an overview of observations dropped in the construction of the analysis samples.

questions, report a 50-50 chance. This may be an expression of lack of knowledge of the probability rather than a belief that it is precisely 0.5 (Fischhoff & Bruine de Bruin, 1999; Bruine de Bruin & Carman, 2012). The figures reveal that the subjective probabilities are close to unbiased for these samples. This is especially true for working full-time after age 62, for which the means of both the subjective probability and the outcome are 0.54. For age 65, the mean subjective probability (0.35) is slightly lower than the objective probability (0.38). Individuals are thus somewhat more likely to work full-time after age 65 than they expected when asked the subjective probability question 3-8 years earlier.

Figure 2, which shows the mean subjective probabilities and the mean outcomes by wave, reveals that the objective probabilities of working past the target ages have increased over time and the mean subjective probabilities have been tracking this trend quite well, particularly so for age 62. For age 65, respondents underestimate the probability in earlier waves, but not in later waves. The graphs also suggest that respondents pay attention to salient information that turns out not very relevant; namely, the mean subjective probabilities show a spike at the time of the financial crisis (wave 9, 2008), with no such increases in the mean outcomes.

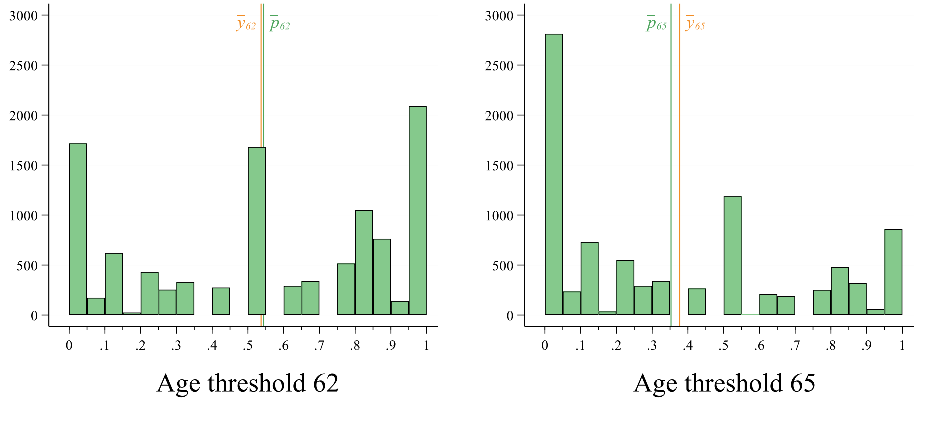
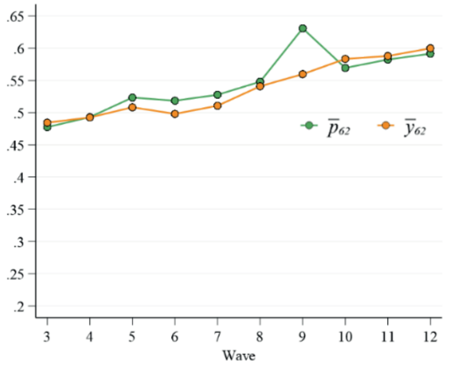
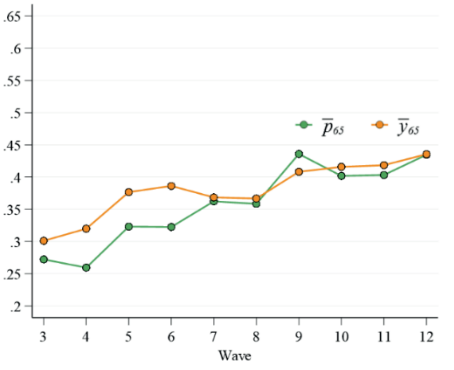


Figure 1. Distribution of subjective probabilities of working past age thresholds

Notes: Bin size equals 0.05. y-axis reports frequencies. Vertical lines show the proportion who work full-time past the threshold ages ($\bar{y}$, where $\bar{y} = {}^1/_n \sum y_i$) and the corresponding means of the subjective probabilities ($\bar{p} = {}^1/_n \sum p_i$). n=10,704 (age threshold 62) and 8,819 (65).



Age threshold 62



Age threshold 65

Figure 2. Mean subjective probabilities of working past age thresholds and mean outcomes by wave

Notes: For each wave, figure shows the mean subjective probabilities, where $\bar{p} = {}^{1}/_{n} \sum p_i$, reported in that wave and the mean outcome, where $\bar{y} = {}^{1}/_{n} \sum y_i$, at the target age over all respondents who report a subjective probability in that wave. n=10,704 (age threshold 62) and 8,819 (65).

3.3 Methodology

We use the sample mean squared error – MSE – of the subjective probabilities to measure their average inaccuracy across all respondents:

$$MSE_T = \frac{1}{n_T} \sum_i (p_{Ti} - y_{Ti})^2 \in [0,1] \tag{1}$$

where p_{Ti} is the reported subjective probability of working full-time past age $T \in \{62, 65\}$ of observation i and y_{Ti} is a binary indicator of the respective outcome, which equals 1 if observation i works full-time past age T, and 0 otherwise (including dying before age T). n_T is sample size for qT. To avoid clutter, we drop the subscript T and describe decomposition of the MSE at any threshold age below.

The MSE increases with outcome variance, $Var(y) = \bar{y}(1 - \bar{y})$ where $\bar{y} = {}^{1}/_{n} \sum y_i$. This captures uncertainty, and so greater difficulty of the prediction task. Inaccuracy also increases with bias of the subjective probabilities, $bias = \bar{p} - \bar{y}$, where $\bar{p} = {}^{1}/_{n} \sum p_i$. Accuracy increases with discriminatory power of the subjective probabilities, i.e., the degree to which they are correlated with the outcome. This can be quantified by the discrimination slope, the difference between the means of the subjective probability for those who work full-time past the threshold age and for those who do not: $\Delta p = \bar{p}_1 - \bar{p}_0$, where $\bar{p}_k =$

${}^{1}/_{n_k} \sum 1(y_i = k)p_i,\ n_k = \sum 1(y_i = k),\ k \in \{0,1\}$. With binary outcomes, the discrimination slope relates to the covariance between the outcomes and predictions in the following way: $Cov(p, y) = \Delta p Var(y)$. Lastly, the MSE increases with the variance of the subjective probabilities, $Var(p)$. Part of this variance is explained by the outcome and so captures the degree to which the probabilities give a signal of the outcome: $signal = \Delta p^2 Var(y)$. The part of the variance that remains unexplained by the outcome is $noise = Var(p) - \Delta p^2 Var(y)$. This can arise from the influence of factors that are unrelated to retirement on the subjective probabilities, from measurement error stemming from inability to accurately express beliefs about future work status as probabilities or from limited comprehension of the probability question or probability laws. These five determinants of the inaccuracy of subjective probabilities correspond to the components of a decomposition of the MSE (Yates, 1982):

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

The discrimination slope can be decomposed further into a part that captures the use of information contained in observed predictors of retirement and a part unrelated to them. The latter captures the use of additional (private) information that respondents have and that is relevant to their retirement (Bago d'Uva & O'Donnell, 2022). To implement this further decomposition, we model both the subjective probabilities and the outcomes as linear functions of observed predictors of retirement ($\boldsymbol{X}$):

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

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

where β_j^P is the partial association of the subjective probabilities with the j^{th} predictor, β_j^Y is the partial association of the outcome with the respective predictor; and ε_i and υ_i are random errors. Application of ordinary least squares (OLS) to models (3) and (4) provides estimates of weights on predictors that give the best linear predictions of the subjective probabilities and the outcomes, respectively. For example, $\hat{\beta}_j^P$ is the OLS estimate of the weight individuals, on average, implicitly give to the predictor X_j when forming their beliefs about the chances they will be working full-time at the threshold age. The counterpart OLS estimate $\hat{\beta}_j^y$ is the objective weight on that predictor that best predicts the outcome. The variation in the OLS residuals $\hat{\varepsilon}_i$ and $\hat{\upsilon}_i$ represents variation in the subjective probabilities and the outcomes, respectively, that is linearly uncorrelated with the observed predictors, $\boldsymbol{X}$. These predictors include the health, financial, job and sociodemographic characteristics listed in section 2 (see also Table A1) plus wave fixed effects.[23]

The resulting decomposition of the discrimination slope is:

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

Where $\Delta z = \bar{z}_1 - \bar{z}_0$, $\bar{z}_k = {}^1/_{n_k} \sum 1(y_i = k) z_i$, $k \in \{0,1\}$, $z_i \in \{p_i, \hat{p}_i, \hat{\varepsilon}_i, \hat{y}_i\}$, $\hat{p}_i = \sum_{j=1}^{J} \hat{\beta}_j^p X_{ji}$, $\hat{y}_i = \sum_{j=1}^{J} \hat{\beta}_j^y X_{ji}$ and $\hat{\beta}_j^y$ and $\hat{\beta}_j^p$ are OLS estimates of the coefficients of (3) and (4), as defined above (Bago d'Uva & O'Donnell, 2022).

The term $\Delta\hat{y}$ quantifies the degree to which retirement can be (linearly) predicted from the observed predictors. Greater predictability increases the potential discriminatory power, and thus the accuracy, of the subjective probabilities (eq. (2)). $\Delta\hat{p}$ measures the extent to which the subjective probabilities are able to realize this potential. $\Delta\hat{y} - \Delta\hat{p}$ is therefore the discrepancy between the discrimination that could be achieved if the predictors were weighted optimally to predict the outcomes and the discrimination that is actually achieved with the subjective weighting of them. This represents the loss of discriminatory power because of suboptimal use of information contained in those predictors in the formation of beliefs about retirement chances. This term can be further decomposed to reveal information extraction from each predictor or set of predictors.[24] Finally, $\Delta\hat{\varepsilon}$ is the part of the discriminatory power of the subjective probabilities that comes from the use of information on future retirement status that is not (linearly) related to the observed predictors.

We estimate model (3) by pooling data from waves 3-12 and using predictors measured in the same wave as the subjective probability for each observation. We estimate model (4) by pooling data on outcomes from waves 4-15 and regressing these on the same values of the predictors that are used for the respective observations to estimate model (3). Since respondents report subjective probabilities in every wave, many have multiple observations in our sample. We compute bootstrap standard errors for the MSE and its components shown in equations (2) and (5) using 100 replications.[25]

[23] These are, for both eq. (3) and eq. (4), baseline levels at the time when the subjective probability question is reported.
[24] $\Delta\hat{y} - \Delta\hat{p} = \sum_{j=1}^{J} (\hat{\beta}_j^y - \hat{\beta}_j^p) \Delta X_j$, where $\Delta X_j = \bar{X}_{j1} - \bar{X}_{j0}$, $\bar{X}_{jk} = {}^1/_{n_k} \sum 1(y_i = k) X_{ji}$, $k \in \{0,1\}$ (Bago d'Uva & O'Donnell, 2022).
[25] For the main results, we also use 1000 replications. This gives virtually the same standard errors.

3.4 Results

3.4.1 Prediction inaccuracy

Table 1, Panel A, presents the MSE and its decomposition using eq. (2) for $q62$ and $q65$. The MSEs are above or equal to 0.25, which would be the value obtained if everyone reported a 50-50 chance and so the subjective probabilities were not at all discriminatory with respect to the outcome. The bias (squared) is small, as seen in Figures 1 and 2. Effectively, it does not contribute to inaccuracy. The decomposition reveals that outcome variance is large, and so prediction of retirement is difficult. There is substantial covariance between the outcomes and subjective probabilities, which contributes to accuracy. This is however (almost) completely offset by inaccuracy generated by noise in the subjective probabilities, which is responsible for more than 40% of the MSE. This suggests that individuals do not have well-formed beliefs about their chances of retiring at 62 and 65 or they struggle to express their beliefs in probabilities. It could also reflect attention paid to factors that are irrelevant to retirement chances when forming and reporting subjective probabilities. The signal about retirement that is contained in the subjective probabilities is only about 11% of their variance.

Panel B shows that the discrimination slope is about 0.24 for both probabilities indicating that those who work full-time past the respective age threshold report, on average, a 24 percentage point (pp) higher probability of working to that age than those who stop working before that age. Less than a quarter of this discriminatory power is derived from appropriate use of information provided by the observed predictors ($\Delta\hat{p} = \Delta p - \Delta\hat{\varepsilon} = 0.05$). The rest ($\Delta\hat{\varepsilon} \approx 0.18$) is from use of information that respondents apparently have about their retirement chances that is not (linearly) correlated with the observed predictors. This low contribution of the observed predictors to the discriminatory power of the subjective probabilities is not mainly due to underweighting of those predictors ($\Delta\hat{y} - \Delta\hat{p}$) – this is no more than one quarter of the potential discriminatory power that would be achieved if the predictors were weighted optimally ($(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$). Rather, it is the low potential predictability of retirement from the observed predictors ($\Delta\hat{y}$) that explains why success in using this information accounts for such a small part of the discriminatory power of the subjective probabilities. If individuals were to use all the information in the observed predictors (optimally) and nothing else to predict their retirement, then there would be a 6-7 pp difference ($\Delta\hat{y}$) between the predictions of those who continued to work after the threshold age and the predictions of those who did not.

Table 1. Decomposition of inaccuracy and discrimination of subjective probabilities of continued work

		Working full-time after age			
		62		65	
		Estimate	SE	Estimate	SE
A. MSE	$\boldsymbol{\frac{1}{n}\sum(p_i - y_i)^2}$	**0.265**	**(0.003)**	**0.249**	**(0.004)**
Decomposition, eq.(2)					
outcome variance	$Var(y)$	0.249	(<0.001)	0.235	(0.001)
bias^2	$(\bar{p} - \bar{y})^2$	<0.001	(<0.001)	0.001	(<0.001)
covariance	$-2(\Delta p)Var(y)$	-0.117	(0.003)	-0.112	(0.004)
signal	$(\Delta p)^2 Var(y)$	0.014	(0.001)	0.013	(0.001)
noise	$Var(p) - (\Delta p)^2 Var(y)$	0.120	(0.001)	0.112	(0.001)
B. Discrimination slope	$\boldsymbol{\Delta p = \bar{p}_1 - \bar{p}_0}$	**0.235**	**(0.007)**	**0.239**	**(0.009)**
Decomposition, eq.(5)					
outcome predictability	$\Delta\hat{y}$	0.064	(0.004)	0.072	(0.005)
inappropriate weighting	$-(\Delta\hat{y} - \Delta\hat{p})$	-0.012	(0.003)	-0.018	(0.004)
	$100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$	18.6%		24.5%	
private information	$\Delta\hat{\varepsilon}$	0.183	(0.006)	0.185	(0.009)
Mean y	$\bar{y}$	0.537		0.377	
Mean p	$\bar{p}$	0.544		0.353	
Sample size	n	10,704		8,819	

Notes: Panel A gives eq.(2) decomposition of MSE of subjective probabilities of working full-time after age 62 and age 65. Panel B gives eq.(5) decomposition of the discrimination slope of the subjective probabilities. For any variable or prediction z, $\Delta z = \bar{z}_1 - \bar{z}_0$, $\bar{z}_k = {}^1/_{n_k} \sum 1(y_i = k) z_i,\ k \in \{0,1\}$. See equations and text for other notation. Bootstrap standard errors (100 replications) in parentheses. Table A3 contains estimates of models (3) and (4) used in $\Delta\hat{y}$ and $\Delta\hat{p}$. Sample includes HRS wave 3-12 respondents aged 54-59 at *q*62 and aged 57-62 at *q*65 with full item response on subjective probabilities and other covariates used in the analyses, and for whom it is possible to determine if they worked full-time after age 62 or 65, respectively.

Table 2 shows contributions of four sets of predictors to discrimination that a) potentially could be achieved if the predictors were weighted optimally (i.e., if $\Delta\hat{p}$ were equal to $\Delta\hat{y}$), b) actually is achieved with subjective weighting of the predictors ($\Delta\hat{p}$), and c) is not achieved ($\Delta\hat{y} - \Delta\hat{p}$). The largest contribution to the shortfall comes from predictors related to health. This is true for age 62, but especially so for age 65. Applying the optimal (outcome model) weights to health differences between those who work full-time past age 65 and those who do not gives a 2.2 pp predicted difference in employment status between the two groups compared to 1.0 pp difference given by application of the subjective weights to these health differences. While this implies that there is insufficient recognition of the degree to which health at the baseline (3 to 8 years prior to the outcome occurrence) is correlated with continued work, the more striking finding is that even if the information in these observed baseline health indicators were to be used fully, it would still be of relatively little use in predicting future retirement. Future retirement is expected to be driven not only by baseline health but also by future health shocks. Indeed, contemporaneous health is a strong predictor of current employment status (De Nardi et al., in press; Dobkin et al., 2018). Our finding thus suggests limited predictability of such future work-limiting health shocks from baseline health status. Individuals do well in weighting baseline job characteristics appropriately in predicting future work. However, again, these characteristics contain little information. There is underweighting of the association between sociodemographics and continued work, which is mostly due to underappreciation of the extent to which the probability of working past the threshold age increases as the individuals gets closer to that age (see Table A3). Observed financial characteristics play an even lesser role in predicting future work, as well as in the formation of the respective probabilities.

The decompositions (eqns. (2) and (5)) are robust to changes in outcome definitions (Appendix B, Table B1). They are also robust to including in the models information on social security and employer-sponsored pension wealth, which is only available in certain waves (Table B3).

Table 2. Contributions of observed predictors to potential and achieved discrimination between workers and non-workers at target age

	Potential $\Delta\hat{y}$	Achieved $\Delta\hat{p}$	Shortfall $\Delta\hat{y}-\Delta\hat{p}$
A. Work full-time to age 62			
Total	**0.064 (0.004)**	**0.052 (0.003)**	**0.012 (0.003)**
Contributions			
Health	0.017 (0.002)	0.010 (0.002)	0.007 (0.002)
Financial	0.002 (0.001)	0.002 (0.001)	0.000 (0.001)
Job	0.022 (0.003)	0.023 (0.002)	-0.001 (0.001)
Sociodemographics	0.022 (0.003)	0.017 (0.002)	0.005 (0.002)
B. Work full-time to age 65			
Total	**0.072 (0.005)**	**0.054 (0.004)**	**0.018 (0.004)**
Contributions			
Health	0.022 (0.003)	0.010 (0.002)	0.013 (0.003)
Financial	0.005 (0.002)	0.002 (0.001)	0.002 (0.001)
Job	0.016 (0.003)	0.017 (0.002)	-0.001 (0.002)
Sociodemographics	0.029 (0.004)	0.025 (0.003)	0.004 (0.002)

Notes: n = 10,704 for panel A and 8,819 for panel B. 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 factors to the measures in the top row. The left-hand column gives, in each row for the set of 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 simulations) in parentheses. See Table A1 for the factors included in each set. See Table A3 for the OLS estimates $\hat{\beta}_j^y$ and $\hat{\beta}_j^p$ for all j. Sample includes HRS wave 3-12 respondents aged 54-59 at *q*62 and aged 57-62 at *q*65 with full item response on subjective probabilities and other covariates used in the analyses, and for whom it is possible to determine if they worked full-time after age 62 or 65, respectively.

3.4.2 Heterogeneity in prediction inaccuracy

Table 3 presents estimates of differences in the MSE by education and cognition, conditional on age, sex, marital status and race.[26] The MSE is generally larger, indicating greater inaccuracy, at lower education and cognition levels. The differences are generally larger for the prediction of work status at age 62 and they are larger by education than by cognition. The MSE of that prediction by the bottom quartile group of the cognition score is almost 3 points larger than that by the top quartile group, controlling for education as well as sociodemographics. The respective difference between high school dropouts and college graduates is 7 points.

The analysis in Table 4 aims at unveiling possible sources of that heterogeneity by education. Panel A presents the unconditional MSE (first row) and its decomposition (eq. (2)). Panel B further decomposes the discriminatory power of the subjective probabilities (eq. (5)). Firstly, the second row of Panel A reveals that the less accurate subjective probabilities reported by the lower education groups are not driven by a greater outcome (employment) variance – a more difficult prediction task – in lower education groups. Rather, these groups make noisier predictions, particularly for $q62$, possibly indicative of poorer understanding of probabilities and/or greater attention given to (salient but) irrelevant factors in forming beliefs about future retirement. The subjective probabilities of the less educated also have much less discriminatory power, which is indicated by lower magnitudes of covariances.

[26] These are OLS estimates of linear regression models of $(p_i - y_i)^2$ on indicators of education and cognition categories (separately and jointly) with controls for age, sex, marital status and race. See Appendix Table A4 for coefficient estimates without controls for age, sex, marital status and race.

Table 3. Education and cognition differences in inaccuracy of subjective probabilities of continued work

	(1)		(2)		(3)	
A. Working full-time at age 62						
Education (ref. College graduate)						
High school dropout or GED	0.078	(0.011)			0.069	(0.012)
High school graduate	0.038	(0.008)			0.033	(0.009)
Some college	0.034	(0.008)			0.031	(0.009)
Cognitive functioning (ref. Top quartile)						
Bottom quartile			0.047	(0.010)	0.028	(0.010)
2nd Bottom quartile			0.014	(0.009)	0.004	(0.009)
2nd Top quartile			0.006	(0.009)	0.001	(0.009)
A. Working full-time at age 65						
Education (ref. College graduate)						
High school dropout or GED	0.049	(0.012)			0.039	(0.013)
High school graduate	0.014	(0.009)			0.009	(0.009)
Some college	0.027	(0.009)			0.024	(0.009)
Cognitive functioning (ref. Top quartile)						
Bottom quartile			0.037	(0.011)	0.028	(0.012)
2nd Bottom quartile			0.017	(0.009)	0.012	(0.010)
2nd Top quartile			0.006	(0.010)	0.004	(0.010)

Notes: n = 10,704 for panel A and 8,819 for panel B. OLS estimates of linear regression models of $(p_i - y_i)^2$ on indicators of educational attainment (column (1)), cognition score quartile group (column (2), and both education and cognition (column (3)). All specifications control for sex, single-year age dummies, marital status (married/partnered), and race (ref. White, Black and other). Robust standard errors in parentheses. The means of the dependent variable, i.e. MSE, of the reference groups are 0.231 and 0.247, for education and cognitive functioning for $q62$ and 0.228 and 0.232 for $q65$, respectively.

Table 4. Decompositions of inaccuracy and discrimination of subjective probabilities of continued work by education

		q62				*q65*			
		High school dropout	High school graduate	Some college	College graduate	High school dropout	High school graduate	Some college	College graduate
A. MSE	$\frac{1}{n}\sum(p_i - y_i)^2$	**0.315**	**0.272**	**0.271**	**0.231**	**0.282**	**0.244**	**0.259**	**0.228**
		(0.010)	**(0.006)**	**(0.006)**	**(0.006)**	**(0.010)**	**(0.006)**	**(0.006)**	**(0.006)**
Decomposition, eq.(2)									
outcome variance	$Var(y)$	0.248	0.250	0.248	0.241	0.227	0.222	0.230	0.248
		(0.001)	(0.000)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)
bias^2	$(\bar{p} - \bar{y})^2$	0.000	0.000	0.001	0.000	0.003	0.001	0.001	0.003
		(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
covariance	$-2(\Delta p)Var(y)$	-0.073	-0.114	-0.107	-0.136	-0.065	-0.096	-0.101	-0.149
		(0.010)	(0.006)	(0.006)	(0.006)	(0.010)	(0.006)	(0.006)	(0.006)
signal	$(\Delta p)^2 Var(y)$	0.005	0.013	0.012	0.019	0.005	0.010	0.011	0.022
		(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
noise	$Var(p) - (\Delta p)^2 Var(y)$	0.134	0.123	0.117	0.107	0.113	0.107	0.119	0.104
		(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
B. Discrimination slope	ΔP	**0.147**	**0.228**	**0.216**	**0.283**	**0.144**	**0.217**	**0.220**	**0.301**
		(0.020)	**(0.013)**	**(0.012)**	**(0.012)**	**(0.020)**	**(0.013)**	**(0.012)**	**(0.012)**
Decomposition, eq.(5)									
outcome predictability	$\Delta\hat{y}$	0.046	0.073	0.084	0.063	0.063	0.069	0.086	0.090
		(0.012)	(0.010)	(0.010)	(0.009)	(0.012)	(0.010)	(0.010)	(0.009)
inappropriate weighting	$-(\Delta\hat{y} - \Delta\hat{p})$	-0.026	-0.021	-0.031	-0.004	-0.045	-0.023	-0.028	-0.021
		(0.012)	(0.007)	(0.008)	(0.005)	(0.012)	(0.007)	(0.008)	(0.005)
	$100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$	57.7%	28.6%	36.2%	6.4%	71.3%	33.2%	33.2%	23.9%
private information	$\Delta\hat{\varepsilon}$	0.127	0.176	0.162	0.224	0.125	0.171	0.162	0.223
		(0.018)	(0.013)	(0.011)	(0.012)	(0.018)	(0.013)	(0.011)	(0.012)
Mean y	$\bar{y}$	0.455	0.512	0.539	0.596	0.347	0.332	0.357	0.455
Mean p	$\bar{p}$	0.477	0.507	0.571	0.586	0.292	0.305	0.381	0.404
Sample size	n	1,474	3,038	2,971	3,221	1,333	2,549	2,337	2,600

Notes: Contents of table, samples, and methods are as Table 1 except here the sample is stratified by education. Models (3) and (4) are estimated separately for each education group. Because we stratify by education, we do not include the education dummies in the regressions. See Table A5 for the inappropriate weighting of sets of factors by these groups.

Panel B further decomposes the discriminatory power of the subjective probabilities of working past each threshold age for the different education groups (eq.(5)). The difference between the average probabilities for those who do work past each age and those who do not (Δp) is about twice as large for college graduates than it is for high school dropouts (first row). In part, this is because retirement of dropouts is more difficult to predict from observed predictors – $\Delta\hat{y}$ is smaller for this group. But this is not the main explanation for the difference. Dropouts also make substantially larger mistakes in weighting those predictors. Namely, they leave unused 58% (71%) of the potential discriminatory power of the predictors when forming expectations to retire by 62 (65). In contrast, college graduates waste only 6% and 24%, respectively, of the information contained in the predictors. Even more striking is the difference in use of private information. For college graduates, there is a 22 pp difference in the mean subjective probability model residuals between those who work past the age thresholds and those who do not ($\Delta\hat{\varepsilon} = 0.22$). For high school dropouts, this is only 13 pp. This means that, after taking into account information extracted from the observed predictors, the less educated either have less extra information to draw on when forming beliefs about when they will retire or are less able to incorporate it in their subjective probabilities.

3.5 Discussion

Inaccurate retirement expectations may result in poor life cycle planning and, consequently, suboptimal savings that impinge on well-being in old age. We show that, while subjective probabilities of working past two retirement age thresholds reported by older US workers align well with the mean objective probabilities, they are highly inaccurate. Many thus make substantial mistakes. In fact, the inaccuracy of the subjective probabilities is approximately equivalent in magnitude to that that would occur if all were to report a retirement probability of 50%. This high level of inaccuracy is largely due to the large variance in full-time employment at the age thresholds, which makes the prediction task harder. But another important contributing factor is that the subjective probabilities are noisy. This may reflect a tendency to pay too much attention to salient but irrelevant factors when forming beliefs about future work status, and/or difficulty in expressing these beliefs on a probability scale.

More positively, we show that US workers are, on average, rather good at extracting relevant information from observed predictors of retirement. They utilize more than three-quarters of the potential that these predictors have to discriminate between those who work full-time past the age thresholds and those who do not. Most of the unused information is related to health. Insufficient weight is given to the association between current health and future work status. This may reflect difficulty in forecasting health itself.

Strong ability, on average, to use relevant information from observed predictors of retirement does not actually help US workers that much in forming accurate retirement expectations because the predictors are rather weak. Even if used optimally, they would predict only a 6-7 pp difference in the objective probability of working past age 62 (or 65) between those who do end up working past that age and those who do not. Retirement, however, is not only influenced by information available at the time of prediction, but also by future shocks. Our findings thus suggest limited predictability of such shocks from our baseline predictors and that accurate retirement expectations depend to a much greater extent on ability to use information that workers appear to have but is not related to observed predictors. This private information accounts for more than three quarters of the discriminatory power of subjective retirement probabilities.

These findings confirm previous evidence that subjective retirement probabilities covary with known predictors of retirement, predict retirement even after conditioning on these predictors, but contain substantial error (e.g., Kézdi & Shapiro, 2023; McGarry, 2004). We extend these findings by quantifying the inaccuracy of these subjective probabilities as well as the contributions of noise and use of information to that inaccuracy. On average, subjective probabilities are very inaccurate. This is surprising since retirement, for most, is not a complete shock. Economic analysis of retirement presumes that it is planned. Rational agents would stop working when the (discounted) returns to continued work drop below the costs. With sufficient information on the benefits and costs to working to each age and with sufficient cognitive capacity to process that information, or with access to expert advice, it would be possible to form accurate retirement expectations that would only be thrown off course by shocks, such as ill-health or redundancy. Our results suggest that many US workers do not hold accurate retirement expectations. Given that reported subjective probabilities of continued work predict consumption and savings behavior (Haider & Stephens, 2007; Romm, 2015), inaccurate retirement expectations would imply that individuals are not planning optimally for their retirement. This is worrying, even more so in recent decades where responsibility for securing retirement wealth has shifted from employers to workers (Lusardi & Mitchell, 2011).

We show substantial heterogeneity in accuracy and identify sources contributing to it. The retirement predictions are much less accurate for the lower educated. Their predictions contain more noise. Further, their predictions are much less able to discriminate between those who work full-time past the age thresholds and those who do not. They make larger mistakes in weighting the observed predictors and use less private information. Those without a high school diploma utilize less than half of the potential discriminatory power contained in the observed predictors, whereas college graduates use at least three quarters of this potential. The predictions of the latter also contain twice as much private information as those of the former. All this indicates that the less educated are less able to plan optimally for their retirement. This large heterogeneity is cause for worry, especially with recent mortality trends. The gap in longevity between higher and lower educated older Americans is becoming even wider (Hudomiet et al., 2021a). Integrating out uncertainty in a rational way, this should lead the lower educated to take up early social security benefits and retire earlier so as to still be able to enjoy their retirement, but they may not do so due to their underappreciation of health risks. Given the large contribution of the difference in private information to the education difference in accuracy, future research should investigate what is the additional information possessed by higher education groups that allows them to better forecast their age of retirement. Perhaps the retirement path of the less educated is harder to predict, e.g., due to more uncertain labor market conditions, or they are not utilizing important information available to them, which if appropriately employed, could lead to more optimal retirement planning.

Like all studies that use data on reported subjective probabilities, we cannot guarantee that these reports represent people's true beliefs. The inaccuracies we measure may arise partly from reporting errors due to difficulty transforming true beliefs to probabilities (Gigerenzer & Hoffrage, 1995), which is consistent with our finding that noise tends to be higher for the less educated. Reporting errors may involve use of focal responses (0, 0.5, 1) and other rounding to one decimal place. Studies that have modelled this reporting behavior generally find that it only has a modest impact on probabilistic beliefs (Basset & Lumsdaine, 2001; Giustinelli et al., 2022; Kleinjans & van Soest, 2014; Manski & Molinari, 2010). Kezdi & Shapiro (2023) also show that it is unlikely that all prediction error can be attributed to reporting errors – they find that approximately 20% of the variance of subjective probabilities of continued work is reporting error.

Another important source of inaccuracy is that, at the time of prediction, information available is different than at the time of retirement. Shocks, such as ill-health or divorce, can result in unrealized retirement expectations. It is thus conceivable that individuals hold inaccurate expectations of retirement because of such shocks. We do not model these, but doing so may give more insight into why retirement expectations are not realized. This may also help explain why predictions of the higher educated contain more private information – perhaps they are able to incorporate, to some extent, in their expectations information on future changes.

Lastly, we should acknowledge that there may be instances when people's retirement expectations are in line with their outcome and at the same time they are not planning optimally for their retirement. People may underweight the correlation between current health and future ability to work and because of this arrive at a high subjective probability of working full-time after age 65. By age 65 they may then be in poor health and would be better off retiring, but they may not be able to so do because they did not plan for this. In such instances reported subjective probabilities of continued work are accurate and still coincide with, and in fact result from, suboptimal planning. Despite this possibility, we observe low accuracy, which indicates that suboptimal planning is likely to occur.

In conclusion, we find that older US workers, especially the least educated, hold inaccurate expectations about continued work beyond standard retirement ages. The predictions contain private information, which helps to reduce inaccuracy, but they are also noisy. The predictions of the lower educated include much less private information, and also less information on observed predictors of retirement, partly because they make larger mistakes when weighting these. This potentially has important consequences for retirement planning and well-being in old age, especially in light of growing inequalities in longevity, and suggests that it is important to investigate what constitutes private information and why the less educated have less of it to help shape policies to combat growing inequalities.

Appendices

Appendix A. Additional Tables and Figures

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

Variable	Definition	Mean: Sample $q62$	Mean: Sample $q65$
Health status			
SAH	Self-assessed health reported on a five-point scale.		
Excellent		0.175	0.164
Very good		0.377	0.383
Good		0.313	0.321
Fair/poor		0.136	0.133
CES-D 8 score	Reliable and valid measure of depression. The 8-item measure asks respondents if: (1) they felt depressed, (2) everything was an effort, (3) their sleep was restless, (4) they were happy, (5) they felt lonely, (6) they felt sad, (7) they could not get going and (8) they enjoyed life in the past week (all 1 if yes and 0 if no). Score is (1) + (2) + (3) - (4) + (5) + (6) + (7) - (8).	1.12	1.07
Subjective probability of living to age 75	Self-reported probability of living to age 75. Reported on a 0-100 scale and divided by 100 to obtain probabilities.	0.663	0.669
Health problem that limits paid work	1 if have health problem that limits paid work, 0 otherwise	0.075	0.071
Number of functional limitations	Count of number of functional limitations that have any difficulty with. Functional limitations included are: (1) walking several blocks, (2) walking a block, (3) walking across the room, (4) sitting for two hours, (5) getting up from a chair, (6) getting in and out of bed, (7) walking up several flights of stairs, (8) walking up a flight of stairs, (9) lifting 10 pounds and (10) picking up a small object (dime). All 1 if have difficulty, 0 if no difficulty.	0.95	1.02
Number of health conditions	Count of number of health conditions that have ever been diagnosed by a doctor. Health conditions are: (1) high blood pressure, (2) diabetes, (3) cancer, (4) lung disease, (5) heart problem, (6) stroke, (7) psychological problem and (8) arthritis. All 1 if have been told by doctor that have the condition, 0 otherwise.	1.15	1.29
Cognition score	Standardized 27- item score (see, e.g., Crimmins et al., 2011) combining immediate word recall (10 points), delayed word recall (10 points), sequentially subtracting 7 from 100 (5 points), counting backwards from 20 (2 points). A higher score indicates better cognitive capacity.	17.21	17.19
Financial characteristics			
Earnings	Individual earnings from last calendar year in dollars. Sums wage/salary income, bonuses/overtime pay/commissions/tips, second job or military reserve earnings, and professional practice or trade income.	49,189	47,949

Household wealth (without IRAs) ($)	Total net household wealth in dollars, excluding individual retirement accounts.	249,259	283,177
IRA wealth ($)	Pension wealth in dollars from individual retirement accounts and Keogh accounts.	45,647	53,111
Job characteristics			
Pension from current job	1 if have pension from current job, 0 otherwise.	0.750	0.735
Covered by a union	1 if covered by a union, 0 otherwise.	0.230	0.220
Years of tenure on current job		13.4	13.8
Total years worked		32.6	35.8
Health insurance from job	1 if covered by health insurance from current or previous employer, 0 otherwise.	0.772	0.778
Health insurance from job covers retirees	1 if covered by health insurance from current or previous employer and this insurance plan covers the respondent up to age 65, 0 otherwise.	0.380	0.417
Sociodemographics			
Age	years	56.5	59.3
Male	1 if male	0.479	0.487
Race	Reported primary race.		
White	/Caucasian	0.763	0.777
Black	/African American	0.170	0.162
Other	Other primary race, e.g., Hispanic.	0.068	0.061
Married (incl. partnered)	1 if reported being married or living with partner, 0 otherwise.	0.737	0.726
Education	Highest level of education based on reported years of education and degrees/diplomas.		
Below high school (incl. GED)		0.138	0.151
High school graduate		0.284	0.289
Some college		0.278	0.265
College graduate		0.301	0.295
n		10,704	8,819

Notes: In models, age is entered as single-year age dummies. Analysis sample for models (3) and (4) estimated using HRS wave 3-12 respondents aged 54-59 at *q*62 and aged 57-62 at *q*65 with full item response on subjective probabilities and other covariates used in the analyses, and for whom it is possible to determine if they worked full-time after age 62 or 65, respectively. CES-D8 is the Center for Epidemiologic Studies Depression (CESD) scale. See also RAND codebook: 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 observations	
	*q*62	*q*65
Full-time workers aged 54-59 for *q*62 or aged 57-62 for *q*65 in waves 3-12	19,720	17,185
Proxy interview or AHEAD cohort	-1,164	-1,181
Not asked subjective probability of working full-time past age threshold	-83	-88
Non-response to subjective probability of working full-time past age threshold	-167	-229
Other missing	-65	-233
Reported subjective probability	18,241	15,545
Cannot determine if worked full-time past age threshold	-3,880	-3,298
Observe if worked full-time past age threshold	14,361	12,247
Missing on retirement predictors	-3,657	-3,428
Item response on all retirement predictors	10,704	8,819

Notes: These probability questions are not asked to the AHEAD cohort. Respondents are not asked to report their subjective probability of working full-time past age 62 or age 65 if they do not give numerical responses to three prior questions about expectations of house values and giving or receiving an inheritance. Non-response includes don't know or refused to answer. Other missing are missing values for other reasons than those mentioned above.

Table A3. OLS estimates of models for subjective probability of working full-time after age 62 and age 65 (*p62* and *p65*) and indicator of actually working full-time after these ages (*y62* and *y65,* respectively)

	*p*62		*y*62		*p*65		*y*65	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Health status								
SAH (ref. Excellent)								
Very good	0.018	(0.010)	0.000	(0.014)	-0.003	(0.011)	-0.045	(0.016)
Good	0.002	(0.011)	-0.029	(0.015)	-0.024	(0.012)	-0.056	(0.017)
Fair/poor	-0.008	(0.015)	-0.082	(0.021)	-0.028	(0.016)	-0.099	(0.022)
CES-D 8 score	-0.001	(0.002)	0.001	(0.003)	0.003	(0.002)	-0.005	(0.003)
Subjective probability of living to age 75	0.172	(0.013)	0.040	(0.013)	0.159	(0.014)	0.059	(0.020)
Health problem that limits paid work	-0.036	(0.014)	-0.035	(0.020)	-0.023	(0.015)	-0.048	(0.020)
Number of functional limitations	-0.004	(0.003)	-0.009	(0.004)	-0.002	(0.003)	-0.001	(0.004)
Number of health conditions	0.004	(0.003)	-0.028	(0.005)	0.002	(0.004)	-0.030	(0.005)
Financial characteristics								
Earnings ($10,000)	0.000	(0.001)	0.002	(0.001)	0.002	(0.001)	0.005	(0.002)
Household wealth (without IRAs) ($10,000)	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
IRA wealth ($10,000)	-0.003	(0.000)	-0.003	(0.000)	-0.002	(0.000)	-0.002	(0.000)
Job characteristics								
Pension from current job	0.007	(0.009)	0.040	(0.013)	-0.014	(0.010)	-0.038	(0.014)
Covered by a union	-0.081	(0.009)	-0.069	(0.012)	-0.050	(0.009)	-0.057	(0.012)
Years of tenure on current job	-0.006	(0.000)	-0.004	(0.000)	-0.006	(0.000)	-0.003	(0.000)
Total years worked	0.001	(0.000)	-0.001	(0.001)	0.001	(0.000)	-0.001	(0.001)
Health insurance from job	0.075	(0.009)	0.091	(0.013)	0.026	(0.010)	0.079	(0.015)
Health insurance from job covers retirees	-0.063	(0.008)	-0.074	(0.011)	-0.041	(0.008)	-0.052	(0.012)
Sociodemographics								
Age (ref. 54 for *q*62, 57 for *q*65)								
55, 58 resp.	0.007	(0.012)	-0.034	(0.017)	-0.003	(0.011)	-0.014	(0.016)
56, 59 resp.	0.022	(0.012)	0.031	(0.017)	0.016	(0.011)	0.050	(0.016)
57, 60 resp.	0.025	(0.012)	0.010	(0.017)	0.031	(0.012)	0.051	(0.017)
58, 61 resp.	0.043	(0.012)	0.095	(0.017)	0.049	(0.012)	0.111	(0.017)
59, 62 resp.	0.081	(0.012)	0.098	(0.018)	0.105	(0.014)	0.156	(0.020)
Male	0.044	(0.007)	0.031	(0.010)	0.045	(0.008)	0.046	(0.011)
Race (ref. White)								
Black	-0.107	(0.010)	0.012	(0.014)	-0.075	(0.010)	0.020	(0.015)
Other	-0.053	(0.014)	0.050	(0.020)	-0.026	(0.015)	0.087	(0.022)
Married (incl. partnered)	-0.050	(0.008)	0.031	(0.011)	-0.073	(0.009)	-0.013	(0.012)
Education (ref. Below high school)								
High school graduate	0.024	(0.012)	0.058	(0.016)	0.011	(0.012)	-0.018	(0.016)
Some college	0.060	(0.012)	0.061	(0.017)	0.051	(0.012)	-0.025	(0.017)
College graduate	0.074	(0.013)	0.105	(0.018)	0.076	(0.013)	0.053	(0.019)

Cognition score	0.005	(0.001)	0.004	(0.001)	0.002	(0.001)	0.007	(0.001)
Year fixed effects (ref. 1996)								
1998	0.021	(0.014)	0.013	(0.019)	-0.018	(0.014)	0.020	(0.020)
2000	0.046	(0.015)	0.017	(0.020)	0.036	(0.015)	0.068	(0.020)
2002	0.039	(0.016)	0.011	(0.022)	0.030	(0.015)	0.070	(0.022)
2004	0.059	(0.015)	0.036	(0.021)	0.073	(0.017)	0.062	(0.023)
2006	0.081	(0.015)	0.063	(0.021)	0.080	(0.016)	0.070	(0.023)
2008	0.152	(0.015)	0.074	(0.021)	0.144	(0.016)	0.100	(0.022)
2010	0.115	(0.014)	0.100	(0.020)	0.122	(0.015)	0.102	(0.021)
2012	0.124	(0.014)	0.100	(0.020)	0.120	(0.015)	0.101	(0.021)
2014	0.121	(0.015)	0.096	(0.022)	0.142	(0.017)	0.093	(0.024)
Constant	0.282	(0.029)	0.361	(0.041)	0.183	(0.030)	0.235	(0.042)
R-squared	0.136		0.064		0.135		0.072	
Mean dep. var.	0.544		0.537		0.353		0.377	
n	10,704		10,704		8,819		8,819	

Notes. Models (3), columns 1 and 3, and (4), columns 2 and 4, estimated using HRS wave 3-12 respondents aged 54-59 at *q*62 and aged 57-62 at *q*65 with full item response on subjective probabilities and shown covariates used in the analyses, and for whom it is possible to determine if they worked full-time after age 62 or 65, respectively.

Table A4. Heterogeneity in prediction inaccuracy (MSE) without controls for sex, age, marital status and race

	(1)		(2)	
A. *q*62				
Education (ref. College graduate)				
High school dropout or GED	0.084	(0.011)		
High school graduate	0.041	(0.008)		
Some college	0.040	(0.008)		
Constant	0.231	(0.006)		
Cognitive functioning (ref. Top quartile)				
Bottom quartile			0.057	(0.010)
2nd Bottom quartile			0.016	(0.009)
2nd Top quartile			0.007	(0.009)
Constant			0.247	(0.006)
A. *q*65				
Education (ref. College graduate)				
High school dropout or GED	0.054	(0.012)		
High school graduate	0.016	(0.009)		
Some college	0.031	(0.009)		
Constant	0.228	(0.006)		
Cognitive functioning (ref. Top quartile)				
Bottom quartile			0.044	(0.011)
2nd Bottom quartile			0.018	(0.009)
2nd Top quartile			0.007	(0.010)
Constant			0.232	(0.007)

Notes: n = 10,704 for panel A and 8,819 for panel B. Columns (1) and (2) show MSE of the subjective probability of *q*62 or *q*65 by educational attainment group and cognition score quartile group, respectively. Robust standard errors in parentheses.

Table A5. Inappropriate weighting of retirement factors by education

	***q*62**				***q*65**			
	Below high school	High school graduate	Some college	College graduate	Below high school	High school graduate	Some college	College graduate
Total $\Delta\hat{y} - \Delta\hat{p}$	**0.026**	**0.021**	**0.031**	**0.004**	**0.045**	**0.023**	**0.028**	**0.021**
	(0.012)	**(0.007)**	**(0.008)**	**(0.005)**	**(0.012)**	**(0.007)**	**(0.008)**	**(0.005)**
Contributions								
Health status	0.000	0.015	0.013	0.005	0.013	0.017	0.018	0.008
	(0.005)	(0.005)	(0.005)	(0.003)	(0.007)	(0.005)	(0.006)	(0.004)
Financial characteristics	0.006	0.001	0.002	-0.001	0.001	0.002	0.005	0.002
	(0.004)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.003)	(0.002)
Job characteristics	0.006	-0.005	0.011	-0.006	0.009	-0.003	0.002	0.007
	(0.005)	(0.003)	(0.005)	(0.003)	(0.007)	(0.003)	(0.004)	(0.005)
Sociodemographics	0.014	0.010	0.005	0.006	0.023	0.007	0.003	0.004
	(0.008)	(0.005)	(0.005)	(0.004)	(0.010)	(0.005)	(0.006)	(0.005)
n	1,474	3,038	2,971	3,221	1,333	2,549	2,337	2,600

Notes: Top row gives $\Delta\hat{y} - \Delta\hat{p}$ for each education group. See notes to 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 predictors included in each set. Because we stratify by education, we do not include the education dummies in the regressions. Bootstrap standard errors (100 simulations) in parentheses.

Appendix B. Robustness analyses

Table B1. Robustness of decomposition of prediction inaccuracy and discrimination to different definitions of outcomes

	*q*62				*q*65			
	Baseline (1)	*y* 3-year window (2)	*y* 7-year window (3)	*y* no deaths (4)	Baseline (1)	*y* 3-year window (2)	*y* 7-year window (3)	*y* no deaths (4)
A. MSE $\frac{1}{n}\sum(p_i - y_i)^2$	**0.265**	**0.265**	**0.265**	**0.261**	**0.249**	**0.247**	**0.250**	**0.250**
Decomposition, eq.(2)								
outcome variance $Var(y)$	0.249	0.249	0.248	0.248	0.235	0.233	0.236	0.238
bias² $(\bar{p} - \bar{y})^2$	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.001
covariance $-2(\Delta p)Var(y)$	-0.117	-0.118	-0.117	-0.119	-0.112	-0.111	-0.112	-0.115
signal $(\Delta p)^2 Var(y)$	0.014	0.014	0.014	0.014	0.013	0.013	0.013	0.014
noise $Var(p) - (\Delta p)^2 Var(y)$	0.120	0.119	0.120	0.118	0.112	0.112	0.112	0.112
B. Discrimination slope ΔP	**0.235**	**0.237**	**0.235**	**0.241**	**0.239**	**0.239**	**0.238**	**0.242**
Decomposition, eq.(5)								
outcome predictability $\Delta\hat{y}$	0.064	0.066	0.064	0.060	0.072	0.071	0.071	0.068
inappropriate weighting $-(\Delta\hat{y} - \Delta\hat{p})$	-0.012	-0.013	-0.012	-0.009	-0.018	-0.017	-0.017	-0.014
$100(\Delta\hat{y} - \Delta\hat{p})/\Delta\hat{y}$	18.6%	20.0%	18.5%	14.5%	24.5%	24.0%	24.1%	20.8%
private information $\Delta\hat{\varepsilon}$	0.183	0.185	0.183	0.189	0.185	0.185	0.184	0.188
$\bar{y}$	0.537	0.526	0.540	0.552	0.377	0.369	0.380	0.392
$\bar{p}$	0.544	0.544	0.544	0.545	0.353	0.353	0.353	0.355
n	10,704	10,704	10,704	10,411	8,819	8,819	8,819	8,476

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 window – up to 3 years and up to 7 years, respectively, after age 62 or 65 – used to define the outcome. Baseline using up to 5 years. Column (4) shows estimates after dropping from the sample respondents who die before reaching the age threshold.

Table B2. Social security and pension wealth variables and means for samples used to estimate models in Table B3

Variable	Definition	*q*62 Mean	*q*62 n	*q*65 Mean	*q*65 n	Wave
Individual social security (SS) wealth	Present value of respondent's predicted SS wealth based on the retirement insurance benefit, calculated based on respondent's own earnings records, assuming a claim age of 62 (for *q*62) or 65 (for *q*65).	158,147	2,097	174,845	1,630	7, 10
Household social security (SS) wealth	Sum of the respondent and spouse's individual SS wealth, including incremental spouse and survivor benefits, all assuming both respondent and spouse claim at age of 62 (for *q*62) or 65 (for *q*65).	245,669	2,097	272,689	1,630	7, 10
Defined benefit (DB) value at age 62 or 65	Present values of pension wealth from DB plans from current jobs if retired at age of 62 (for *q*62) or 65 (for *q*65). Present values are calculated using the HRS Pension Estimation Program using information on respondents matched DB plans or imputed values for respondents without a matched plan from the employer survey but who do indicate to have a DB plan.	82,453	2,097	79,803	1,630	7, 10
Defined benefit (DB) value at *t*	Values of pension wealth from DB plans from current jobs if retire at age *t*. Values are calculated using the HRS Pension Estimation Program using information on respondents matched DB plans or imputed values for respondents without a matched plan from the employer survey but who do indicate to have a DB plan.	60,825	1,224	62,650	977	10
Defined contribution (DC) value	Reported, or imputed for those respondents indicating to have DC wealth but no values are given in the employer survey, DC values from all DC and/or combination accounts from respondent's current job.	60,535	2,097	59,484	1,630	7, 10
Household income	Total income for the last calendar year. It is the sum of respondent and spouse earnings, pensions and annuities, Supplemental Security Income and Social Security Disability, Social Security retirement, unemployment and workers compensation, other government transfers, household capital income, and other income.	93,754	2,097	92,316	1,630	7, 10

Notes: Notes as in Table A1, unless otherwise specified below or in the Table above. All amounts are in US dollars. See https://hrsdata.isr.umich.edu/data-products/cross-wave-prospective-social-security-wealth-measures-pre-retirees and https://hrsdata.isr.umich.edu/data-products/cross-wave-prospective-social-security-benefit-wealth-measures-pre-retirees-waves-10-13 for detailed data description and codebook of the social security estimates used. See https://hrsdata.isr.umich.edu/data-products/employer-sponsored-pension-wealth-current-jobs-2004 and https://hrsdata.isr.umich.edu/data-products/employer-sponsored-pension-wealth-current-jobs-2010 for detailed data description and codebook of the pension wealth variables used.

Table B3. Robustness of decomposition of prediction inaccuracy and discrimination to including information on social security (SS) and pension wealth

		q62					*q65*				
		If retire at 62			If retire at t		If retire at 65			If retire at t	
		Baseline	Ind. SS and pension wealth	HH SS and ind. pension wealth	Baseline	Ind. SS wealth and pension wealth if retire at t	Baseline	Ind. SS and pension wealth	HH SS and ind. pension wealth	Baseline	Ind. SS wealth and pension wealth if retire at t
		(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
A. MSE	$\frac{1}{n}\sum(p_i - y_i)^2$	**0.255**	**0.255**	**0.255**	**0.240**	**0.240**	**0.258**	**0.258**	**0.258**	**0.260**	**0.260**
Decomposition, eq.(2)											
outcome variance	$Var(y)$	0.247	0.247	0.247	0.243	0.243	0.239	0.239	0.239	0.243	0.243
bias^2	$(\bar{p}-\bar{y})^2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
covariance	$-2(\Delta p)Var(y)$	-0.115	-0.115	-0.115	-0.124	-0.124	-0.107	-0.107	-0.107	-0.105	-0.105
signal	$(\Delta p)^2 Var(y)$	0.013	0.013	0.013	0.016	0.016	0.012	0.012	0.012	0.011	0.011
noise	$Var(p)-(\Delta p)^2 Var(y)$	0.110	0.110	0.110	0.105	0.105	0.114	0.114	0.114	0.111	0.111
B. Discrimination slope	ΔP	**0.233**	**0.233**	**0.233**	**0.255**	**0.255**	**0.223**	**0.223**	**0.223**	**0.217**	**0.217**
Decomposition, eq.(5)											
outcome predictability	$\Delta\hat{y}$	0.079	0.094	0.093	0.092	0.122	0.080	0.084	0.083	0.097	0.108
inappropriate weighting	$-(\Delta\hat{y}-\Delta\hat{p})$	-0.025	-0.031	-0.031	-0.031	-0.044	-0.035	-0.037	-0.036	-0.043	-0.048
	$100(\Delta\hat{y}-\Delta\hat{p})/\Delta\hat{y}$	31.3%	33.2%	33.2%	33.7%	36.5%	43.7%	44.4%	43.7%	43.6%	44.7%
private information	$\Delta\hat{\varepsilon}$	0.179	0.170	0.171	0.194	0.178	0.178	0.176	0.176	0.162	0.158
$\bar{y}$		0.554	0.554	0.554	0.583	0.583	0.397	0.397	0.397	0.414	0.414
$\bar{p}$		0.555	0.555	0.555	0.571	0.571	0.390	0.390	0.390	0.406	0.406
n		2,097	2,097	2,097	1,224	1,224	1,630	1,630	1,630	977	977

Notes: Column (1) gives baseline estimates for sample in (2) and (3). In columns (2) and (3), we add social security (SS) and pension wealth variables to the baseline specification of models (3) and (4). Column (3) adds present values of individual SS wealth using a claim age of 62 for *q62* and 65 for *q65*, present values of defined benefit (DB) pension wealth if retire at age 62 for *q62* and 65 for *q65* and defined contribution (DC) pension wealth values at time t. Column (3) is similar to column (2), except that it adds present values of household SS wealth and includes total household income instead of individual earnings. Column (4) gives baseline estimates for sample in (5). Column (5) is again similar to column (3), except that it adds values of defined benefit (DB) pension wealth if retire at t. Table B2 presents definitions and means of these variables. Please note that the sample sizes are smaller because the variables included are not observed in all waves (see Table B2).

Part 2: Health behaviors

Chapter 4

Does changing health behavior explain the falling gender gap in mortality in Russia?

Joint work with Teresa Bago d'Uva

Abstract

Around the world, women live longer than men, particularly in Russia, where there was a 13-year difference between female and male life expectancy in the early 2000s. This gender gap in life expectancy fell by 2 years over the next decade that saw implementation of tobacco and alcohol control policies. We use longitudinal data and nonlinear Oaxaca-Blinder decomposition to estimate contributions of health behaviors to the gender gap in 5-year mortality and how these contributions changed between 2000-2003 and 2010-2013. We show that gender differences in smoking, drinking, diet and exercise explain 49% of the gender gap in mortality in 2000-2003 but only 29% in 2010-2013. Differences in smoking contribute most to the mortality gap but this contribution fell from 41% to 28%. The gender difference in alcohol use contributes only about a fifth of that of smoking to the gender gap in each period. These results are consistent with men's declining tobacco and alcohol consumption, which may possibly be due to the control policies implemented, being the main behaviors that explain the narrowing of the gender mortality gap.

4.1 Introduction

Women outlive men virtually everywhere (Barford et al., 2006; Crimmins et al., 2019; Kolip & Lange, 2018; Luy & Gast, 2014), although large differences exist across countries and over time (Cullen et al., 2015; Oksuzyan et al., 2008). Explanations cite biological and behavioral differences between men and women (Luy, 2003; Rieker & Bird, 2005; Okzusyan et al., 2008). Men and women differ in many health behaviors (Luy, 2003; Nathanson, 1984; Schünemann et al., 2017). Men are more likely to drink alcohol, eat unhealthy food, and to engage in risky behaviors, such as dangerous driving (Luy, 2003). They are also less likely to invest in their health and to seek health care (Green & Pope, 1999), and they are exposed to greater physical health risks at work (Emslie & Hunt, 2008). Case & Paxson (2005) show that the gender gap in mortality in the US can largely be explained by differences in the distribution of chronic conditions and the greater mortality risks associated with these conditions for men. Particularly, men are more likely to die from smoking-related conditions, due to their higher rates of lifetime smoking. Schünemann et al. (2017) estimate that almost 80% of the gender gap in US life expectancy can be explained by differences in health behavior. Ross et al. (2012) show that the gender gap in US mortality is highest at low levels of education and, importantly, that the education gradient in mortality is steeper for males. This latter difference is especially pronounced for deaths related to destructive and risky behaviors, such as deaths from respiratory disease, lung cancer, stroke, suicide and accidents.

This paper estimates contributions of gender differences in unhealthy behaviors to the gender gap in mortality in Russia, a country where the male-female gap in life expectancy is particularly pronounced and where unhealthy behaviors, such as drinking alcohol and smoking, are prevalent and especially high among men. Figure 1 shows the striking gender difference in life expectancy in Russia, which has remained above 10 years since the beginning of the 21st century. Until very recently, Russians, particularly men, were among the largest consumers of alcohol in Europe (WHO, 2023). Hard liquor accounts for over half of Russian alcohol consumption and a substantial part of this is produced and sold in the informal market (Leitzel, 2022). Russia also has one of the highest smoking rates in the world, especially among men (Shkolnikov et al., 2020). In 2018, 45% of Russian males aged 15+ reported to smoke daily (OECD, 2023). Notably, Russia also has a high prevalence of young smokers (Inchley et al., 2016).

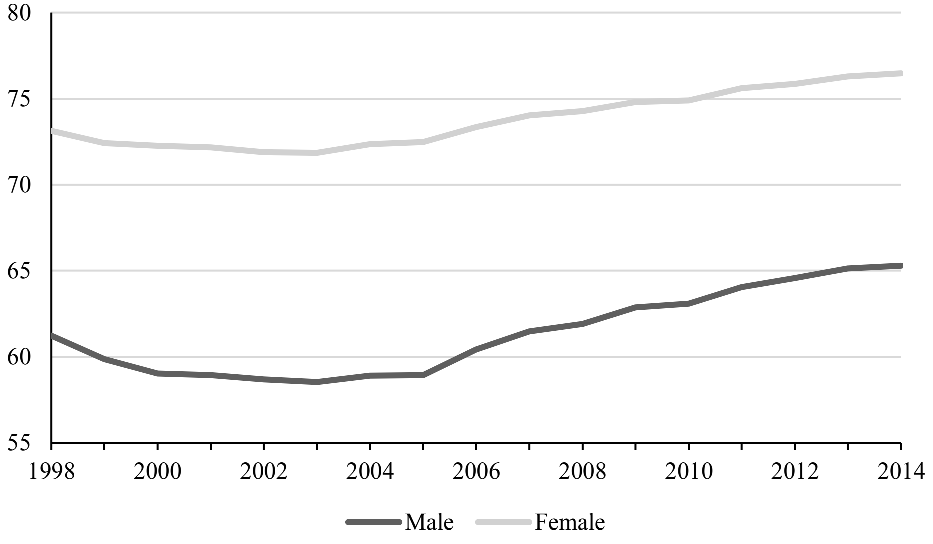


Figure 1. Male and female life expectancy at birth in the Russian Federation
Note: Data obtained from the World Bank (2023).

The prevalence of unhealthy behaviors has led the Russian government to start introducing in 2003 a battery of alcohol and tobacco control regulations. Excise taxes on alcohol have steadily increased, and minimum prices for alcoholic beverages were introduced, starting in 2003 with vodka, and were raised gradually over the years (Gil et al., 2016; WHO, 2019). Additionally, it started monitoring the production and sale of alcohol and introduced hours-of-sale restrictions and advertising bans (Leitzel, 2022; WHO, 2019). In the late 2000s, it started implementing tobacco control policy. Russia ratified the Framework Convention on Tobacco Control[27] in 2008, and since then has increased excise taxes on tobacco, banned advertising and enforced public places to be 100% smoke-free (Leitzel, 2022; Shkolnikov et al., 2020; WHO, 2017). Figure 1 shows a reduction in the gender gap in life expectancy of about 2 years from 2000 to after 2010. This is driven by the steeper rise in life expectancy for men after 2005, which may possibly be explained by the implemented policies. Indeed, smoking prevalence and alcohol consumption have declined following the implementation of the policies (Rehm & Ferreira-Borges, 2018).

This paper is the first to assess the contribution of unhealthy behaviors – smoking, drinking, diet and exercise – to the gender gap in mortality in Russia in the early 2000s (2000-2003), when there were practically no anti-alcohol and tobacco measures in place, and in the early

[27] Treaty of the World Health Organisation (WHO) to reduce the consumption of tobacco.

2010s (2010-2013), a period marked by various alcohol and tobacco control policies. Using data from the Russian Longitudinal Monitoring Survey (RLMS) from 2000-2008 and 2010-2018, we show that the gender gap in 5-year mortality has narrowed considerably, from 6.4 percentage points in the early 2000s to 3.8 percentage points in the early 2010s. Along with it, there has been a noticeable decline in gender differences in health behaviors. Using nonlinear Oaxaca-Blinder decomposition, we find that gender differences in health behaviors explain a substantial part of the gender mortality gap in the early 2000s (49%) and the early 2010s (29%). Gender differences in smoking behavior account for most of the contribution of behaviors to the gender mortality gap – 41% of the mortality gap in the early 2000s, which fell to 28% in the next decade. Perhaps somewhat surprisingly, men's heavy drinking seems to explain less of the gender gap – differences in alcohol consumption explain 9% of the gap in the early period, and 3% in the later period. Our results are consistent with men's declining tobacco and alcohol consumption being the main behaviors that explain the narrowing of the gender mortality gap in this period. These declines may possibly be explained by the implemented policies.

Our results add to previous literature investigating the gender mortality gap in Russia (Kossova et al., 2019; Luy & Wegner-Siegmundt, 2014; Trias-Llimós & Janssen, 2018). Luy & Siegmundt (2014) find that from 1980-2009 gender differences in smoking-related deaths explain around 40% (more than 4 years) of the gender gap in life expectancy. This contribution has been relatively stable over that period. Trias-Llimós & Janssen (2018) focus instead on alcohol consumption and show that differences in alcohol-attributable mortality explain 17% (2 years) of the gender gap in life expectancy from 1990-2012. Importantly, they find a declining contribution of alcohol since 2005. Few papers, however, investigate the contribution of both behaviors simultaneously. This is important since excessive drinkers are likely to smoke (Cockerham et al., 2012), and causes of death (partly) attributable to smoking or alcohol often overlap (e.g., ischemic heart disease) (Luy & Wegner-Siegmundt, 2014; Trias-Llimós & Janssen, 2018).

Kossova et al. (2020) investigate the relationship of the gender gap in life expectancy both with drinking (measured by alcohol sales) and smoking (proxied by deaths from respiratory diseases). They find a positive relationship of the gender gap with both behaviors in 1998-2015. Their analysis does not, however, enable a direct comparison of the relative contribution of each. McCartney et al. (2011) analyzed the contribution of smoking-related and alcohol-related deaths to the gender gap in mortality in several European countries. Their study did not include Russia. However, for arguably comparable countries such as Latvia, Lithuania and Ukraine, the gender differences in smoking-attributable mortality account for more than 40% of the gender gap in the mid-2000s; alcohol-attributable mortality explains about half of that.

We address the shortcomings of previous studies on Russia by decomposing the gender gap in 5-year mortality using self-reported health behaviors – smoking, drinking, diet and exercise. This adds to the existing evidence by providing the first comprehensive analysis of the contribution of these different health behaviors to the gender gap in Russia, and thus revealing the conditional importance of each for explaining the gender gap.

This paper also relates to the literature which studies declines in Russian mortality following the implementation of the alcohol and tobacco regulations. Grigoriev et al. (2014) show that the mortality decline from 2003-2011 can be largely attributed to a decline in cardiovascular mortality. Based on the sharp drop in deaths from accidental alcohol poisonings, they argue that the decrease in overall alcohol consumption, and the even steeper decrease in the consumption of hard liquor, is for a large part responsible for this so-called 'cardiovascular revolution' and especially for the drop in premature mortality. In contrast, due to the small decline of tobacco use between 2003 and 2011, they hypothesize that reductions in the prevalence of smoking contributed little to the mortality decline. Similarly, Danilova et al. (2020) find that a substantial part of the gain in life expectancy between 2003-2017 is related to a decline in the prevalence of harmful alcohol consumption, although its influence on life expectancy has declined in more recent years, with other factors becoming more important. Shkolnikov et al. (2020) show that since 2008 there has been a steady decline in, especially male, smoking prevalence. Although these declines are unlikely to have contributed to the decrease in female mortality, they can explain a small part of the decline in male mortality since the mid-2000s (Shkolnikov et al., 2020). Although these reductions coincide with the implementation of anti-alcohol and tobacco measures in Russia, it is not clear that these measures are driving the declines. Grigoriev & Andreev (2015) find that the alcohol control regulations contributed to the decline in male mortality, but cannot fully explain it.

We add to this literature by decomposing the gender gap in mortality for two periods a decade apart; 2000-2003, when there were practically no anti-alcohol and tobacco measures in place, and 2010-2013, a period marked by various alcohol and tobacco control policies. This quantifies the change in the gender gap and the contribution of health behaviors to the gap over time.

The contribution of this paper is thus twofold. First, by using data on self-reported health behaviors and nonlinear Oaxaca-Blinder decomposition, we are able to show that gender differences in smoking explain a larger part of the gender gap in 5-year mortality in Russia than gender differences in alcohol consumption. This is the first paper to quantify conditional contributions of health behaviors to the gender mortality gap in Russia. Second, we show that that the gender gap in mortality has declined from the early 2000s to the early 2010s, and with it the contributions of gender differences in smoking and alcohol to the gap. These changes over time may possibly be explained by the implemented alcohol and tobacco control policies during this decade.

4.2 Data & Methodology

4.2.1 Data

We use data from 18 rounds (2000-2008 and 2010-2018) of the Russian Longitudinal Monitoring Survey (RLMS, 2019). The RLMS is a nationally representative survey designed to collect information on the health and economic welfare of households and individuals in the Russian Federation.[28] The data are obtained via face-to-face interviews using two main questionnaires – an individual questionnaire for all adults and children and a household questionnaire. Each year the data include two types of samples, a longitudinal and a cross-sectional sample. This paper uses data from the longitudinal follow-up sample, which provides information on the deaths of household members.[29] This sample is supplemented each year by households in the cross-sectional sample who move or split, and by replenishments over time of the cross-sectional sample. The cross-sectional sample is nationally representative, the panel study is not necessarily so.

The data include a broad range of socioeconomic and health information, such as on health status and health behaviors. Mortality data is collected via the household questionnaire. In each round, respondents are asked whether any household members have died since the previous round. The deceased household members constitute the recorded deaths for our mortality analysis. We analyze 5-year mortality, i.e., an indicator of whether or not an individual died within 5 years following a survey round. Individuals in our sample may be present in multiple years across both periods; the 5-year follow-up period moves with each observation. We exclude individuals who lived in a single household at any point during the relevant 5-year period, as in principle there is no one to report their deaths.[30] It is common procedure in papers analyzing mortality data from the RLMS to exclude single households (e.g., Brainerd & Cutler, 2005; Perlman & Bobak, 2008). Our sample thus consists of panel observations from 2000-2003 and 2010-2013 for which we can observe mortality over the next 5 years and for which we can establish whether individuals lived alone at any point

[28] Detailed information about the sampling design and implementation is available at http://www.cpc.unc.edu/projects/rlms-hse/project/sampling.
[29] More detailed information can be found in Appendix A.
[30] Our results are robust to excluding only observations in the given year(s) that individuals are single, rather than the full 5-year period in case they are single at any point herein (see Appendix C, Table C.3).

during this period. Mortality trends in the RLMS are similar to national Russian mortality trends, albeit at lower levels (Brainerd & Cutler, 2005; Perlman & Bobak, 2008).

We decompose gender differences in 5-year mortality as a function of differences in health behaviors and socioeconomic control variables. Health behaviors include alcohol consumption, smoking, physical exercise and Body Mass Index (BMI). Frequency of alcohol consumption is defined in the following response categories of the question "How often have you consumed alcoholic beverages in the last 30 days?": none in the last 30 days (never or infrequent drinkers), 1-3 times in the last 30 days (occasional drinkers), 1-3 times a week (moderate drinkers) and 4-7 times a week (frequent or excessive drinkers). We also measure binge drinking, based on information on the quantity of alcohol usually consumed in one sitting. We follow the WHO definition that defines binge drinking as consuming at least 60 grams of pure alcohol on at least one occasion during the last 30 days.[31] Individual smoking behavior is measured in the following categories: never smokers, past smokers and current smokers. Leisure-time physical exercise (respondents are specifically asked not to count physical activities at work) is coded into three categories: no exercise, light physical exercise for relaxation fewer than three times a week and the last category which includes moderate and intensive exercise (specifically, medium and intensive physical exercise fewer than three times a week, intensive physical exercise at least three times a week for 15 minutes or more, and daily exercise not less than 30 minutes a day). Lastly, BMI is constructed using self-reported height and weight, and is used to classify individuals as either underweight, normal weight, overweight, or obese.[32] BMI proxies eating behavior, especially fat or calorie intake, and physical exercise not captured by the leisure-time exercise variable, such as work-related physical activity.

Age, education, household income and area of residence are added as sociodemographic control variables. This is done to ensure that the contributions of health behaviors do not capture the contribution of other (health) behaviors correlated with the socioeconomic variables. We model the relationship between age and mortality in a flexible way, using 5-year age intervals (except for 18-24). Education is a categorical variable containing the highest level of education achieved. Some low frequency educational levels were grouped, resulting in the following three categories: less than secondary education (primary education

[31] The RLMS includes detailed information on which types of alcohol (beer, wine, hard liquor, other alcohol) respondents consumed in the last 30 days and on how many grams of this type of alcohol they usually consumed in a day. We use alcohol percentages reported in Kuznetsova (2020) to obtain grams of pure alcohol consumed from beer, wine, hard liquor, and other alcohol consumption. From 2006 onward, the RLMS asks how many days per month individuals usually consume the different types of alcohol. With this additional information, we can estimate total pure alcohol consumption per month for the period 2010-2013. Using this more complete measure of alcohol consumption leaves our results for this period unchanged (Appendix C, Tables C.4 and C.5).

[32] BMI is calculated by dividing weight in kilograms by height in meters squared.

and incomplete secondary education), secondary education (general or vocational) and higher education (Bachelor, Master or Doctoral degree). Real monthly equivalized household income includes all sources of income received by household members (and uses as equivalence scale the square root of the household size[33]). Household income per household member is subsequently categorized into quartiles. Area of residence is measured by two variables: an indicator of whether the individual lives in an urban or rural area; and the respective economic region (Central, Ural, North Caucasus, Volga, West Siberian, East Siberian, Volga-Vyatka, Northwestern, Central Black Earth, Far Eastern or Northern). Lastly, we include year fixed effects. The samples used in the analyses result from dropping observations with missing values on any of the variables used and contain 8,635 (6,851) observations of women (men) for 2000-2003 and 13,535 (10,593) observations of women (men) for 2010-2013.

Table 1 shows the sample means of the variables used in the analyses for 2000-2003 and 2010-2013 by gender. There are substantial differences between men and women, in line with the existing literature: men have a higher mortality rate than women, but this gender gap in mortality has decreased over time. In the early 2000s, men were 6.4 percentage points more likely to die within 5 years than women. By the early 2010s, the male-female mortality gap dropped to 3.8 percentage points. Interestingly, this decrease seems to be mostly driven by a reduction in male mortality over time. Table 1 also shows unhealthier behaviors of men. Men are more likely to drink alcohol at least once per week and much more likely to binge drink, which is mostly driven by their hard liquor consumption. Men are also substantially more likely to smoke. These differences have however decreased over time, and most of the decline is due to a reduction in men's alcohol and tobacco use (while women's tobacco use increased in this period). Men decreased their frequency of alcohol use, but mostly their binge drinking. Interestingly, we also see that men in the later period were less likely to binge drink hard liquor, but more likely to binge drink beer. This indicates a shift in men's alcohol consumption towards less alcohol-dense beverages.

The reduction in men's unhealthy behaviors could explain part of the decrease in male mortality and the gender gap. On the other hand, men seem to exercise more frequently and intensively than women, although the vast majority of both men and women do not engage in leisure-time exercise. Over time, both men and women engage more in leisure-time exercise, with a larger increase among women. We also observe a healthier BMI for men than women, which is best reflected in their lower obesity rate. The incidence of obesity, however, has increased more steeply for men, while at the same time men's prevalence of

[33] Equivalence scale used in OECD publications (http://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf).

normal weight has increased. It thus seems unlikely that changes in diet and exercise can explain a significant part of the decrease in the gender gap in mortality, in contrast with the gender-specific changes observed in smoking and drinking.

Table 1. Sample means by gender in early 2000s and early 2010s

	Early 2000s			*Early 2010s*			*Early 2010s – Early 2000s*	
	Women	Men	Men – Women	Women	Men	Men – Women	Δ Women	Δ Men
5-year mortality	0.042	0.106	0.064**	0.048	0.086	0.038**	0.006*	-0.020**
No alcohol in the last month	0.487	0.287	-0.200**	0.558	0.364	-0.193**	0.071**	0.077**
Alcohol 1-3 times last month	0.386	0.315	-0.071**	0.326	0.280	-0.046**	-0.060**	-0.035**
Alcohol 1-3 times a week	0.120	0.336	0.216**	0.109	0.308	0.199**	-0.010*	-0.028**
Alcohol 4-7 times a week	0.007	0.062	0.055**	0.006	0.047	0.041**	-0.001	-0.014**
Binge drinker	0.171	0.553	0.381**	0.134	0.438	0.304**	-0.038**	-0.114**
Beer	0.010	0.078	0.069**	0.015	0.107	0.092**	0.005**	0.029**
Wine	0.019	0.027	0.008**	0.029	0.031	0.002	0.010**	0.003
Hard liquor	0.148	0.524	0.377**	0.101	0.378	0.276**	-0.047**	-0.147**
Other alcohol	0.012	0.014	0.003	0.003	0.006	0.003**	-0.009**	-0.008**
Never smoker	0.795	0.191	-0.605**	0.781	0.251	-0.529**	-0.014*	0.061**
Past smoker	0.079	0.188	0.109**	0.077	0.197	0.120**	-0.002	0.008
Current smoker	0.126	0.621	0.495**	0.142	0.552	0.410**	0.016**	-0.069**
No exercise	0.842	0.803	-0.040**	0.807	0.786	-0.020**	-0.036**	-0.016**
Light exercise	0.092	0.093	0.001	0.109	0.086	-0.023**	0.018**	-0.006
Moderate or intensive exercise	0.066	0.105	0.039**	0.084	0.127	0.043**	0.018**	0.023**
BMI	26.7	25.1	-1.560**	27.1	26.1	-1.023**	0.428**	0.964**
Underweight	0.032	0.022	-0.010**	0.034	0.015	-0.019**	0.002	-0.007**
Normal weight	0.291	0.315	0.024**	0.297	0.372	0.075**	0.006	0.057**
Overweight	0.415	0.541	0.126**	0.382	0.443	0.060**	-0.033**	-0.098**
Obese	0.262	0.121	-0.140**	0.287	0.170	-0.117**	0.025**	0.049**
Age	43.4	44.1	0.729**	45.7	44.5	-1.230**	2.278**	0.319
Less than secondary education	0.157	0.208	0.051**	0.159	0.212	0.053**	0.002	0.004
Secondary education	0.660	0.628	-0.032**	0.577	0.586	0.009	-0.082**	-0.042**
Higher education	0.184	0.165	-0.019**	0.264	0.202	-0.062**	0.081**	0.038**
Lowest income quartile	0.259	0.240	-0.020**	0.253	0.251	-0.001	-0.007	0.012†
2nd income quartile	0.247	0.257	0.010	0.250	0.246	-0.004	0.003	-0.012†
3rd income quartile	0.249	0.248	-0.001	0.247	0.256	0.009	-0.002	0.008
Highest income quartile	0.245	0.255	0.010	0.251	0.247	-0.004	0.006	-0.008
Rural	0.294	0.315	0.021**	0.301	0.322	0.021**	0.007	0.008
N	8,635	6,851		13,535	10,593			

Notes: Early 2000s (2010s) corresponds to 2000-2003 (2010-2013). Binge drinking (beer) is defined as consuming at least 60 grams of pure alcohol (from beer consumption) on at least one occasion in the last 30 days. Detailed definitions of measures given in paper. †$p < .10$; *$p < .05$; **$p < .01$

4.2.2 Methodology

We use nonlinear Oaxaca-Blinder decomposition analysis to assess to what extent health behaviors explain the gender gap in 5-year mortality. Five-year mortality is modelled as a function of health behaviors and socioeconomic control variables, x, using probit models, which results in the following predicted probability of dying within 5 years of year *t* for individual *i* with gender *j* (male, female):

$$\hat{y}_{ijt} = \widehat{Pr}\left(died_{ijt} \mid x_{ijt}\right) = \Phi\left(x_{ijt}\hat{\beta}_j\right) \quad (1)$$

and where $\Phi(.)$ represents the cumulative distribution function of the standard normal distribution, and $\hat{\beta}$ is a vector of gender-specific estimated model coefficients.

An extension of the Blinder-Oaxaca decomposition method for nonlinear models (Yun, 2004) is used to estimate the contribution of the explanatory variables to the gender difference in predicted probabilities of dying. The difference between the mean predicted probability for men ($\overline{\hat{y}_m}$) and women ($\overline{\hat{y}_w}$) can be decomposed as follows:

$$\overline{\hat{y}_m} - \overline{\hat{y}_w} = \overline{\Phi\left(x_m\hat{\beta}_m\right)} - \overline{\Phi\left(x_w\hat{\beta}_w\right)} = \underbrace{\left\{\overline{\Phi\left(x_m\hat{\beta}_m\right)} - \overline{\Phi\left(x_w\hat{\beta}_m\right)}\right\}}_{E} + \underbrace{\left\{\overline{\Phi\left(x_w\hat{\beta}_m\right)} - \overline{\Phi\left(x_w\hat{\beta}_w\right)}\right\}}_{C} \quad (2)$$

where $\overline{\Phi\left(x_j\hat{\beta}_j\right)} = {}^{1}/_{n_j} \sum_{i \in j} \Phi\left(x_{ij}\hat{\beta}_j\right)$, with j = male, female and x_j are covariates for gender *j*, and $\overline{\Phi\left(x_w\hat{\beta}_m\right)} = {}^{1}/_{n_w} \sum_{i \in w} \Phi\left(x_{iw}\hat{\beta}_m\right)$; subscript *t* is suppressed to avoid clutter. E denotes the overall endowment effect, the estimated change in the gender gap associated with men having women's endowments, and *C* the overall coefficient effect, the estimated change in the gender gap associated with men having women's estimated coefficients. A positive endowment (coefficient) effect represents a reduction in the gender 5-year mortality gap related to men and women having the same characteristics (coefficients). As shown in equation (2), we assess the endowment (coefficient) effect by fixing the coefficients (characteristics) to those of men (women).[34]

Eq. (2) decomposes the gender mortality gap in aggregate contributions of differences in characteristics and coefficients. We are however also interested in the individual contribution of each explanatory variable, particularly that of health behaviors, which requires detailed decomposition analysis. This gives individual endowment and coefficient

[34] Appendix C, Table C.6 shows the decomposition of the negative female-male mortality differential which fixes the coefficients (characteristics) to those of women (men). This is the reverse of the fixing of our baseline estimates. Our main conclusions remain in that reversed decomposition.

contributions of each explanatory variable x_k, E_k and C_k, respectively. One way to calculate E_k and C_k is by sequentially replacing one group's characteristics and coefficients, respectively, with those of the other group. However, in the case of nonlinear models, the results are sensitive to the order in which the explanatory variables are substituted. Yun (2004) proposed a simple solution to this path dependency problem by using weights derived from a first-order Taylor linearization of the aggregate coefficient (*C*) and endowment (*E*) effects around the means of the explanatory variables, $\overline{x}_w\hat{\beta}_w$ and $\overline{x}_m\hat{\beta}_m$.[35] The resulting decomposition formula, invariant to the order of sequential substitution, is as follows:

$$\overline{\hat{y}_m} - \overline{\hat{y}_w} = \Sigma_{k=1}^{K} W_{\Delta x_k} \left\{ \overline{\Phi(x_m\hat{\beta}_m)} - \overline{\Phi(x_w\hat{\beta}_m)} \right\} + \Sigma_{k=1}^{K} W_{\Delta\hat{\beta}_k} \left\{ \overline{\Phi(x_w\hat{\beta}_m)} - \overline{\Phi(x_w\hat{\beta}_w)} \right\} \tag{3}$$

with the following kth individual weight for *E*:

$$W_{\Delta x_k} = \frac{\hat{\beta}_{m_k}(\bar{x}_{m_k} - \bar{x}_{w_k})}{\Sigma_{k=1}^{K} \hat{\beta}_{m_k}(\bar{x}_{m_k} - \bar{x}_{w_k})} \tag{4}$$

and the following kth individual weight for *C*:

$$W_{\Delta\hat{\beta}_k} = \frac{\bar{x}_{m_k}(\hat{\beta}_{m_k} - \hat{\beta}_{w_k})}{\Sigma_{k=1}^{K} \bar{x}_{m_k}(\hat{\beta}_{m_k} - \hat{\beta}_{w_k})} \tag{5}$$

where $\Sigma_{k=1}^{K} W_{\Delta x_k} = \Sigma_{k=1}^{K} W_{\Delta\hat{\beta}_k} = 1$.

The gender gap in mortality can then be expressed as the sum of the unique contributions of the explanatory variables:

$$\overline{\hat{y}_m} - \overline{\hat{y}_w} = E + C = \Sigma_{k=1}^{K} W_{\Delta x_k} E + \Sigma_{k=1}^{K} W_{\Delta\hat{\beta}_k} C = \Sigma_{k=1}^{K} E_k + \Sigma_{k=1}^{K} C_k \tag{6}$$

Without further adjustments, in this detailed decomposition the coefficient effects of categorical variables would not be invariant to the choice of reference category of categorical explanatory variables. Specifically, changing the reference category of a given variable would change the sum of the coefficient effects across its categories due to a redistribution of coefficient effects between the constant and the categories. This is undesirable, given that the choice of the reference categories of explanatory variables is arbitrary. We use the simple and intuitive solution proposed by Yun (2005) to overcome this identification problem, which involves normalizing the probit regression coefficients of eq. (1). The normalized coefficient of each category of a given variable is defined as its deviation from the mean of the coefficients across all respective categories. This is obtained as follows:

[35] Yun's (2004) method only works for nonlinear models with once differentiable functional form.

$$\overline{\hat{\beta}_k} = \frac{\Sigma_{m_k=1}^{M_k} \hat{\beta}_{m_k}}{M_k} \tag{7}$$

where $\hat{\beta}_{m_k}$ is the estimated coefficient of category m_k of categorical variable k, with $m_k = 1, \dots, M_k$ and $\hat{\beta}_{m_k}$= 0 for the reference category. The normalized coefficients and constant term are defined as:

$$\hat{\beta}^*_{mk} = \hat{\beta}_{m_k} - \overline{\hat{\beta}_k} \tag{8}$$

$$\hat{\alpha}^* = \hat{\alpha} + \Sigma_{k=1}^{K} \overline{\hat{\beta}_k} \tag{9}$$

where subscript j is omitted for ease of presentation. The means of the coefficients of the categorical variables are added to the constant to sustain mathematical consistency. We use these normalized coefficients in our decomposition analyses, equations (2)-(6). Standard errors are clustered at the individual level throughout.

4.3 Results

Table 2 shows the decomposition of the gender gap in 5-year mortality in the early 2000s (columns 1 and 2) and the early 2010s (columns 3 and 4). Probit model estimation results for men and women and the respective normalized coefficients which are used in the decomposition are shown in Table B.1 in Appendix B. In the presentation of decomposition results, we follow the common terminology, namely endowment and coefficients effects, although we note that these should not be interpreted as causal effects.

In the early 2000s, men were predicted to be 6.4 percentage points more likely to die in the next 5 years than women. The overall decomposition shows that differences in endowments can explain 3.4 percentage points, 53%, of this gap. Smoking behavior, as indicated by the detailed decomposition, makes the largest contribution to the overall endowment effect: men adopting women's smoking behavior, while keeping constant the other covariates, is associated with a reduction of 2.6 percentage points (41%) in the male excess mortality. Perhaps surprisingly, the contribution of alcohol consumption is less than a fourth of that of smoking: men having women's drinking behavior is correlated with a decrease in the gender gap of 0.5 percentage points (9%). A substantial part of this contribution is explained by binge drinking; if this were the same for men and women, then the reduction in the gender gap related to it would be 0.7 percentage points (10%) (Appendix C, Table C.7). Leisure-time exercise does not appear to explain the gender gap in mortality, as expected from the relatively small gender difference in exercise frequencies (Table 1). The detailed endowment effect for leisure-time exercise is small and statistically insignificant at a 10% significance level. The overall effect of BMI is statistically significant at 5%, but the size is negligible.

Overall, differences in health behaviors explain half of the gender gap in the early 2000s, and most of this contribution is captured by gender differences in smoking behavior, and to a lesser extent by gender differences in alcohol consumption.

Differences in coefficients account for the remainder, namely 3 percentage points, of the gender difference in 5-year mortality in the early 2000s (47%). The detailed decomposition indicates that most of this effect (82%) is not explained by the coefficients of the explanatory variables included in this analysis but rather captured in the contribution of the constant term – it is therefore unrelated to the covariates in our model. Gender differences in leisure-time exercise coefficients are in the advantage of men: if men had women's adverse associations of not engaging in exercise (Table B.1), then this would be correlated with an increase in the gender mortality gap of 3 percentage points (49%). This large advantage of men is, however, not visible a decade later and the coefficient effect across all health behaviors is insignificant at 10%.

Table 2. Decomposition of gender difference in predicted 5-year mortality in early 2000s and early 2010s

	Early 2000s		*Early 2010s*	
	Absolute	Percent	Absolute	Percent
Male mortality	0.1061**		0.0858**	
Female mortality	0.0424**		0.0481**	
Male excess mortality	0.0637**		0.0377**	
Overall decomposition results				
Endowment effect	0.0336**	52.72	0.0081	21.34
Coefficient effect	0.0301**	47.28	0.0297**	78.66
Detailed endowment effects				
Alcohol consumption	0.0054*	8.53	0.0010	2.75
Smoking behavior	0.0261**	40.92	0.0104†	27.66
Exercise	-0.0001	-0.17	-0.0004	-1.04
BMI	-0.0002*	-0.26	-0.0002	-0.63
All health behaviors	0.0312**	49.02	0.0108	28.74
Other variables	0.0024**	3.70	-0.0028	-7.40
Detailed coefficient effects				
Alcohol consumption	-0.0045	7.00	0.0036	9.55
Smoking behavior	0.0052	8.16	0.0017	4.59
Exercise	-0.0283**	-44.37	-0.0039	-10.33
BMI	0.0051	7.98	0.0010	2.60
All health behaviors	-0.0224	-35.23	0.0024	6.41
Other variables	0.0006	1.02	0.0072	19.20
Constant	0.0519**	81.49	0.0200*	53.05
N	15,486		24,128	

Notes: As in Table 1. Decomposition based on differences in characteristics shown in Table 1 and normalized coefficients shown in Table B.1. Contribution of other variables is the sum of those of age, education, income, area (rural or urban), economic region and year. They are shown separately in Table C.7 in the Appendix. †$p < .10$; *$p < .05$; **$p < .01$

Decomposition results for the early 2010s are shown in the last two columns of Table 3. The gender gap in 5-year mortality has decreased considerably from 6.4 percentage points in the early 2000s to 3.8 percentage points in the early 2010s. While 3.4 (53%) of the 6.4 percentage point gap was explained by endowment differences in the early 2000s, this decreased substantially to 0.8 (21%) of the 3.8 gap a decade later. Moreover, this endowment effect is no longer significant. This suggests that, along with a decline in the mortality gap, gender differences in characteristics have become less important for explaining it. Specifically, this decline in the total endowment effect is almost fully captured by that of health behaviors – 3.1 percentage points (49%) in 2000-2003 to 1.1 (29%) of the gender gap in 2010-2013. This represents more than three-quarters of the decline in the mortality gap (6.4-3.8). While not representing causal effects, these results are consistent with a narrowing of the gender gap in health behaviors during that decade contributing to narrowing the gender gap in mortality.

Changes in smoking behavior and alcohol consumption between the early 2000s and the early 2010s explain most of that decrease in the endowment effect of health behaviors. In the early 2010s, gender differences in smoking behavior contributed 1 percentage point (28%) to the gender mortality gap, less than half the contribution a decade earlier, 2.61 percentage points (41%). The part of the mortality gender gap explained by differences in alcohol consumption decreased more than fivefold, from 0.54 to 0.10 percentage points (respectively, 9% and 3% of the 6.4 and 3.8 percentage point gender gaps), and became insignificant at 10%.

These declines in the contribution of smoking behavior and alcohol consumption are explained by different trends in these behaviors by gender. In this period, smoking rates declined among men, and slightly increased among women (Table 1). Men reduced their excessive alcohol consumption by more than women. These trends result in a narrower gender difference in these unhealthy behaviors. Additionally, in the early 2010s the adverse association of male smoking with mortality is smaller than in the early 2000s (Table B.1), which could be explained by a reduction of tobacco consumption among male smokers. Similarly, the association of binge drinking with mortality has also reduced among men, possibly reflecting a decrease in pure alcohol consumed by male binge drinkers. This is supported by Table 1, which shows a shift toward binge drinking less alcohol-dense beverages which could lead to a reduction in the amount of pure alcohol consumed in a drinking session. The overall endowment effects of diet and exercise remain in the late 2010s very small and both statistically insignificant at 10%.

Differences in coefficients explain the majority of the gender gap in the early 2010s (79%), but more than two-thirds of this is unrelated to the covariates in our model (2.0/3.0). Overall, gender differences in health behaviors make the largest discernible contribution to the mortality gap, and these contributions are substantially smaller than in the early 2000s. Men's reduced tobacco and alcohol consumption appear to explain the decrease in male excess mortality and, along with it, the narrowing of the gender gap.

Our results are insensitive to a battery of robustness checks (Appendix C). We decompose a simple model containing only health behaviors, age dummies and year fixed effects (Table C.2) to alleviate potential concerns with blocking part of the association of health behaviors and mortality by conditioning on determinants of health behaviors, such as education. We also use linear Oaxaca-Blinder decomposition analysis (Table C.1). Both alternatives leave our main findings unchanged. The former estimates slightly larger contributions of gender differences in health behaviors in both periods; the latter does so only for the later period; both robustness checks still show a decline in the absolute contribution of health behaviors across the decade studied. In our baseline analysis we exclude individuals who are single at any point during the full 5-year follow-up period; the results obtained are very similar when excluding only observations in the given year(s) that individuals are single (Table C.3). From 2006 the RLMS includes detailed information (quantity and frequency) on type of alcohol consumed, which makes it possible to estimate total pure alcohol consumption for 2010-2013 (Table C.4). Using this more complete measure for this period does not change the contribution of gender differences in alcohol consumption, nor that of the other health behaviors (Table C.5). Lastly, we decompose the negative mortality differential $(\overline{\hat{y}_w} - \overline{\hat{y}_m})$, where we assess the endowment (coefficient) effect by fixing the coefficients (characteristics) to those of women (men) (Table C.6). The weighting of the endowment and coefficients effects is thus reversed compared to the baseline estimates. Again, we observe the same overall pattern in the contribution of gender differences in health behaviors. Using women's coefficients, the contribution of alcohol use is, however, larger in the later period than a decade earlier, but in both periods that contribution is insignificant at 10%.

4.4 Discussion

Life expectancy in Russia is characterized by a gender difference of more than 10 years. At the same time, unhealthy behaviors are prevalent, especially among Russian males. The Russian government has therefore started implementing in 2003 a battery of alcohol and tobacco control policies to reduce alcohol use and tobacco consumption. The aim of this paper is to estimate the contribution of several unhealthy behaviors – smoking, drinking, exercise and diet – to the gender gap in 5-year mortality in Russia in two relevant periods a decade apart: 2000-2003 and 2010-2013. By doing so, we shed light on the contribution of those health behaviors to the gender gap in mortality, and how a decade of changes may be related to it.

We find that the gender gap in mortality has reduced considerably from 6.4 percentage points in the early 2000s to 3.8 percentage points in the early 2010s. Gender differences in health behaviors explain a substantial part of that gap in the early 2000s (3.1 percentage points, 49%) and less in the early 2010s (1.1 percentage points, 29%). Our results suggest that most of this contribution is explained by gender differences in smoking behavior – 2.6/6.4 (41%) of the gap in the early 2000s and 1.0/3.8 (28%) a decade later. Alcohol consumption explains substantially less – 0.5/6.4 (9%) in the earlier period and 0.1/3.8 (3%) in the later period. Declining tobacco and excessive alcohol use of men appear to be the main behaviors explaining the narrowing of the mortality gender gap in Russia.

One limitation of this paper is that alcohol use is known to be underestimated in the RLMS (Kuznetsova, 2020). The contribution of alcohol use to the gender mortality gap that we find is smaller than could be expected from previous papers (Trias-Llimós & Janssen, 2018). Nemtsov (2003) argues that alcohol use is lower in the RLMS in part because it is a household survey. As such it excludes certain demographic groups, such as Russian servicemen, which may be more prone to excessive alcohol use. Additionally, respondents of the RLMS may understate their alcohol consumption – in light of Russia's history with alcohol abuse there may be more stigma on honestly reporting drinking behavior, and respondents may also misreport for fear of repercussions from the government (Kuznetsova, 2020). Heavy drinkers, mostly male, may underreport their alcohol use or refuse to answer entirely. On the other hand, we have no evidence that underreporting of alcohol use in the RLMS has changed over time – alcohol consumption trends from the RLMS and from national statistics on accounted and unaccounted consumption are similar (Kuznetsova, 2020). This suggests that we likely estimate a lower bound for the contribution of alcohol to the gender mortality gap in both periods.

Nevertheless, we argue that smoking may still explain a larger share of the gender mortality gap in Russia than alcohol use. Indeed, this is the case for other former Soviet Union countries (McCartney et al., 2011). A potential reason for this larger contribution of smoking to the mortality gap is that smoking is 'stickier' – people tend to smoke for large periods of their life, whereas excessive alcohol use is mostly concentrated among certain age groups, in Russia specifically among middle-aged males (Kuznetsova, 2020; Zaridze, 2014). Furthermore, even in more representative samples, male excessive self-reported alcohol use is much less widespread in Russia than male smoking (Zaridze, 2014).

Another limitation of the current study is the potential underestimation of the gender mortality gap. We exclude individuals who live alone at any point during the relevant 5-year follow-up period, since mortality is underestimated for single households in the RLMS. We thus only include individuals who do not live alone. Those who do are likely to be older, and may also be men with a shorter lifespan, especially those who consume excessive amounts of alcohol. The latter could contribute to the underestimated alcohol consumption levels discussed above. Attrition may also make our sample less representative. Individuals who are ill and die due to their condition may drop out of the survey some years before they die. These individuals are then excluded from the analysis for certain years, which leads to an underestimation of 5-year mortality rates. The gender mortality gap we decompose may therefore be smaller than the national gender mortality gap. Lastly, since BMI is constructed using self-reported height and weight, it is likely that it contains measurement error – women (men) tend to underreport (overreport) their weight (height) (Burke & Carman., 2017; Cawley et al., 2015), which may lead to an underestimation of this gender difference and its contribution to the gender gap.

The Russian government introduced both alcohol and tobacco reducing measures in the last two decades, starting with alcohol policies in 2003 and tobacco policies in 2009. These policies may partly explain the observed reductions in drinking, smoking and subsequently mortality. Alcohol and tobacco consumption decreased from the early 2000s to the early 2010s, and the contribution of these health behaviors to the gender mortality gap also decreased substantially in this period, from 3.1 (49% of the 6.4 percentage point gap) to 1.1 (28% of 3.8). On the other hand, tobacco policies were still relatively new in the early 2010s, and some of the more severe regulations, such as prohibiting smoking in public places, fully came into force only in 2014. The observed reductions in drinking and smoking may have also been driven partly by a general declining trend in unhealthy lifestyles in Russia since the mid-2000s (Danilova et al., 2020; Shkolnikov et al., 2020). This paper shows that the Russian population has decreased its alcohol and tobacco consumption, especially males, and that this is associated with a decline in mortality rates and, along with it, with that of the gender mortality gap.

Appendices

Appendix A – Data construction

Data from 18 rounds (2000-2008 and 2010-2018) of the RLMS are used. The main dataset is the longitudinal individual dataset, which covers the years 1994-2018 and includes all necessary information, except information on mortality, household income and household size. This information is obtained from the longitudinal household dataset in the years 2000-2003 and 2010-2013. Household data from 2001-2008 and 2011-2018 are used to construct a variable which indicates whether the individual died within 5 years of year t in the relevant time periods, 2000-2003 and 2010-2013. For example, to determine 5-year mortality in the year 2000 we only use information from the household survey from 2001-2005.

The household data also includes movers which means that individuals can have multiple entries for the mortality variable, reported by different household members at different points in time. The vast majority of the duplicate entries do not lead to conflicting mortality results (either all households reply that the individual has died, or all households reply that the individual is still alive) and thus duplicate entries can be easily removed. However, a few duplicate entries do lead to conflicting results; certain household members indicate that the individual has died, while others indicate that the individual is still alive. These conflicting mortality results, which concern 134 individuals (out of more than 435,000 individuals with duplicate mortality data), were dropped.

Appendix B – Estimated and normalized coefficients

Table B.1. Estimated coefficients and normalized coefficients of health behaviors from probit models in early 2000s and early 2010

	Early 2000s				*Early 2010s*			
	Women		Men		Women		Men	
	$\hat{\beta}$	$\hat{\beta}^*$	$\hat{\beta}$	$\hat{\beta}^*$	$\hat{\beta}$	$\hat{\beta}^*$	$\hat{\beta}$	$\hat{\beta}^*$
No alcohol in last month	0.215*	0.114	0.078	-0.001	0.235**	-0.012	0.269**	0.076
Alcohol 1-3 times in last month	0.000	-0.101	0.000	-0.079	0.000	-0.247	0.000	-0.192
Alcohol 1-3 times a week	0.018	-0.083	-0.042	-0.121	0.252*	0.005	0.071	-0.121
Alcohol 4-7 times a week	0.172	0.071	0.281*	0.202	0.501†	0.254	0.430**	0.237
Binge drinker	0.113	0.057	0.142†	0.071	0.132	0.066	0.114†	0.057
Never smoker	0.000	-0.283	0.000	-0.180	0.000	-0.176	0.000	-0.147
Past smoker	0.283	0.000	0.136	-0.044	0.307*	0.131	0.119	-0.028
Current smoker	0.566**	0.283	0.404**	0.224	0.220†	0.044	0.321**	0.175
No exercise	0.000	0.488	0.000	-0.012	0.000	0.155	0.000	0.063
Light exercise	-0.555**	-0.068	0.061	0.049	-0.199†	-0.045	-0.040	0.023
Moderate or intensive exercise	-0.908**	-0.420	-0.024	-0.037	-0.264†	-0.110	-0.148	-0.085
Underweight	0.406*	0.385	0.091	0.134	0.199	0.262	0.185	0.172
Normal weight	0.000	-0.021	0.000	0.043	0.000	0.063	0.000	-0.013
Overweight	-0.225*	-0.246	-0.228**	-0.184	-0.237**	-0.174	-0.081	-0.094
Obese	-0.096	-0.118	-0.036	0.007	-0.215**	-0.152	-0.053	-0.066
Constant	-2.773**	-1.925	-2.395**	-1.217	-2.863**	-1.648	-2.121**	-1.317
N	8,635		6,851		13,535		10,593	

Notes: As in Table 1. Outcome is 5-year mortality. All regressions include controls for age, education, income, area (rural or urban), economic region and year as explained in text. Standard errors are clustered at the individual level. Normalized coefficients obtained according to equations (7), (8) and (9). †$p < .10$; *$p < .05$; **$p < .01$

Appendix C – Robustness analyses

Table C.1. Linear decomposition of the gender difference in 5-year mortality in early 2000s and early 2010s

	Early 2000s		*Early 2010s*	
	Absolute	Percent	Absolute	Percent
Male excess	0.0637**		0.0377**	
Overall decomposition results				
Endowment effect	0.0328**	51.41	0.0067	17.74
Coefficient effect	0.0310**	48.59	0.0310**	82.26
Detailed endowment effects				
Alcohol consumption	0.0065*	10.22	0.0012**	3.10
Smoking behavior	0.0241**	37.44	0.0169**	44.83
Exercise	-0.0002	-0.31	-0.0005	-1.39
BMI	0.0007**	1.02	-0.0001	-0.15
Total lifestyles	0.0311**	48.81	0.0175**	46.38
Other variables	0.0017**	2.59	-0.0108**	-28.64
Detailed coefficient effects				
Alcohol consumption	-0.0127	-19.95	-0.0014	-3.64
Smoking behavior	-0.0058	-9.17	-0.0071**	-18.95
Exercise	-0.0106	-16.66	-0.0023	-5.99
BMI	-0.0017	-2.69	-0.0020	-5.22
Total lifestyles	-0.0309	-48.47	-0.0127†	-33.81
Other variables	-0.0263**	-41.24	-0.0056**	-14.77
Constant	0.0881**	138.30	0.0493**	130.84
N	15,486		24,128	

Notes: As in Table 2. Linear Oaxaca-Blinder decomposition. Contribution of other variables is the sum of those of age, education, income, area (rural or urban), economic region and year. $^{\dagger}p < .10$; $*p < .05$; $**p < .01$.

Table C.2. Simple decomposition of the gender difference in 5-year mortality in early 2000s and early 2010s

	Early 2000s		*Early 2010s*	
	Absolute	Percent	Absolute	Percent
Male excess	0.0638**		0.0376**	
Overall decomposition results				
Endowment effect	0.0382**	59.91	0.0117*	31.19
Coefficient effect	0.0256**	40.09	0.0259**	68.81
Detailed endowment effects				
Alcohol consumption	0.0059*	9.26	0.0015**	3.98
Smoking behavior	0.0289**	45.27	0.0155**	41.29
Exercise	-0.0003	-0.50	-0.0007	-1.79
BMI	0.0003*	0.42	-0.0001	-0.36
Total lifestyles	0.0347**	54.44	0.0162**	43.12
Other variables	0.0035**	5.46	-0.0045**	-11.93
Detailed coefficient effects				
Alcohol consumption	-0.0032	-5.06	0.0039	10.31
Smoking behavior	0.0051	8.05	0.0008	2.20
Exercise	-0.0276**	-43.33	-0.0047	-12.42
BMI	0.0043	6.76	0.0008	2.17
Total lifestyles	-0.0214	-33.59	0.0009	2.27
Other variables	-0.0025	-3.95	0.0037	9.89
Constant	0.0495**	77.64	0.0213*	56.66
N	15,486		24,128	

Notes: As in Table 2. Contribution of other variables is the sum of those of age and year. †$p < .10$; *$p < .05$; **$p < .01$.

Table C.3. Decomposition of the gender difference in 5-year mortality in early 2000s and early 2010s excluding individuals who are single in a given year

	Early 2000s		*Early 2010s*	
	Absolute	Percent	Absolute	Percent
Male excess	0.0615**		0.0334**	
Overall decomposition results				
Endowment effect	0.0281**	45.71	0.0061	18.36
Coefficient effect	0.0334**	54.29	0.0273**	81.64
Detailed endowment effects				
Alcohol consumption	0.0049*	8.00	0.0006	1.84
Smoking behavior	0.0235**	38.30	0.0089	26.61
Exercise	0.0000	-0.04	-0.0004	-1.18
BMI	-0.0002**	-0.38	-0.0004	-1.28
Total lifestyles	0.0282**	45.88	0.0087	26.00
Other variables	-0.0001**	-0.17	-0.0026	-7.64
Detailed coefficient effects				
Alcohol consumption	-0.0042	-6.89	0.0039	11.53
Smoking behavior	0.0053	8.65	0.0010	2.99
Exercise	-0.0298**	-48.43	-0.0037	-10.95
BMI	0.0043	7.00	0.0008	2.41
Total lifestyles	-0.0244†	-39.67	0.0020	5.99
Other variables	0.0009	1.53	0.0065	19.45
Constant	0.0568**	92.43	0.0188*	56.20
N	16,611		27,746	

Notes: As in Table 2. Contribution of other variables is the sum of those of age, education, income, area (rural or urban), economic region and year. †$p < .10$; *$p < .05$; **$p < .01$.

Table C.4. Sample means of total pure alcohol consumption by gender

	Early 2010s		
	Women	Men	Women – Men
Total pure alcohol	68.0	341.3	273.3**
No alcohol	0.559	0.366	-0.192**
0-100 grams of pure alcohol	0.273	0.142	-0.131**
100-200 grams of pure alcohol	0.089	0.120	0.031**
200-400 grams of pure alcohol	0.050	0.152	0.102**
400-1000 grams of pure alcohol	0.022	0.142	0.121**
> 1000 grams of pure alcohol	0.007	0.076	0.069**
N	13,516	10,535	

Notes: We define categories of total pure alcohol consumption following Wood et al. (2018). **$p < .01$

Table C.5. Decomposition of the gender difference in 5-year mortality in early 2010s using total alcohol consumption

	Early 2010s	
	Absolute	Percent
Male excess	0.0375**	
Overall decomposition results		
Endowment effect	0.0079	20.93
Coefficient effect	0.0297**	79.07
Detailed endowment effects		
Alcohol consumption	0.0008	2.19
Smoking behavior	0.0106	28.20
Exercise	-0.0004	-1.08
BMI	-0.0003	-0.69
Total lifestyles	0.0107†	28.62
Other variables	-0.0029	-7.68
Detailed coefficient effects		
Alcohol consumption	0.0031*	8.27
Smoking behavior	0.0015	3.92
Exercise	-0.0040	-10.77
BMI	0.0010	2.61
Total lifestyles	0.0015	4.03
Other variables	0.0072	19.21
Constant	0.0210*	55.83
N	24,051	

Notes: We define categories of total pure alcohol consumption as in Table C.4. Other variables comprise age, education, income, area (rural or urban), economic region and year. †$p < .10$; *$p < .05$; **$p < .01$.

Table C.6. Reverse decomposition of the gender difference in 5-year mortality in early 2000s and early 2010s

	Early 2000s		*Early 2010s*	
	Absolute	Percent	Absolute	Percent
Male excess	-0.0637**		-0.0377**	
Overall decomposition results				
Endowment effect	-0.0270**	42.46	-0.0083	21.93
Coefficient effect	-0.0366**	57.54	-0.0295**	78.07
Detailed endowment effects				
Alcohol consumption	-0.0011	1.80	-0.0034	8.96
Smoking behavior	-0.0265*	41.65	-0.0066	17.40
Exercise	0.0031**	-4.79	0.0004	-0.94
BMI	-0.0003*	0.53	-0.0002	0.50
Total lifestyles	-0.0250**	39.19	-0.0098	25.92
Other variables	-0.0021**	3.27	0.0015	-3.98
Detailed coefficient effects				
Alcohol consumption	0.0029	-4.56	-0.0006	1.57
Smoking behavior	0.0025	-3.89	-0.0032	8.41
Exercise	0.0346**	-54.39	0.0042	-11.11
BMI	-0.0063	9.93	-0.0006	1.63
Total lifestyles	0.0337*	-52.91	-0.0002	0.49
Other variables	-0.0004	0.61	-0.0073*	19.33
Constant	-0.0700**	109.85	-0.0220*	58.25
N	15,486		24,128	

Notes: As Table 2, unless otherwise specified below. We decompose the female-male mortality differential by fixing the coefficients (characteristics) to those of women (men). This fixing is the reverse of that in our baseline estimates. Other variables is the sum of the contributions of age, education, income, area (rural or urban), economic region and year. $^{\dagger}p < .10$; $^{*}p < .05$; $^{**}p < .01$

Table C.7. Decomposition of the gender difference in 5-year mortality in early 2000s and early 2010s – complete results

	Early 2000s Absolute	Percent	*Early 2010s* Absolute	Percent
Male excess	0.0637**		0.0377**	
Overall decomposition results				
Endowment effect	0.0336**	52.72	0.0081	21.34
Coefficient effect	0.0301**	47.28	0.0297**	78.66
Detailed endowment effects				
Alcohol consumption				
No alcohol in the last month	0.0000	0.04	-0.0011	-2.79
Alcohol 1-3 times in the last month	0.0007	1.07	0.0006*	1.68
Alcohol 1-3 times a week	-0.0032*	-5.00	-0.0017*	-4.56
Alcohol 4-7 times a week	0.0013*	2.10	0.0007*	1.84
Binge drinker	0.0066†	10.32	0.0025	6.58
Smoking behavior				
Never smoked	0.0132*	20.72	0.0056	14.73
Past smoker	-0.0006	-0.92	-0.0002	-0.63
Current smoker	0.0135**	21.12	0.0051*	13.57
Exercise				
No exercise	0.0001	0.09	-0.0001	-0.24
Light exercise	0.0000	0.01	0.0000	-0.10
Medium, intensive or daily exercise	-0.0002	-0.27	-0.0003	-0.70
BMI				
Underweight	-0.0002	-0.26	-0.0002	-0.61
Normal weight	0.0007	1.04	-0.0001	-0.14
Overweight	-0.0005**	-0.85	-0.0005	-1.34
Obese	-0.0001	-0.19	0.0006	1.46
Age	0.0032**	5.00	-0.0035	-9.36
Education	0.0003	0.42	0.0009	2.44
Income	-0.0005**	-0.72	0.0000	0.06
Rural/Urban area	-0.0002	-0.33	0.0000	-0.09
Region	-0.0004†	-0.64	-0.0002	-0.42
Year	0.0000	-0.03	0.0000	-0.04
Detailed coefficient effects				
Alcohol consumption				
No alcohol in the last month	-0.0041	-6.44	0.0030	7.88
Alcohol 1-3 times in the last month	0.0006	0.97	0.0011	2.84
Alcohol 1-3 times a week	-0.0003	-0.53	-0.0008	-2.21
Alcohol 4-7 times a week	0.0001	0.11	0.0000	-0.02
Binge drinker	-0.0007	-1.11	0.0004	1.05
Smoking behavior				
Never smoked	0.0060	9.41	0.0014	3.60
Past smoker	-0.0003	-0.40	-0.0007	-1.96
Current smoker	-0.0005	-0.85	0.0011	2.96
Exercise				
No exercise	-0.0309**	-48.50	-0.0045	-11.84
Light exercise	0.0008	1.23	0.0004	1.18
Medium, intensive or daily exercise	0.0018†	2.90	0.0001	0.33
BMI				
Underweight	-0.0006	-0.93	-0.0002	-0.49
Normal weight	0.0020	3.08	-0.0018	-4.65
Overweight	0.0013	2.07	0.0014	3.78
Obese	0.0024	3.76	0.0015	3.95
	-0.0039	-6.10	0.0031	8.12

Age				
Education	0.0038	6.02	0.0018†	4.70
Income	0.0000*	0.06	0.0000	0.00
Rural/Urban area	-0.0004	-0.71	0.0010	2.59
Region	0.0012	1.84	0.0012	3.24
Year	-0.0001	-0.09	0.0002	0.56
Constant	0.0519**	81.49	0.0200*	53.05
N	15,486		24,128	

Notes: As Table 2. †$p < .10$; *$p < .05$; **$p < .01$

Chapter 5

When the sun goes down: Effects of sunset time on adolescent sleep, mental health and education

Abstract

Short sleep duration is associated with depressive symptoms and lower cognitive performance in adolescents. Causal evidence of effects of chronic sleep deprivation, however, is lacking. I estimate effects of long-term exposure to sunset time across the United States on the sleep duration of adolescents and, subsequently, on the risk of depression and on educational attainment by young adulthood. I find that adolescents exposed to later sunset times go to bed later, sleep less, and spend more time on sedentary activities. They are more likely to have symptoms of depression in adolescence and to be diagnosed with that condition by young adulthood. While they are not less likely to graduate from high school or to attend college, they are less likely to graduate from college. The estimated effects are large. Exposure to a 15-minute (1 standard deviation) delay in sunset time persistently during adolescence is estimated to reduce sleep duration by 5 minutes each weeknight, increase the probability of being diagnosed with depression by 2 percentage points, and reduce the probability of being a college graduate by 3.5 percentage points. These findings are consistent with persistent sleep deprivation in adolescence causing adverse effects on mental health and education.

5.1 Introduction

Average sleep duration is shorter now than in previous decades (Matricciani et al., 2012; Roenneberg, 2013). This has raised concerns about adverse consequences of sleep deprivation for health and cognitive performance (Capuccio et al., 2010; Gregory & Sadeh, 2012; Lim & Dinges, 2010). Nevertheless, there are substantial gaps in the existing evidence. In particular, little is known about the impact of chronic exposure to sleep loss on adolescents' human capital formation. I use long-term variation in sunset time across the United States to provide the first evidence of the impact of adolescents' chronic exposure to later sunsets - which reduces sleep - on depression and educational attainment by young adulthood. The estimated effect on depression is the first in any setting, while the estimated effect on educational attainment is the first in a high-income country.

Insufficient sleep is common among US adolescents. More than two-thirds of high school students sleep less than the minimum recommended eight hours on school nights (Wheaton et al., 2018). Early school start times and evening screen time contribute to poor adolescent sleep (e.g. Avery et al., 2022; Hale & Guan, 2015; Wheaton et al., 2016).[36] At the same time, depressive symptoms are prevalent, and increasing, among US adolescents. In 2015, 30% of high school students had experienced persistent feelings of sadness or hopelessness (Centers for Disease Control and Prevention, 2023). In 2019, this increased to 37%. Association studies link short adolescent sleep to poor mood and emotion regulation (e.g. Owens, 2014; Short et al., 2020) and a causal basis for these correlations is provided by experiments that manipulate sleep (Baum et al., 2014; Lo et al., 2016). Experiments also show that non-pharmacological sleep interventions decrease depressive symptoms (see Freeman et al., 2017; Freeman et al., 2020; Gee et al., 2019). Sleep deprivation could therefore be an important driver of depressive symptoms in adolescents and addressing it may be effective in both preventing and treating depression (Freeman et al., 2020). Experimental studies also find that sleep deprivation impairs cognitive functioning (Beebe et al., 2017; Lo et al., 2016). Sleep deprivation during adolescence may thus be especially harmful to health and productivity later in life by causing persistent health problems and by interfering with educational performance and subsequent attainment (Currie, 2020).

This paper exploits detailed adolescents' time-use data from the US Panel Study of Income Dynamics (PSID) to estimate effects of chronic exposure to later sunset time on the sleep duration of middle and high school students and on their educational attainment and risk of experiencing depression by young adulthood. I exploit variation in annual average sunset time from eastern to western locations *within* time zones to estimate effects on sleep loss in

[36] Approximately 40% of US public high schools start before 8 AM (National Center for Education Statistics, 2020).

adolescence. I find that later sunsets induce later bedtimes and shorter sleep duration. The latter result is plausibly due to rigid school start times. A 15-minute (1 standard deviation) delay in annual average sunset time reduces adolescent sleep by approximately 5 minutes on each typical weeknight, and it increases the probability of having insufficient (< 8 hours) sleep by 3 percentage points. Sleep-deprived adolescents increase time spent on sedentary leisure. Chronic exposure to later sunset time increases the probability of having symptoms of depression during adolescence and of being diagnosed with that condition by young adulthood. It increases the probability of college graduation, although no impact is found on high school graduation nor college attendance. The estimated effects are large. A 15-minute delay in annual average sunset time increases the likelihood of being diagnosed with depression by 2 percentage points and reduces the likelihood of graduating from college by 3.5 percentage points. These results suggest that chronic sleep deficits can have a lasting impact on human capital formation.

This paper is related to the vast experimental literature which examines the effect of adult sleep on cognition. Most of these studies look at the impact of short-term sleep loss (Lim & Dinges, 2010). Relatively few studies investigate effects of cumulative sleep loss. Van Dongen et al. (2003) is one of the few. They conducted a two-week experiment in a sleep laboratory and find very large effects of cumulative sleep deprivation on cognitive performance.

Recent studies build on these findings by estimating the impact of cumulative adult sleep on health and economic outcomes using quasi-experimental designs. Gibson & Schrader (2018), using annual average sunset time as an instrumental variable (IV) for sleep, show that sleep duration increases earnings for US workers. It does so to an extent that is comparable in importance to education (Gibon & Schrader, 2018). Using the same IV, Giuntella et al. (2017) find that sleep duration increases cognitive performance and decreases depressive symptoms in older, Chinese workers. Giuntella & Mazzonna (2019) show that sleep also impacts physical health. They exploit discontinuities in sunset time at time zone borders in the US and find that sleep influences various health outcomes – sleep-deprived adults are more likely to be obese, to have diabetes, to suffer from cardiovascular diseases, and are also more likely to have certain forms of cancer, such as breast cancer. In contrast, a three-week field experiment in India finds that increasing nighttime sleep among the urban poor has no detectable impact on short-term cognition, productivity or well-being (Bessone et al., 2021). The authors argue that this may be explained by large differences in sleep efficiency between contexts. Nighttime sleep in their sample is highly interrupted, much more so than in high-income countries.

Evidence on the causal effects of chronic sleep loss in adolescents is much more limited. Many experiments that focus on adolescents look at effects of short-run sleep loss or extreme sleep deprivation (De Bruin et al., 2017; Short et al., 2020). Two experimental studies address these shortcomings and show that realistic doses of cumulative sleep loss impair adolescents' cognitive functioning and mood (Beebe et al., 2017; Lo et al., 2016). Lo et al. (2016) find declines in attention and memory for each night of sleep loss, whereas mood showed no further declines near the end of the sleep manipulation period.

Only one paper has investigated the impact of chronic sleep loss on children's educational attainment. Jagnani (2022) exploits variation in annual average sunset time across India and finds that chronic exposure to delayed sunset time, which decreases child sleep, reduces years of education and primary and middle school completion rates among school-age children. The effects are substantial. A 15-minute delay in annual average sunset time decreases the probability of primary school completion by 3 percentage points and middle school completion by 2 percentage points. However, as evidenced by the difference between the findings of Bessone et al. (2021) and those from developed countries on short-term effects of sleep loss (e.g., Van Dongen et al., 2003), one should avoid extrapolating Jagnani's (2022) results to high-income contexts. Living conditions in India are very different from those in the US and are likely to impair sleep (quality). Jagnani (2022) also exploits variation in daily sunset times across the year to show that school-age children who sleep less at night increase time spent on indoor leisure and daytime napping, and allocate less time to studying. These sleep-deprived children also perform worse on math tests in the short-run, which likely contributes to their lower education attainment.

Several other recent studies have looked at the impact of adolescent sleep on short-run academic performance by exploiting variation in school start times (see Carrell et al., 2011; Edwards, 2012; Heissel & Norris, 2018; Hinrichs, 2011; Groen et al., 2019) or daylight savings time (DST) transitions (see Gaski & Sagarin, 2011; Herber et al., 2017). These papers find mixed results. Some show detrimental effects on students' test scores (e.g. Carrell et al., 2011; Gaski & Sagarin, 2011). Others find no evidence of effects (Herber et al., 2017; Hinrichs, 2011). These studies seldom measure adolescents' sleep or how adolescents reallocate their time in response to changes in sleep.

The contribution of this paper is threefold. First, by using detailed time-use data for US adolescents, I am able to show that adolescents who experience later sunsets sleep less as a result of delaying their bedtime, and they increase time spent on sedentary leisure. This paper is the first to investigate such time-use responses of adolescents in a high-income context. Second, I present the first evidence of the impact of adolescents' chronic exposure to later sunset time on depression. I estimate that it increases the likelihood of having depressive symptoms during adolescence and of being diagnosed with depression by young adulthood. Third, this paper provides the first estimates of the impact of adolescents' chronic exposure to later sunset time on educational attainment in a high-income setting. I find no evidence for effects on high school graduation or college attendance, but exposure to later sunset time during adolescence decreases the likelihood of graduating from college. These findings are consistent with persistent sleep deprivation in adolescence having adverse effects on mental health and education.

5.2 Data

The Panel Study of Income Dynamics (PSID) is a longitudinal survey of a representative sample of US individuals and families. I use data from the PSID Child Development Supplement (CDS) and the Transition into Adulthood Supplement (TAS) (PSID, 2021).[37] These studies follow PSID children over time and collect extensive information on their health, well-being and educational attainment. The original CDS cohort of PSID children (0-12 years) was interviewed in 1997, and again in 2002 and 2007. In 2014 a new cohort covering all children in PSID families (0-17 years) was interviewed. This new cohort was interviewed again in 2019. The original CDS cohort was followed as young adults (17-28 years) into the TAS, first in 2005 and then every two years until 2015, when all participants of the original CDS cohort reached adulthood and were eligible to be included in at least one TAS wave.[38] In 2017, the TAS was expanded to include all young adults in the PSID.

A distinctive feature of the CDS is that children are asked to fill out detailed weekday and weekend time diaries in each wave. The time diary is mailed ahead of the scheduled CDS interview; respondents are asked to complete it prior to the interview for one randomly selected weekday and weekend day. Adolescents complete the diaries themselves, with help from their primary caregiver, if needed. During the in-person or by-telephone interview, the CDS interviewer reviews the time diary with the child and the primary caregiver, and edits it if necessary.[39] The diary includes information on the duration, start time and ending time of various (dis)aggregated activities. The CDS interviewer also records how typical the diary day was.

The CDS groups time-use activities across some broad categories, the most important of which are active leisure, education, passive leisure, and personal care. I define eight time-use activities; namely, sleep, exercise, mentally-active sedentary leisure, passive sedentary

---

[37] Because of an initial oversampling of low-income families, the CDS oversamples Black children.

[38] CDS children are eligible for inclusion in the TAS when they turn at least 18 in the calendar year of the TAS wave and they no longer attend high school. They are no longer eligible once they are older than 28.

[39] See https://psidonline.isr.umich.edu/CDS/Guide/Documents.aspx for detailed information about the sampling design and data collection of the CDS and the TAS.

leisure, naps, school and study. I distinguish between passive and mentally-active sedentary leisure because of their potentially differential effects on mental health and cognition (Hallgren et al., 2018; Johnson et al., 2007; Kühn et al., 2014; Wanders et al., 2021; Werneck at al., 2021a; Werneck et al., 2021b). Sleep is a separate category under personal care activities and is defined as total nighttime sleep duration.[40] Exercise and mentally-active sedentary leisure are mostly created by grouping two sets of separate active leisure activities. Exercise includes physically active leisure, such as sports and outdoor activities. Mentally-active sedentary leisure includes mentally-active, mostly seated leisure, such as arts, crafts games and reading (of books, magazines, newspapers).[41, 42] Passive sedentary leisure includes, among other things, face-to-face and telephone conversations, watching television and relaxing.[43] Naps (including rest periods) are a subcategory of personal care in the time diary. These naps, which can also occur in the evening, are recorded separately from nighttime sleep. School and study are subcategories of education and include, respectively, time spent attending school and time spent on studying or homework.[44]

The CDS also collects information on children's depressive symptoms. Since the second wave (2002), adolescents aged 12 and over are asked to complete the Children's Depression Inventory Short Form (CDI:S; Kovacs, 1992). This is a validated diagnostic instrument which measures depressive symptoms in children. The short form consists of choosing 10 statements which best reflect the child's feelings in the last two weeks (see Appendix A for the full list of statements). A higher CDI:S score indicates higher depressiveness and this instrument is also a valid screening tool for depression (Allgaier et al., 2012). I classify adolescents as being at risk of depression if they have a score equal to or higher than 3, and 0 otherwise (Allgaier et al., 2012).

The TAS contains information on depression diagnosis. During every TAS wave, young adults aged 17 to 28 are asked whether they have ever been told by a doctor or other health professional that they have an emotional or psychiatric disorder, and, if so, in a consecutive question, what the disorder was. Respondents who report *depression* in response to the

[40] It includes time spent in bed trying to sleep, but not asleep. It excludes time in bed spent on leisure activities.
[41] Games under active leisure includes video games, but excludes computer games. The subcategory video games is only available until 2007, after which video games are registered separately under computer use as computer games. I therefore add the subcategory computer games to mentally-active sedentary leisure.
[42] I include reading in mentally-active sedentary leisure, following the literature on this distinction (see e.g. Wanders et al., 2021; Werneck at al., 2021a), even though it is registered under the passive leisure category in the CDS time diary.
[43] Watching television includes watching movies or shows via streaming services or on laptops, phones or other electronic devices. I also include in the passive sedentary leisure category the following activities registered in the time diary under computer use: texting, video communication or social media use.
[44] Until 2007 using a computer for homework or studying was coded separately under computer use, since 2014 it is included under homework and studying in the category education. I therefore include the subcategory homework of computer use in the time-use activity study. Attending school includes recess but excludes lunch breaks.

second question in any wave are coded as having been diagnosed with depression. Since I observe TAS respondents at varying ages from 17 to 28, and some only once, I capture diagnoses up until age 17 at the earliest to diagnoses up until age 28 at the latest.

The TAS also collects detailed information about education. From 2009 onwards, the data includes a carefully constructed measure of highest education level. I use this information to construct three measures of educational attainment: high school graduation, some college attendance and college graduation. High school graduation takes value 1 if adolescents have obtained a high school diploma, and 0 otherwise.[45] Some college attendance takes value 1 if adolescents have started an academic degree program at a college or university or have obtained at least an undergraduate two-year academic degree at a college or university, and 0 otherwise.[46] College graduation takes value 1 if adolescents have obtained at least a four-year bachelor degree, and 0 otherwise. I measure high school graduation and some college attendance at age 20 or over and college graduation at age 23 or over. Figure 1 shows an age line of measurement of the different outcomes. CDS outcomes (time-uses and depressive symptoms) vary across waves per adolescent, while TAS outcomes do not (depression diagnosis and education outcomes by a certain age).

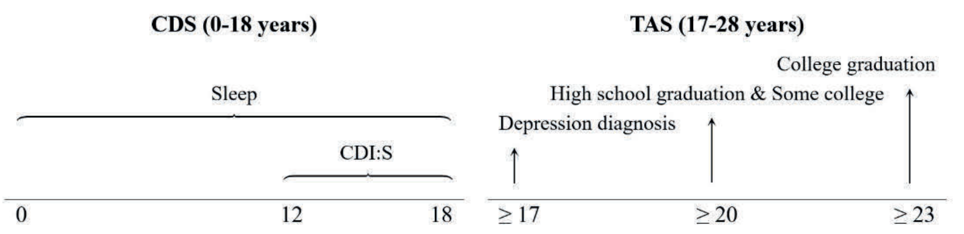


Figure 1. Age line (y-axis) of CDS and TAS outcomes. *Notes*: CDS outcomes vary across waves per adolescent (brackets), TAS outcomes have only one value per adolescent (arrows). At which age during young adulthood that value is measured differs across adolescents and TAS outcomes.

Location information of adolescents at the county level is obtained via restricted-use PSID family level data (1997, 2001, 2007, 2013, 2019). Annual average sunset time for each county in the US, based on the county's population centroid, is obtained from the National Ocean and Atmospheric Administration (NOAA) solar calculator (NOAA, 2020). It is computed by first calculating daily sunset times, taking into account DST transitions, for the centroid coordinates, and then adding these up per coordinate pair and dividing by the

[45] It takes value 0 when adolescents have obtained a GED instead of a high school diploma.
[46] Undergraduate academic degrees at colleges and universities in the US include associate, typically two-year, degrees and bachelor, typically four-year, degrees.

number of days in the year. County population centroids are obtained from the Unites States Census Bureau (US Census Bureau, 2021).

The sample used in this paper includes observations from all CDS waves of adolescents aged 11 years or older, who reside in the Eastern, Central or Pacific time zone and are not home-schooled.[47] It thus only includes adolescents living in the contiguous US. The Mountain time zone is excluded because it contains too few observations. Home-schooled children are excluded because they may have more flexible start times which may vary with solar cues. Adolescents who live in counties which span multiple time zones are not included, nor are those who live in Indiana since this state does not observe DST in some periods under study.[48] Time diary observations are additionally restricted to typical weekday entries.[49] I work with different sample sizes when examining the different outcomes, due to non-overlapping missing (or set to missing) outcomes.[50]

Most of the adolescents included in the analysis sample are spread across the Eastern (52%) or Central (34%) time zone, and some live in the Pacific (14%) time zone, mostly near the coast. Table B in the Appendix reports summary statistics of main variables used in the analyses. Half of the sample is female, and Black adolescents are overrepresented. Adolescents experience an annual average sunset time around 7 PM, with one standard deviation corresponding to about 15 minutes around that time.[51] On average, adolescents go to bed shortly before 10 PM, wake up around 6:30 AM and sleep around 8 hours and 30 minutes on typical weekdays. Still, 40% obtains insufficient sleep for their age, following the consensus on the recommended minimum amount of sleep for children by the American Academy of Sleep Medicine; 9 hours if aged 12 or younger and 8 hours if aged 13 or over (Paruthi et al., 2016). Adolescents spend, on average on a typical weekday, 6 hours and 30 minutes attending school, 45 minutes on homework and studying, and about 10 minutes napping. In their leisure time, they spend around 35 minutes on exercise, 1 hour on mentally-active sedentary leisure, and over 2 hours on passive sedentary leisure. Approximately 35%

---

[47] The bulk of my observations come from adolescents in the initial CDS cohort, and specifically from CDS 2002 and 2007. In these waves most of the initial CDS cohort had turned at least 11, although some observations also stem from adolescents 11 or older in CDS 1997. To a lesser extent, the sample includes observations of adolescents 11 or older in the new CDS cohort from CDS 2014 and 2019. Roughly 10% are follow-up observations of the same adolescents across different CDS waves.

[48] Arizona does not observe DST in all periods under study, but this state lies in the Mountain time zone and is therefore already excluded.

[49] Typicality of the diary day is measured on a 1-5 scale, with higher values indicating less typicality. Only time diary observations with typicality 1-3 are included. Weekdays are Monday-Friday. Since the diary day starts at midnight, this includes Sunday-Monday until Thursday-Friday sleep. The resulting time diary observations stem mostly from the fall and winter months and do not include the summer holiday months July and August.

[50] Sociodemographic and location characteristics are very similar across the different sample sizes.

[51] This holds within the Eastern and Central time zones. Annual average sunset time is somewhat earlier in the Pacific time zone (6:45 PM), and contains less variation (1 standard deviation equals about 10 minutes). See also Figure 1 below.

of the observed adolescents are at risk of being depressed and 13% is diagnosed with depression by young adulthood. Around 90% of adolescents graduate from high school and 70% attends at least some college by age 20 or over, and 30% graduates from college with a bachelor degree by age 23 or over. These percentages are in line with US national statistics.[52] On average, adolescents are 14 years old at the time of a CDS interview and they are 24 years old at the time of their last TAS interview.

5.3 Empirical strategy

Comparing outcomes of adolescents with different sleep durations would not give estimates of causal effects because sleep duration may be influenced by depressive feelings and the effort exerted to attain education, and there is scope for unobserved confounders of sleep duration, such as stress. I therefore follow others (Gibson & Shrader, 2018; Giuntella et al., 2017; Giuntella & Mazzonna, 2019; Jagnani, 2022) in using variation in annual average sunset time to estimate effects on sleep loss in adolescence. Different from others, I estimate longer-run effects of sunset time in adolescence on depression and education.

The relationship between sunset time and sleep derives from the biological link between sleep and sunlight (Roenneberg & Merrow, 2007). Our circadian rhythm listens to cues from our environment and sunlight is the most established and recurrent signal that we receive (Roenneberg & Merrow, 2007; Walker, 2017). As the sun sets and the environment becomes darker, our brain starts releasing melatonin, marking the timing of sleep onset (Walker, 2017).[53] Hence, when the sun sets at a later hour, individuals tend to go to bed later. This later bedtime could, in principle, be compensated by a similar shift in wake time. However, social constraints, such as school start times and work schedules, largely dictate wake times and do not vary much with solar cues (Hamermesh, 2008).

I exploit variation in annual average sunset time across counties within US time zones to look at effects on adolescent sleep and, subsequently, the risks of depression and educational attainment by young adulthood. Within a time zone, the sun sets earlier in eastern counties than in western counties, leading individuals in more western counties to go to bed later and sleep less. By construction, there is roughly a one-hour difference in sunset time within each US time zone. The initial placement of the time zone boundaries was done by railroad companies in 1883, and the width of the zones was chosen to ensure a one-hour difference

[52] See e.g., Centers for Disease Control and Prevention (2023) and Goodwin et al. (2022) for estimates on depression, and Ryan & Bauman (2016) for statistics on educational attainment.
[53] Melatonin itself does not generate sleep, but it helps govern the timing of when sleep happens. It is also good to mention that most adolescents experience a shift in their circadian rhythm, a so-called sleep-phase delay (Walker, 2017). For most adolescents, then, melatonin production and the instruction of sleep in response to waning daylight is delayed by about two hours.

in solar time (Library of Congress, 2010). These time standards were then quickly enacted by the US government (Allen, 2000). There have only been minor changes to the boundaries since then. The time zone boundaries do not always align with state borders. A quarter of the contiguous US states cover two time zones (Hamermesh et al., 2008). Figure 2 delineates the time zones and shows variation in annual average sunset time across counties in the contiguous US. It shows that the Eastern, Central and Mountain time zones have similar variation in annual average sunset time (1 standard deviation equals at least 15 minutes), whereas the Pacific time zone has less (1 standard deviation equals less than 10 minutes). As mentioned above, most of my observations stem from the Eastern (52%) and Central (34%) time zone.

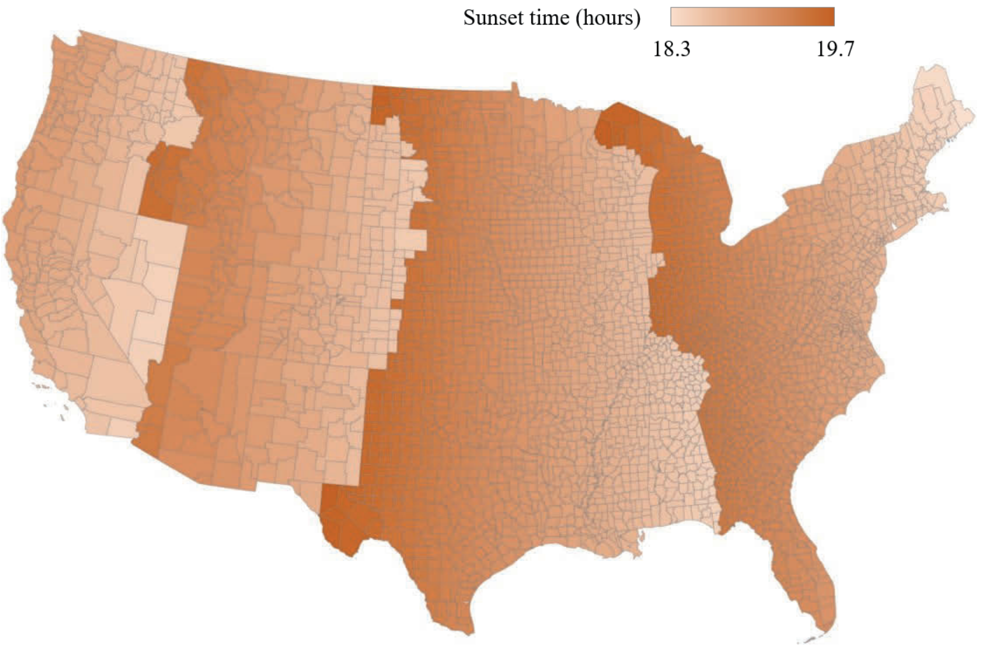


Figure 2. Annual average sunset time and time zones in the contiguous US. *Notes*: From left to right, the map delineates the Pacific, Mountain, Central and Eastern time zone. Annual average sunset time was calculated using coordinates of the population centroids of US counties and the NOAA solar calculator. The darker the colour, the later the annual average sunset time. For the purpose of this graph, it is assumed that all counties in the US observe DST and counties spanning multiple time zones were allocated to the time zone most of the county resided in. This graph was created using Excel maps.

I use the following empirical specification to estimate effects on short-run outcomes in adolescence:

$$Y_{ict} = \beta_0 + \beta_1 Sunset_{ct} + \beta_2 X_{ict} + \beta_3 K_{ct} + \mu_z + \eta_w + \lambda_m + \varepsilon_{ict} \quad (1)$$

where Y_{ict} is sleep, time-use or risk of depression for adolescent i in county c in wave t, $Sunset_{ct}$ is annual average sunset time in county c in wave t, and X_{ict} includes sociodemographic controls, namely, single-year age dummies and dummies for gender, race and attendance at a public school. K_{ct} is a set of geographic characteristics of county c in wave t and linearly controls for annual average temperature, total annual precipitation, minimum distance to coast, population density and land area, and includes latitude and its square.[54] μ_z are time zone fixed effects, and η_w and λ_m indicate interview wave and month fixed effects, respectively.[55] To increase power, the sample used to estimate regression (1) includes some (< 23%) follow-up observations of the same adolescents across different waves.

To estimate effects on longer-run outcomes measured in young adulthood in the TAS, which do not vary within adolescents, I use sunset time and covariates from one wave of CDS.[56] The empirical specification closely follows equation (1):

$$Y_{icT} = \beta_0 + \beta_1 Sunset_{ct} + \beta_2 X_{ict} + \beta_3 K_{ct} + \mu_z + \eta_w + \varepsilon_{ict} \quad (2)$$

Now, Y_{icT} is depression diagnosis or educational attainment by time T in young adulthood for adolescent i in county c, who experienced annual average sunset in wave t during adolescence, where $T > t$. Everything else is as defined above.

I thus compare outcomes between adolescents across eastern and western counties within a time zone.[57] The underlying identification assumption is that, conditional on the controls, annual average sunset time between counties within a time zone is unrelated to other factors that influence adolescent outcomes. The geographical controls are important in ensuring the

[54] Climate data is obtained from the National Centers for Environmental Information (NCEI) (NCEI, 2023). Minimum distance to coast includes distance to the Great Lakes. Population density is calculated by dividing total population estimates by land area. The former are obtained from the US Census Bureau (US Census Bureau, 2023) and the latter from RAND State Statistics (RAND State Statistics, 2021). A linear control for land area is included to account for the fact that for larger counties, which are mostly concentrated in the Pacific time zone, the population centroid may be a noisy approximation of the annual average sunset time that residents of those counties experience.
[55] For time-use outcomes the interview month equals the month in which the recorded diary day took place.
[56] This entails selecting one CDS wave observation for adolescents whom I observe in multiple CDS waves. For the initial CDS cohort, this requires choosing between CDS 2002 and 2007. The baseline estimates include the latest available CDS wave observation for this cohort, i.e., CDS 2007, and are not sensitive to changing this to CDS 2002 (see Results). For the new CDS cohort, I can only observe certain longer-run outcomes and only for older adolescents in CDS 2014 who participated in TAS 2017 or 2019.
[57] There is too little variation in annual average sunset time within counties, hence why I measure sunset time at the county level and no county fixed effects are added in equations (1) and (2).

validity of the conditional exogeneity assumption. I include time zone indicators to account for coordination within time zones (Hamermesh, 2008). Within a time zone, annual average sunset time is linearly related to longitude and hence associated with coastal distance and economic activity (Gibson & Schrader, 2018). It is therefore important to control for coastal distance and, likewise, for population density. I also add controls for temperature and precipitation and flexibly control for latitude to account for (other) differences in climate between counties. Locations with later sunset times are exposed to more sunlight in the evening, and sunlight, through vitamin D production, may directly impact mood (Kjaergaard et al., 2012). However, annual average sunlight duration varies very little across locations in the contiguous US, so it is unlikely to be a confounder.[58]

If the conditional independence assumption holds, then the β_1 estimates in equations (1) and (2) are likely to capture the impact of chronic exposure to sunset time, given that annual average sunset time is a fixed characteristic of locations and adolescents are likely to reside in the same county for an extended period of time.[59] Moreover, given the biological relationship between sleep and sunlight, the β_1 estimates for outcomes on depression and education are likely to capture the cumulative impact of sunset-induced sleep loss and subsequent time-use responses of adolescents.

Standard errors in equations (1) and (2) are clustered at the county level to account for potential correlation of adolescents living in the same county. Although the geographic controls help to mitigate concerns about the validity of the identification assumption, I cannot exclude the possibility that there are other factors correlated with annual average sunset time that influence adolescents' outcomes. Nor can I rule out, for the depression and education outcomes, that mechanisms other than adolescent sleep are driving the effects. I therefore investigate these possibilities in Section 4 and fail to find evidence that potential confounders or other mechanisms are driving the results.

5.4 Results

5.4.1 Sleep and other time-uses

Table 1 shows estimates of the impact of later sunsets on adolescents' sleep schedules. I show effects per 15 minutes (≈ 1 standard deviation) of annual average sunset time to ensure that the effect size reflects the identifying variation. A 15-minute delay in annual average sunset time is estimated to decrease adolescent sleep, on typical weekdays, by about 5

[58] Still, later sunsets may increase adolescents' daily exposure to sunlight. I discuss the potential implications of this in the Conclusion.

[59] There is some, but very little, variation in annual average sunset time of counties over the years. Small differences arise due to changes in DST transitions and leap years.

minutes and increases the likelihood of obtaining insufficient sleep by 3 percentage points (pp) (or by 7.5% of the mean). Later sunsets reduce sleep by delaying adolescents' bedtimes. They have no discernible impact on wake times. These results are insensitive to the inclusion of additional interview characteristics, different latitude specifications and potential sleep outliers (see Table C.1). They are also in line with estimates from previous studies (e.g. Giuntella & Mazzonna, 2019; Jagnani, 2022).

Table 1. Effect of later sunset time on adolescent sleep duration and schedules

	(1) Sleep (minutes)	(2) Insufficient sleep (0,1)	(3) Bedtime (minutes)	(4) Wake time (minutes)
Sunset (15 minutes)	-5.219**	0.030**	4.219**	0.452
	(2.204)	(0.013)	(1.981)	(1.460)
R^2	0.09	0.07	0.15	0.06
Mean dep. variable	8 hrs+35 mins	0.40	9:50 PM	6:33 AM
N	2,636	2,636	2,289	2,289

Notes: Sample of adolescents in the CDS, 1997-2019. All models follow equation (1) and include sociodemographic characteristics (single-year age dummies and dummies for gender, race and public school), geographic characteristics (linear controls for annual average temperature, total annual precipitation, minimum distance to coast, population density and land area, and time zone indicators and latitude and its square) and interview characteristics (interview month and wave dummies). Observations with sleep < 4 and > 16 hours per day are not included. Insufficient sleep is defined as < 9 hours of sleep per day for adolescents aged 12 or younger and as < 8 hours for adolescents aged 13 or older. Bedtimes and wake times are retrieved from the disaggregated CDS time diary files, which timestamp nighttime sleep episodes. Observations of adolescents who experience fragmented sleep and thus report multiple bedtimes and wake times per diary day are excluded. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

Table 2 presents the effects of later sunset time on other time-uses of adolescents to investigate how adolescents make use of the increased wake time resulting from sunset-induced sleep loss. It shows that adolescents exposed to later sunsets spend more time on mentally-active sedentary leisure activities. A 15-minute delay in annual average sunset time increases time spent on mentally-active sedentary leisure by an estimated 4 minutes. This appears to be driven by its largest subcategory: playing games, which includes video gaming (see Table C.2, column 1). I find no evidence for changes in time allocated to exercise, passive sedentary leisure or napping. Passive sedentary leisure consists mostly of watching television. I therefore consider this activity also separately, but I find no detectable impact (see Table C.2, column 2). Furthermore, given the correlation between social media use and depression in adolescents (Lin et al., 2016), I also investigate separately the effect on time allocated to social media use, available from 2014. I find no observable impact on it (Table C.2, column 3), possibly because adolescents in the sample spent little time on social media.

Lastly, sleep-deprivation seems to reduce time spent by adolescents on homework and studying, although the point estimate is statistically insignificant.

The results are largely in line with Jagnani (2022). He finds that school-age children in India who experience later sunsets allocate more time to indoor leisure, study less and nap more. I fail to find evidence for the latter, which may be explained by differences in context and sample. Indeed, time allocated to napping is higher in Jagnani's (2022) sample, which could be due to his inclusion of primary school children or due to differences in the daily schedules of US and Indian students.[60]

Table 2. Effect of later sunset time on adolescent time-uses (minutes)

	(1) Exercise	(2) Mentally-active sedentary leisure	(3) Passive sedentary leisure	(4) Nap	(5) Study
Sunset (15 minutes)	-0.519	4.187**	-1.908	1.420	-2.673
	(1.496)	(1.914)	(2.967)	(1.001)	(1.645)
R^2	0.05	0.08	0.06	0.06	0.04
Mean dep. Variable	36.02	59.10	131.67	12.17	45.84
N	2,636	2,636	2,636	2,636	2,636

Notes: Sample of adolescents in the CDS, 1997-2019. Controls and sample as in column 1 of Table 1. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

5.4.2 Depression

Table 3 reports results on depression outcomes. I find that a 15-minute delay in annual average sunset time increases the likelihood of being at risk of depression in adolescence by 2.1 pp and the likelihood of being diagnosed with depression by young adulthood by 2 pp (which correspond to 6% and 15% of the respective means). These point estimates are robust to accounting for additional interview characteristics, different latitude specifications and changes in sample size to address non-overlapping missing outcomes or use of covariates from one wave (see Table D.1 and D.2). Although there are no estimates available in the literature for the impact of chronic sleep deprivation on depression for adolescents, my point estimate for the CDI:S score (see Table D.1, column 6) is of similar magnitude to that found

[60] The latter spend less than 4 hours per day attending school, which could give more flexibility to nap in the afternoon, and they may also be more likely to do so because of heat during the day in India. In fact, Jagnani (2022) finds that the increase in nap duration takes place mostly between 2 and 4 PM. Between these times US adolescents are more than likely still in school, attending after school activities or travelling home.

for the CES-D score, a validated depression scale, for older Chinese workers in Giuntella et al. (2017). That study finds that a 10-minute delay in annual average sunset time, which reduces sleep in their sample by approximately 5 minutes, increases the depression score by 4% with respect to the mean. My point estimate represents a 3% increase of the mean CDI:S score per 5 minutes of sleep loss (15-minute delay in sunset time).

Table 3. Effect of later sunset time in adolescence on depression in adolescence and young adulthood

	(1) At risk of depression during adolescence (0, 1)	(2) Diagnosed with depression by young adulthood (0, 1)
Sunset (15 minutes)	0.021*	0.020**
	(0.012)	(0.008)
R^2	0.04	0.07
Mean dep. variable	0.36	0.13
N	3,806	2,785

Notes: Sample of adolescents in the CDS, 2002-2019 (column 1) and 2002-2014 (column 2). Estimates follow equations (1) and (2) and include sociodemographic characteristics (single-year age dummies and dummies for gender, race and public school), geographic characteristics (linear controls for annual average temperature, total annual precipitation, minimum distance to coast, population density and land area, and time zone indicators and latitude and its square) and interview characteristics (interview month and wave dummies for column 1, and wave dummies for column 2). At risk of depression is defined as a CDI:S score ≥ 3. Depression diagnosis is retrieved from the TAS, 2015-2019, measured at age 17-28. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

5.4.3 Educational attainment

Table 4 presents results for educational attainment. There is virtually no impact of annual average sunset time on the likelihood of graduating from high school or attending at least some college. The point estimates are small and statistically insignificant. In contrast, I find that a 15-minute delay in annual average sunset time reduces college graduation by 3.5 pp. These point estimates are relatively insensitive to changes in latitude specification, addressing potential selection bias introduced by varying sample sizes, and measuring college graduation at age 25 or over (see Tables E.1, E.2 and E.3). These results suggest that chronic sleep deprivation induced by later sunsets does not ultimately hinder adolescents' completion of high school or admittance to at least a two-year college but it does impair chances of college graduation. The former education levels may possibly be achieved even when sleep-deprived, as evidenced by the high proportion of US adolescents who graduate from high school and attend at least some college and do not get enough sleep (see Table B). On the other hand, even if this does not reflect on attaining said degrees, I do find some suggestive evidence that sleep-deprived adolescents perform worse on standardized tests taken during high school (see Table E.4).

The magnitude of the effect I find on college graduation seems in line with Jagnani's (2022) findings for primary and middle school completion rates of school-age children in India. Per 10-minute delay in annual average sunset time, which corresponds to 5 minutes of sleep loss in his sample, he finds respective increases of roughly 4% and 6% of the respective means. My estimate implies an increase of roughly 10% of the mean college graduation rate per 5 minutes of sleep loss (15-minute delay in sunset time). It is worth noting that in his sample 48% of students have completed primary school and only 21% have completed middle school. The somewhat higher magnitude found here could result from my focus on tertiary education attainment, which is a longer-run outcome and may possibly be more affected by chronic sleep loss, or on a high-income context where sleep efficiency is likely higher.

Table 4. Effect of later sunset time in adolescence on educational attainment

	(1) Graduated high school at age 20 or over (0, 1)	(2) Attended some college at age 20 or over (0, 1)	(3) Graduated college at age 23 or over (0, 1)
Sunset (15 minutes)	-0.001	-0.009	-0.035**
	(0.008)	(0.012)	(0.016)
R^2	0.04	0.08	0.15
Mean dep. variable	0.88	0.70	0.30
N	2,236	2,236	1,766

Notes: Sample of adolescents in the CDS, 2002-2014 (columns 1 and 2) and 2002-2007 (column 3). Estimates follow equation (2) and include sociodemographic characteristics (single-year age dummies and dummies for gender, race and public school), geographic characteristics (linear controls for annual average temperature, total annual precipitation, minimum distance to coast, population density and land area, and time zone indicators and latitude and its square) and interview characteristics (wave dummies). High school graduation and some college attendance are measured at age 20 or over in the TAS, 2005-2019. College graduation is measured at age 23 or over in the TAS, 2002-2015, and indicates having obtained at least a 4-year bachelor degree. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

5.4.4 Robustness checks and alternative mechanisms

I perform robustness checks to investigate potential confounding factors and to rule out alternative mechanisms, besides sleep, that may drive the results. First, following Giuntella & Mazzonna (2019), I exploit the sharp discontinuity in annual average sunset time at the

Eastern-Central time zone border (see Figures F.1 and F.2).[61] This is motivated by the potential concern that the baseline estimates, which are largely based on observations from the Eastern time zone, are driven, even after controlling for coastal distance, by longitude and its correlation with economic activity. Table F.1 reports regression discontinuity estimates. These are imprecisely estimated given the small sample size. However, the point estimates are similar to the baseline point estimates, once multiplied by four to obtain estimates per hour delay in annual average sunset time. These quantitatively similar results mitigate the potential concern that the baseline estimates are driven by economic activity through early sunset observations on the East coast.

Second, I conduct a placebo test using my baseline identification strategy. It is likely that the effects of sunset time on depression and education are driven by sleep loss, but I cannot rule out that there are other differences between eastern and western counties within time zones that are correlated with these outcomes. I therefore use visual or auditory impairment as a placebo outcome, as this has not been directly associated with sleep loss in adolescents. Indeed, I find no impact of later sunsets on serious seeing or hearing difficulties (see Table F.2).

Possible mechanisms at play, other than effects of sunset time on adolescents' own sleep, could be effects on sleep of their teachers, parents and neighbours. They are exposed to the same sunsets and thus are also likely to experience sleep loss. Such sleep loss may in turn influence their performance and health, which may then influence adolescents' outcomes. Indeed, previous studies have shown that sunset time influences adult wages and health (e.g. Gibson & Schrader, 2018; Giuntella & Mazzonna, 2019). I therefore follow Jagnani (2022) in examining these potential pathways by controlling for socioeconomic characteristics (family income, median household income at the county level and mothers' years of education) and school quality (pupil-teacher ratio, which relates to teacher absenteeism and performance).[62] The resulting point estimates (Table F.3) are mostly similar to the baseline estimates. The estimates for college graduation when controlling for socioeconomic characteristics, and in particular family income, are somewhat lower in magnitude (from 3.5 pp to around 3 pp) but remain large and significant. It thus appears unlikely that effects on parents and teachers are driving the results found on adolescents' outcomes. I cannot

[61] This is the only time zone border I can exploit, given that I do not have sufficient observations in the Mountain time zone.

[62] I am aware that some of these may be bad controls when trying to estimate the effect of sunset time. However, the objective here is to investigate possible mechanisms through which sunset time may impact outcomes. A reduction in magnitude of the estimated coefficient when controlling for these possible mechanisms may then indicate these mechanisms being at play. Furthermore, it is worth noting that residential sorting may be another concern. Controlling for family income and median household income should address this also. See Eagle & Glenn (2018) for a study which links teacher absences with pupil-teacher ratios.

however rule out that later sunset time may, to some extent, also operate through parental sleep loss and its impact on family income. Finally, while peers are also exposed to the same sunset time, it seems unlikely that peer effects in education, which when correctly estimated tend to be small (Angrist, 2014; Feld & Zölitz, 2017), are driving the effects of sunset time on adolescents' own outcomes. Jagnani (2022) also fails to find evidence that these alternative mechanisms drive his effects.

5.5 Discussion

In the last two decades, much research has been dedicated to investigating the harmful impacts of sleep loss on a variety of outcomes. Substantial gaps nevertheless remain in the literature. In particular, evidence on the effects of chronic sleep loss on adolescents' human capital formation is lacking. This paper uses variation in annual average sunset time across the United States to estimate effects on the sleep duration of adolescents, and the subsequent risks of depression and educational attainment by young adulthood.

I find that later sunset time delays adolescent bedtimes but, due to school start times which do not vary much with solar cues, has no discernible impact on wake times. As a result, adolescents exposed to later sunsets sleep, on average, 5 minutes less on each typical weeknight, are 3 pp more likely to obtain insufficient sleep, and instead spent more time on mentally-active sedentary leisure. Chronic sunset-induced sleep deprivation appears to have significant impacts on adolescents' mental health and educational attainment. Later sunsets increase the likelihood of adolescents being at risk of depression and of them being diagnosed with depression by young adulthood. Later sunset time also significantly reduces the likelihood of graduating from college but appears to have no impact on graduating from high school or attending some college, which are both substantially more prevalent than college graduation.

Various robustness checks leave my baseline estimates qualitatively unchanged. I also fail to find evidence for later sunset time operating primarily through mechanisms other than adolescent sleep, such as family income. However, I cannot completely rule out such alternative mechanisms. Adolescents who sleep less increase time allocated to mentally-active sedentary leisure. The effects on mental health and education are likely to also reflect this increase in leisure time. It appears unlikely however that it drives the results. In fact, association studies find that such leisure time is not associated with depression and may even be protective of developing it (Hallgren et al., 2018; Werneck at al., 2021). Playing video games has been associated with cognitive improvements (Kühn et al., 2014). Increased mentally-active sedentary leisure may thus even positively impact adolescent outcomes.

A potential limitation includes the lack of control for the possible influence of sunlight duration. Annual average sunlight duration is very similar across all locations in the contiguous US. However, later sunsets do imply more sunlight duration in the evening instead of in the morning. This may increase adolescents' daily exposure to sunlight and so may have a direct positive impact on mood (Kjaergaard et al., 2012). If so, this would introduce a downward bias in my estimates since adolescents who experience later sunsets, and so are potentially exposed to more sunlight, sleep less. The results would then represent lower bound estimates.

Lastly, it is good to note that, although location information and sleep duration is observed during adolescence, the effects may also contain the impact of cumulative sleep loss incurred during school-age childhood (as well as during young adulthood, for the long-run outcomes). This is because annual average sunset time captures long-term variation in sunset time, and adolescents tend to reside in the same location for an extended period of time. Overall, the effects should be interpreted as capturing the impact of cumulative sleep loss, incurred at least partly during adolescence.

The results suggest that chronic sleep deficits can have a lasting impact on the human capital formation of adolescents. Policies targeted at delaying school start times may contribute to increasing adolescent sleep duration by enabling later wake times and may thus reduce sleep deficits and their harmful consequences. Broader research into why adolescents do not sleep enough is also warranted. Sleep duration is not fixed, it is to some extent the result of individual choices (Biddle & Hamermesh, 1990) and adolescents may not make them optimally, partly due to time-inconsistent preferences (Kroese et al., 2014; Avery et al., 2022). Investigating time-uses that cause adolescents to postpone bedtimes and identifying the behavioral mechanisms underlying them can help create effective interventions to combat insufficient sleep.

Appendices A-F

Appendix A. The Children's Depression Inventory Short Form (CDI:S).

The 10 statements which are included in the short form are shown below.[63] The raw CDI:S score is calculated by summing the scores of each statement, where a, c, d, f and i are recoded as 1 = 0, 2 = 1 and 3 = 2, and items b, e, g, h and j as 1 = 2, 2 = 1, and 3 = 0.[64]

a. For the next 10 questions, select the sentence from each group that best describes your feelings during the last two weeks.

I am sad once in a while	1
I am sad many times	2
I am sad all the time	3

b. Select the sentence that best describes your feelings during the last 2 weeks.

Nothing will ever work out for me	1
I am not sure if things will work out for me	2
Things will work out for me O.K.	3

c. Select the sentence that best describes your feelings during the last 2 weeks.

I do most things O.K.	1
I do many things O.K.	2
I do everything wrong	3

d. Select the sentence that best describes your feelings during the last 2 weeks.

I hate myself	1
I do not like myself	2
I like myself	3

e. Select the sentence that best describes your feelings during the last 2 weeks.

I feel like crying every day	1
I feel like crying many days	2
I feel like crying once in awhile	3

f. Select the sentence that best describes your feelings during the last 2 weeks.

Things bother me all the time	1
Things bother me many times	2
Things bother me once in awhile	3

[63] See https://psidonline.isr.umich.edu/cds/questionnaires/cds-ii/english/cdsii_child_assess.pdf (pages 60-61).
[64] See https://simba.isr.umich.edu/CDS/DC/s.aspx and https://simba.isr.umich.edu/cb.aspx?vList=CDI_02.

g. Select the sentence that best describes your feelings during the last 2 weeks.

I look O.K	1
There are some bad things about my looks	2
I look ugly	3

h. Select the sentence that best describes your feelings during the last 2 weeks.

I do not feel alone	1
I feel alone many times	2
I feel alone all the time	3

i. Select the sentence that best describes your feelings during the last 2 weeks.

I have plenty of friends	1
I have some friends, but I wish I had more	2
I do not have any friends	3

j. Select the sentence that best describes your feelings during the last 2 weeks.

Nobody really loves me	1
I am not sure if anybody loves me	2
I am sure that somebody loves me	3

Appendix B. Summary statistics Table**.**

Table B. Summary statistics.

	Mean	SD	N
Annual average sunset time (hours 0-24)	18.93	0.27	4,719
Age CDS	14.08	2.01	4,719
Age TAS	23.49	3.23	2,785
Female	0.50		4,719
Race			
White	0.45		4,719
Black	0.45		4,719
Other	0.10		4,719
Time zone			
Eastern	0.52		4,719
Central	0.34		4,719
Pacific	0.14		4,719
Time-use			
Sleep (hours)	8.59	1.45	2,636
Insufficient sleep (0,1)	0.40		2,636
Bedtime (hours 0-24)	21.84	0.97	2,289
Wake time (hours 0-24)	6.55	0.76	2,289
Exercise (hours)	0.60	1.06	2,636
Mentally-active sedentary leisure (hours)	0.98	1.37	2,636
Passive sedentary leisure (hours)	2.19	1.86	2,636
Naps (hours)	0.20	0.71	2,636
Attend school (hours)	6.56	1.86	2,636
Study (hours)	0.76	0.96	2,636
Depression			
CDI:S score (0-8)	2.57	2.82	3,806
At risk of depression during adolescence (0,1)	0.36		3,806
Depression diagnosis by young adulthood (0,1)	0.13		2,785
Educational attainment			
Graduated high school at age 20 or over (0,1)	0.88		2,236
Attended some college at age 20 or over (0,1)	0.70		2,236
Graduated college at age 23 or over (0,1)	0.30		1,766

Notes: Sample of adolescents in the CDS, 1997-2019. Sample sizes vary due to non-overlapping missing outcomes. Summary statistics for variables used in all analyses are shown for all observations which are included in at least one of the analyses (N = 4,719). Sleep observations < 4 and > 16 hours per day are not included. Insufficient sleep for age is defined as < 9 hours of sleep per day for adolescents aged 12 or younger and as < 8 hours for adolescents aged 13 or older. Bedtimes and wake times are retrieved from the disaggregated CDS time diary files, which timestamp nighttime sleep episodes. Here, the sample size is smaller because I exclude observations of adolescents who experience fragmented sleep and thus report multiple bedtimes and wake times per diary day. At risk of depression is defined as a CDI:S score ≥ 3. Age TAS is the age of CDS adolescents at their last TAS interview. Age TAS, depression diagnosis and educational attainment are retrieved from the TAS.

Appendices C-E. Alternative specifications and additional results for sleep and other time-uses (C), depression (D) and educational attainment (E).

Table C.1. Effect of later sunset time on adolescent sleep (minutes) for alternative specifications

	(1) Sleep	(2) Sleep	(3) Sleep	(4) Sleep
Sunset (15 minutes)	-5.260**	-4.341*	-5.453**	-4.678**
	(2.164)	(2.310)	(2.315)	(2.225)
R^2	0.11	0.10	0.09	0.09
Mean dep. variable	08:35[a]	08:35[a]	08:35[a]	08:36[a]
N	2,636	2,636	2,636	2,641
Interview characteristics	Yes	No	No	No
Latitude spline	No	Yes	No	No
Latitude parallels	No	No	Yes	No
Sleep outliers	No	No	No	Yes

Notes: Controls and sample as in column 1 of Table 1, unless otherwise specified below. Column 1 includes day-of-the-week dummies and two dummies which indicate if the sleep observation took place in the week following a DST transition or took place during covid (i.e., in March of 2020 or later, concerns < 1% of observations). Column 2 follows Gibson & Shrader's (2018) specification and includes a ten-piece linear spline in latitude instead of latitude and latitude squared. Column 3 follows Giuntella & Mazzonna (2019) and includes indicators for 9 cells constructed using the three time zones and three latitude parallels (< 34th parallel, 34th to 40th parallel, > 40th parallel) instead of time zone indicators and latitude and latitude squared. Column 4 excludes sleep observations < 2 and > 20 hours per day instead of < 4 and > 16 hours per day. Standard errors are shown in parentheses and are clustered at the county level. [a] hours : minutes *** p < 0.01, ** p < 0.05, * p < 0.1

Table C.2. Effect of later sunset time on adolescent leisure activities (minutes)

	(1) Mentally-active sedentary leisure: Games	(2) Passive sedentary leisure: Watch TV	(3) Passive sedentary leisure: Social media
Sunset (15 minutes)	3.095*	-1.616	-0.817
	(1.724)	(2.301)	(1.025)
R^2	0.08	0.04	0.08
Mean dep. variable	46.53	93.62	4.10
N	2,636	2,636	590

Notes: Controls and sample as in column 1 of Table 1, unless otherwise indicated below. The sample size in column 3 is smaller since social media use is only recorded since 2014. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

Table D.1. Effect of later sunset time on risk of depression in adolescence for alternative specifications

	(1) At risk of depression	(2) At risk of depression	(3) At risk of depression	(4) At risk of depression	(5) At risk of depression	(6) Depression score
Sunset (15 minutes)	0.021*	0.021	0.023*	0.018	0.020	0.081
	(0.012)	(0.013)	(0.012)	(0.012)	(0.014)	(0.066)
R^2	0.04	0.04	0.04	0.04	0.04	0.05
Mean dep. variable	0.36	0.35	0.36	0.36	0.36	2.57
N	3,806	2,986	3,806	3,806	2,433	3,806
Covid dummy	Yes	No	No	No	No	No
Exclude CDS 2019	No	Yes	No	No	No	No
Latitude spline	No	No	Yes	No	No	No
Latitude parallels	No	No	No	Yes	No	No
Overlapping sample	No	No	No	No	Yes	No

Notes: Controls, sample and outcome as in column 1 of Table 3, unless otherwise specified below. Column 1 includes a dummy for covid (i.e., if the interview took place in March of 2020 or later). Column 2 excludes CDS 2019 observations to check robustness to covid. Latitude spline and parallels as in Table C.1. Column 5 restricts the analysis to adolescents in the CDS, 2002-2014, for whom I observe depression diagnosis. Column 6 shows the impact on the raw CDI:S score. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

Table D.2. Effect of later sunset time on depression diagnosis by young adulthood for alternative specifications

	(1) Diagnosed with depression	(2) Diagnosed with depression	(3) Diagnosed with depression	(4) Diagnosed with depression	(5) Diagnosed with depression	(6) Diagnosed with depression
Sunset (15 minutes)	0.020**	0.017**	0.021**	0.021**	0.019**	0.019**
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)
R^2	0.07	0.06	0.07	0.07	0.07	0.07
Mean dep. variable	0.13	0.11	0.13	0.13	0.12	0.13
N	2,785	2,478	2,785	2,785	2,433	2,791
Age TAS	Yes	No	No	No	No	No
Exclude TAS 2019	No	Yes	No	No	No	No
Latitude spline	No	No	Yes	No	No	No
Latitude parallels	No	No	No	Yes	No	No
Overlapping sample	No	No	No	No	Yes	No
Other CDS sample	No	No	No	No	No	Yes

Notes: Controls, sample and outcome as in column 2 of Table 3, unless otherwise specified below. Column 1 includes single-year age dummies for the latest age at which I observe adolescents in the TAS, 2005-2019. Column 2 measures depression diagnosis using the 2005-2017 TAS observations to check robustness to covid. Latitude spline and parallels as in Table C.1. Column 5 restricts the analysis to adolescents for whom I observe CDI:S scores. Column 6 includes covariates from CDS 2002, instead of CDS 2007, for adolescents whom I observe in both waves. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

Table E.1. Effect of later sunset time on high school graduation at age 20 or over for alternative specifications

	(1) Graduated high school	(2) Graduated high school	(3) Graduated high school	(4) Graduated high school
Sunset (15 minutes)	-0.007	-0.006	-0.000	-0.001
	(0.008)	(0.007)	(0.009)	(0.008)
R^2	0.05	0.05	0.05	0.04
Mean dep. variable	0.88	0.88	0.88	0.88
N	2,236	2,236	1,766	2,241
Latitude spline	Yes	No	No	No
Latitude parallels	No	Yes	No	No
Overlapping sample	No	No	Yes	No
Other CDS sample	No	No	No	Yes

Notes: Controls, sample and outcome as in column 1 of Table 4, unless otherwise specified below. Latitude spline and parallels as in Table C.1. Column 3 restricts the analysis to adolescents for whom I observe college graduation by age 23 or over. Column 4 includes covariates from CDS 2002, instead of CDS 2007, for adolescents whom I observe in both waves. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

Table E.2. Effect of later sunset time on college attendance at age 20 or over for alternative specifications

	(1) Attended some college	(2) Attended some college	(3) Attended some college	(4) Attended some college
Sunset (15 minutes)	-0.017	-0.016	-0.006	-0.010
	(0.013)	(0.012)	(0.013)	(0.012)
R^2	0.08	0.08	0.08	0.07
Mean dep. variable	0.70	0.70	0.71	0.69
N	2,236	2,236	1,766	2,241
Latitude spline	Yes	No	No	No
Latitude parallels	No	Yes	No	No
Overlapping sample	No	No	Yes	No
Other CDS sample	No	No	No	Yes

Notes: Controls, sample and outcome as in column 2 of Table 4, unless otherwise specified below. Latitude spline and parallels as in Table C.1. Column 3 restricts the sample to adolescents for whom I observe college graduation by age 23 or over. Column 4 includes covariates from CDS 2002, instead of CDS 2007, for adolescents whom I observe in both waves. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

Table E.3. Effect of later sunset time on college graduation at age 23 or over for alternative specifications

	(1) Graduated college	(2) Graduated college	(3) Graduated college	(4) Graduated college
Sunset (15 minutes)	-0.038**	-0.031*	-0.042**	-0.033**
	(0.016)	(0.016)	(0.019)	(0.016)
R^2	0.16	0.15	0.15	0.14
Mean dep. variable	0.30	0.30	0.33	0.29
N	1,766	1,766	1,261	1,773
Latitude spline	Yes	No	No	No
Latitude parallels	No	Yes	No	No
At age 25 or over	No	No	Yes	No
Other CDS sample	No	No	No	Yes

Notes: Controls, sample and outcome as in column 3 of Table 4, unless otherwise specified below. Latitude spline and parallels as in Table C.1. Column 3 measures college graduation at age 25 or over in the TAS, 2005-2015. Column 4 includes covariates from CDS 2002, instead of CDS 2007, for adolescents whom I observe in both waves. Standard errors are shown in parentheses and are clustered at the county level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table E.4. Effect of later sunset time on standardized test scores

	(1) ACT score	(2) SAT score
Sunset (15 minutes)	-0.185	-6.186
	(0.270)	(16.03)
R^2	0.17	0.10
Mean dep. variable	22.34	1084.61
N	812	677

Notes: Controls and sample as in column 3 of Table 4, unless otherwise specified below. ACT and SAT scores are composite scores and are measured at age 17 or over in the TAS. Not all adolescents take these tests (or report a valid test score), hence the smaller sample size. Standard errors are shown in parentheses and are clustered at the county level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix F. Alternative explanations.

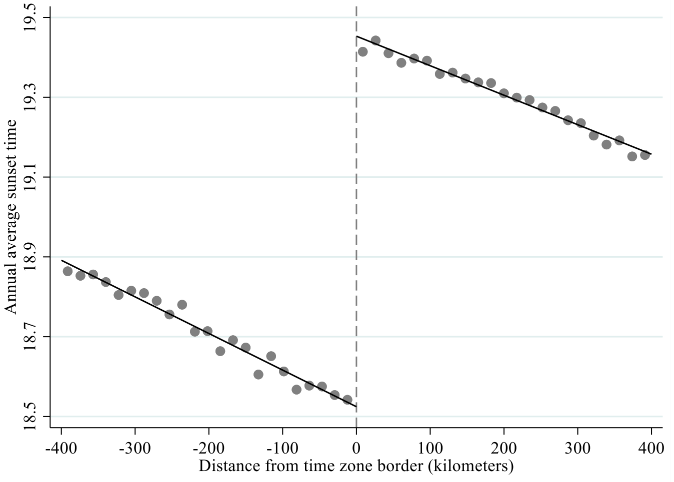


Figure F.1. Discontinuity in average annual sunset time at the Eastern-Central time zone border. *Notes*: The Eastern (late) side is depicted on the right side of the border (cutoff), the Central (early) side is depicted on the left side of the border (cutoff). The number of bins is automatically determined by the cmogram command.

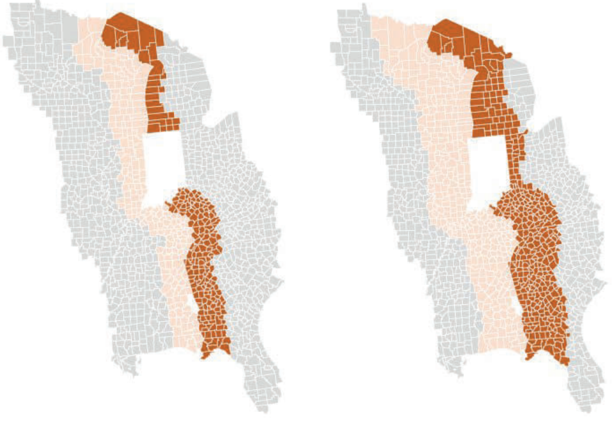


Figure F.2. Counties within 150 and 250 kilometers from the Eastern-Central time zone border. *Notes*: These figures illustrate the Eastern-Central time zone border and indicate counties within 150 kilometers (left figure) and within 250 kilometers (right figure) from the time zone border, on a map which shows counties within 500 kilometers from the border. Light colored counties are on the early side of the time zone border, dark colored on the late side. Indiana, which does not observe DST for some period under study, and counties which lie in two time zones are excluded. This map was generated using Excel maps, which does not depict the water body of the Great Lakes above Indiana.

Table F.1. Regression discontinuity estimates from the Eastern-Central time zone border

	(1) Sleep (hours)	(2) At risk of depression	(3) Diagnosed with depression	(4) Graduated high school	(5) Attended some college	(6) Graduated college
Late sunset side	-0.351	0.092	0.067	0.007	-0.033	-0.136
	(0.316)	(0.131)	(0.086)	(0.079)	(0.126)	(0.136)
Bandwidth (kilometers)	244	130	200	237	237	197
N	433	398	361	342	342	223

Notes: Sample of adolescents in the CDS, 1997-2019. Table reports conventional local linear regression discontinuity estimates using the rdrobust command (Calonico et al., 2017). The estimates are constructed using the triangular kernel and optimal bandwidths are chosen by the built-in optimal bandwidth selector (mserd), except for columns 2 and 3. Here, the optimal bandwidths are large (> 350), and given the bias-variance trade-off inherent in the optimization, I manually override these bandwidths to reduce bias. I choose smaller bandwidths while still maintaining a sample size similar to that used for the other estimates. Following Giuntella & Mazzonna (2019), all estimates include a linear control for latitude and indicators for the three latitude parallels (< 34th parallel, 34th to 40th parallel, > 40th parallel). Standard errors are shown in parentheses and are clustered at the county level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table F.2. Effect of later sunset time on visual or auditory impairment as a placebo outcome

	(1) Visual or auditory impairment
Sunset (15 minutes)	0.002
	(0.006)
R^2	0.02
Mean dep. variable	0.06
N	3,806

Notes: Controls and sample as in column 1 of Table 3. Visual or auditory impairment is an indicator which takes value 1 when a doctor or other health professional has ever told the adolescents' primary caregiver that the adolescent has a serious hearing difficulty or deafness or has serious difficulty seeing or blindness, 0 otherwise. Standard errors are shown in parentheses and are clustered at the county level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table F.3. Effect of later sunset time on depression and educational attainment controlling for socioeconomic status (SES) and school quality

Panel A

	(1) At risk of depression	(2) At risk of depression	(3) At risk of depression	(4) At risk of depression	(5) At risk of depression
Sunset (15 minutes)	0.021*	0.021*	0.020	0.018	0.024*
	(0.012)	(0.012)	(0.014)	(0.014)	(0.014)
R^2	0.04	0.04	0.04	0.04	0.04
Mean dep. variable	0.36	0.36	0.36	0.36	0.36
N	3,806	3,806	2,553	2,553	3,153

Panel B

	(1) Diagnosed with depression	(2) Diagnosed with depression	(3) Diagnosed with depression	(4) Diagnosed with depression	(5) Diagnosed with depression
Sunset (15 minutes)	0.018**	0.019**	0.021**	0.019**	0.025***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
R^2	0.07	0.07	0.07	0.07	0.07
Mean dep. variable	0.13	0.13	0.13	0.13	0.12
N	2,785	2,785	2,703	2,703	2,322

Panel C

	(1) Graduated high school	(2) Graduated high school	(3) Graduated high school	(4) Graduated high school	(5) Graduated high school
Sunset (15 minutes)	0.002	0.000	-0.001	0.001	0.003
	(0.007)	(0.008)	(0.007)	(0.008)	(0.009)
R^2	0.06	0.05	0.07	0.08	0.05
Mean dep. variable	0.88	0.88	0.88	0.88	0.88
N	2,236	2,236	2,181	2,181	1,901

Panel D

	(1) Attended some college	(2) Attended some college	(3) Attended some college	(4) Attended some college	(5) Attended some college
Sunset (15 minutes)	-0.002	-0.006	-0.007	-0.002	-0.004
	(0.011)	(0.011)	(0.011)	(0.010)	(0.013)
R^2	0.10	0.09	0.13	0.14	0.07
Mean dep. variable	0.70	0.70	0.70	0.70	0.69
N	2,236	2,236	2,181	2,281	1,901

Panel E

	(1) Graduated college	(2) Graduated college	(3) Graduated college	(4) Graduated college	(5) Graduated college
Sunset (15 minutes)	-0.026*	-0.031**	-0.036***	-0.028**	-0.036**
	(0.014)	(0.013)	(0.013)	(0.012)	(0.018)
R^2	0.19	0.18	0.23	0.26	0.13
Mean dep variable	0.30	0.30	0.30	0.30	0.27
N	1,766	1,766	1,724	1,724	1,513
Family income	Yes	No	No	No	No
Median household income	No	Yes	No	No	No
Education mother	No	No	Yes	No	No
All SES characteristics	No	No	No	Yes	No
School quality	No	No	No	No	Yes

Notes: Controls, sample and outcome, unless otherwise specified below, as in column 1 of Table 3 for Panel A, as in column 2 of Table 3 for Panel B, as in column 1 of Table 4 for Panel C, as in column 2 of Table 4 for Panel D and as in column 3 of Table 4 for Panel E. Column 1 adds a linear control for family income, which is retrieved from the PSID family-level data. It represents income from the previous tax year. Column 2 adds a linear control for median household income at the county level. This data is retrieved from the US Census Bureau.[65] Column 3 linearly controls for the years of education of the mother. This is obtained from the TAS and could not be established for everyone, which is why some observations are missing. Column 4 controls for family income, median household income and mothers' years of education. Column 5 controls for the pupil-teacher ratio at the school the adolescent attends. This information is retrieved from the National Center for Education Statistics (NCES), and linked with restricted-use PSID school codes.[66] The pupil-teacher ratios can only be linked for adolescents who attend public schools, which is why the sample size is smaller. Standard errors are shown in parentheses and are clustered at the county level. *** p < 0.01, ** p < 0.05, * p < 0.1

[65] United States Census Bureau. (2021). *Small Area Income and Poverty Estimates (SAIPE) Datasets*. https://www.census.gov/programs-surveys/saipe/data/datasets.html

[66] National Center for Education Statistics. (n.d.). *Elementary/Secondary Information System (ELSi)*. https://nces.ed.gov/ccd/elsi/tableGenerator.aspx

Chapter 6

Discussion

The four empirical studies in this thesis examine health- and age-related expectations and health behaviors. They do so mostly independently. On the one hand, investigating the accuracy of expectations which could influence behavior, and, on the other, examining health behaviors and some of their consequences.

Chapters 2 and 3 show that older Americans hold inaccurate expectations of long-term care risk and of when they will retire. Sources of inaccuracy differ. In the case of long-term care risk, individuals severely underweight risk factors that providers of long-term care insurance observe. This may lead them not to take insurance that they would have purchased had their risk perceptions been accurate. In contrast, US workers do rather well in using information contained in objective predictors of retirement when forming their retirement expectations. Inaccuracy here mainly derives from uncertainty surrounding work status at standard retirement ages. But the expectations are also noisy, which could indicate attention paid to irrelevant factors when forming and reporting expectations. The latter is also true for expectations of long-term care risk. Finally, both chapters indicate that the lower educated and the least cognitively able hold the least accurate expectations. This makes them the most vulnerable to detrimental consequences resulting from suboptimal behavior.

Chapters 4 and 5 investigate the potential impact of health behaviors. Together, the evidence generated in these chapters suggests that preventing unhealthy behavior can raise well-being and reduce inequalities. Chapter 4 finds that gender differences in smoking and drinking together can explain a large share of the gender mortality gap in Russia. This joint contribution, as well as the gender gap itself, declined over time as tobacco and alcohol control policies were introduced. These findings, while not causal, are consistent with men's reduced tobacco and alcohol consumption – possibly driven by the control policies – explaining most of the decrease in the gender mortality gap. Chapter 5 finds that persistent exposure to later sunset times reduces sleep duration of US adolescents, increases their likelihood of experiencing depressive symptoms and of subsequently being diagnosed with depression by young adulthood. That exposure has no discernable impact on high school graduation or college attendance but is found to decrease the likelihood of graduating from college. These results suggest that chronic sleep deprivation in adolescence can have lasting adverse effects on human capital formation.

Expectations are, presumably, an important determinant of behavior. We would expect, for example, that the perceived health and education consequences of insufficient sleep would influence the decision to go to bed on time. This thesis has, for the most part, not explored the link between expectations and behavior. Chapter 2 does investigate the impact of reported long-term care risk perceptions on insurance behavior of older Americans. Although it is difficult to determine whether individuals act on their expectations, the various estimates are at least all consistent with expectations influencing take-up of long-term care insurance. An important avenue for future research is to investigate the relation between expectations and health behavior.

Inaccurate beliefs of health risk can lead to suboptimal health behavior. Individuals may not hold accurate beliefs because they are unable to acquire full information and they may lack the ability to process this information rationally. This thesis does not formally investigate the formation of beliefs. Yet doing so may help understand why individuals seem to do well in incorporating observed information into retirement expectations, but not into long-term care expectations. The latter is also true for survival expectations (Bago d'Uva & O'Donnell, 2022). Relatedly, how individuals revise their beliefs in response to new information or health shocks can influence their health behavior. Understanding the updating of beliefs, and the subsequent impact on health behavior, is also a fruitful direction for further research. Particularly, standard neoclassical economics assumes that individuals separate their beliefs and preferences when making decisions, yet beliefs may be motivated by preferences and, in some instances, beliefs may even adjust to rationalize the decisions we make (Bénabou & Tirole, 2016; Caplin & Leahy, 2019; Manski, 2023; Prati & Saucet, 2024; Zimmerman, 2020). Having long-term care insurance may raise the subjective expectation of moving to a nursing home, and the preference for not moving to a nursing home in old age may lower the subjective risks associated with it. Retirement, which is typically not perceived as a severe, adverse event, may not be subject to the same adjustment in subjective risks. Further research on the separability of beliefs and preferences may elucidate this.

This thesis contributes to two mostly separate strands of the economics literature but with important connections that merit more attention – expectations and health behaviors. It shows that individuals hold inaccurate expectations of important health- and age-related events which may influence their behavior and lead to adverse consequences in old age. It also shows that various health behaviors, likely influenced by the (possibly also inaccurate) perceived risks of engaging in them, can explain health inequalities and be detrimental to human capital formation. Finally, the evidence presented suggests that the least educated and least cognitively able are even more prone to holding inaccurate expectations; this can in turn lead to their suboptimal decision making with negative consequences for well-being over the life cycle. Future research on the relation between expectations and health behaviors, and on the instrumental roles of beliefs and preferences – and the extent to which they may not be separable – may be helpful to inform policies aiming at preventing suboptimal behavior, increasing well-being over the life cycle and reducing inequalities in it.

Bibliography

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.

Allen, N. (2000). The times they are a-changing: the influence of railroad technology on the adoption of standard time zones in 1883. *The History Teacher, 33*(2), 241-256.

Allgaier, A. K., Frühe, B., Pietsch, K., Saravo, B., Baethmann, M., & Schulte-Körne, G. (2012). Is the Children's Depression Inventory Short version a valid screening tool in pediatric care? A comparison to its full-length version. *Journal of Psychosomatic Research, 73*(5), 369-374.

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. 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*.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics, 30*, 98-108.

Avery, M., Giuntella, O., & Jiao, P. (2022). Why Don't We Sleep Enough? A Field Experiment Among College Students. *Review of Economics and Statistics*, 1-45.

Ayyagari, P. (2019). Health insurance and early retirement plans: Evidence from the affordable care act. *American Journal of Health Economics, 5*(4), 533-560.

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.

Barford, A., Dorling, D., Smith, G. D., & Shaw, M. (2006). Life expectancy: women now on top everywhere. *BMJ, 332*(7545), 808.

Bassett, W. F., & Lumsdaine, R. L. (2001). Probability limits: Are subjective assessments adequately accurate? *Journal of Human Resources, 36*(2), 327-363.

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

Bavafa, H., Mukherjee, A., & Welch, T. Q. (2023). Inequality in the golden years: Wealth gradients in disability-free and work-free longevity in the United States. *Journal of Health Economics, 92*, 102820.

Benitez-Silva, H., & Dwyer, D. S. (2005). The rationality of retirement expectations and the role of new information. *Review of Economics and Statistics, 87*(3), 587-592.

Bernheim, B. D. (1989). The timing of retirement: A comparison of expectations and realizations. In Wise, D.A. (Ed.), *The economics of aging* (pp. 335-358). University of Chicago Press.

Bessone, P., Rao, G., Schilbach, F., Schofield, H., & Toma, M. (2021). The economic consequences of increasing sleep among the urban poor. *The Quarterly Journal of Economics, 136*(3), 1887-1941.

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.

Boerma, T., Hosseinpoor, A. R., Verdes, E., & Chatterji, S. (2016). A global assessment of the gender gap in self-reported health with survey data from 59 countries. *BMC Public Health, 16*(1), 675.

Bottazzi, R., Jappelli, T., & Padula, M. (2006). Retirement expectations, pension reforms, and their impact on private wealth accumulation. *Journal of Public Economics, 90*(12), 2187-2212.

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.

Brainerd, E., & Cutler, D. M. (2005). Autopsy on an Empire: Understanding Mortality in Russia and the Former Soviet Union. *Journal of Economic Perspectives, 19*(1), 107–130.

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.

Bruine de Bruin, W., Chin, A., Dominitz, J., & van der Klaauw, W. (2023). Household surveys and probabilistic questions. In R. Bachmann, G. Topa & W. van der Klaauw (Eds.), *Handbook of Economic Expectations* (pp. 3-31). Academic Press.

Burke, M. A., & Carman, K. G. (2017). You can be too thin (but not too tall): Social desirability bias in self-reports of weight and height. *Economics & Human Biology, 27*, 198-222.

Caliendo, F. N., Casanova, M., Gorry, A., & Slavov, S. (2023). Retirement timing uncertainty: Empirical evidence and quantitative evaluation. *Review of Economic Dynamics, 51*, 226-266.

Cappuccio, F. P., D'Elia, L., Strazzullo, P., & Miller, M. A. (2010). Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep, 33*(5), 585-592.

Case, A., & Paxson, C. H. (2005). Sex Differences in Morbidity and Mortality. *Demography, 42*(2), 189–214.

Chan, S., & Stevens, A. H. (2004). Do changes in pension incentives affect retirement? A longitudinal study of subjective retirement expectations. *Journal of Public Economics, 88*(7-8), 1307-1333.

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

Crimmins, E. M., Kim, J. K., Langa, K. M., & Weir, D. R. (2011). Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66*(suppl_1), i162-i171.

Crimmins, E. M., Shim, H., Zhang, Y. S., & Kim, J. K. (2019). Differences between men and women in mortality and the health dimensions of the morbidity process. *Clinical Chemistry, 65*(1), 135-145.

Cullen, M. R., Baiocchi, M., Eggleston, K., Loftus, P., & Fuchs, V. (2015). The weaker sex? Vulnerable men, resilient women, and variations in sex differences in mortality since 1900 (No. w21114). *National Bureau of Economic Research*.

Danilova, I., Shkolnikov, V. M., Andreev, E., & Leon, D. A. (2020). The changing relation between alcohol and life expectancy in Russia in 1965–2017. *Drug and Alcohol Review, 39*(7), 790–796.

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.

De Bruin, E. J., van Run, C., Staaks, J., & Meijer, A. M. (2017). Effects of sleep manipulation on cognitive functioning of adolescents: A systematic review. *Sleep Medicine Reviews, 32*, 45-57.

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

De Nardi, M., Pashchenko, S., & Porapakkarm, P. (in press). The lifetime costs of bad health. *The Review of Economic Studies*.

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. (2023). Expectations about the productivity of effort and academic outcomes: Evidence from a randomized information intervention. *Working paper*. http://dx.doi.org/10.2139/ssrn.4483970

Dobkin, C., Finkelstein, A., Kluender, R., & Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *American Economic Review, 108*(2), 308-352.

Dwyer, D. S., & Hu, J., (2000). Retirement expectations and realizations: The role of health shocks and economic factors. In O. S. Mitchell, P. B. Hammond, & A. M. Rappaport (Eds.), *Forecasting retirement needs and retirement wealth* (pp. 274–287). Philadelphia: University of Philadelphia Press.

Eagle, D.L., & Glenn, W.J. (2018). Teacher Absences in the Commonwealth of Virginia: An Analysis of Patterns and Predictors and Implications for Policy. *Journal of School Administration Research and Development, 3*(1), 32-41.

Edwards, F. (2012). Early to rise? The effect of daily start times on academic performance. *Economics of Education Review, 31*(6), 970-983.

Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics, 35*(2), 387-428.

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*.

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L. M., Nickless, A., Harrison, P. J., ... & Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. *The Lancet Psychiatry, 4*(10), 749-758.

Freeman, D., Sheaves, B., Waite, F., Harvey, A. G., & Harrison, P. J. (2020). Sleep disturbance and psychiatric disorders. *The Lancet Psychiatry, 7*(7), 628-637.

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.

Green, C. A., & Pope, C. R. (1999). Gender, psychosocial factors and the use of medical services: a longitudinal analysis. *Social Science & Medicine, 48*(10), 1363-1372.

Grigoriev, P., & Andreev, E. M. (2015). The huge reduction in adult male mortality in Belarus and Russia: Is it attributable to anti-alcohol measures? *PLoS one, 10*(9), e0138021.

Haider, S. J., & Stephens Jr, M. (2007). Is there a retirement-consumption puzzle? Evidence using subjective retirement expectations. *The Review of Economics and Statistics, 89*(2), 247-264.

Hale, L., & Guan, S. (2015). Screen time and sleep among school-aged children and adolescents: a systematic literature review. *Sleep Medicine Reviews, 21*, 50-58.

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

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

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

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

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 (2023). *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.

Health and Retirement Study (2023). *HRS Biennial Core and Exit Files, 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.

Health and Retirement Study (2023). *RAND HRS Biennial Files, 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.

Hudomiet, P., Hurd, M. D., & Rohwedder, S. (2021a). Forecasting mortality inequalities in the US based on trends in midlife health. *Journal of Health Economics, 80*, 102540.

Hudomiet, P., Hurd, M. D., Parker, A. M., & Rohwedder, S. (2021b). The effects of job characteristics on retirement. *Journal of Pension Economics & Finance, 20*(3), 357-373.

Hurd, M. (1999). Labor market transitions in the HRS: effects of the subjective probability of retirement and of pension eligibility. In J.P. Smith & R.J. Willism (Eds.), *Wealth, work and health: Innovations in measurement in the social sciences* (pp. 267-290). Michigan: University of Michigan Press.

Hurd, M. D. (2009). Subjective probabilities in household surveys. *Annual Review of Economics, 1*(1), 543-562.

Hurd, M.D., & McGarry, K. (1995). Evaluation of the subjective probabilities of survival in the health and retirement study. *The Journal of Human Resources, 30*, S268-S292.

Hurd, M.D., & McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal, 112*(482), 966-985.

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

Idler, E. L. (2003). Discussion: Gender Differences in Self-Rated Health, in Mortality, and in the Relationship Between the Two. *The Gerontologist, 43*(3), 372–375.

Inchley, J., Currie, D., & Young, T. (2016). Growing up unequal: gender and socioeconomic differences in young people's health and well-being. *Health Behaviour in School-aged Children (HBSC) study: international report from the 2013/2014 survey* (No. 7). World Health Organization.

Jagnani, M. (2022). Children's Sleep and Human Capital Production. *Review of Economics and Statistics*, 1-45.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263–291.

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.

Kolip, P., & Lange, C. (2018). Gender inequality and the gender gap in life expectancy in the European Union. *European Journal of Public Health, 28*(5), 869–872.

Kroese, F. M., De Ridder, D. T., Evers, C., & Adriaanse, M. A. (2014). Bedtime procrastination: introducing a new area of procrastination. *Frontiers in Psychology, 5*, 611.

Kossova, T., Kossova, E., & Sheluntcova, M. (2020). Gender Gap in Life Expectancy in Russia: The Role of Alcohol Consumption. *Social Policy and Society, 19*(1), 37–53.

Kuznetsova, P. O. (2020). Alcohol mortality in Russia: assessment with representative survey data. *Population and Economics, 4*, 75.

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

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

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

Library of Congress (2010). Time!. http://memory.loc.gov/ammem/today/nov18.html

Lillard, L.A., & Willis, R.J. (2001). Cognition and wealth: the importance of probabilistic thinking. *Michigan Retirement Research Center Research Paper No. WP 2001-007*.

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.

Lindeboom, M., & Montizaan, R. (2020). Disentangling retirement and savings responses. *Journal of Public Economics, 192*, 104297.

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

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy and retirement planning in the United States. *Journal of Pension Economics & Finance, 10*(4), 509-525.

Luy, M. (2003). Causes of Male Excess Mortality: Insights from Cloistered Populations. *Population and Development Review, 29*(4), 647–676.

Luy, M., & Gast, K. (2014). Do women live longer or do men die earlier? Reflections on the causes of sex differences in life expectancy. *Gerontology, 60*(2), 143-153.

Luy, M., & Wegner-Siegmundt, C. (2014). The impact of smoking on gender differences in life expectancy: more heterogeneous than often stated. *The European Journal of Public Health, 25*(4), 706–710.

Manski, C. F. (2004). Measuring expectations. *Econometrica, 72*(5), 1329-1376.

Manski, C. F. (2023). Looking ahead to research enhancing measurement of expectations. In R. Bachmann, G. Topa & W. van der Klaauw (Eds.), *Handbook of Economic Expectations* (pp. 809-822). Academic Press.

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

Matricciani, L., Olds, T., & Petkov, J. (2012). In search of lost sleep: secular trends in the sleep time of school-aged children and adolescents. *Sleep Medicine Reviews, 16*(3), 203-211.

McGarry, K. (2004). Health and retirement: do changes in health affect retirement expectations? *Journal of Human Resources, 39*(3), 624-648.

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.

Medicare. (n.d.). Medicare Part A coverage – nursing home care. https://www.medicare.gov/what-medicare-covers/what-part-a-covers/medicare-part-a-coverage-nursing-home-care

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

National Center for Education Statistics (2020). Start time for US public high schools. https://nces.ed.gov/pubs2020/2020006/index.asp

National Centers for Environmental Information (2023). Climate at a Glance: County Mapping. https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/mapping

National Oceanic and Atmospheric Administration (2020). Solar Calculator: Sunrise, Sunset, Solar Noon and Solar Position for Any Place on Earth. https://gml.noaa.gov/grad/solcalc/

OECD. (2023). Daily smokers, % of population aged 15+. https://data.oecd.org/healthrisk/daily-smokers.htm

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.

Owens, J., Adolescent Sleep Working Group, Committee on Adolescence, Au, R., Carskadon, M., Millman, R., ... & O'Brien, R. F. (2014). Insufficient sleep in adolescents and young adults: an update on causes and consequences. *Pediatrics, 134*(3), e921-e932.

Panel Study of Income Dynamics. (2021). *Public and restricted-use dataset*. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

Paruthi, S., Brooks, L. J., D'Ambrosio, C., Hall, W. A., Kotagal, S., Lloyd, R. M., ... & Wise, M. S. (2016). Recommended amount of sleep for pediatric populations: a consensus statement of the American Academy of Sleep Medicine. *Journal of Clinical Sleep Medicine, 12*(6), 785-786.

Perlman, F., & Bobak, M. (2008). Socioeconomic and Behavioral Determinants of Mortality in Posttransition Russia: A Prospective Population Study. *Annals of Epidemiology, 18*(2), 92–100.

Prati, A., & Saucet, C. (2024). The causal effect of a health treatment on beliefs, stated preferences and memories. *Journal of Health Economics, 94*, 102864.

RAND State Statistics (2021). Land & Water Area. https://randstatestats.org/us/stats/land---water-area.html

Rehm, J., & Ferreira-Borges, C. (2018). Risk factor policies, morbidity, and mortality in Russia. *The Lancet, 392*(10153), 1094-1095.

Rieker, P. P., & Bird, C. E. (2005). Rethinking Gender Differences in Health: Why We Need to Integrate Social and Biological Perspectives. *The Journals of Gerontology: Series B, 60*(Special_Issue_2), S40–S47.

Romm, A. T. (2015). The effect of retirement date expectations on pre-retirement wealth accumulation: The role of gender and bargaining power in married US households. *Journal of Family and Economic Issues, 36*, 593-605.

Roenneberg, T. (2013). The human sleep project. *Nature, 498*(7455), 427-428.

Roenneberg, T., & Merrow, M. (2007, January). Entrainment of the human circadian clock. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 72, pp. 293-299). Cold Spring Harbor Laboratory Press.

Ross, C. E., Masters, R. K., & Hummer, R. A. (2012). Education and the Gender Gaps in Health and Mortality. *Demography, 49*(4), 1157–1183.

Russia Longitudinal Monitoring survey, RLMS-HSE (2019). Conducted by National Research University Higher School of Economics and OOO Demoscope together with Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences. (RLMS-HSE websites: https://rlms-hse.cpc.unc.edu, https://www.hse.ru/org/hse/rlms).

Ryan, C.L., & Bauman, K. (2016). Eduational Attainment in the United States: 2015. *Current Population Reports*. https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf

Schünemann, J., Strulik, H., & Trimborn, T. (2017). The gender gap in mortality: How much is explained by behavior? *Journal of Health Economics, 54*, 79–90.

Short, M. A., Booth, S. A., Omar, O., Ostlundh, L., & Arora, T. (2020). The relationship between sleep duration and mood in adolescents: A systematic review and meta-analysis. *Sleep Medicine Reviews, 52*, 101311.

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.

Trias-Llimós, S., & Janssen, F. (2018). Alcohol and gender gaps in life expectancy in eight Central and Eastern European countries. *European Journal of Public Health, 28*(4), 687–692.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297-323.

United States Census Bureau (2021). Centers of Population, 2010 Census. https://www.census.gov/geographies/reference-files/time-series/geo/centers-population.2010.html#list-tab-1319302393

United States Census Bureau (2023). Population and Housing Unit Estimates Datasets. https://www.census.gov/programs-surveys/popest/data/data-sets.html

Van Dongen, H. P., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep, 26*(2), 117-126.

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

Walker, M. (2017). *Why we sleep: The new science of sleep and dreams*. Penguin UK.

Wanders, L., Bakker, E.A., van Hout, H.P., Eijsvogels, T.M., Hopman, M.T., Visser, L.N., Wouters, H., & Thijssen, D.H. (2021). Association between sedentary time and cognitive function: A focus on different domains of sedentary behavior. *Preventive Medicine, 153*, 106731.

Weiner, J. S., & Tishkovskaya, S. (2018). Improving Bayes: Effects of multiple sources and format on Bayesian reasoning. *The Quarterly Journal of Experimental Psychology, 71*(9), 1917-1934.

Werneck, A. O., Hoare, E., Stubbs, B., van Sluijs, E. M., & Corder, K. (2021a). Associations between mentally-passive and mentally-active sedentary behaviours during adolescence and psychological distress during adulthood. *Preventive Medicine, 145*, 106436.

Werneck, A. O., Hoare, E., Stubbs, B., van Sluijs, E. M., & Corder, K. (2021b). Association of mentally-active and mentally-passive sedentary behaviour with depressive symptoms among adolescents. *Journal of Affective Disorders, 294*, 143-150.

Wheaton, A. G., Chapman, D. P., & Croft, J. B. (2016). School start times, sleep, behavioral, health, and academic outcomes: a review of the literature. *Journal of School Health, 86*(5), 363-381.

Wheaton, A. G., Jones, S. E., Cooper, A. C., & Croft, J. B. (2018). Short sleep duration among middle school and high school students—United States, 2015. *Morbidity and Mortality Weekly Report, 67*(3), 85.

Wood, A. M., Kaptoge, S., Butterworth, A. S., Willeit, P., Warnakula, S., Bolton, T., Paige, E., Paul, D.S., Sweeting, M., Burgess, S., Bell, S., Astle, W., Stevens, D., Koulman, A., Selmer, R.M., Verschuren, W.M.M, Sato, S., Njølstad, I., Woodward, M., ... & Danesh, J. (2018). Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *The Lancet, 391*(10129), 1513-1523.

Woodruff (2020). Age-specific retirement effects of the ACA exchanges. *Department of Economics, University of Colorado Boulder, Working Paper No. 20-07.* https://www.colorado.edu/economics/sites/default/files/attached-files/20-07_-_woodruff.pdf

World Bank (2023). Life expectancy at birth, total (years). https://data.worldbank.org/indicator/SP.DYN.LE00.IN

World Health Organization. (2017). New survey shows over 21% decline in tobacco use among Russian adults. https://www.euro.who.int/en/countries/russian-federation/news/news/2017/10/new-survey-shows-over-21-decline-in-tobacco-use-among-russian-adults

World Health Organization. (2019). Alcohol-related deaths drop in Russian Federation due to strict alcohol control measures, new report says. https://www.euro.who.int/en/health-topics/disease-prevention/alcohol-use/news/news/2019/10/alcohol-related-deaths-drop-in-russian-federation-due-to-strict-alcohol-control-measures,-new-report-says

World Health Organization. (2023). Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol). https://www.who.int/data/gho/data/indicators/indicator-details/GHO/alcohol-recorded-per-capita-(15-)-consumption-(in-litres-of-pure-alcohol)

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

Yun, M.-S. (2004). Decomposing differences in the first moment. *Economics Letters, 82*(2), 275–280.

Yun, M.-S. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry, 43*(4), 766–772.

Zaridze, D., Lewington, S., Boroda, A., Scélo, G., Karpov, R., Lazarev, A., Konobeevskaya, I., Igitov, V., Terechova, T., Boffetta, P., Sherliker, P., Kong, X., Whitlock, G., Boreham, J., Brennan, P., & Peto, R. (2014). Alcohol and mortality in Russia: prospective observational study of 151 000 adults. *The Lancet, 383*(9927), 1465-1473.

Zimmerman, F. (2020). The dynamics of motivated beliefs. *American Economic Review, 110*(2), 337-363.

Data acknowledgments

Chapters 2 and 3 use data from the Health and Retirement Study (HRS) which is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The analysis uses Early Release data (RAND HRS Longitudinal File 2020). These data have not been cleaned and may contain errors that will be corrected in the Final Public Release version of the dataset.

Chapter 4 uses data from the Russia Longitudinal Monitoring Survey (RLMS-HSE) which is conducted by the National Research University Higher School of Economics and OOO Demoscope together with the Carolina Population Center, University of North Carolina at Chapel Hill, and the Institute of Sociology of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences.

Chapter 5 uses data from the Panel Study of Income Dynamics (PSID). The collection of data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and R01 AG040213, and the National Science Foundation under award numbers SES 1157698 and 1623684.

Summary

This thesis is a collection of empirical studies on health-related expectations and behaviors. These shape outcomes over the life cycle, influencing not only health but also education, retirement and well-being in old age.

Chapter 2 shows that older Americans hold inaccurate expectations of long-term care risk. Accuracy is measured by comparing the subjective probabilities of moving to a nursing home with realizations of that event. Decomposing our measure of accuracy reveals that individuals severely underestimate the association between risk factors that insurers observe and future nursing home use. This is especially pronounced for the least cognitively able. This may lead them not to take insurance that they would have purchased had their risk perceptions been accurate.

Chapter 3 measures the accuracy of American workers' subjective probabilities of continuing to work full-time past the ages of 62 and 65. There is substantial inaccuracy, which mainly derives from uncertainty surrounding work status at ages 62 and 65. The expectations are also noisy, which could indicate attention paid to irrelevant factors when forming and reporting expectations. The least educated hold the least accurate retirement expectations and may therefore be the most vulnerable to adverse impacts on lifetime welfare resulting from poor retirement planning.

Chapter 4 decomposes the large male-female gap in mortality in Russia using health behaviors. It finds that gender differences in health behaviors – particularly smoking and, to a lesser extent, drinking – explain a large share of the gender mortality gap. This contribution, as well as the gender gap itself, declined over time as tobacco and alcohol control policies were introduced. These findings, while not causal, are consistent with men's reduced tobacco and alcohol consumption explaining most of the decrease in the gender mortality gap.

Chapter 5 uses variation in sunset time across the United States to estimate effects of long-term exposure to sunset time on adolescents' sleep duration, mental health and educational attainment. Persistent exposure to later sunset times during adolescence reduces sleep duration and increases the likelihood of being diagnosed with depression by young adulthood. It has no discernable impact on high school graduation or college attendance, but is found to reduce the likelihood of graduating from college. These findings suggest that chronic sleep deprivation during adolescence can have a lasting impact on human capital formation.

Overall, this thesis contributes to two mostly separate strands of the economics literature but with important connections that merit more attention – expectations and health behaviors. Future research on the relation between expectations and health behaviors may be helpful to inform policies aiming at preventing suboptimal behavior, increasing well-being over the life cycle and reducing inequalities in it.

Samenvatting

Deze thesis bevat empirische studies over gezondheidsgerelateerde verwachtingen en gedrag. Deze beïnvloeden uitkomsten gedurende de levenscyclus, zoals gezondheid, pensioen en welzijn op latere leeftijd.

Hoofdstuk 2 toont aan dat oudere Amerikanen inaccurate verwachtingen hebben over het risico op langdurige zorg. Inaccuraatheid wordt gemeten door de subjectieve kansen van het verhuizen naar een verpleegtehuis te vergelijken met de daadwerkelijke uitkomsten. Het ontleden van inaccuraatheid onthult dat individuen de correlatie tussen risicofactoren die verzekeraars observeren en een toekomstig verpleegtehuisbezoek sterk onderschatten. Dit geldt vooral voor mensen met een lage cognitie. Dit kan ertoe leiden dat mensen geen verzekering afsluiten die zij wel zouden hebben genomen als hun risicopercepties accuraat waren geweest.

Hoofdstuk 3 meet de accuraatheid van de subjectieve kansen van Amerikaanse werkenden om voltijd te blijven werken na 62 en 65 jaar. Er is aanzienlijke inaccuraatheid, voornamelijk door onzekerheid over werkstatus op die leeftijden. De verwachtingen bevatten ook veel ruis, wat kan aangeven dat er veel aandacht wordt besteed aan irrelevante factoren bij het vormen en rapporteren van de verwachtingen. De laagst opgeleide mensen hebben de meest inaccurate pensioenverwachtingen en hebben daardoor mogelijk de hoogste kans op het slecht plannen voor hun pensioen met nadelige gevolgen voor later welzijn.

Hoofdstuk 4 ontleedt de grote man-vrouw kloof in sterfte in Rusland met behulp van gezondheidsgedrag. Het blijkt dat genderverschillen in gezondheidsgedrag, met name roken en in mindere mate alcohol drinken, een groot deel van de gendersterftekloof kunnen verklaren. Deze bijdrage, evenals de genderkloof zelf, is in de loop van de tijd, terwijl beleidsmaateregelen voor tabak- en alcoholvermindering werden ingevoerd, afgenomen. Hoewel niet causaal, zijn deze resultaten wel consistent met een afname van de gendersterftekloof door een verminderde consumptie van tabak en alcohol door mannen.

Hoofdstuk 5 gebruikt variatie in zonsondergangstijden voor de Verenigde Staten om de effecten hiervan op de slaapduur van adolescenten, hun mentale gezondheid en opleidingsniveau te schatten. Consistent een latere zonsondergang meemaken gedurende adolescentie vermindert de slaapduur en vergroot de kans op een depressiediagnosis voor vroege volwassenheid. Het heeft geen waarneembaar effect op het afronden van de middelbare school of het volgen van hoger onderwijs, maar het blijkt wel de kans op het afstuderen aan de universiteit te verkleinen. Dit suggereert dat een chronisch slaaptekort tijdens de adolescentie een blijvend effect kan hebben op de vorming van menselijk kapitaal.

Deze thesis draagt bij aan twee grotendeels gescheiden gebieden van de economische literatuur – verwachtingen en gezondheidsgedrag – maar met belangrijke verbanden die meer aandacht verdienen. Toekomstig onderzoek naar de relatie tussen verwachtingen en gezondheidsgedrag kan helpen bij het informeren van beleid gericht op het voorkomen van suboptimaal gedrag, het bevorderen van welzijn over de levenscyclus en het verminderen van ongelijkheden daarin.

About the author

Lisa Voois was born on January 27, 1995, in Dordrecht, the Netherlands. In 2013, she began her bachelor's studies at Erasmus University Rotterdam, initially in Economics, but later, in 2015, also in Philosophy. In 2017, she obtained her master's degree in Health Economics, also from Erasmus University Rotterdam.

In 2018, after completing her Philosophy degree, she began her PhD in Health Economics at the Erasmus School of Economics and Tinbergen Institute under the supervision of Teresa Bago d'Uva and Owen O'Donnell. Her research interests lie at the intersection of health economics and applied microeconometrics. Her PhD research focuses on the consequences of health behaviors and the accuracy of health-related subjective expectations. Her latest research interests include the economics of sleep and mental health.

In September 2024, she will start as an Assistant Professor at the Erasmus School of Health Policy and Management, where she will continue her research.

Portfolio

Research papers

Long-term care risk perceptions, information, and insurance. *Working paper*, with Teresa Bago d'Uva and Owen O'Donnell.

Ready to retire? Accuracy of retirement expectations. *Working paper*, with Teresa Bago d'Uva and Owen O'Donnell.

Does changing health behavior explain the falling gender gap in mortality in Russia? *Working paper*, with Teresa Bago d'Uva.

When the sun goes down: Effects of sunset time on adolescent sleep, mental health and education. *Working paper.*

Education

PhD in Economics
Erasmus School of Economics and Tinbergen Institute 2018-2024

TI Research Qualification
Tinbergen Institute 2018-2020

MSc in Health Economics
Erasmus School of Economics 2016-2017

BA in Philosophy
Erasmus School of Philosophy 2015-2018

BSc in Economics
Erasmus School of Economics 2013-2016

Teaching

Tutorial instructor, Statistics (pre-master course), Erasmus School of Health Policy and Management, 2023-2024

Tutorial instructor, Regulation of Healthcare Markets (bachelor course), Erasmus School of Health Policy and Management, 2023-2024

Thesis supervision (bachelor and master), Erasmus School of Economics and Erasmus School of Health Policy and Management, 2020-2021, 2022-2023, 2023-2024

Teaching assistant, Research Project (bachelor course), Erasmus School of Economics, 2020-2021, 2021-2022, 2022-2023

Teaching Assistant, Applied Microeconometrics (master course), Erasmus School of Economics, 2017-2018, 2019-2020

Tutorial instructor, Microeconomics (bachelor course), Erasmus School of Economics, 2016-2017, 2017-2018

Tutorial instructor, Macroeconomics (bachelor course), Erasmus School of Economics, 2015-2016, 2016-2017, 2017-2018

Tutorial instructor, Introduction to Econometrics (bachelor course), Erasmus School of Economics, 2015-2016, 2016-2017, 2017-2018

Conferences

2024 Essen Health Conference (Essen, Germany), European Society for Population Economics (ESPE) conference (Rotterdam, the Netherlands), European Health Economics Association (EuHEA) conference (Vienna, Austria), Economics of Mental Health Workshop (Minneapolis, U.S.)

2023 Lowlands Health Economics Study Group (lolaHESG) conference (Egmond aan Zee, the Netherlands)

2022 Lowlands Health Economics Study Group (lolaHESG) conference (Maastricht, the Netherlands), American Society of Health Economists (ASHEcon) conference (Austin, U.S.), European Health Economics Association (EuHEA) conference (Oslo, Norway), Netspar International Pension Workshop (online)

2021 Russia Longitudinal Monitoring Survey of Higher School of Economics (RLMS-HSE) User conference (online), International Health Economics Association (iHEA) congress (online), European Health Economics Association (EuHEA) PhD conference (online), Tinbergen PhD Jamboree (online)

2020 Health Economics Rotterdam PhD Symposium (online)

2019 American Society of Health Economists (ASHEcon) conference (on scholarschip, no presentation, Washington D.C., U.S.)

Other activities

Refereed for Journal of Health Economics, Health Economics

Organized the yearly Health Economics Rotterdam PhD Symposiums and the monthly Health Economics Rotterdam PhD meetings

Tinbergen publication list

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. For a full list of PhD theses that appeared in the series we refer to List of PhD Theses – Tinbergen.nl. The following books recently appeared in the Tinbergen Institute Research Series:

799 Y. XIAO, *Fertility, parental investments and intergenerational mobility*
800 X. YU, *Decision Making under Different Circumstances: Uncertainty, Urgency, and Health Threat*
801 G. ANTONECCHIA, *Productivity and Strategies of Multiproduct Firms*
802 H. KWEON, *Biological Perspective of Socioeconomic Inequality*
803 D.K. DIMITROV, *Three Essays on the Optimal Allocation of Risk with Illiquidity, Intergenerational Sharing and Systemic Institutions*
804 J.B. BLOOMFIELD, *Essays on Early Childhood Interventions*
805 S. YU, *Trading and Clearing in Fast-Paced Markets*
806 M.G. GREGORI, *Advanced Measurement and Sampling for Marketing Research*
807 O.C. SOONS, *The Past, Present, and Future of the Euro Area*
808 D. GARCES URZAINQUI *The Distribution of Development. Essays on Economic Mobility, Inequality and Social Change*
809 A.C. PEKER, *Guess What I Think: Essays on the Wisdom in Meta-predictions*
810 A. AKDENIZ, *On the Origins of Human Sociality*
811 K. BRÜTT, *Strategic Interaction and Social Information: Essays in Behavioural Economics*
812 P.N. KUSUMAWARDHANI, *Learning Trends and Supply-side Education Policies in Indonesia*
813 F. CAPOZZA, *Essays on the Behavioral Economics of Social Inequalities*
814 D.A. MUSLIMOVA, *Complementarities in Human Capital Production: The Role of Gene-Environment Interactions*
815 J.A. DE JONG, *Coordination in Market and Bank Run Experiments*
816 Y. KIM, *Micro studies of macroprudential policies using loan-level data*
817 S.R. TER MEULEN, *Grade retention, ability tracking, and selection in education*
818 A.G.B. ZIEGLER, *The Strategic Role of Information in Markets and Games: Essays in Behavioral Economics*

819 I. VAN DER WERVE, *Panel data model for socioeconomic studies in crime and education*
820 Y. GU, *Roads, Institutions and the Primary Sector in West Africa*
821 Y. LI, *Share Repurchases in the US: An extensive study on the data, drivers, and consequences*
822 R. DIAS PEREIRA, *What Makes us Unique? Genetic and Environmental Drivers of Health and Education Inequalities*
823 H.P. LETTERIE, *Essays on the regulation of long-term care in the Netherlands*
824 D.D. PACE, *Essays on the cognitive foundations of human behavior and on the behavioral economics of climate change*
825 J.N. VAN BRUMMELEN, *On the estimation of parameters in observation-driven time series models*
826 Z. CSÁFORDI, *Essays on Industry Relatedness*
827 B. VAN OS, *On Dynamic Models: Optimization-Based Methods and Practical Applications*
828 D.T. Ó CEALLAIGH, *Self-control Failures and Physical Inactivity: Measuring, Understanding and Intervening*
829 S.B. DONOVAN, *Ties that bind and fray: Agglomeration economies and location choice*
830 A. SOEBHAG, *Essays in Empirical Asset Pricing*
831 H. YUAN, *Essays in Behavioral Economics*
832 A. LENGYEL, *Essays on Government Bond Markets and Macroeconomic Stabilization*
833 S. KÜTÜK, *Essays on Risk Creation in the Banking Sector*
834 E. VLADIMIROV, *Essays on the Econometrics of Option Pricing*
835 R.E.K. PRUDON, *From the onset of illness to potential recovery. Empirical economic analysis of health, disability and work*
836 K. MOUSSA, *Signal Extraction by the Extremum Monte Carlo Method*
837 D. FAVOINO, *The Adaptation of Firms to Institutional Change*
838 B. WACHE, *Information Frictions in Financial Flows*
839 A. FEHÉR, *Essays in Law and Economics*
840 Q. WIERSMA, *Dynamic Models for Multi-Dimensional Time Series*
841 R. SILVESTRINI, *On the Importance of Firm Heterogeneity, Business Dynamism, and Market Power Dynamics in the Macroeconomy*
842 E.S.R. DIJK, *Innovative Start-Ups and Competition Policy – How to Reign in Big*
843 T.D. SCHENK, *Essays in Causal Inference with Panel Data*
844 S. TYROS, *Workers' Skills and (green) Technology Adoption*
845 C.J. GRASER, *Mechanisms for the Evolution of Prosociality*

846 K. IOANNIDIS, *On the role of information in strategic and individual decision making*

847 G.M. Miyazato Szini, *Advances in Panel and Network Econometrics*

This thesis is a collection of empirical studies on health-related expectations and behaviors. Part 1 examines expectations of important later-life events – retirement and nursing home admission. These expectations may influence saving and insurance decisions, with consequences for well-being in old age and at other points in the life cycle. Part 2 investigates health behaviors, including smoking, drinking and sleeping. These behaviors influence future health and may also impact later economic outcomes, such as education and retirement. Overall, this thesis contributes to two mostly separate strands of the economics literature, but shows important connections that merit more attention.

tinbergen institute

Tinbergen Institute

Erasmus University Rotterdam
Erasmus School of Economics

W: www.eur.nl/ese
www.tinbergen.nl