Dissertation submitted for the degree of
PhD in Economic and Social Sciences
at the Vienna University of Economics and Business

Health measures and healthcare utilisation in ageing populations: Demographic and economic perspectives

by Sonja Spitzer

March 2020

Primary supervisor:
Prof. Dr. Wolfgang Lutz
Department of Demography
Faculty of Social Sciences
University of Vienna, Austria

Secondary supervisor:
Prof. Dr. Alexia Fürnkranz-Prskawetz
Institute of Statistics and Mathematical Methods in Economics
Vienna University of Technology, Austria

Abstract

Population ageing substantially reshapes the demand for healthcare. It is important that efforts to adapt healthcare systems and to enhance the health of older adults be based on reliable information about health and healthcare behaviour. This thesis answers three research questions concerned with health measures and healthcare utilisation from economic and demographic viewpoints. It investigates (i) the reliability of popular health indicators against the backdrop of survey errors, (ii) the accuracy of perceived and reported health by individuals, and (iii) the effects of individual health perception on healthcare behaviour. Analyses are based on demographic and econometric methods employing data from the Survey of Health, Ageing and Retirement in Europe for the population 50+ in up to 19 European countries. Results highlight the importance of socio-economic heterogeneities in survey participation and health perception for the reliability of health measures on a micro and macro level as well as for healthcare utilisation. These findings are of utmost importance when successfully preparing healthcare systems for future demographic change.

Extended summary

Population ageing is largely attributable to improvements in social and economic development. Although a success story, it does pose substantial challenges for social and policy institutions. The increasing number of older adults requires intergenerational agreements to be reconsidered and has far-reaching implications for the labour market as well as for public pension and healthcare systems. Reliable information about population health and the drivers of healthcare utilisation is thus essential to facilitate evidence-based policies that respond to the demands caused by demographic change.

First and foremost, monitoring population health and adapting healthcare systems require accurate measures of health that capture all relevant population groups. Data describing population health, however, are frequently based on surveys and thus subject to bias. Most prominently, distortions in survey data are caused by non-observation and measurement errors. This dissertation provides an extensive investigation into the reliability of measures of health against the backdrop of such survey errors from a macro and micro perspective. Moreover, the thesis contributes to the literature by illustrating the use of appropriate survey methods to adjust for bias. While previous work has considered individual domains concerned with the limitations of survey data in terms of representation and measurement, this thesis provides a more comprehensive understanding of the relationship between distortions in survey data and popular health indicators in the context of ageing societies. As an additional key contribution, this thesis links measurement errors to healthcare utilisation. It explores individual health perception biases as a potential driver of healthcare utilisation—a relation that was previously ignored in the literature and has important implications for public spending.

More concretely, the thesis answers three research questions from economic and demographic viewpoints. It investigates (i) the reliability of popular health indicators against the backdrop of survey errors, (ii) the accuracy of perceived and reported health by individuals and (iii) the effects of individual health perception on healthcare behaviour. The

research questions are answered employing demographic as well as econometric methods based on cross-sectional and longitudinal data from the Survey of Health, Ageing and Retirement in Europe for the 50+ population in up to 19 European countries.

The first dissertation publication considers non-observation errors by investigating bias in health expectancies due to educational differences in survey representation. To that end, calibrated weights that consider the education structure in the general population according to censuses are applied to measures of activity limitations from survey data. Findings show that health expectancies are substantially biased because low-educated individuals in most European countries are underrepresented in surveys. The publication also demonstrates how the flawed education structure in survey data can be adjusted for, which is especially important for indicators such as health expectancies that have high political impact.

The second publication analyses bias in health measures due to individual health misperception and thus covers measurement errors. It explores which demographic characteristics bias self-reported physical and cognitive health status of older Europeans. Matching performance measures and their self-reported equivalents allows individuals that over- and underestimate their health to be differentiated. Results based on relative importance analysis show that differences in reporting behaviour due to cultural background, age and education result in a large bias in self-assessed health, while gender plays a minor role. These findings are crucial given that self-assessed data are often the only information available to researchers and policymakers when asking health-related questions.

The third publication of the thesis investigates health misperception as a potential driver of doctor visits and concomitant out-of-pocket expenditure. It shows that individuals who overestimate their health visit the doctor less often and have lower out-of-pocket spending than individuals who correctly assess their health, which is crucial for preventive care such as screenings. In contrast, individuals who underestimate their health have higher healthcare utilisation. Projections suggest that increased doctor visits due to underestimation of health will cost the average European country 81 million international dollars per year by 2060. Given the persistent inequalities in healthcare access among socio-economic groups, these results serve as a potential starting point for a reconciliation between increasing healthcare for those in need while considering resource limitations.

In summary, the thesis provides important insights both for scholars engaged with health-related survey data and for health authorities concerned with the sustainability of healthcare systems. It highlights the importance of socio-economic heterogeneities in survey participation and health perception for the reliability of health measures and for healthcare utilisation, which are of the utmost importance in preparing effective and efficient healthcare systems for future demographic change.

Kurzfassung

Das Altern der Bevölkerung stellt soziale und politische Institutionen – und nicht zuletzt auch das Gesundheitssystem – vor neue Herausforderungen. Daher ist es wichtig, dass der Gesundheitszustand älterer Menschen anhand verlässlicher Daten erfasst wird und Gesundheitssysteme, basierend auf ebenso verlässlichen Informationen, umgestaltet werden. Diese Dissertation beantwortet drei Forschungsfragen zu den Themen Gesundheitsindikatoren und Gesundheitsleistungen aus volkswirtschaftlicher sowie demographischer Perspektive. Sie untersucht (i) die Verlässlichkeit weitverbreiteter Gesundheitsindikatoren vor dem Hintergrund von Stichproben- und Messfehlern, (ii) die Genauigkeit des selbst eingeschätzten Gesundheitszustandes und (iii) den Einfluss des subjektiven Gesundheitszustandes auf die Inanspruchnahme von Gesundheitsleistungen. Die Analysen stützen sich auf Daten der SHARE-Erhebung über Gesundheit, Älterwerden und das Leben im Ruhestand in Europa für die Bevölkerungsgruppe 50 plus in bis zu 19 europäischen Ländern. Die Ergebnisse zeigen auf der Mikro- sowie Makroebene, wie wichtig sozioökonomische Unterschiede in der Einschätzung der eigenen Gesundheit sowie im Teilnahmeverhalten bei Umfragen sind; einerseits für die Verlässlichkeit von Gesundheitsindikatoren und andererseits dafür, wie häufig Gesundheitsleistungen in Anspruch genommen werden. Die daraus resultierenden Schlussfolgerungen sind von außerordentlicher Wichtigkeit, da sie wegweisend dafür sind, wie das Gesundheitswesen auf den demographischen Wandel vorbereitet werden kann.

Contents

1 Introduction

1.1 Motivation

Europeans are living longer than ever before. Since 1960 life expectancy has increased by an average of two years each decade (Eurostat, 2019). Women born in Europe in 2020 can expect to live to the age of 83 and men to 76. By 2060 life expectancy at birth is projected to be 91 and 85 years, respectively (Wittgenstein Centre for Demography and Global Human Capital, 2018). This substantial increase in longevity, along with persistent low fertility and sometimes reinforcing migration patterns (Christensen et al., 2009), has increased the median age from 35 in 1990 to 43 in 2020 and will increase it further to 48 in 2060 (Wittgenstein Centre for Demography and Global Human Capital, 2018). The United Nations considers current population ageing to be unprecedented in the history of humanity and profound in its consequences and implications for human life. The development is likely to endure and will gain in speed in future years (Eurostat, 2019; United Nations, 2007).

Increased longevity is a success story largely attributed to the reduction in child mortality, advances in medical technology and improvement in lifestyle, living conditions and labour conditions (Eurostat, 2019). It does, however, pose substantial challenges for social and policy institutions (Christensen et al., 2009; Kluge et al., 2019; Lee & Mason, 2019; Lutz et al., 2003). Between 2018 and 2050, the number of Europeans aged 75–84 will increase by 64%, and the number of those aged 85 and above will more than double. In contrast, the under-55 population will drop by 9.6% (Eurostat, 2019). Although in the future older individuals are likely to be healthier than they are today (Christensen et al., 2009; Sanderson & Scherbov, 2008, 2010, 2013), these demographic changes will still affect economic growth and labour supply and place public finances under substantial pressure (Kluge et al., 2019; Prskawetz et al., 2008; Prskawetz & Sambt, 2014). This new landscape will not only require a reconsideration of intergenerational agreements within families and in the population at large but will also have far-reaching implications for the labour market, specifically regarding the inclusion of older employees in the workforce and retirement age regulations.

Moreover, population ageing has an enormous impact on the healthcare system. The growing number of older adults increases the demand on health and long-term care because the risk of disease and disability grows with age. Most European countries provide substantial public healthcare coverage where three-quarters of health spending is financed through compulsory schemes (OECD & European Commission, 2018), which is why the increasing number of older people poses a particular challenge for the sustainability of public healthcare systems. Between 2008 and 2017, health spending in the European Union increased from 8.8% of GDP to 9.6% of GDP (European Commission, 2018) and it will further grow by 0.9 percentage points until 2070 (OECD & European Commission, 2018). The public healthcare system will come under additional pressure due to the reducing number of contributors and potential informal caregivers. The working-age population (aged 15–64) in the European Union is projected to decrease from 333 million in 2016 to 292 million in 2070. In 2016 there were 3.3 working-age individuals for every person over 65; by 2070 this number will reduce to 2.0 (European Commission, 2018).

To meet the fiscal challenges posed by ageing societies, authorities regularly call for greater efficiency in healthcare provision and a reduction in wasteful healthcare spending (OECD, 2019a; OECD & European Commission, 2018). Furthermore, it is frequently suggested that the employability of older workers should be promoted and their skills development encouraged and also that their retirement ages should be adjusted in line with their increasing life expectancy (OECD, 2019a). Both efforts, however, have strong distributional implications. Inequalities in healthcare access among income groups are persistent in the European Union and OECD countries. For example, individuals with higher incomes visit the doctor more often, particularly specialists, and take part in preventive screenings more regularly (Devaux, 2015; Devaux & de Looper, 2012; OECD, 2019b) than their lower-income counterparts. What is more, low-income groups, along with women and migrants, more frequently report unmet healthcare needs (Baeten et al., 2018). With respect to pension reforms, it is frequently shown that increasing the retirement age has the potential to undermine the progressivity of pension systems because life expectancy is lower for lower socio-economic groups (Sánchez-Romero et al., 2019; Sánchez-Romero & Prskawetz, 2017). Hence, any efficiency efforts need to be balanced against distributional concerns. Preparing public and private healthcare services for future demographic change while ensuring high quality universal health care for all population groups is thus one of the biggest challenges of our time.

1.2 Problem definition

When approaching challenging reforms related to the healthcare system, it is vital to make decisions based on reliable information about population health and healthcare utilisation. Accurate measures of health are necessary information for public and private healthcare providers when planning, adapting and monitoring present and future health coverage and care services with an eye to healthcare expenditure (Thacker et al., 2006). In the context of pension systems and the aspiration to raise the retirement age, it is also important to know how many additional life years are going to be healthy life years that allow individuals to remain in the labour force and also how these healthy life years will vary by socio-economic characteristics.

Whenever feasible, "allocation of public health resources should be based [...] on objective assessments of health status, burden of disease, injury, and disability, their preventability, and related costs" (Thacker et al., 2006, p.14). In reality, however, information on population health is frequently based on self-reports from survey data, as this is often the only information available when researchers and policy makers ask health-related questions concerning large cross-national populations. Furthermore, other data sources such as administrative data do not usually provide the rich information needed to capture the multidimensional aspects of health as defined by the World Health Organization (2020), where "health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity". Surveys are a flexible tool for capturing a rich set of health and related dimensions while being relatively cost- and time-efficient. A major problem with survey-based statistics, however, is that they are subject to bias unless two conditions are met (Groves et al., 2004):

(i) Representation: The subset of individuals participating in the survey must have similar characteristics to those of the larger population that the survey seeks to describe. If not, statistics based on the survey suffer from non-observation error.

(ii) Measurement: The answers given by the survey participants must accurately describe the characteristics of the respondents. If not, the statistics are subject to errors of observation, also called measurement errors.

Previous studies have emphasised that surveys frequently do not meet these conditions. For example, survey participation is often selective and differs by individual characteristics such as gender, age and socio-economic status (Boshuizen et al., 2006; Cheung et al., 2017; Demarest et al., 2013; Korkeila et al., 2001; Reinikainen et al., 2018; Sousa-Poza & Henneberger, 2000; Tolonen et al., 2015). In studies on health, bias occurs because (i) individuals with low socio-economic status are less likely to participate in surveys and (ii) individuals with low socio-economic status are also more likely to suffer from poor health; thus, the variable of interest differs by the same characteristics that cause differences in the survey representation. If inferences about the general population are made based on unweighted statistics from such flawed data, the study is subject to error. A related concern is the more general misrepresentation challenges, such as the gender data gap which has recently received significant attention. The gender data gap describes the phenomenon of data frequently not being disaggregated by gender and thus tending to ignore heterogeneities between men and women. Furthermore, conclusions relevant for both genders are often based on information about men only, which has substantial implications for women's health (Buvinic et al., 2014; Buvinic & Levine, 2016; Temin & Roca, 2016; The World Bank, 2020).

It is also repeatedly reported in the literature that health surveys are affected by measurement errors. For example, individuals frequently over- or underestimate their own health status (Bago d'Uva et al., 2008; Beaudoin & Desrichard, 2011; Coman & Richardson, 2006; Furnham, 2001; Jürges, 2007). Additionally, health perception differs by sociodemographic characteristics such as age (Crossley & Kennedy, 2001; Srisurapanont et al., 2017), gender (Merrill et al., 1997; Schneider et al., 2012), country of residence (Capistrant et al., 2014; Jürges, 2007), education (Bago d'Uva et al., 2008; Choi & Cawley, 2017), and ethnicity (Jackson et al., 2017). Against this backdrop, it is essential to better understand the reliability of the self-reported health measures frequently used to evaluate population ageing.

Health misperception does not only affect the accuracy of health measures but might also influence individual health behaviour. Previous literature shows that people who overestimate their health and abilities are more likely to have a fall (Sakurai et al., 2013) or have traffic accidents (Preston & Harris, 1965). Evidence on the effect of health perception on health behaviour is scarce, in particular, no study has considered health misperception as a potential driver of healthcare utilisation. Enhancing the efficiency of healthcare provision is often mentioned as an important tool in managing the costs of population ageing (OECD, 2019a). Hence, understanding the link between health misperception and healthcare utilisation is an important aspect to consider when future healthcare provision is being planned.

1.3 Research objectives and questions

This thesis fills research gaps within two important topics related to health in ageing populations, namely, (i) health measures and (ii) health behaviour, in particular, healthcare utilisation. The first contribution is a comprehensive investigation of the reliability of health measures against a backdrop of representation and measurement errors in survey data. The thesis also explores the potential drivers of measurement errors and determines if these drivers further affect the healthcare behaviour of older individuals—a link that has not been studied directly to date.

The dissertation starts by exploring the effect of non-observation errors on the reliability of health measures that are frequently used to analyse healthy ageing in Europe. More concretely, the first part of the thesis considers the heterogeneities in survey representation by individual characteristics and investigates the effect of non-response-bias (one of the largest concerns for correct representation in survey data (Groves et al., 2004)) on health expectancy (one of the most commonly used health indicators in Europe (Bogaert et al., 2018; European Commission, 2010; Jagger, 2015; Jagger et al., 2013; Jagger et al., 2011; Robine & Cambois, 2013)). Despite the immense interest in health expectancy among scholars and health authorities, previous work has not addressed whether the indicator is distorted by educational differences in survey participation. Thus, the first publication of the thesis answers the following research question:

RQ 1: *How are health measures biased by heterogeneities in survey representation?*

Moreover, the first part of the dissertation contributes to the literature by illustrating how bias can be adjusted for using appropriate survey methods. Following this, the thesis investigates the second condition defined by Groves et al. (2004) and analyses the impact of observation errors on health indicators that are based on self-reported information. The second publication explores biased beliefs about health status as a potential source of measurement error. Specifically, it investigates how health misperception—that is the over- or underestimation of one's health—affects health measures. Hence, the second publication of the thesis answers the following research question:

RQ 2: *How accurate is self-assessed health status and how are health measures biased by individual health misperception?*

To that end, self-assessed measures of physical and cognitive health are matched with their performance-tested equivalents, a method that has previously been used only for small-scale studies. The thesis contributes to the literature by applying this approach to a large cross-country dataset that allows country comparisons of health perception biases for different health dimensions simultaneously. Moreover, the dissertation employs a novel approach that enables the bias in self-reported health status to be decomposed into its contributing determinants.

Biased beliefs not only affect the reliability of survey statistics but are also a strong predictor of behaviour in several life domains. They have substantial implications in areas such as labour market decisions and outcomes, education (Reuben et al., 2017) as well as savings and investment choices (Anderson et al., 2017; Malmendier & Tate, 2005). Beliefs are particularly relevant for health as they can directly affect the risk of accidents (Preston & Harris, 1965) and consequently can have serious long-lasting effects on health and mortality. It is hypothesised that individuals' (mis)perception of their own health can also alter health-seeking behaviour and their subsequent utilisation of healthcare services, such as timely screenings, immunisations, annual health checks and doctor visits. Furthermore, prior research has already shown that non-monetary barriers to healthcare, such as long waiting times and geographic distance, along with attitudes towards and knowledge of health and healthcare, can explain the underuse of preventive care (Carrieri & Bilger, 2013). Health misperception, especially overestimating health, could provide an additional explanation for the persistent underuse of preventive care in Europe. No previous work has analysed the effect of biased beliefs concerning health on healthcare utilisation; thus, the final research question covered in the thesis is:

RQ 3: *How does individual health misperception affect healthcare utilisation?*

It is important to note that in the context of the aforementioned research questions, it is almost impossible for survey users to distinguish measurement errors from processing errors, the latter referring to, for example, error introduced after the survey by the staff responsible for coding answers. Furthermore, it is often impossible to differentiate between misreporting and misperception; thus, this thesis does not differentiate between measurement errors and processing errors or between health misreporting and health misperception. For similar reasons, coverage and sampling errors are not the focus of this thesis.

Topic	Research question	Publication	Relevant dimensions
Health measures	**RQ 1:** How are health measures biased by heterogeneities in survey representation?	1^{st}	• Country • Gender • Age • Socio-economic characteristics
Health measures; Healthcare utilisation	**RQ 2:** How accurate is self-assessed health status and how are health measures biased by individual health misperception?	2^{nd}	
Healthcare utilisation	**RQ 3:** How does individual health misperception affect healthcare utilisation?	3^{rd}	

Table 1: Thesis structure

1.4 Thesis structure

The dissertation is structured into three parts according to the research questions. Table 1 illustrates this structure and indicates the topics covered by each part. Each of the three research question is answered in a separate, self-contained publication that establishes an important aspect of the overall contribution. Relevant related work, limitations and future work are discussed in each article respectively. All three publications consider heterogeneities among the most important demographic groups and thus include separate analyses by country, gender and age. Furthermore, socio-economic characteristics such as educational attainment are considered whenever relevant throughout the dissertation.

References

Anderson, A., Baker, F. & Robinson, D. T. (2017). Precautionary savings, retirement planning and misperceptions of financial literacy. *Journal of Financial Economics*, *126*(2), 383–398.

Baeten, R., Spasova, S., Vanhercke, B. & Coster, S. (2018). *Inequalities in access to healthcare. A study of national policies* (tech. rep.). European Commission.

Bago d'Uva, T., O'Donnell, O. & Van Doorslaer, E. (2008). Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology*, *37*(6), 1375–1383.

Beaudoin, M. & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*(2), 211–241.

Bogaert, P., Van Oyen, H., Beluche, I., Cambois, E. & Robine, J. M. (2018). The use of the global activity limitation indicator and healthy life years by member states and the European Commission. *Archives of Public Health*, *76*(1), 1–7.

Boshuizen, H. C., Viet, A. L., Picavet, H. S., Botterweck, A. & van Loon, A. J. (2006). Non-response in a survey of cardiovascular risk factors in the Dutch population: Determinants and resulting biases. *Public Health*, *120*, 297–308.

Buvinic, M., Furst-Nichols, R. & Koolwal, G. (2014). *Mapping gender data gaps* (tech. rep.). Data2x: UN Foundation.

Buvinic, M. & Levine, R. (2016). Closing the gender data gap. *Significance*, *13*(2), 34–37.

Capistrant, B. D., Glymour, M. M. & Berkman, L. F. (2014). Assessing mobility difficulties for cross-national comparisons: Results from the WHO study on AGEing and Adult Health. *Journal of the American Geriatrics Society*, *62*(2), 329–335.

Carrieri, V. & Bilger, M. (2013). Preventive care: Underused even when free. Is there something else at work? *Applied Economics*, *45*(2), 239–253.

Cheung, K. L., Ten Klooster, P. M., Smit, C., De Vries, H. & Pieterse, M. E. (2017). The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health*, *17*.

Choi, A. & Cawley, J. (2017). Health disparities across education: The role of differential reporting error. *Health Economics*, *27*(3), e1–e29.

Christensen, K., Doblhammer, G., Rau, R. & Vaupel, J. W. (2009). Ageing populations: the challenges ahead. *Lancet*, *374*(9696), 1196–1208.

Coman, L. & Richardson, J. (2006). Relationship between self-report and performance measures of function: a systematic review. *Canadian journal on aging*, *25*(3), 253–270.

Crossley, T. F. & Kennedy, S. (2001). The reliability of self-assessed health status. *Journal of Health Economics*, *21*, 643–658.

Demarest, S., Van Der Heyden, J., Charafeddine, R., Tafforeau, J., Van Oyen, H. & Van Hal, G. (2013). Socio-economic differences in participation of households in a Belgian national health survey. *European Journal of Public Health*, *23*(6), 981–985.

Devaux, M. (2015). Income-related inequalities and inequities in health care services utilisation in 18 selected OECD countries. *The European Journal of Health Economics*, *16*(1), 21–33.

Devaux, M. & de Looper, M. (2012). Income-related inequalities in health service utilisation in 19 OECD countries , 2008-2009. *OECD Health Working Papers*, (58).

European Commission. (2010). Turning Europe into a true Innovation Union. Retrieved August 14, 2018, from http://europa.eu/rapid/press-release_MEMO-10-473_en.htm

European Commission. (2018). *The 2018 ageing report: Economic & budgetary projections for the 28 EU member states (2016-2070)* (tech. rep.).

Eurostat. (2019). *Ageing Europe: Looking at the lives of older people in the EU*. Luxembourg, Publications Office of the European Union.

Furnham, A. (2001). Self-estimates of intelligence: culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, *31*, 1381–1405.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2004). *Survey Methodology* (R. M. Groves, G. Kalton, J. N. K. Rao, N. Schwarz & C. Skinner, Eds.). Hoboken, New Jersey, John Wiley & Sons, Inc.

Jackson, J. D., Rentz, D. M., Aghjayan, S. L., Buckley, R. F., Meneide, T. F., Sperling, R. A. & Amariglio, R. E. (2017). Subjective cognitive concerns are associated with objective memory performance in Caucasian but not African-American persons. *Age and Ageing*, *46*(6), 988–993.

Jagger, C. (2015). Trends in life expectancy and healthy life expectancy. *Government Office for Science*.

Jagger, C., McKee, M., Christensen, K., Lagiewka, K., Nusselder, W., Van Oyen, H., Cambois, E., Jeune, B. & Robine, J. M. (2013). Mind the gap - Reaching the European target of a 2-year increase in healthy life years in the next decade. *European Journal of Public Health*, *23*(5), 829–833.

Jagger, C., Weston, C., Cambois, E., Van Oyen, H., Nusselder, W., Doblhammer, G., Rychtarikova, J. & Robine, J. M. (2011). Inequalities in health expectancies at older ages in the European Union: Findings from the Survey of Health and Retirement in Europe (SHARE). *Journal of Epidemiology and Community Health*, *65*(11), 1030–1035.

Jürges, H. (2007). True health vs response style: exploring cross-country differences in self-reported health. *Health economics*, *16*, 163–178.

Kluge, F. A., Goldstein, J. R. & Vogt, T. C. (2019). Transfers in an aging European Union. *Journal of the Economics of Ageing*, *13*, 45–54.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H. & Koskenvuo, M. (2001). Non-response and related factors in a nation-wide health survey. *European Journal of Epidemiology*, *17*(11), 991–999.

Lee, R. & Mason, A. (2019). Analyzing population ageing and intergenerational redistribution: NTA and AGENTA. *Vienna Yearbook of Population Research*, *17*, 1–5.

Lutz, W., O'Neill, B. C. & Scherbov, S. (2003). Europe's population at a turning point. *Science*, *299*, 1991–1992.

Malmendier, U. & Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*, *60*(6), 2661–2700.

Merrill, S. S., Seeman, T. E., Kasl, S. V. & Berkman, L. F. (1997). Gender differences in the comparison of self-reported disability and performance measures. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *52*(1), 19–26.

OECD. (2019a). Fiscal challenges and inclusive growth in ageing societies. Retrieved December 11, 2019, from https://www.oecd.org/economy/ageing-inclusive-growth/

OECD. (2019b). *Health for everyone? Social inequalities in health and health systems*. Paris, OECD Publishing.

OECD & European Commission. (2018). *Health at a glance: Europe 2018: State of health in the EU cycle*. Paris, OECD Publishing.

Preston, C. E. & Harris, S. (1965). Psychology of drivers in traffic accidents. *Journal of Applied Psychology*, *49*(4), 284–288.

Prskawetz, A., Bloom, D. E. & Lutz, W. (2008). Introduction. In A. Prskawetz, D. E. Bloom & W. Lutz (Eds.), *Population aging, human capital accumulation, and productivity growth*. New York, Population Council.

Prskawetz, A. & Sambt, J. (2014). Economic support ratios and the demographic dividend in Europe. *Demographic Research*, *30*, 963–1010.

Reinikainen, J., Tolonen, H., Borodulin, K., Härkänen, T., Jousilahti, P., Karvanen, J., Koskinen, S., Kuulasmaa, K., Männistö, S., Rissanen, H. & Vartiainen, E. (2018). Participation rates by educational levels have diverged during 25 years in Finnish health examination surveys. *European Journal of Public Health*, *28*(2), 237–243.

Reuben, E., Wiswall, M. & Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *Economic Journal*, *127*(604), 2153–2186.

Robine, J.-M. & Cambois, E. (2013). Healthy life expectancy in Europe. *Population & Societies*, (499).

Sakurai, R., Fujiwara, Y., Ishihara, M., Higuchi, T., Uchida, H. & Imanaka, K. (2013). Age-related self-overestimation of step-over ability in healthy older adults and its relationship to fall risk. *BMC Geriatrics*, *13*(44), 1–9.

Sánchez-Romero, M., Lee, R. D. & Prskawetz, A. (2019). Redistributive effects of different pension systems when longevity varies by socioeconomic status. *NBER Working Paper Series*, *25944*.

Sánchez-Romero, M. & Prskawetz, A. (2017). Redistributive effects of the US pension system among individuals with different life expectancy. *The Journal of the Economics of Ageing*, *10*, 51–74.

Sanderson, W. C. & Scherbov, S. (2008). Rethinking age and aging. *Population Bulletin*, *63*(4), 3–16.

Sanderson, W. C. & Scherbov, S. (2010). Remeasuring aging. *Science*, *329*(5997), 1287–1288.

Sanderson, W. C. & Scherbov, S. (2013). The characteristics approach to the measurement of population aging. *Population and Development Review*, *39*(4), 673–685.

Schneider, U., Pfarr, C., Schneider, B. S. & Ulrich, V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *European Journal of Health Economics*, *13*(3), 251–265.

Sousa-Poza, A. & Henneberger, F. (2000). Wage data collected by telephone interviews: an empirical analysis of the item nonresponse problem and its implications for the estimation of wage functions. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, *136*(1), 79–98.

Srisurapanont, M., Suttajit, S., Eurviriyanukul, K. & Varnado, P. (2017). Discrepancy between objective and subjective cognition in adults with major depressive disorder. *Scientific Reports*, *7*(1), 1–7.

Temin, M. & Roca, E. (2016). Filling the gender data gap. *Studies in Family Planning*, *47*(3), 264–269.

Thacker, S. B., Stroup, D. F., Carande-Kulis, V., Marks, J. S., Kakoli, R. & Gerberding, J. L. (2006). Measuring the public's health. *Public health reports*, *121*, 14–22.

The World Bank. (2020). Filling data gaps. Retrieved February 29, 2020, from http://datatopics.worldbank.org/gender/fillingdatagaps

Tolonen, H., Ahonen, S., Jentoft, K., Susieand Kuulasmaa & Heldal, J. (2015). Differences in participation rates and lessons learned about recruitment of participants - The

European Health Examination Survey Pilot Project. *Scandinavian Journal of Public Health*, *43*(2), 212–219.

United Nations. (2007). *World population ageing 2007* (tech. rep.). Department of Economic and Social Affairs, Population Division.

Wittgenstein Centre for Demography and Global Human Capital. (2018). Wittgenstein Centre Data Explorer Version 2.0 (Beta). Retrieved February 22, 2020, from http://www.wittgensteincentre.org/dataexplorer

World Health Organization. (2020). Constitution. Retrieved February 15, 2020, from https://www.who.int/about/who-we-are/constitution

2 Data

The research questions posed in this thesis are mainly answered based on the Survey of Health, Ageing and Retirement in Europe (SHARE). This is a representative cross-country panel study of non-institutionalised individuals aged 50 and older as well as their younger spouses (Börsch-Supan, 2019a, 2019b, 2019c, 2019d, 2019e; Börsch-Supan et al., 2013; Börsch-Supan et al., 2008; Malter & Börsch-Supan, 2013, 2015, 2017).[1] SHARE provides rich information on health, socio-economic background, employment and social networks based on around 380,000 interviews from about 140,000 individuals. The survey was launched in 2004/2005 in 11 European countries and grew considerably to 27 European countries and Israel in its latest wave in 2017. SHARE is particularly well suited to conducting research that compares countries in Europe, as the data are ex ante harmonised. It also focuses on older individuals, which makes it an ideal data source for this thesis.

SHARE is based on probability samples with close to full target population coverage for all countries; however, details regarding the sample design, specifically, the sampling frame, vary by country (for an overview, see Bergmann et al. (2017), De Luca and Rossetti (2018), Lynn et al. (2013)). To be included in SHARE, individuals must regularly live in one of the survey countries and speak its language(s). Respondents are usually surveyed in their homes by interviewers using computer-assisted personal interviews. For details on response rates, consult Bergmann et al. (2017). Table 2 provides an overview on the survey waves, country datasets and variables utilised in each publication of the dissertation.

[1]This dissertation uses data from SHARE Waves 1, 2, 4, 5, and 6 (DOIs: 10.6103/SHARE.w1.700, 10.6103/SHARE.w2.700, 10.6103/SHARE.w4.700, 10.6103/SHARE.w5.700, 10.6103/SHARE.w6.700). The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982) and Horizon 2020 (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged(see www.share-project.org).

Paper	Waves	Countries	Relevant variables
1st	4 (2010–2012)	AT, BE, CZ, DK, EE, FR, DE, HU, IT, PL, PT, SI, ES	Country, gender, age, educational attainment, activity limitations, hand grip strength
2nd	2 (2006 & 2007) 4 (2010–2012) 5 (2013)	AT, BE, CZ, DK, EE, FR, DE, GR, HU, IE, IT, LU, NL, PL, PT, SI, ES, SE, CH	Country, gender, age, educational attainment, wave, self-reported and tested ability to stand up from a chair, self-reported and tested memory (immediate and delayed), frailty, indicators for employment, parenthood and partnership status
3rd	1 (2004 & 2005) 2 (2006 & 2007) 4 (2010-2012) 5 (2013) 6 (2015)	AT, BE, CZ, DK, EE, FR, DE, IT, LU, NL, PL, SI, ES, SE, CH	Country, gender, age, educational attainment, wave, self-reported and tested ability to stand up from a chair, self-reported and tested memory, self-reported and tested walking ability, doctor visits, out-of-pocket health expenditure, chronic diseases, activity limitations, Alzheimer's disease, indicators for retirement and partnership status, household income, household wealth, health access, supplementary health insurance, household size

Table 2: Overview of relevant SHARE data

For the first publication on health measures and survey representation, SHARE data are complemented with data from two additional sources. First, the European Population and Housing Censuses from 2011 are used (Eurostat, 2018). The data are retrieved from Eurostat, which, along with the National Statistical Institutes, has combined national censuses from 2011 for 32 European countries and has the data structured in a comparable manner. The second additional data source is the life tables provided by Eurostat for 2011 (Eurostat, 2011).

In the third publication on health perception and healthcare utilisation, the total public cost of health misperception is estimated. To that end, predicted costs per outpatient visit from the World Health Organization (2011) is utilised along with population predictions from the Wittgenstein Centre for Demography and Global Human Capital (2018).

References

Bergmann, M., Kneip, T., Luca, G. D. & Scherpenzeel, A. (2017). *Survey participation in the Survey of Health , Ageing and Retirement in Europe (SHARE), Wave 1-6.Based on Release 6.0.0 (March 2017).*

Börsch-Supan, A. (2019a). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 7.0.0. SHARE-ERIC. Data set. 10.6103/SHARE.w1.700

Börsch-Supan, A. (2019b). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 7.0.0. SHARE-ERIC. Data set. 10.6103/SHARE.w2.700

Börsch-Supan, A. (2019c). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 7.0.0. SHARE-ERIC. Data set. 10.6103/SHARE.w4.700

Börsch-Supan, A. (2019d). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 7.0.0. SHARE-ERIC. Data set. 10.6103/SHARE.w5.700

Börsch-Supan, A. (2019e). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 7.0.0. SHARE-ERIC. Data set. 10.6103/SHARE.w6.700

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S. & Zuber, S. (2013). Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001.

Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A., Mackenbach, J., Siegrist, J. & Weber, G. (Eds.). (2008). *First results from the Survey of Health, Ageing and Retirement in Europe (2004-2007). Starting the longitudinal dimension*. Mannheim, Mannheim Research Institute for the Economics of Aging (MEA).

De Luca, G. & Rossetti, C. (2018). Computing calibrated weights. *SHARE Manual*.

Eurostat. (2011). Life table. Retrieved April 25, 2019, from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_mlifetable&lang=en

Eurostat. (2018). 2011 Census Hub. Retrieved June 27, 2018, from https://ec.europa.eu/eurostat/web/population-and-housing-census/census-data/2011-census

Lynn, P., De Luca, G. & Ganninger, M. (2013). Sample Design in SHARE Wave Four. In F. Malter & A. Börsch-Supan (Eds.), *Share wave 4: Innovations & methodology* (pp. 74–123).

Malter, F. & Börsch-Supan, A. (Eds.). (2013). *SHARE Wave 4: Innovations and Methodology*. Munich, Max Planck Institute for Social Law; Social Policy.

Malter, F. & Börsch-Supan, A. (Eds.). (2015). *SHARE Wave 5: Innovations & Methodology*. Munich, MEA, Max Planck Institute for Social Law; Social Policy.

Malter, F. & Börsch-Supan, A. (Eds.). (2017). *SHARE Wave 6: Panel innovations and collecting Dried Blood Spots*. Munich, Max Planck Institute for Social Law; Social Policy.

Wittgenstein Centre for Demography and Global Human Capital. (2018). Wittgenstein Centre Data Explorer Version 2.0 (Beta). Retrieved February 22, 2020, from http://www.wittgensteincentre.org/dataexplorer

World Health Organization. (2011). Country-specific inpatient and outpatient estimates in 2010 currency. Retrieved February 10, 2020, from https://www.who.int/choice/cost-effectiveness/inputs/country_inpatient_outpatient_2010.pdf?ua=1

3 Dissertation publications

3.1 Health measures and survey representation (1st Publication)

The first publication of the dissertation answers RQ 1: "*How are health measures biased by heterogeneities in survey representation?*". It was published on January 27th 2020 as

Spitzer, S. (2020). Biases in health expectancies due to educational differences in survey participation of older Europeans: It's worth weighting for. *The European Journal of Health Economic*. https://doi.org/10.1007/s10198-019-01152-0

Abstract: Health expectancies are widely used by policymakers and scholars to analyse the number of years a person can expect to live in good health. Their calculation requires life tables in combination with prevalence rates of good or bad health from survey data. The structure of typical survey data, however, rarely resembles the education distribution in the general population. Specifically, low-educated individuals are frequently under-represented in surveys, which is crucial given the strong positive correlation between educational attainment and good health. This is the first study to evaluate if and how health expectancies for 13 European countries are biased by educational differences in survey participation. To this end, calibrated weights that consider the education structure in the 2011 censuses are applied to measures of activity limitation in the Survey of Health, Ageing and Retirement in Europe. The results show that health expectancies at age 50 are substantially biased by an average of 0.3 years when the education distribution in the general population is ignored. For most countries, health expectancies are overestimated; yet remarkably, the measure underestimates health for many Central and Eastern European countries by up to 0.9 years. These findings highlight the need to adjust for distortion in health expectancies, especially when the measure serves as a base for health-related policy targets or policy changes.

The European Journal of Health Economics
https://doi.org/10.1007/s10198-019-01152-0

ORIGINAL PAPER

Biases in health expectancies due to educational differences in survey participation of older Europeans: It's worth weighting for

Sonja Spitzer[1]

Received: 9 August 2019 / Accepted: 17 December 2019

Abstract
Health expectancies are widely used by policymakers and scholars to analyse the number of years a person can expect to live in good health. Their calculation requires life tables in combination with prevalence rates of good or bad health from survey data. The structure of typical survey data, however, rarely resembles the education distribution in the general population. Specifically, low-educated individuals are frequently underrepresented in surveys, which is crucial given the strong positive correlation between educational attainment and good health. This is the first study to evaluate if and how health expectancies for 13 European countries are biased by educational differences in survey participation. To this end, calibrated weights that consider the education structure in the 2011 censuses are applied to measures of activity limitation in the Survey of Health, Ageing and Retirement in Europe. The results show that health expectancies at age 50 are substantially biased by an average of 0.3 years when the education distribution in the general population is ignored. For most countries, health expectancies are overestimated; yet remarkably, the measure underestimates health for many Central and Eastern European countries by up to 0.9 years. These findings highlight the need to adjust for distortion in health expectancies, especially when the measure serves as a base for health-related policy targets or policy changes.

Keywords Activity limitations · Education and inequality · Health expectancies · Survey participation · Iterative proportional fitting (IPF) · Survey of Health, Ageing and Retirement in Europe (SHARE)

JEL Classification C83 · I18 · I31 · J14

Introduction

Life expectancy continues to increase in Europe. We live longer, but do we live healthier? Answering this question is of utmost importance in the presence of demographic change. How long and how healthy we live is necessary information for public and private healthcare providers to plan health coverage and care services. Furthermore, policymakers are interested in the employability of older generations when adapting pension systems, in particular, when adjusting the retirement age. Whether we spend our additional life years in good or bad health is frequently analysed via health expectancy (HEX), an indicator that captures the number of years a person can expect to live in good health. This concept was developed half a century ago [1, 2] and has garnered increasing attention from both scholars and policymakers. For example, the European Commission aims to add 2 years of healthy life for the average European between 2010/2011 and 2020 to improve the sustainability of the European social and healthcare systems [3, 4]. Furthermore, many European governments use HEX to set health-related targets and make policy changes based on this measure [5].

HEX usually combines information on mortality with prevalence rates of good or bad health from survey data; therefore, it captures both the quantity and quality of additional life years. A key problem with this approach, however, is that survey participation is often selective and differs by individual characteristics such as gender, age and socio-economic status. A common deviation is that highly educated individuals are more likely to participate in surveys than low-educated individuals, leading to an overrepresentation of the

✉ Sonja Spitzer
sonja.spitzer@iiasa.ac.at

1 Wittgenstein Centre for Demography and Global Human Capital (Univ. Vienna, IIASA, VID/ÖAW), International Institute for Applied Systems Analysis (IIASA), Schloßplatz 1, 2361 Laxenburg, Austria

Published online: 27 January 2020

highly educated among the respondents [6–8]. This mismatch is crucial given the strong positive correlation between educational attainment and good health [9–12]. Overrepresenting healthy, well-educated individuals in surveys makes countries appear to have healthier populations than is actually the case.

The aim of this study is to explore if and how HEX differs when the education structure in the general population is considered. For this purpose, prevalence rates of bad health from the Survey of Health, Ageing and Retirement in Europe (SHARE) for 13 European countries are adjusted with calibrated weights based on auxiliary information from censuses. Although there has been vast research on HEX, to the best of my knowledge, no previous work has addressed whether biases in the education composition distort the measure. Given the widespread use of HEX among scholars and health authorities, knowing the reliability of the indicator in the context of flawed survey data is pivotal. Moreover, this study contributes to the literature by illustrating how bias can be adjusted for when auxiliary information on the true population structure is available.

Background

Educational attainment affects health

The positive correlation between educational attainment and good health is well established [9]. For example, the average life expectancy at birth of well-educated Europeans is 7 years higher than that of low-educated individuals [13]. Furthermore, low-educated persons report higher activity limitations [14] and higher levels of bodily pain [12]. This can be partially explained with economic rationales, such as the positive link between education and income or correlations between education and occupational choice [11]. Additionally, differences in health behaviour are potential drivers of the education gradient in health. On one hand, low-educated persons are more likely to smoke, drink heavily, and be obese than highly educated persons. On the other hand, they are less likely to use preventive care, drive safely, and live in safe houses [15]. While the positive relationship between socio-economic advantages and health is found throughout Europe, the magnitude of that correlation varies by gender and country. First, the education gradient is larger for men than for women in life expectancy [16] as well as in HEX [17]. Second, in Central and Eastern European (CEE) countries, highly educated individuals are much healthier than low-educated individuals; whereas the difference is small in, for example, Denmark [18]. While most social health inequalities among older Europeans are driven by current socio-economic conditions, childhood circumstances also add to the health differences between socio-economic groups [19].

Educational attainment affects survey participation

Educational attainment is associated not only with health but also with survey participation. Low-educated persons are frequently underrepresented in health surveys, for example, in Belgium [7, 20], Denmark [21], and Finland, where the gap in survey participation between low- and well-educated individuals has substantially widened over time [6]. This violation of the “missing at random” assumption can be attributed to coverage errors, sampling errors, and non-response errors [22]. Coverage errors stem from the mismatch between the survey’s target population and its sampling frame, for example, when phone registers serve as sampling frames, although low-educated persons are less likely to own phones than the highly educated. Sampling errors denote the gap between sampling frame and the sample, which emerges because not all individuals in the sampling frame can be surveyed due to time and money constraints. To account for the unequal selection probabilities of sample units, surveys frequently provide sampling weights. Finally, non-response errors stem from differences between the invited sample and the actual respondents.

The strong association between non-response and low education [23] can be explained by three channels [22]. First, low-educated persons are harder to contact due to their socio-demographic and social–environmental attributes. For example, they might have unstable life paths and are consequently more likely to change their address. Second, participation in surveys is usually voluntary and low-educated persons are more likely to refuse to participate than the highly educated. Finally, low-educated individuals may be less likely to provide the requested survey data for reasons such as being too sick to participate or because they are less aware of certain domains such as their health or financial situation.

Education is not the only characteristic corresponding with lower response rates. Gender and age also impact survey participation, which is why these variables are commonly considered in survey weights. Furthermore, characteristics such as race [24] and relationship status [8] are associated with non-response. This study, however, only focuses on education-related biases. First and foremost, education is a common proxy for socio-economic status that is rather stable over lifetime with relatively low measurement error. Furthermore, the education gradient in response behaviour is well established. Finally, register or census data on the education structure in the general population are more readily available than auxiliary information on other socio-economic characteristics, making it more possible to compare the education distribution in the general population to that in the survey data.

Educational differences in survey participation bias the prevalence of good and bad health

In summary, highly educated individuals are, on average, healthier than low-educated individuals and are more likely to participate in surveys. Thus, both the variable of interest (health) and the likelihood to participate in a survey are influenced by educational attainment. When inferences about the health of the general population are made based on unweighted prevalence rates from such flawed surveys, the general population appears healthier than what is true in reality. For example, Van Der Heyden et al. [20] found that the prevalence of people with diabetes and asthma in Belgium is underestimated when the actual education distribution in the general population is not considered. In the Netherlands, education-related non-response leads to negative biases in the prevalence of low self-assessed health, smoking, alcohol intake, and low physical activity [25].

Prevalence of good or bad health is needed to calculate HEX

Prevalence rates of good or bad health are one of the main components needed when calculating HEX, which makes the education-related bias in survey data a major concern. Similar to life expectancy, HEX varies substantially among European countries and is particularly low in CEE countries [26]. Around 2010, HEX at birth was 70.1 years for Swedish men but only 52.6 years for Slovakian men. For women, HEX at birth ranged from 71.5 years in Malta to 52.7 years in Slovakia [27]. Overall, women live a larger proportion of their life disabled than men [28, 29].

While life expectancy has clearly increased throughout Europe, evidence on HEX is less conclusive. The outcome depends on the health dimension that is considered [30] as well as the survey utilised [31]. Analysing 25 European countries between 2005 and 2010, [30] show that years in poor general self-rated health at age 65 decreased by 0.5 (1.1) years for men (women). By contrast, years with chronic morbidity increased at the same time and years without activity limitations remained stable. Analysing the latter separately for different countries, Jagger [32] found that HEX increased in some countries but decreased in others. In addition to differences in health measures, surveys, subpopulations and the relationship between mortality and morbidity, the lack of a consistent time trend in HEX might be partly explained by the small number of observations in the surveys utilised. Analysing prevalence by country, gender, and age requires sufficient numbers of observations in each country–gender–age cell. This is often not the case, especially at older ages. Consequently, prevalence rates based on these small cells are often not reliable and have large confidence intervals: the small cell sizes make it difficult to separate the signal from the noise.

Regardless of the evidence on the inadequate representation of the low-educated persons in surveys, studies typically do not adjust for prevalence rates of HEX. One explanation for this might be that auxiliary information on the actual education distribution in the general population is not readily available. Register data are only accessible for some European countries and censuses are only conducted with long time intervals. Yet whenever available, auxiliary data on the actual education distribution in the general population can be utilised to calibrate weights so that they account for deviations between the true distribution and the survey distribution.

Data

The following sections describe analyses of whether adjusting for the education structure in the general population changes the prevalence of bad health and consequently the HEX for European countries. The analyses rely on three different data sources. Auxiliary information that is expected to capture the actual education distribution in the general population is taken from Eurostat's Census database, which provides Population and Housing Censuses for Europe. These census data are used to generate calibrated weights via iterative proportional fitting (IPF). In addition, life tables from Eurostat [33] along with prevalence of bad health from SHARE are taken to compute HEX with Sullivan's method [2, 34]. Analyses and comparisons of HEX in Europe are frequently based on SHARE [26, 35, 36] as well as on the European Statistics on Income and Living Conditions (EU-SILC) and on the European Health Interview Survey (EHIS). This analysis utilises SHARE, because its sampling and weighting procedure is well documented, thus enabling an exact replication of the calibration approach employed [37, 38].

The Survey of Health, Ageing and Retirement in Europe (SHARE)

Prevalence rates of bad health are extracted from the fourth wave of SHARE, which was mainly conducted in 2011, and consequently corresponds with the census data [39–42]. Although some interviews took place in 2010 and 2012, 94% of all observations stem from 2011. In total, 16 European countries participated in the fourth wave; however, 3 of these countries do not provide reliable census data via Eurostat for the requested year (see "Eurostat data for post-stratification weights and life tables"). Therefore, the analysis is restricted to 13 countries including Austria, Belgium, Czechia, Denmark, Estonia, France, Germany, Hungary, Italy, Poland, Portugal, Slovenia, and Spain.

The target population of SHARE consists of all non-institutionalised individuals aged 50 and older who regularly live in the respective survey country and speak its language(s). Spouses of target individuals are included in the data regardless of their age; however, for this study, all individuals younger than 50 years old are excluded [42–44]. The remaining number of respondents lies between 1615 in Germany and 6754 in Estonia. Some countries only provide small numbers of observations per gender–age–education cell, especially at higher ages. Respondent numbers for Germany, Poland, and Portugal are particularly small: all three countries provide less than 2000 observations. Germany and Poland also have small respondent numbers at ages 50–54, because their panel was not refreshed since Wave 2 in 2007. Details on the number of respondents for each country are summarised in Appendix 1.1. All numbers for SHARE refer to the final set of respondents used for the calculations in this paper.

The survey is based on probability samples with close to full target population coverage for all countries, yet details regarding the sample design, in particular the sampling frame, vary by country (for an overview, see [38, 43, 44]). Respondents were surveyed in their homes by interviewers using computer-assisted personal interviews. For details on response rates, consult [44].

For the calibration of weights, information on the proportions of respondents by country, gender, age, and educational attainment is required. Educational attainment is split into three groups in accordance with the International Standard Classification of Education [45]. The "low-educated" group includes individuals whose educational attainment is lower secondary education and less. The "medium-educated" group includes individuals with upper secondary or post-secondary non-tertiary education. The "high-educated" group includes all individuals with higher than post-secondary non-tertiary education. A fourth category was added to capture all individuals with missing values in their education variable (2.2%). The education categories are directly comparable to the categories in the census data. By construction, country information has no missing values in SHARE. The gender variable also has no missing values. Age information is available for all observations save four individuals in Czechia, who are subsequently excluded. To calculate proportions in SHARE for IPF, age is grouped into 10-year age groups with 90 + serving as an open-ended category. Details regarding the survey proportions by country, gender, age, and education are presented in Appendix 1.1.

HEX in Europe is most commonly calculated based on the Global Activity Limitation Indicator (GALI) [5, 27, 46, 47], making the health measure the obvious choice for this analysis. Moreover, evaluations show that GALI similarly measures function and disability across European countries [48, 49], allowing cross-country comparisons. In particular, GALI is based on the reply to the following survey question: "For the past 6 months at least, to what extent have you been limited because of a health problem in activities people usually do?" The question is answered by each survey participant based on three categories: "severely limited", "limited but not severely", and "not limited". For the purpose of this study, GALI is dichotomised into a binary variable with (1) "severely limited" and (0) "not severely limited". Prevalence of bad health π is calculated by country, gender, and 5-year age group; 85 years of age serves as an open-ended category. In the final set of respondents, GALI has missing values for only 0.58% of the survey participants. Because there is no evidence of an education-related pattern in item non-response concerning GALI, this study only focuses on unit non-response.

GALI is a self-assessed health measure, and as such, is likely biased depending on the respondent's individual characteristics [50–53] and cultural background [54–57]. Low-educated survey respondents are particularly prone to misreporting their health. Some evidence suggests that low-educated individuals have the tendency to overestimate their physical health; whereas, highly educated persons tend to underestimate their physical health [57]. If that is the case, the bias in HEX that is associated with underrepresentation of low education could appear smaller than it actually is, because low-educated individuals are overstating their physical abilities. Furthermore, self-assessed measures are often upward biased at older ages [57, 58], presumably due to peer effects [59]. Thus, as a robustness analysis, the prevalence of bad health is also estimated based on grip strength, a tested measure that is expected to be less biased by systematic misreporting. Despite GALI and grip strength measuring different health domains, additional calculations based on grip strength are expected to reveal if self-reported and tested health measures are equally biased by educational differences in survey participation.

Grip strength is primarily used to measure sarcopenia, the age-related decrease in muscle mass [60]. Furthermore, it is a strong predictor of mortality [61], mobility, and cognition [62]. While GALI only captures activity limitations, grip strength is often considered a proxy for overall health. In SHARE, grip strength is ascertained twice per hand for each participant via a handheld Smedley dynamometer (for details, see Ref. [63]). In accordance with the literature, the maximum of these four measurements is used for robustness analysis [61, 63, 64]. Grip strength is measured in kilograms, yet the calculation of HEX requires a binary outcome variable. Consequently, thresholds have to be applied, dividing the participants into groups of impaired and unimpaired. The European Working Group on Sarcopenia in Older People (EWGSOP) suggests cut-off values < 20 kg for women and

< 30 kg for men to determine the onset of sarcopenia [60]. More recent evidence, however, suggests that such pragmatic thresholds do not fully capture critically weak hand grip [61]. Moreover, grip strength varies by factors such as body height and country of residence [63], implying that thresholds should be adapted accordingly. Because the purpose of this study is not to analyse grip strength as such, the pragmatic approach suggested by EWGSOP is deemed satisfactory. If the thresholds are indeed inaccurate, they would affect both the adjusted and unadjusted prevalence rates and, therefore, would not distort the results.

Eurostat data for post-stratification weights and life tables

The calibration of weights requires auxiliary information on the actual population structure. To this end, it is assumed that the auxiliary information captures the true structure in the population with respect to certain characteristics such as gender, age, and education. For this study, the European Population and Housing Censuses are utilised as auxiliary data [65]. Along with the National Statistical Institutes, Eurostat combined national censuses from 2011 for 32 European countries and structured them in a comparable manner. Sixteen of these countries overlap with the countries from SHARE Wave 4. Because the Netherlands, Sweden, and Switzerland show irregularities in the census data provided by Eurostat, these countries are not included in the current analysis, leaving a sample of 13 countries.

For each country, population totals by gender, age, and education for individuals over 50 years of age are extracted from the censuses. The totals are used as control totals when calibrating weights. Some countries have missing information on educational attainment, which is why four education categories are constructed. The education groups "low educated", "medium educated", and "high educated" are based on the same criterion as adopted in SHARE, which are described in "The Survey of Health, Ageing and Retirement in Europe (SHARE)". In addition, an education category denoted "unknown education" is created. Regarding gender and age, missing values are negligible, which is why this analysis is only based on the known population, and census cells for unknown gender and age are excluded. The census does not differentiate between institutionalised and non-institutionalised persons, which is why it is assumed that both groups are comparable. For details regarding the population proportions by country, gender, age, and education based on the censuses, consult Appendix 1.1.

In addition to prevalence rates, the calculation of HEX with Sullivan's method relies on life tables provided by Eurostat for 2011 [33]. They are prepared to resemble standard abridged period life tables by country, gender, and 5-year age group, with 85 + considered an open-ended category.

Education distribution in SHARE versus that in the censuses

By comparing the education distribution of participants in SHARE with that in the respective censuses, three country groups can be differentiated: countries for which SHARE data fit the education distribution in the population, country data in which highly educated individuals are overrepresented and low-educated individuals are underrepresented, and remarkably, country data in which this trend is reversed. Tables comparing the distributions can be found in Appendix 1.1.

The only two SHARE datasets resembling the education distribution in the population are those for Italy and Spain. The fit for Italy is close to perfect (Table 9). Spain shows slight deviations in the younger age groups, but overall achieves concordance between SHARE and the census (Table 13). Both countries have little variation in education within age groups. For example, the vast majority of the 70 + population is low educated. This pattern might explain the good fit with respect to the education distribution. However, Portugal also has little variation in education within age groups, but the education distribution in SHARE varies strongly from that in the census (Table 11). Hence, non-complex education distributions do not guarantee concordance between the education structure in surveys and the general population.

For most countries, high-educated individuals are overrepresented and low-educated individuals are underrepresented in SHARE. This pattern is in line with the literature discussed in "Background". The countries belonging to that category are Austria, Belgium, Denmark, Germany, Hungary, Portugal, and to a lesser extent France and Slovenia. The deviation is particularly strong in Denmark, where the proportions in SHARE differ from those in the census on average by 51% for men and 52% for women in the age group of 50–89 (Table 4).

Interestingly, three CEE countries show the opposite pattern. In Czechia, Estonia, and Poland, low-educated individuals are overrepresented in the survey. Deviations are minor for Estonia (Table 5) and Poland (Table 10). For Czechia, however, SHARE proportions deviate from the census by 95% for men and 38% for women on average (Table 3). While high-educated individuals are underrepresented in the Estonian and Polish data, medium-educated individuals are underrepresented in the Czech data. Overall, the findings presented in this subsection suggest a need for education-adjusted weights (EW) when making inferences based on survey data.

Method

To determine if distortions in the education distribution of survey data affect HEX, SHARE sampling design weights are adjusted via IPF so that the education structure in SHARE would match the education structure in the respective census. Following that, two sets of prevalence rates of severe activity limitations are computed. The first set π^{EW} is calculated using EW; whereas the control set π^{RW} uses standard weights without adjustment. Finally, Sullivan's method is applied to calculate HEX^{EW} with education-adjusted prevalence rates and HEX^{RW} with the unadjusted rates. Comparing the two sets of HEX reveals if and how the measure is biased by educational differences in survey participation.

Generating calibrated weights via IPF

Frequently, the proportions of certain characteristics in survey data deviate from the proportions of the same characteristics in the general population. Assuming that the distribution in the general population is known, calibrated weights can be generated for each survey respondent to account for these discrepancies. Calibrated weights are usually based on sampling design weights, which compensate for unequal selection probabilities of sample units, and in the case of SHARE, are provided with the survey data. They are defined as the inverse of the probability of being included in the sample. These design weights account for the unequal selection of sample units, but not for unit non-response [43].

A common method for calibrating sampling design weights is IPF, also known as raking. For this approach, marginal totals for each variable on which the weights are calibrated are taken from an auxiliary source that is assumed to capture the true distribution in the general population. When applying IPF, sampling design weights are iteratively modified by a multiplicative factor until convergence is achieved and the marginal totals of the adjusted weights conform to the corresponding marginal totals from the auxiliary source [66, 67]. After the adjustment, groups that were formerly underrepresented have relatively larger weights; whereas groups that were formerly overrepresented have relatively smaller weights. Importantly, the original information provided by the sampling design weights is maintained, since the weights within a group increase proportionally. The empirical strategy of this study evolves around three different sets of calibrated weights, which are discussed in more detail below.

SHARE weights (SW)

SHARE provides its own set of calibrated weights to account for differences in response behaviour. However, their weights do not consider the education structure in the general population [38]. For the remainder of this paper, these weights are referred to as SHARE weights (SW). The SW are generated based on a calibration approach by Deville and Särndal [68], which is implemented using Stata's sreweight command by [69]. Control totals for the SW stem from the Eurostat regional database. The weights are calculated separately for each country, considering NUTS 1 regions as well as eight gender–age groups, with cutoffs at 50–59 years, 60–69 years, 70–79 years, and an open-ended category of 80 + years. In some countries, finer partitions are made below age 59 [37, 38].

Replicated weights (RW)

In a first step, the SW are replicated; this second set of weights is referred to as replicated weights (RW). Using RW instead of SW ensures that differences between estimates with and without education-adjusted weights do not stem, for example, from methodological differences applied for SW and EW. The goal is for RW to be as close as possible to the SW. However, some amendments in the method are made, so that later, education could be added as an additional control total. First, control totals are used for each calibration variable separately, instead of cross-classification. For example, instead of using age–gender totals, separate totals for age and gender are applied. The rationale behind this modification in the method is that calibrated weights are generally less stable and less likely to converge when observations are thinly spread over the calibration cells [66]. Using separate totals increases the number of observations by calibration cell. As a second amendment, Stata's survwgt rake algorithm by [67] is used to generate the RW because it appears more robust than the sreweight command [70]. Third, control totals for NUTS 1 regions are not considered in this study, again, to increase the weight's stability. The control total was included for a robustness analysis but did not alter the results. Fourth, an additional age category of 80–89 years is included, making 90 + the open-ended category. Finally, the Eurostat regional database does not provide information by education, which is why the 2011 census is used for this paper instead. Although these five changes are made, prevalence rates calculated based on the SW are almost identical to those calculated based on the RW, which confirms the approach.

Education-adjusted weights (EW)

Following the replication of SW, the EW are calculated. They are identical to the RW, except that an additional control total for education is considered for the calibration. Hence, EW vary for each individual observation, depending on the individual's sampling design weight, gender, age, and educational attainment. In addition, the 2.2% of individuals with missing values for education receive a calibrated weight, since both the prevalence rates by education and the control totals include a category for "unknown education".

Weighted prevalence rates of bad health π are calculated based on RW (π^{RW}) and EW (π^{EW}). In particular, the prevalence rates for the main analysis are based on the binary GALI measures, and prevalence rates for the robustness analysis are based on dichotomised grip strength. The means are calculated separately by country, gender, and 5-year age group, which follows the most common approach to calculate HEX in Europe. Prevalence rates π^{RW} and π^{EW} based on GALI along with the confidence intervals are presented in Appendix 1.2.

Calculating HEX with Sullivan's method

HEX is computed by applying Sullivan's method [2, 34]. According to the standard life table notation (e.g. [71]), let

l_x = number of survivors at exact age x (beginning of age interval i)

L_i = number of person-years lived in age interval i

π_i = prevalence of severe activity limitations in age interval i.

Then HEX at age x is calculated separately by country and gender as follows:

$$\text{HEX}_x = \frac{1}{l_x} \sum_{i=x}^{A} (1 - \pi_i) \times L_i,$$

where the 5-year age groups range from $i = 0$ to A. More specifically, prevalence rates π_i are used to divide person-years lived according to the Eurostat life tables into years with and without severe activity limitations. Following that, HEX is calculated by dividing the number of individuals surviving to a certain age x by the total years lived healthily from age x onwards. Two sets of HEX are calculated. HEX^{EW} is based on π^{EW}, the prevalence of severe activity limitations in age interval i weighted with EW. HEX^{RW} is based on π^{RW}, the prevalence of severe activity limitations in age interval i weighted with RW. The bias in HEX due to the misrepresentation of educational groups in the survey is computed as the difference between HEX^{RW} and HEX^{EW} and denoted as ΔHEX. Confidence intervals around HEX^{RW}, HEX^{EW} and ΔHEX are approximated using the method suggested by [72].

An alternative to calculating HEX via Sullivan's method is the multistate life table method, which is sometimes said to be more accurate [73, 74]; however, Mathers and Robine [75] report that differences between the two methods are small. Furthermore, Sullivan's method is the most common approach to calculate HEX in Europe for both health authorities and scholars, which makes the results of this study comparable.

Results

Prevalence of bad health with and without adjusted weights

The differences between adjusted (π^{EW}) and unadjusted (π^{RW}) prevalence rates correspond to the deviation in education structure in SHARE from the census (see tables in Appendix 1.2). For Italy and Spain, π^{RW} and π^{EW} are rather similar. For all country datasets in which high-educated individuals are overrepresented and low-educated individuals are underrepresented, π^{RW} is smaller than π^{EW}, indicating a downward bias in mean activity limitation. This finding is in line with the evidence that education and good health are positively correlated. The size of the bias depends on the deviation between SHARE data and the census. It is minor for countries such as France, where the deviation is small: π^{RW} at age 50 is 0.095 (0.097) for men (women) and π^{EW} at age 50 is 0.105 (0.107) for men (women). Yet the bias is severe for countries such as Denmark, where the deviation is large: π^{RW} at age 50 is 0.074 (0.076) for men (women) and π^{EW} at age 50 is 0.107 (0.110) for men (women).

For the three countries in which low-educated individuals are overrepresented, π^{RW} is larger than π^{EW}, indicating an upward-bias in mean activity limitation. Consequently, these countries appear healthier once the education structure in the general population is considered. The countries concerned are Czechia, Estonia, and Poland. The shift is most pronounced for Czechia, which is in line with the finding that the Czech SHARE data are particularly distorted.

Confidence intervals of π^{EW} and π^{RW} are mostly overlapping due to the small numbers of observations in the age–gender–education cells. For example, the male age group 90+ in Germany only consists of five men, and that in Slovenia consists of four men only. In Austria, the male age group 90+ consisted of 20 men, of which 7 are low educated, 6 are medium educated, 6 are high educated, and 1 has unknown education. While the aggregated data show a clear positive link between educational attainment and good health, the direction of the relationship between education and health in these small gender–age cells is sometimes the opposite. For example, the seven low-educated men in the Austrian 90+ group reported on average better health than the six high-educated men. Due to the reversal, prevalence of bad health is slightly lower for that group, once EW are applied. Given the small number of observations in certain cells and the subsequently large confidence intervals, HEX as well as differences in HEX have to be interpreted cautiously, especially for Portugal and Germany, where confidence intervals are particularly large and no clear age gradient in severe activity limitations for men is visible.

Comparing prevalence rates based on grip strength measures with those based on GALI leads to similar findings as described above. Yet for most countries, the age gradient in bad health is steeper when measured via grip strength, so the prevalence of bad health at old age is usually higher. This finding could be explained with the evidence that participants rate their health relatively better at old age than at young age (see "The Survey of Health, Ageing and Retirement in Europe (SHARE)"). Most notably, Portuguese and German men show a clear age gradient in education when health is tested with grip strength, while no such age gradient is visible when health is measured with GALI.

Bias in HEX

Figure 1 shows how HEX at age 50 is biased because of educational differences in survey participation. The bias is given in absolute years and the countries are ranked based on the average bias in all age groups. Results for German as well as Polish men are not shown, because small numbers of observations at young ages and subsequent large confidence intervals prevent a meaningful illustration and interpretation of the difference in HEX for those countries at ages 50–54. In addition to Fig. 1, HEX^{RW} and HEX^{EW} are presented in Appendix 1.2 for all age groups, along with the respective bias in absolute years denoted as ΔHEX and the proportional bias denoted as $\Delta\%$. Confidence intervals for HEX^{RW}, HEX^{EW} and ΔHEX are also provided in Appendix 1.2.

On average, HEX at age 50 is biased by 0.3 years, yet the deviation varies substantially between countries and genders. It is larger for women (0.4 years) than for men (0.2 years), presumably due to the higher life expectancy of women in general. For most parts, the bias resembles the deviations between SHARE and the census, and consequently, the deviation between π^{RW} and π^{EW}. As a result, HEX^{RW} and HEX^{EW} are similar for Italy and Spain, since SHARE mimics the censuses in those countries. At age 50, ΔHEX for Spanish men (women) is only −0.04 (0.00) years. For Italian men (women), the bias is only −0.07 (−0.06) years. Overall, the deviations are even smaller at older ages.

By contrast, HEX at age 50 is upward-biased in countries for which high-educated persons are overrepresented in the SHARE data. This is the case for Belgium, Denmark, Austria, Germany, Hungary, France, and Slovenia. Without EW, these countries appear to have a healthier population than is actually the case. At age 50, the upward bias is largest for women in Belgium, where HEX is overestimated by 0.87 years or 3.5%. The opposite is true for Estonia, Czech Republic, and Poland, where low-educated individuals are overrepresented in the SHARE data. Consequently, these countries appear unhealthier than they actually are. At age 50, the downward bias is largest for Czech women, whose HEX is 0.85 years or 3.2% lower when the education structure in the general population is ignored. Since the bias has different magnitudes, and more importantly, different directions, it affects the country ranking of HEX. For example, Danish men aged 50 appear to have relatively high HEX without the EW (rank 4 of 13) but drop to the lower middle field (rank 7 of 13) when adjustments are made.

ΔHEX mostly decreases with age, since life expectancy decreases with age. The proportional bias $\Delta\%$, however, remains stable over all age groups or decreases only slightly

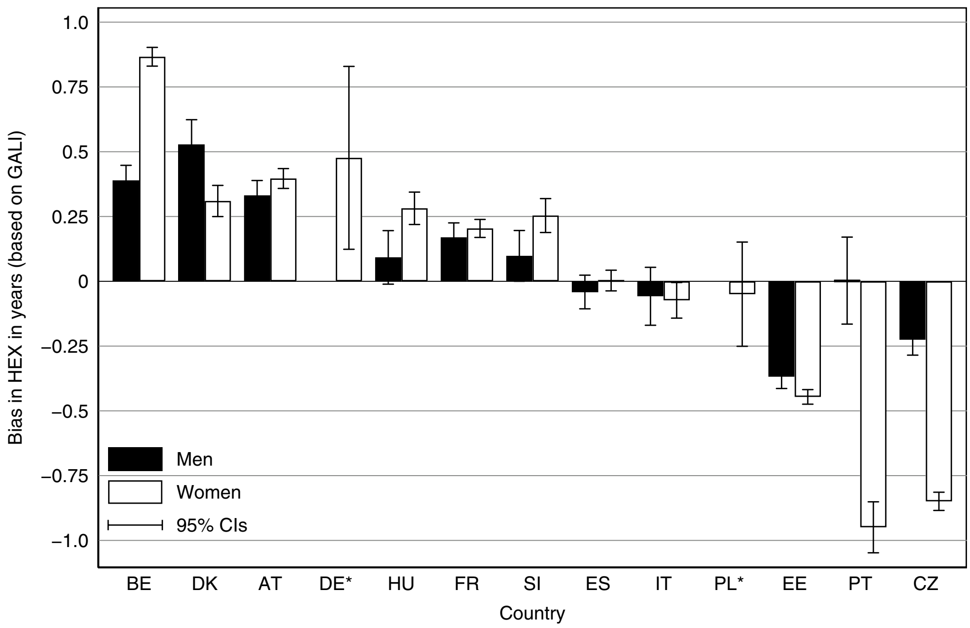


Fig. 1 Bias in HEX based on GALI at age 50 in 2011. The bias is given in absolute years, i.e. ΔHEX is calculated as the difference between HEX^{RW} and HEX^{EW}. *Results for German as well as Polish men are not shown, because small numbers of observations at ages 50–54 and subsequent large confidence intervals prevent a meaningful illustration and interpretation of the difference between HEX for those countries

for the most part. Overall, the country and gender differences described for age 50 also hold for older age groups. Due to uncertainty in the data, however, some age groups in some countries (e.g., male age group 90+ in Austria) do not show the expected sign for ΔHEX. As indicated in the previous sections, the results for Germany and Portugal have to be treated especially carefully due to the small cell sizes. HEX at age 50 for Portuguese men appears to be severely underestimated, although the data clearly show that high-educated men are overrepresented in the Portuguese SHARE data (Table 11).

As a robustness analysis, HEX based on grip strength is also provided (Fig. 2). The overall bias appears smaller when the tested indicator is applied: average ΔHEX at age 50 is reduced to 0.17 years but is still larger for women (0.23 years) than for men (0.11 years). Even though the overall level of the bias is lower when grip strength is utilised, the general findings are confirmed. The bias is still negligible for Italy and Spain. The countries showing an upward bias based on GALI also show an upward bias based on grip strength; the same holds for all countries showing downward biases. Moreover, the inconsistencies in the Portuguese data disappear once grip strength is used. HEX at age 50 for both Portuguese men and women appears to be overestimated without the EW, just as expected when comparing the Portuguese SHARE data with the census. By contrast, the results for German women suggest an unexpected downward bias of HEX, albeit with a large confidence interval, which indicates once again that results based on small numbers of respondents must be handled with care.

Discussion

This study is the first to evaluate if HEX in Europe is biased by educational differences in survey participation. The analysis showed that SHARE data for 11 of the 13 countries analysed did not resemble the education structure in the general population. In most countries, high-educated individuals were overrepresented, leading to an upward bias in HEX by up to 0.87 years, because of the positive correlation between educational attainment and good health. Contrary to what is suggested in the literature, most CEE countries analysed showed the opposite pattern that high-educated individuals were less likely to participate in surveys. As a consequence, HEX was underestimated by up to 0.85 years in those countries. These biases are crucially important, especially since HEX is frequently used by health authorities to assess population health and to make comparisons between countries. Future studies could fruitfully explore this issue further by exploring the non-response related bias in HEX for other surveys such as EHIS and EU-SILC. Investigating EU-SILC is considered particularly relevant since the data are used to monitor the European Commission's aim to add 2 years of healthy life for the average European by 2020.

Related literature suggests that the biases are in fact larger and that the results ascertained in this study constitute a lower bound. First and foremost, this is because the low-educated individuals who participate in surveys are most likely healthier than the low-educated individuals who are not captured. Studies have shown that low-educated respondents have lower mortality [76], better self-reported health [77–79], and suffer less from psychosis [80] than

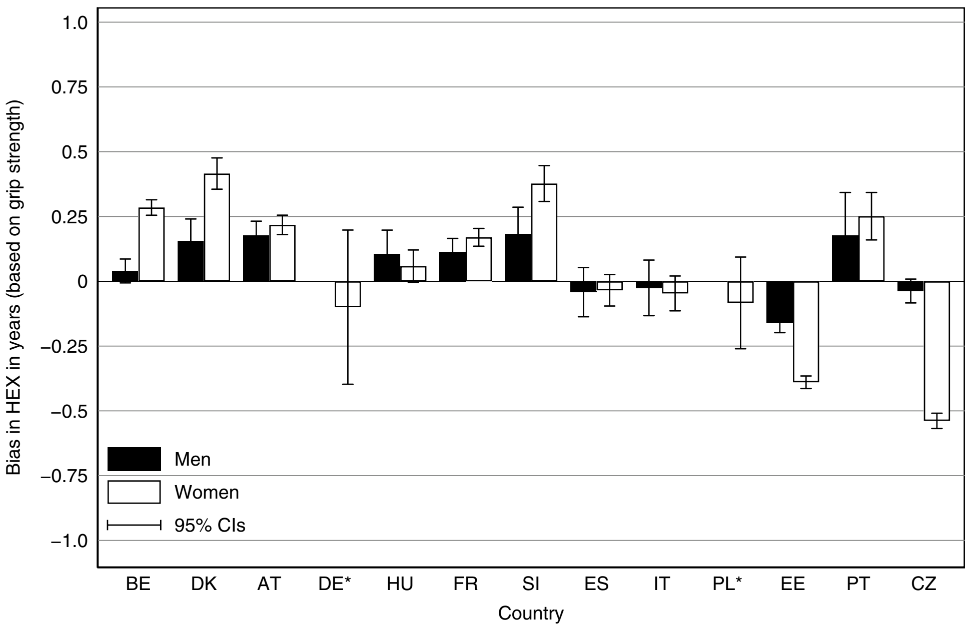


Fig. 2 Bias in HEX based on grip strength at age 50 in 2011. The bias is given in absolute years, i.e. ΔHEX is calculated as the difference between HEX^{RW} and HEX^{EW}. *Results for German as well as Polish men are not shown, because small numbers of observations at ages 50–54 and subsequent large confidence intervals prevent a meaningful illustration and interpretation of the difference between HEX for those countries

low-educated non-respondents. Thus, being included in the survey is a collider that creates an artificial negative correlation between educational attainment and health. Importantly, this collider bias introduces an even larger bias for all countries in which high-educated persons are overrepresented. In addition, measurement errors in education might increase the biases. For example, [81] found that a substantial proportion of Danish SHARE participants exaggerated their level of education, especially when they were low educated. If unhealthy low-educated individuals exaggerate their level of education, they artificially narrow the health gap between low- and high-educated participants, adding to the bias. Finally, the survival bias might increase the bias in HEX if unhealthier low-educated persons have higher mortality and consequently do not appear in the survey.

An important finding of this study was that, in contrast to common results from the literature, low-educated individuals are not necessarily more likely to be underrepresented in surveys than the highly educated. The education structures in the Italian and Spanish SHARE are almost identical to those in the respective censuses. Consequently, HEX appears to be unbiased for these countries. Potentially, this is because educational attainment hardly varies within age groups in both nations, making it easier to survey the "correct" distribution. However, Portugal has similar education patterns across age but a still highly biased HEX. What could also explain the good fit for Italy and Spain is that the effect of education on health appears to be weaker than that for other countries. Both nations are among the countries with the highest life expectancy in Europe [33], even though their overall level of education is low compared to Western and Northern European countries [65]. Moreover, the education gradient in life expectancy is very pronounced in most of Europe; yet interestingly, Italy was the only country in the sample in which life expectancy at age 50 was slightly lower for the highly educated (34.6 years) than for the medium educated (35.2 years) [13]. Unfortunately, Eurostat does not provide life expectancy by education for Spain, thereby preventing a comparison. [16] found similar results for Italian women during the 1990s, although not for men. The evidence suggests that the association between education and health might be weaker in both countries than in other European countries. If the link between education and survey participation is weaker too, this would be an additional explanation for their unbiased HEX.

The CEE countries Czechia, Estonia, and Poland also did not follow the expected pattern in terms of educational differences in survey participation. Contrary to what is generally found in the literature, high-educated individuals were underrepresented in all three countries, most profoundly so in Czechia. One explanation for this curious finding is that in all three countries, high-educated individuals are much more likely to keep working at older ages, presumably due to low pension replacement rates. This pattern holds for both men and women. For the age group of 65–74, Estonian academics had the highest employment rate in the sample (26.9%), followed by the highly educated in Czechia (20.5%), Italy (19.7%), and Poland (18.6%) [82]. As a result, the highly educated might be less likely to participate in surveys due to time constraints: when an interviewer knocks on their doors, they might simply be at work. A second, somewhat speculative, explanation for the low participation of high-educated individuals in Czechia, Estonia, and Poland could be related to trust or the lack thereof. It is well established that post-communist societies in Europe have, on average, lower levels of trust in institutions [83] and lower levels of social trust [84]. If the highly educated were more distrustful than low-educated individuals, this could explain the participation pattern in the three countries. What contradicts this speculation is the fact that Slovenia is also a CEE country with a similar history. However, the Slovenian SHARE data follow the common pattern of too few low-educated respondents.

HEX is calculated by combining the prevalence of good and bad health from survey data with life tables. This study analysed how distortion in the education structure of surveys affects HEX via biases in prevalence rates. In addition, one could analyse whether educational differences in life expectancy also add to the bias. Due to data restrictions, it is commonly assumed that all educational groups share the same life expectancies when applying Sullivan's method. However, Eurostat data for a small sample of European countries show that all countries but Italy have a clear education gradient in life expectancy. The educational differences are most pronounced in the CEE countries, save Slovenia, and are weakest in the Nordic countries [13]. If and how these differences bias HEX in the context of distorted surveys cannot be said a priori, as the bias depends on the interactions between the education distribution in the general population and the education-related response behaviour in the respective country. Thus, this study only focused on distortions due to prevalence rates to stay within scope. Furthermore, this study evaluated HEX in its most common form, which is without education-specific mortality. However, future studies should explore how educational differences in life expectancy affect the bias in HEX, especially in countries where the education gradient in mortality is strong.

The main limitation of this paper is data driven. For most countries, SHARE captures non-institutionalised persons only. Since the census does not differentiate between institutionalised and non-institutionalised persons, it was assumed that both groups are comparable. If this assumption is violated due to educational differences between the two groups, prevalence rates based on EW might deviate from the prevalence rates for the general population.

Overall, the findings of this study highlight the need to account for distortions in the education structure of survey data. First and foremost, this can be achieved by preventing the misrepresentation of certain educational groups in the first place, and if prevention does not lead to accurate representation, by adjusting for deviations with survey methods such as calibrated weights. Literature has shown that survey modes [23], recruitment methods [85], interviewer experience, and the number of attempted contacts [22] affect survey participation and consequently might be helpful for counteracting heterogeneities in survey representation. However, past evidence has also revealed that response rates have declined over time [22], and that the gap in response behaviour between high- and low-educated individuals has increased [6]. If this pattern continues, survey methods that adjust for misrepresentation will become even more important in the future. Although auxiliary information on the education structure in the general population is not available for each European country at any given year, censuses might still allow for the calibration of weights since the education structure at old age changes slowly [86], or as Schumacher [87] puts it: "education does not 'jump'".

Conclusion

Survey participation differs substantially among educational groups, which leads to biased health expectancy (HEX) when the discrepancies are not accounted for. This study was the first to explore the magnitude and direction of the bias in HEX for 13 European countries based on the Survey of Health, Ageing and Retirement in Europe (SHARE) for 2011. To this end, calibrated weights were generated so that the education structure in SHARE would resemble that of the respective Population and Housing Census.

The analysis revealed that SHARE did not accurately resemble the education structure in the general population for 11 of the 13 countries investigated, which lead to substantial biases in HEX. In most of the datasets, high-educated individuals were overrepresented. Due to the positive correlation between educational attainment and good health, HEX was upward-biased for these countries by as much as 0.87 years. Remarkably, most CEE countries showed the opposite pattern that high-educated individuals were underrepresented. As a result, HEX was underestimated for these countries by up to 0.85 years.

Understanding the sensitivity of HEX measures is crucial because of their immense scientific and political influence. In the context of ever decreasing survey response rates, it is of utmost importance that the flawed education structure in survey data is prevented and adjusted for. Only then, it is possible to accurately assess policy targets based on HEX.

Acknowledgements I am very grateful to Sergei Scherbov and Warren Sanderson whose guidance and comments greatly improved this work. Furthermore, I want to thank Vanessa Di Lego, Simone Ghislandi, Anne Goujon, Bernhard Hammer, Wolfgang Lutz, Nadia Steiber, and the participants of the Austrian Health Economics Association Workshop 2018 for their valuable input.

Funding Parts of this research were developed in the Young Scientists Summer Programme at IIASA with financial support from the Austrian National Member Organisation. Futhermore, this work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 741105). Open access funding was provided by the International Institute for Applied Systems Analysis (IIASA). The funders had no role in the design and execution of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, and approval of the manuscript. Data from the Survey of Health, Ageing and Retirement in Europe (SHARE) are used in this paper (https://doi.org/10.6103/share.w4.611). The SHARE data collection has primarily been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARE-LIFE: CIT4-CT-2006-028812), and FP7 (SHARE-PREP: No. 211909, SHARE-LEAP: No. 227822, SHARE M4: No. 261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see http://www.share-project.org).

Appendix

1.1 Proportions in SHARE versus those in the censuses

See Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13.

Table 1 Austria

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	63	9.5	86,887	15.4	206	23.9	170,957	29.6
	Medium	405	61.3	367,802	65.0	394	45.8	326,967	56.6
	High	187	28.3	111,220	19.7	240	27.9	79,609	13.8
	Unknown	6	0.9	0	0.0	21	2.4	0	0.0
	Total	661	100.0	565,909	100.0	861	100.0	577,533	100.0
60–69	Low	98	13.0	79,259	18.8	255	25.4	176,335	38.1
	Medium	416	55.2	263,463	62.6	519	51.7	249,273	53.9
	High	230	30.5	78,097	18.6	218	21.7	37,067	8.0
	Unknown	10	1.3	0	0.0	11	1.1	0	0.0
	Total	754	100.0	420,819	100.0	1003	100.0	462,675	100.0
70–79	Low	92	16.5	86,735	29.0	316	43.3	215,302	57.6
	Medium	284	51.0	164,705	55.1	272	37.3	143,121	38.3
	High	176	31.6	47,386	15.9	132	18.1	15,268	4.1
	Unknown	5	0.9	0	0.0	10	1.4	0	0.0
	Total	557	100.0	298,826	100.0	730	100.0	373,691	100.0
80–89	Low	47	25.1	41,385	33.6	152	50.5	151,359	63.9
	Medium	81	43.3	64,003	51.9	103	34.2	77,106	32.6
	High	51	27.3	17,831	14.5	41	13.6	8221	3.5
	Unknown	8	4.3	0	0.0	5	1.7	0	0.0
	Total	187	100.0	123,219	100.0	301	100.0	236,686	100.0
90+	Low	7	35.0	4742	36.4	20	58.8	29,223	66.7
	Medium	6	30.0	6016	46.2	11	32.4	12,972	29.6
	High	6	30.0	2262	17.4	2	5.9	1647	3.8
	Unknown	1	5.0	0	0.0	1	2.9	0	0.0
	Total	20	100.0	13,020	100.0	34	100.0	43,842	100.0

Table 2 Belgium

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	298	35.7	295,514	39.9	329	31.1	296,759	40.0
	Medium	217	26.0	210,435	28.4	339	32.0	213,803	28.8
	High	297	35.6	180,721	24.4	364	34.4	183,135	24.7
	Unknown	23	2.8	54,628	7.4	26	2.5	48,576	6.5
	Total	835	100.0	741,298	100.0	1,058	100.0	742,273	100.0
60–69	Low	299	38.4	264,576	48.0	331	40.4	315,593	54.4
	Medium	203	26.1	122,045	22.2	240	29.3	117,672	20.3
	High	265	34.0	121,519	22.1	236	28.8	102,593	17.7
	Unknown	12	1.5	42,791	7.8	13	1.6	44,314	7.6
	Total	779	100.0	550,931	100.0	820	100.0	580,172	100.0
70–79	Low	213	46.1	223,675	59.3	294	53.0	312,619	66.1
	Medium	103	22.3	58,576	15.5	131	23.6	64,268	13.6
	High	142	30.7	56,867	15.1	122	22.0	44,972	9.5
	Unknown	4	0.9	37,802	10.0	8	1.4	51,189	10.8
	Total	462	100.0	376,920	100.0	555	100.0	473,048	100.0
80–89	Low	140	56.5	106,684	61.5	247	69.0	217,454	69.8
	Medium	50	20.2	25,946	14.9	59	16.5	34,466	11.1
	High	54	21.8	20,467	11.8	50	14.0	18,623	6.0
	Unknown	4	1.6	20,457	11.8	2	0.6	41,186	13.2
	Total	248	100.0	173,554	100.0	358	100.0	311,729	100.0
90 +	Low	16	64.0	9905	61.3	42	73.7	35,935	69.7
	Medium	6	24.0	2155	13.3	6	10.5	4791	9.3
	High	2	8.0	2004	12.4	8	14.0	3018	5.9
	Unknown	1	4.0	2087	12.9	1	1.8	7835	15.2
	Total	25	100.0	16,151	100.0	57	100.0	51,579	100.0

Table 3 Czechia

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	284	45.4	60,953	8.8	373	42.4	143,319	20.0
	Medium	244	39.0	495,476	71.2	397	45.1	468,487	65.5
	High	93	14.9	108,342	15.6	98	11.1	82,322	11.5
	Unknown	5	0.8	31,312	4.5	12	1.4	20,992	2.9
	Total	626	100.0	696,083	100.0	880	100.0	715,120	100.0
60–69	Low	423	46.0	62,905	10.4	545	43.8	180,716	25.9
	Medium	360	39.1	443,380	73.0	558	44.8	441,352	63.3
	High	117	12.7	84,381	13.9	122	9.8	59,052	8.5
	Unknown	20	2.2	16,975	2.8	20	1.6	16,155	2.3
	Total	920	100.0	607,641	100.0	1,245	100.0	697,275	100.0
70–79	Low	219	41.5	47,015	16.4	372	53.6	173,996	42.4
	Medium	205	38.8	190,935	66.6	249	35.9	202,787	49.4
	High	94	17.8	41,874	14.6	62	8.9	22,715	5.5
	Unknown	10	1.9	6933	2.4	11	1.6	11,118	2.7
	Total	528	100.0	286,757	100.0	694	100.0	410,616	100.0
80–89	Low	76	39.4	23,055	20.0	181	63.7	120,760	50.6
	Medium	69	35.8	69,424	60.3	77	27.1	100,546	42.1
	High	44	22.8	19,280	16.7	19	6.7	8,445	3.5
	Unknown	4	2.1	3399	3.0	7	2.5	8,933	3.7
	Total	193	100.0	115,158	100.0	284	100.0	238,684	100.0
90+	Low	4	33.3	1816	23.0	14	51.9	13,684	54.6
	Medium	3	25.0	4571	57.9	11	40.7	9,393	37.5
	High	4	33.3	1158	14.7	1	3.7	736	2.9
	Unknown	1	8.3	352	4.5	1	3.7	1,242	5.0
	Total	12	100.0	7897	100.0	27	100.0	25,055	100.0

Table 4 Denmark

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	40	10.5	86,106	24.0	58	13.1	100,625	28.2
	Medium	177	46.3	172,014	47.9	126	28.4	131,424	36.8
	High	158	41.4	91,671	25.5	255	57.6	117,706	32.9
	Unknown	7	1.8	9572	2.7	4	0.9	7650	2.1
	Total	382	100.0	359,363	100.0	443	100.0	357,405	100.0
60–69	Low	33	9.6	92,455	27.4	54	14.7	124,807	36.1
	Medium	168	48.8	155,927	46.3	130	35.3	135,091	39.1
	High	136	39.5	82,314	24.4	179	48.6	80,054	23.1
	Unknown	7	2.0	6145	1.8	5	1.4	5932	1.7
	Total	344	100.0	336,841	100.0	368	100.0	345,884	100.0
70–79	Low	36	17.8	67,694	37.9	77	35.3	112,258	54.0
	Medium	101	50.0	72,763	40.8	77	35.3	60,975	29.3
	High	64	31.7	33,064	18.5	61	28.0	29,855	14.3
	Unknown	1	0.5	4901	2.7	3	1.4	4969	2.4
	Total	202	100.0	178,422	100.0	218	100.0	208,057	100.0
80–89	Low	16	16.8	35,204	48.7	74	50.0	78,481	66.6
	Medium	41	43.2	23,873	33.0	48	32.4	25,763	21.9
	High	33	34.7	11,782	16.3	25	16.9	11,554	9.8
	Unknown	5	5.3	1437	2.0	1	0.7	2045	1.7
	Total	95	100.0	72,296	100.0	148	100.0	117,843	100.0
90 +	Low	4	30.8	335	3.5	15	60.0	1263	4.4
	Medium	5	38.5	166	1.7	8	32.0	309	1.1
	High	3	23.1	278	2.9	1	4.0	190	0.7
	Unknown	1	7.7	8912	92.0	1	4.0	26,913	93.9
	Total	13	100.0	9691	100.0	25	100.0	28,675	100.0

Table 5 Estonia

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	156	19.5	6936	8.5	137	12.6	5282	5.5
	Medium	481	60.1	47,118	57.8	628	57.9	46,585	48.3
	High	162	20.3	26,085	32.0	318	29.3	43,609	45.2
	Unknown	1	0.1	1425	1.7	1	0.1	921	1.0
	Total	800	100.0	81,564	100.0	1,084	100.0	96,397	100.0
60–69	Low	278	31.2	9704	17.0	232	19.8	11,609	14.4
	Medium	419	47.0	29,786	52.3	696	59.4	40,115	49.8
	High	193	21.7	16,698	29.3	242	20.7	28,206	35.0
	Unknown	1	0.1	779	1.4	1	0.1	688	0.9
	Total	891	100.0	56,967	100.0	1171	100.0	80,618	100.0
70–79	Low	318	41.6	11,188	28.9	476	39.6	24,889	33.4
	Medium	281	36.7	16,107	41.6	483	40.1	28,996	38.9
	High	165	21.6	10,877	28.1	243	20.2	19,706	26.5
	Unknown	1	0.1	509	1.3	1	0.1	882	1.2
	Total	765	100.0	38,681	100.0	1203	100.0	74,473	100.0
80–89	Low	147	52.9	5698	42.8	295	57.4	20,559	51.9
	Medium	75	27.0	4154	31.2	157	30.5	11,561	29.2
	High	55	19.8	3230	24.3	61	11.9	6599	16.6
	Unknown	1	0.4	220	1.7	1	0.2	916	2.3
	Total	278	100.0	13,302	100.0	514	100.0	39,635	100.0
90+	Low	7	53.8	441	48.3	31	67.4	2893	62.3
	Medium	3	23.1	277	30.3	11	23.9	1114	24.0
	High	2	15.4	163	17.9	3	6.5	411	8.9
	Unknown	1	7.7	32	3.5	1	2.2	222	4.8
	Total	13	100.0	913	100.0	46	100.0	4640	100.0

Table 6 France

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	181	22.5	1,303,815	31.3	304	30.2	1,703,720	38.8
	Medium	402	49.9	1,959,813	47.1	414	41.1	1,716,270	39.1
	High	203	25.2	895,551	21.5	262	26.0	969,392	22.1
	Unknown	20	2.5	144	0.0	28	2.8	113	0.0
	Total	806	100.0	4,159,323	100.0	1008	100.0	4,389,495	100.0
60–69	Low	284	34.4	1,264,695	40.0	406	41.8	1,748,789	51.3
	Medium	315	38.2	1,277,057	40.4	320	32.9	1,106,511	32.5
	High	201	24.4	617,162	19.5	220	22.6	552,731	16.2
	Unknown	25	3.0	51	0.0	26	2.7	29	0.0
	Total	825	100.0	3,158,965	100.0	972	100.0	3,408,060	100.0
70–79	Low	271	50.6	1,182,924	57.0	461	67.7	1,910,878	70.9
	Medium	166	31.0	645,923	31.1	130	19.1	576,136	21.4
	High	90	16.8	247,312	11.9	70	10.3	207,284	7.7
	Unknown	9	1.7	0	0.0	20	2.9	0	0.0
	Total	536	100.0	2,076,159	100.0	681	100.0	2,694,298	100.0
80–89	Low	194	69.5	712,663	68.2	368	79.7	1,476,693	78.0
	Medium	52	18.6	220,702	21.1	52	11.3	291,174	15.4
	High	27	9.7	111,301	10.7	30	6.5	125,780	6.6
	Unknown	6	2.2	0	0.0	12	2.6	0	0.0
	Total	279	100.0	1,044,666	100.0	462	100.0	1,893,647	100.0
90+	Low	15	53.6	80,282	67.6	60	85.7	277,819	74.4
	Medium	7	25.0	23,167	19.5	4	5.7	59,599	16.0
	High	5	17.9	15,255	12.9	5	7.1	35,760	9.6
	Unknown	1	3.6	0	0.0	1	1.4	0	0.0
	Total	28	100.0	118,704	100.0	70	100.0	373,178	100.0

Table 7 Germany

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	5	4.8	662,600	11.6	22	11.6	1,061,130	18.2
	Medium	53	51.0	3,137,380	54.7	103	54.5	3,164,500	54.4
	High	41	39.4	1,936,590	33.8	54	28.6	1,590,890	27.4
	Unknown	5	4.8	0	0.0	10	5.3	0	0.0
	Total	104	100.0	5,736,570	100.0	189	100.0	5,816,520	100.0
60–69	Low	13	4.4	531,050	12.4	41	12.7	1,184,640	26.0
	Medium	160	54.2	2,256,210	52.8	176	54.5	2,468,540	54.1
	High	106	35.9	1,486,110	34.8	98	30.3	907,790	19.9
	Unknown	16	5.4	0	0.0	8	2.5	0	0.0
	Total	295	100.0	4,273,370	100.0	323	100.0	4,560,970	100.0
70–79	Low	10	3.7	609,250	16.7	56	23.4	1,936,480	43.3
	Medium	152	55.9	1,983,600	54.2	141	59.0	2,023,110	45.2
	High	100	36.8	1,064,890	29.1	38	15.9	513,770	11.5
	Unknown	10	3.7	0	0.0	4	1.7	0	0.0
	Total	272	100.0	3,657,740	100.0	239	100.0	4,473,360	100.0
80–89	Low	5	6.0	246,230	20.1	39	41.9	1,278,640	54.4
	Medium	47	56.6	656,190	53.5	36	38.7	884,140	37.6
	High	29	34.9	325,090	26.5	15	16.1	189,760	8.1
	Unknown	2	2.4	0	0.0	3	3.2	0	0.0
	Total	83	100.0	1,227,510	100.0	93	100.0	2,352,540	100.0
90+	Low	1	20.0	21,300	19.7	3	25.0	225,740	55.8
	Medium	2	40.0	56,130	52.0	6	50.0	149,430	37.0
	High	1	20.0	30,450	28.2	2	16.7	29,180	7.2
	Unknown	1	20.0	0	0.0	1	8.3	0	0.0
	Total	5	100.0	107,880	100.0	12	100.0	404,350	100.0

Table 8 Hungary

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	52	12.3	120,662	17.8	152	27.3	217,215	28.6
	Medium	309	72.9	453,647	66.8	323	58.1	406,335	53.5
	High	62	14.6	104,882	15.4	80	14.4	135,941	17.9
	Unknown	1	0.2	0	0.0	1	0.2	0	0.0
	Total	424	100.0	679,191	100.0	556	100.0	759,491	100.0
60–69	Low	93	17.9	125,036	24.3	200	33.3	271,885	41.1
	Medium	318	61.3	293,669	57.0	296	49.3	297,272	44.9
	High	107	20.6	96,653	18.8	104	17.3	92,447	14.0
	Unknown	1	0.2	0	0.0	1	0.2	0	0.0
	Total	519	100.0	515,358	100.0	601	100.0	661,604	100.0
70–79	Low	79	29.5	177,620	63.8	203	55.9	352,237	73.9
	Medium	133	49.6	52,768	18.9	117	32.2	88,451	18.6
	High	55	20.5	48,165	17.3	42	11.6	35,676	7.5
	Unknown	1	0.4	0	0.0	1	0.3	0	0.0
	Total	268	100.0	278,553	100.0	363	100.0	476,364	100.0
80–89	Low	39	41.1	68,943	64.7	118	77.1	212,204	84.8
	Medium	37	38.9	17,325	16.3	25	16.3	25,654	10.3
	High	18	18.9	20,313	19.1	9	5.9	12,365	4.9
	Unknown	1	1.1	0	0.0	1	0.7	0	0.0
	Total	95	100.0	106,581	100.0	153	100.0	250,223	100.0
90 +	Low	4	44.4	7092	67.5	12	60.0	27,893	87.4
	Medium	2	22.2	1606	15.3	6	30.0	2657	8.3
	High	2	22.2	1806	17.2	1	5.0	1374	4.3
	Unknown	1	11.1	0	0.0	1	5.0	0	0.0
	Total	9	100.0	10,504	100.0	20	100.0	31,924	100.0

Table 9 Italy

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	169	46.8	1,896,312	49.5	280	55.2	2,072,038	51.3
	Medium	156	43.2	1,453,862	37.9	167	32.9	1,462,737	36.2
	High	32	8.9	484,544	12.6	51	10.1	502,340	12.4
	Unknown	4	1.1	0	0.0	9	1.8	0	0.0
	Total	361	100.0	3,834,718	100.0	507	100.0	4,037,115	100.0
60–69	Low	346	60.6	2,079,003	63.3	516	73.6	2,586,617	72.4
	Medium	171	29.9	874,563	26.6	135	19.3	711,707	19.9
	High	40	7.0	333,239	10.1	41	5.8	275,036	7.7
	Unknown	14	2.5	0	0.0	9	1.3	0	0.0
	Total	571	100.0	3,286,805	100.0	701	100.0	3,573,360	100.0
70–79	Low	384	78.9	1,972,475	78.6	413	81.1	2,684,196	86.1
	Medium	69	14.2	374,245	14.9	68	13.4	336,083	10.8
	High	30	6.2	161,577	6.4	19	3.7	95,823	3.1
	Unknown	4	0.8	0	0.0	9	1.8	0	0.0
	Total	487	100.0	2,508,297	100.0	509	100.0	3,116,102	100.0
80–89	Low	144	83.7	936,638	82.8	165	93.2	1,778,669	89.4
	Medium	14	8.1	125,891	11.1	9	5.1	161,484	8.1
	High	11	6.4	68,965	6.1	2	1.1	48,485	2.4
	Unknown	3	1.7	0	0.0	1	0.6	0	0.0
	Total	172	100.0	1,131,494	100.0	177	100.0	1,988,638	100.0
90+	Low	18	85.7	110,847	83.4	27	87.1	354,613	91.5
	Medium	1	4.8	12,692	9.5	2	6.5	24,650	6.4
	High	1	4.8	9432	7.1	1	3.2	8174	2.1
	Unknown	1	4.8	0	0.0	1	3.2	0	0.0
	Total	21	100.0	132,971	100.0	31	100.0	387,437	100.0

Table 10 Poland

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	29	16.7	421,166	15.0	57	20.9	478,116	16.1
	Medium	115	66.1	1,981,997	70.8	156	57.1	2,018,930	67.8
	High	11	6.3	330,327	11.8	21	7.7	423,912	14.2
	Unknown	19	10.9	67,063	2.4	39	14.3	57,925	1.9
	Total	174	100.0	2,800,553	100.0	273	100.0	2,978,883	100.0
60–69	Low	73	22.6	420,733	24.7	144	38.9	672,145	32.7
	Medium	161	49.8	1,019,057	59.9	190	51.4	1,116,799	54.3
	High	39	12.1	230,425	13.5	15	4.1	238,273	11.6
	Unknown	50	15.5	31,166	1.8	21	5.7	29,409	1.4
	Total	323	100.0	1,701,381	100.0	370	100.0	2,056,626	100.0
70–79	Low	80	46.0	395,289	40.8	136	66.0	843,444	55.3
	Medium	57	32.8	432,775	44.7	51	24.8	543,307	35.6
	High	19	10.9	125,120	12.9	7	3.4	113,995	7.5
	Unknown	18	10.3	14,640	1.5	12	5.8	23,721	1.6
	Total	174	100.0	967,824	100.0	206	100.0	1,524,467	100.0
80–89	Low	47	60.3	199,977	53.3	79	75.2	619,859	73.2
	Medium	21	26.9	120,999	32.3	10	9.5	170,244	20.1
	High	5	6.4	47,888	12.8	2	1.9	32,531	3.8
	Unknown	5	6.4	6312	1.7	14	13.3	24,220	2.9
	Total	78	100.0	375,176	100.0	105	100.0	846,854	100.0
90+	Low	3	50.0	17,756	62.4	13	81.3	73,860	77.2
	Medium	1	16.7	7120	25.0	1	6.3	14,091	14.7
	High	1	16.7	2691	9.5	1	6.3	2219	2.3
	Unknown	1	16.7	891	3.1	1	6.3	5478	5.7
	Total	6	100.0	28,458	100.0	16	100.0	95,648	100.0

Table 11 Portugal

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	184	69.7	517,091	77.4	285	74.4	558,254	76.3
	Medium	35	13.3	79,694	11.9	46	12.0	79,177	10.8
	High	40	15.2	71,558	10.7	48	12.5	94,237	12.9
	Unknown	5	1.9	0	0.0	4	1.0	0	0.0
	Total	264	100.0	668,343	100.0	383	100.0	731,668	100.0
60–69	Low	258	78.2	469,350	85.1	299	77.5	556,689	87.7
	Medium	40	12.1	38,466	7.0	33	8.5	29,058	4.6
	High	30	9.1	43,734	7.9	37	9.6	49,145	7.7
	Unknown	2	0.6	0	0.0	17	4.4	0	0.0
	Total	330	100.0	551,550	100.0	386	100.0	634,892	100.0
70–79	Low	158	78.6	364,241	90.9	181	86.2	493,050	93.8
	Medium	16	8.0	16,569	4.1	6	2.9	12,310	2.3
	High	23	11.4	19,782	4.9	15	7.1	20,192	3.8
	Unknown	4	2.0	0	0.0	8	3.8	0	0.0
	Total	201	100.0	400,592	100.0	210	100.0	525,552	100.0
80–89	Low	49	77.8	155,428	92.0	92	82.9	279,326	95.2
	Medium	5	7.9	6162	3.6	8	7.2	6897	2.4
	High	4	6.3	7370	4.4	7	6.3	7061	2.4
	Unknown	5	7.9	0	0.0	4	3.6	0	0.0
	Total	63	100.0	168,960	100.0	111	100.0	293,284	100.0
90+	Low	4	57.1	18,068	91.4	6	60.0	48,108	95.8
	Medium	1	14.3	748	3.8	1	10.0	1109	2.2
	High	1	14.3	952	4.8	1	10.0	990	2.0
	Unknown	1	14.3	0	0.0	2	20.0	0	0.0
	Total	7	100.0	19,768	100.0	10	100.0	50,207	100.0

Table 12 Slovenia

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	87	20.8	39,279	25.3	152	29.1	51,986	34.8
	Medium	270	64.4	92,682	59.7	263	50.4	71,200	47.6
	High	61	14.6	23,315	15.0	106	20.3	26,313	17.6
	Unknown	1	0.2	0	0.0	1	0.2	0	0.0
	Total	419	100.0	155,276	100.0	522	100.0	149,499	100.0
60–69	Low	61	16.0	26,630	25.5	167	36.9	51,794	45.9
	Medium	239	62.7	60,974	58.3	204	45.0	46,809	41.5
	High	79	20.7	17,011	16.3	81	17.9	14,298	12.7
	Unknown	2	0.5	0	0.0	1	0.2	0	0.0
	Total	381	100.0	104,615	100.0	453	100.0	112,901	100.0
70–79	Low	91	32.5	20,867	31.6	206	59.2	59,259	63.2
	Medium	134	47.9	35,849	54.3	108	31.0	28,520	30.4
	High	52	18.6	9365	14.2	33	9.5	6036	6.4
	Unknown	3	1.1	0	0.0	1	0.3	0	0.0
	Total	280	100.0	66,081	100.0	348	100.0	93,815	100.0
80–89	Low	42	38.2	8192	36.2	114	63.7	36,409	67.1
	Medium	45	40.9	10,734	47.4	55	30.7	15,386	28.4
	High	22	20.0	3729	16.5	9	5.0	2434	4.5
	Unknown	1	0.9	0	0.0	1	0.6	0	0.0
	Total	110	100.0	22,655	100.0	179	100.0	54,229	100.0
90+	Low	1	25.0	608	36.4	17	85.0	4361	67.1
	Medium	1	25.0	751	45.0	1	5.0	1877	28.9
	High	1	25.0	310	18.6	1	5.0	266	4.1
	Unknown	1	25.0	0	0.0	1	5.0	0	0.0
	Total	4	100.0	1669	100.0	20	100.0	6504	100.0

Table 13 Spain

Age	Education	Men				Women			
		SHARE		Census		SHARE		Census	
		N	%	*N*	%	*N*	%	*N*	%
50–59	Low	252	62.1	1,644,040	55.9	347	63.2	1,807,090	60.4
	Medium	77	19.0	585,055	19.9	105	19.1	555,465	18.6
	High	63	15.5	711,115	24.2	71	12.9	627,870	21.0
	Unknown	14	3.4	0	0.0	26	4.7	0	0.0
	Total	406	100.0	2,940,210	100.0	549	100.0	2,990,425	100.0
60–69	Low	370	72.5	1,522,130	68.2	467	82.8	1,900,160	78.8
	Medium	52	10.2	279,630	12.5	32	5.7	241,585	10.0
	High	53	10.4	428,610	19.2	38	6.7	268,510	11.1
	Unknown	35	6.9	0	0.0	27	4.8	0	0.0
	Total	510	100.0	2,230,370	100.0	564	100.0	2,410,255	100.0
70–79	Low	401	84.6	1,253,700	80.2	458	88.8	1,763,050	89.3
	Medium	26	5.5	115,365	7.4	19	3.7	105,125	5.3
	High	28	5.9	193,660	12.4	17	3.3	106,470	5.4
	Unknown	19	4.0	0	0.0	22	4.3	0	0.0
	Total	474	100.0	1,562,725	100.0	516	100.0	1,974,645	100.0
80–89	Low	209	87.1	663,570	85.5	292	91.0	1,185,560	92.4
	Medium	5	2.1	41,485	5.3	3	0.9	49,605	3.9
	High	15	6.3	70,815	9.1	11	3.4	48,465	3.8
	Unknown	11	4.6	0	0.0	15	4.7	0	0.0
	Total	240	100.0	775,870	100.0	321	100.0	1,283,630	100.0
90+	Low	25	83.3	80,655	84.0	54	94.7	226,135	91.9
	Medium	2	6.7	6185	6.4	1	1.8	9610	3.9
	High	1	3.3	9170	9.6	1	1.8	10,450	4.2
	Unknown	2	6.7	0	0.0	1	1.8	0	0.0
	Total	30	100.0	96,010	100.0	57	100.0	246,195	100.0

1.2 Prevalence rates and HEX based on GALI by weighting strategy

See Tables 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26.

Table 14 Austria

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.11	0.07	0.15	25.88	25.15	26.60	0.13	0.08	0.17	25.54	24.80	26.29	0.33	0.27	0.39	1.30
	55–59	0.14	0.10	0.19	21.99	21.27	22.71	0.17	0.11	0.22	21.73	20.99	22.46	0.26	0.20	0.32	1.20
	60–64	0.12	0.08	0.15	18.45	17.73	19.18	0.13	0.09	0.17	18.29	17.56	19.03	0.16	0.11	0.21	0.86
	65–69	0.10	0.07	0.13	15.07	14.31	15.82	0.10	0.07	0.13	14.97	14.20	15.73	0.10	0.04	0.16	0.66
	70–74	0.17	0.13	0.21	11.75	10.95	12.56	0.18	0.13	0.22	11.64	10.82	12.47	0.11	0.05	0.17	0.94
	75–79	0.14	0.09	0.19	8.87	7.98	9.76	0.16	0.10	0.22	8.79	7.88	9.70	0.08	−0.02	0.17	0.86
	80–84	0.23	0.15	0.30	6.20	5.12	7.27	0.23	0.15	0.31	6.23	5.14	7.33	−0.04	−0.17	0.10	−0.61
	85+	0.28	0.18	0.39	4.39	2.85	5.94	0.27	0.16	0.37	4.48	2.92	6.05	−0.09	−0.34	0.16	−2.02
Women	50–54	0.06	0.03	0.08	29.34	28.81	29.87	0.06	0.03	0.09	28.94	28.40	29.48	0.40	0.36	0.43	1.37
	55–59	0.11	0.07	0.14	24.95	24.42	25.47	0.12	0.08	0.16	24.58	24.04	25.11	0.37	0.33	0.40	1.50
	60–64	0.09	0.06	0.11	20.91	20.40	21.43	0.09	0.07	0.12	20.60	20.07	21.12	0.32	0.28	0.35	1.53
	65–69	0.09	0.06	0.12	16.90	16.38	17.42	0.10	0.07	0.13	16.61	16.08	17.14	0.29	0.26	0.33	1.76
	70–74	0.17	0.13	0.20	13.03	12.51	13.56	0.18	0.14	0.21	12.76	12.22	13.29	0.28	0.24	0.31	2.17
	75–79	0.24	0.19	0.30	9.63	9.09	10.16	0.25	0.19	0.31	9.37	8.82	9.91	0.26	0.21	0.31	2.77
	80–84	0.26	0.20	0.33	6.84	6.30	7.37	0.29	0.22	0.36	6.61	6.07	7.15	0.23	0.17	0.28	3.42
	85+	0.37	0.28	0.45	4.60	4.00	5.20	0.39	0.30	0.48	4.45	3.84	5.05	0.15	0.08	0.23	3.41

Table 15 Belgium

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.08	0.04	0.13	24.82	24.05	25.58	0.15	0.04	0.25	24.43	23.64	25.21	0.39	0.33	0.45	1.59
	55–59	0.16	0.12	0.19	20.77	20.00	21.53	0.17	0.13	0.21	20.68	19.90	21.46	0.08	0.03	0.13	0.41
	60–64	0.16	0.12	0.19	17.30	16.52	18.08	0.17	0.12	0.21	17.27	16.48	18.06	0.03	−0.02	0.08	0.16
	65–69	0.14	0.10	0.18	14.01	13.20	14.82	0.14	0.10	0.18	14.05	13.23	14.88	−0.04	−0.11	0.02	−0.29
	70–74	0.15	0.10	0.19	10.78	9.92	11.65	0.14	0.09	0.18	10.85	9.98	11.73	−0.07	−0.15	0.01	−0.65
	75–79	0.25	0.19	0.30	7.76	6.80	8.72	0.24	0.18	0.30	7.78	6.81	8.76	−0.03	−0.12	0.06	−0.37
	80–84	0.28	0.21	0.36	5.46	4.30	6.62	0.28	0.21	0.36	5.45	4.27	6.64	0.01	−0.13	0.14	0.10
	85+	0.39	0.30	0.48	3.71	2.02	5.40	0.39	0.29	0.50	3.70	1.97	5.43	0.01	−0.21	0.24	0.32
Women	50–54	0.19	0.13	0.25	25.56	24.97	26.14	0.26	0.16	0.36	24.69	24.09	25.29	0.87	0.83	0.90	3.51
	55–59	0.20	0.17	0.24	21.84	21.26	22.41	0.22	0.17	0.27	21.33	20.75	21.91	0.51	0.47	0.54	2.38
	60–64	0.19	0.16	0.23	18.32	17.76	18.88	0.22	0.17	0.27	17.87	17.31	18.44	0.45	0.41	0.48	2.50
	65–69	0.23	0.18	0.27	14.83	14.28	15.37	0.25	0.19	0.30	14.50	13.95	15.05	0.32	0.28	0.36	2.23
	70–74	0.26	0.21	0.31	11.57	11.04	12.10	0.28	0.22	0.34	11.34	10.81	11.88	0.23	0.18	0.27	2.00
	75–79	0.32	0.26	0.38	8.62	8.11	9.13	0.34	0.27	0.41	8.45	7.94	8.96	0.17	0.12	0.21	1.97
	80–84	0.35	0.29	0.41	6.16	5.67	6.64	0.36	0.29	0.42	6.10	5.61	6.58	0.06	0.02	0.11	1.00
	85+	0.43	0.36	0.50	4.23	3.71	4.74	0.43	0.36	0.51	4.19	3.67	4.71	0.03	−0.02	0.09	0.82

Table 16 Czechia

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.13	0.06	0.19	22.03	21.37	22.70	0.10	0.04	0.16	22.26	21.62	22.90	−0.22	−0.28	−0.17	−1.01
	55–59	0.19	0.15	0.24	18.34	17.68	18.99	0.20	0.14	0.25	18.44	17.81	19.08	−0.11	−0.15	−0.06	−0.57
	60–64	0.15	0.11	0.19	15.25	14.58	15.91	0.15	0.11	0.20	15.37	14.74	16.01	−0.13	−0.17	−0.09	−0.84
	65–69	0.15	0.11	0.19	12.21	11.51	12.91	0.13	0.09	0.17	12.36	11.69	13.04	−0.15	−0.20	−0.10	−1.21
	70–74	0.15	0.11	0.20	9.39	8.60	10.17	0.15	0.10	0.20	9.47	8.72	10.23	−0.09	−0.15	−0.02	−0.90
	75–79	0.24	0.18	0.31	6.67	5.75	7.59	0.24	0.17	0.31	6.76	5.87	7.65	−0.09	−0.17	0.00	−1.32
	80–84	0.34	0.24	0.45	4.59	3.38	5.80	0.32	0.20	0.44	4.68	3.52	5.85	−0.09	−0.24	0.06	−2.01
	85+	0.37	0.22	0.51	3.30	1.34	5.25	0.37	0.19	0.54	3.30	1.43	5.18	−0.01	−0.31	0.30	−0.16
Women	50–54	0.12	0.07	0.17	25.90	25.42	26.38	0.10	0.05	0.15	26.75	26.29	27.21	−0.85	−0.88	−0.81	−3.17
	55–59	0.15	0.11	0.19	21.84	21.39	22.30	0.14	0.09	0.18	22.58	22.14	23.02	−0.74	−0.77	−0.71	−3.27
	60–64	0.11	0.08	0.14	18.08	17.64	18.52	0.09	0.06	0.11	18.75	18.33	19.18	−0.67	−0.70	−0.65	−3.60
	65–69	0.15	0.11	0.18	14.24	13.80	14.68	0.13	0.10	0.17	14.83	14.40	15.25	−0.59	−0.61	−0.56	−3.95
	70–74	0.20	0.15	0.24	10.75	10.31	11.20	0.19	0.14	0.24	11.31	10.87	11.74	−0.55	−0.58	−0.52	−4.87
	75–79	0.28	0.22	0.34	7.67	7.22	8.12	0.24	0.18	0.30	8.24	7.81	8.67	−0.57	−0.61	−0.54	−6.95
	80–84	0.32	0.23	0.41	5.25	4.78	5.71	0.28	0.18	0.37	5.71	5.26	6.16	−0.46	−0.51	−0.41	−8.07
	85+	0.44	0.34	0.55	3.44	2.91	3.97	0.38	0.27	0.49	3.83	3.32	4.35	−0.39	−0.46	−0.33	−10.25

Table 17 Denmark

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.07	0.03	0.11	25.93	25.04	26.83	0.11	0.05	0.17	25.41	24.48	26.33	0.53	0.43	0.62	2.08
	55–59	0.09	0.05	0.14	21.88	20.99	22.77	0.13	0.06	0.20	21.50	20.59	22.42	0.37	0.28	0.47	1.74
	60–64	0.06	0.02	0.10	18.12	17.22	19.03	0.08	0.02	0.13	17.95	17.02	18.87	0.18	0.07	0.28	0.98
	65–69	0.07	0.04	0.11	14.34	13.40	15.28	0.11	0.05	0.17	14.25	13.30	15.21	0.09	−0.01	0.19	0.64
	70–74	0.12	0.06	0.19	10.85	9.84	11.86	0.12	0.06	0.19	10.92	9.90	11.94	−0.07	−0.20	0.07	−0.60
	75–79	0.22	0.13	0.30	7.81	6.69	8.94	0.21	0.12	0.31	7.88	6.74	9.01	−0.06	−0.24	0.11	−0.81
	80–84	0.21	0.10	0.31	5.59	4.22	6.96	0.16	0.07	0.26	5.66	4.28	7.04	−0.07	−0.33	0.19	−1.23
	85+	0.38	0.24	0.51	3.50	1.52	5.48	0.40	0.25	0.56	3.34	1.32	5.37	0.16	−0.24	0.56	4.79
Women	50–54	0.08	0.04	0.11	29.18	28.54	29.81	0.11	0.05	0.17	28.87	28.22	29.52	0.31	0.25	0.37	1.07
	55–59	0.08	0.04	0.12	24.97	24.35	25.59	0.10	0.05	0.14	24.83	24.20	25.46	0.14	0.08	0.20	0.58
	60–64	0.09	0.05	0.13	20.95	20.34	21.55	0.11	0.05	0.17	20.87	20.25	21.48	0.08	0.02	0.14	0.38
	65–69	0.06	0.03	0.10	17.10	16.50	17.70	0.07	0.02	0.11	17.11	16.51	17.70	0.00	−0.07	0.06	−0.02
	70–74	0.12	0.06	0.17	13.26	12.66	13.87	0.10	0.05	0.16	13.30	12.70	13.90	−0.03	−0.11	0.04	−0.26
	75–79	0.10	0.04	0.16	9.95	9.36	10.54	0.10	0.03	0.16	9.92	9.32	10.51	0.03	−0.05	0.12	0.32
	80–84	0.20	0.12	0.28	6.90	6.27	7.53	0.21	0.12	0.31	6.86	6.23	7.49	0.04	−0.05	0.14	0.64
	85+	0.32	0.22	0.43	4.61	3.91	5.32	0.32	0.21	0.43	4.65	3.94	5.35	−0.03	−0.15	0.08	−0.75

Table 18 Estonia

		Replicated weights						Education-adjusted weights						Bias			
Gender	Age	π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.14	0.10	0.18	19.55	18.93	20.17	0.13	0.09	0.17	19.92	19.31	20.52	−0.37	−0.41	−0.32	−1.84
	55–59	0.18	0.14	0.21	16.12	15.50	16.74	0.16	0.12	0.20	16.46	15.85	17.07	−0.34	−0.38	−0.30	−2.09
	60–64	0.19	0.16	0.23	13.13	12.49	13.78	0.18	0.14	0.21	13.41	12.78	14.05	−0.28	−0.32	−0.24	−2.10
	65–69	0.22	0.17	0.26	10.42	9.72	11.12	0.20	0.16	0.24	10.64	9.96	11.33	−0.22	−0.27	−0.18	−2.11
	70–74	0.27	0.22	0.31	8.04	7.25	8.84	0.25	0.21	0.30	8.22	7.43	9.00	−0.18	−0.23	−0.12	−2.13
	75–79	0.31	0.26	0.37	5.92	4.95	6.88	0.30	0.25	0.36	6.05	5.10	7.00	−0.13	−0.21	−0.06	−2.19
	80–84	0.45	0.38	0.52	4.11	2.81	5.42	0.43	0.36	0.50	4.24	2.95	5.53	−0.13	−0.25	0.00	−2.97
	85+	0.40	0.28	0.51	3.33	1.13	5.53	0.38	0.26	0.50	3.43	1.27	5.60	−0.11	−0.47	0.26	−3.13
Women	50–54	0.10	0.08	0.13	24.91	24.47	25.34	0.10	0.07	0.13	25.35	24.92	25.78	−0.45	−0.47	−0.42	−1.76
	55–59	0.15	0.12	0.18	20.80	20.38	21.22	0.13	0.10	0.16	21.23	20.81	21.64	−0.43	−0.45	−0.40	−2.01
	60–64	0.18	0.15	0.21	16.97	16.56	17.38	0.17	0.14	0.20	17.34	16.93	17.74	−0.36	−0.39	−0.34	−2.10
	65–69	0.17	0.14	0.21	13.41	13.02	13.81	0.16	0.13	0.20	13.74	13.34	14.13	−0.32	−0.35	−0.30	−2.35
	70–74	0.24	0.20	0.27	9.95	9.56	10.34	0.22	0.18	0.25	10.24	9.85	10.62	−0.29	−0.31	−0.27	−2.82
	75–79	0.37	0.33	0.42	6.92	6.53	7.31	0.36	0.32	0.40	7.14	6.75	7.53	−0.22	−0.24	−0.19	−3.05
	80–84	0.43	0.38	0.49	4.77	4.37	5.18	0.41	0.36	0.47	4.96	4.55	5.36	−0.18	−0.21	−0.15	−3.70
	85+	0.53	0.46	0.60	3.18	2.71	3.64	0.51	0.44	0.59	3.29	2.83	3.76	−0.12	−0.16	−0.07	−3.52

Table 19 France

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.10	0.06	0.13	25.71	24.98	26.43	0.10	0.07	0.14	25.54	24.80	26.27	0.17	0.11	0.23	0.66
	55–59	0.10	0.07	0.13	21.85	21.12	22.58	0.11	0.08	0.14	21.73	20.99	22.47	0.13	0.08	0.17	0.58
	60–64	0.11	0.08	0.14	18.21	17.46	18.96	0.12	0.09	0.15	18.11	17.35	18.87	0.10	0.05	0.15	0.55
	65–69	0.12	0.08	0.15	14.70	13.92	15.48	0.12	0.08	0.16	14.63	13.84	15.41	0.07	0.02	0.13	0.51
	70–74	0.18	0.13	0.23	11.29	10.47	12.12	0.19	0.14	0.23	11.25	10.41	12.08	0.05	−0.02	0.12	0.42
	75–79	0.19	0.14	0.24	8.31	7.41	9.21	0.20	0.15	0.25	8.29	7.39	9.20	0.02	−0.06	0.10	0.24
	80–84	0.36	0.29	0.43	5.58	4.53	6.63	0.36	0.29	0.43	5.57	4.51	6.63	0.00	−0.10	0.11	0.07
	85+	0.43	0.33	0.53	3.92	2.51	5.33	0.43	0.33	0.53	3.94	2.51	5.36	−0.02	−0.20	0.17	−0.40
Women	50–54	0.10	0.06	0.13	30.20	29.67	30.73	0.11	0.07	0.14	29.99	29.46	30.53	0.20	0.17	0.24	0.68
	55–59	0.10	0.08	0.13	26.05	25.54	26.57	0.11	0.08	0.14	25.90	25.38	26.42	0.15	0.12	0.19	0.60
	60–64	0.08	0.06	0.10	22.01	21.50	22.53	0.08	0.06	0.11	21.89	21.37	22.40	0.13	0.10	0.16	0.59
	65–69	0.11	0.08	0.14	17.91	17.40	18.42	0.12	0.09	0.16	17.79	17.27	18.30	0.12	0.09	0.16	0.68
	70–74	0.15	0.11	0.19	14.02	13.51	14.52	0.16	0.12	0.20	13.96	13.45	14.47	0.06	0.02	0.10	0.42
	75–79	0.20	0.16	0.25	10.43	9.93	10.93	0.20	0.16	0.25	10.39	9.89	10.89	0.04	0.00	0.08	0.38
	80–84	0.27	0.22	0.33	7.28	6.78	7.78	0.28	0.22	0.34	7.24	6.74	7.75	0.04	0.00	0.08	0.52
	85+	0.45	0.39	0.52	4.78	4.25	5.32	0.45	0.39	0.52	4.78	4.25	5.32	0.00	−0.05	0.05	0.01

Table 20 Germany

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.55	−0.13	1.24	21.47	17.78	25.16	0.52	−0.17	1.21	21.46	17.74	25.19	0.00	−3.70	3.71	0.02
	55–59	0.18	0.10	0.27	19.76	18.30	21.22	0.18	0.10	0.26	19.58	18.06	21.10	0.18	−0.03	0.39	0.91
	60–64	0.17	0.10	0.23	16.41	14.94	17.88	0.18	0.10	0.25	16.20	14.66	17.73	0.21	0.04	0.39	1.31
	65–69	0.14	0.08	0.20	13.13	11.60	14.67	0.15	0.08	0.22	12.96	11.36	14.56	0.17	−0.01	0.35	1.31
	70–74	0.18	0.12	0.24	9.86	8.21	11.51	0.17	0.11	0.23	9.71	7.99	11.44	0.15	−0.04	0.33	1.49
	75–79	0.22	0.14	0.31	6.90	5.03	8.76	0.22	0.14	0.31	6.66	4.71	8.61	0.24	−0.02	0.50	3.58
	80–84	0.42	0.29	0.56	4.33	2.00	6.66	0.46	0.32	0.60	4.01	1.58	6.45	0.32	−0.12	0.75	7.86
	85+	0.50	0.30	0.71	2.93	−0.51	6.38	0.55	0.34	0.75	2.68	−0.94	6.30	0.25	−0.71	1.21	9.37
Women	50–54	0.13	−0.01	0.27	26.03	24.86	27.19	0.18	−0.01	0.37	25.55	24.32	26.78	0.48	0.12	0.83	1.86
	55–59	0.20	0.13	0.28	22.01	21.06	22.96	0.22	0.14	0.30	21.77	20.81	22.73	0.24	0.14	0.35	1.12
	60–64	0.14	0.09	0.19	18.45	17.53	19.37	0.16	0.09	0.23	18.28	17.35	19.21	0.18	0.08	0.27	0.97
	65–69	0.22	0.14	0.29	14.68	13.77	15.60	0.22	0.14	0.29	14.60	13.67	15.52	0.09	−0.02	0.20	0.59
	70–74	0.22	0.15	0.29	11.36	10.47	12.25	0.25	0.16	0.33	11.28	10.38	12.18	0.08	−0.02	0.18	0.70
	75–79	0.25	0.16	0.35	8.17	7.27	9.07	0.25	0.15	0.36	8.21	7.31	9.10	−0.04	−0.17	0.10	−0.45
	80–84	0.37	0.22	0.51	5.34	4.44	6.25	0.37	0.21	0.54	5.41	4.50	6.31	−0.06	−0.25	0.12	−1.18
	85+	0.51	0.37	0.65	3.35	2.46	4.24	0.49	0.34	0.64	3.48	2.59	4.37	−0.13	−0.30	0.04	−3.74

Table 21 Hungary

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.13	0.05	0.22	18.34	17.48	19.20	0.15	0.05	0.24	18.24	17.37	19.12	0.09	−0.01	0.20	0.51
	55–59	0.25	0.14	0.37	15.07	14.20	15.94	0.25	0.14	0.37	15.03	14.14	15.91	0.04	−0.03	0.12	0.29
	60–64	0.20	0.12	0.28	12.67	11.75	13.59	0.21	0.13	0.28	12.64	11.71	13.57	0.03	−0.04	0.11	0.25
	65–69	0.22	0.11	0.32	10.20	9.17	11.22	0.21	0.12	0.31	10.20	9.16	11.24	0.00	−0.10	0.10	−0.02
	70–74	0.18	0.10	0.26	7.84	6.65	9.02	0.19	0.10	0.27	7.83	6.63	9.03	0.01	−0.13	0.14	0.10
	75–79	0.37	0.18	0.56	5.37	3.89	6.86	0.36	0.18	0.54	5.40	3.90	6.91	−0.03	−0.24	0.17	−0.58
	80–84	0.58	0.37	0.80	3.96	1.88	6.04	0.59	0.36	0.83	3.92	1.81	6.03	0.04	−0.31	0.38	1.00
	85+	0.29	0.09	0.49	4.09	0.52	7.65	0.29	0.09	0.49	4.08	0.46	7.70	0.01	−0.95	0.97	0.17
Women	50–54	0.15	0.05	0.24	23.64	23.01	24.27	0.18	0.05	0.30	23.35	22.71	24.00	0.28	0.22	0.34	1.21
	55–59	0.16	0.08	0.24	19.93	19.33	20.53	0.16	0.09	0.23	19.80	19.20	20.40	0.13	0.09	0.18	0.66
	60–64	0.14	0.05	0.23	16.42	15.83	17.01	0.16	0.06	0.27	16.29	15.70	16.89	0.13	0.09	0.18	0.80
	65–69	0.19	0.06	0.31	12.95	12.36	13.55	0.19	0.08	0.29	12.91	12.32	13.51	0.04	−0.01	0.09	0.31
	70–74	0.24	0.16	0.32	9.77	9.17	10.36	0.24	0.16	0.32	9.72	9.13	10.32	0.04	−0.02	0.10	0.43
	75–79	0.26	0.16	0.36	7.03	6.43	7.63	0.27	0.17	0.37	6.95	6.35	7.55	0.09	0.02	0.16	1.23
	80–84	0.48	0.34	0.62	4.58	3.94	5.21	0.48	0.36	0.61	4.54	3.90	5.17	0.04	−0.05	0.13	0.88
	85+	0.42	0.25	0.59	3.61	2.88	4.34	0.43	0.26	0.60	3.55	2.82	4.28	0.06	−0.07	0.18	1.66

Table 22 Italy

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.01	−0.01	0.02	26.85	25.95	27.75	0.01	−0.01	0.03	26.91	26.01	27.81	−0.06	−0.17	0.05	−0.22
	55–59	0.07	0.03	0.10	22.31	21.40	23.22	0.06	0.03	0.10	22.37	21.47	23.28	−0.06	−0.15	0.02	−0.27
	60–64	0.07	0.03	0.10	18.20	17.28	19.13	0.06	0.03	0.09	18.26	17.34	19.18	−0.06	−0.13	0.02	−0.31
	65–69	0.13	0.09	0.17	14.25	13.30	15.21	0.13	0.09	0.17	14.31	13.36	15.26	−0.06	−0.14	0.02	−0.40
	70–74	0.12	0.08	0.16	10.83	9.82	11.84	0.11	0.07	0.15	10.88	9.88	11.89	−0.05	−0.14	0.03	−0.48
	75–79	0.22	0.16	0.28	7.48	6.37	8.60	0.22	0.16	0.27	7.52	6.41	8.63	−0.04	−0.15	0.07	−0.49
	80–84	0.30	0.21	0.38	4.84	3.50	6.19	0.29	0.20	0.38	4.88	3.54	6.22	−0.03	−0.20	0.13	−0.71
	85+	0.54	0.40	0.67	2.79	0.86	4.72	0.53	0.40	0.67	2.80	0.88	4.72	−0.01	−0.37	0.34	−0.51
Women	50–54	0.09	0.03	0.15	28.76	28.05	29.48	0.08	0.03	0.14	28.84	28.12	29.55	−0.07	−0.14	0.00	−0.25
	55–59	0.09	0.05	0.12	24.47	23.77	25.17	0.08	0.05	0.12	24.52	23.82	25.22	−0.05	−0.11	0.01	−0.19
	60–64	0.09	0.06	0.12	20.24	19.55	20.93	0.09	0.06	0.12	20.29	19.60	20.98	−0.05	−0.10	0.00	−0.23
	65–69	0.10	0.07	0.14	16.12	15.42	16.81	0.10	0.07	0.14	16.16	15.47	16.86	−0.05	−0.10	0.01	−0.29
	70–74	0.16	0.12	0.21	12.16	11.46	12.85	0.16	0.12	0.21	12.21	11.51	12.90	−0.05	−0.11	0.01	−0.41
	75–79	0.26	0.19	0.32	8.60	7.89	9.31	0.25	0.18	0.31	8.66	7.95	9.37	−0.05	−0.12	0.02	−0.61
	80–84	0.36	0.27	0.45	5.71	4.98	6.45	0.36	0.27	0.45	5.73	5.00	6.46	−0.02	−0.11	0.08	−0.27
	85+	0.52	0.40	0.63	3.62	2.82	4.41	0.52	0.40	0.64	3.61	2.81	4.40	0.01	−0.11	0.13	0.24

Table 23 Poland

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.00	–	–	21.12	19.95	22.28	0.00	–	–	21.37	20.20	22.53	−0.25	−1.41	0.92	−1.17
	55–59	0.18	0.11	0.24	17.04	15.82	18.26	0.16	0.10	0.23	17.30	16.08	18.53	−0.26	−0.39	−0.13	−1.51
	60–64	0.18	0.13	0.24	14.06	12.78	15.34	0.18	0.12	0.24	14.27	12.98	15.55	−0.21	−0.35	−0.08	−1.47
	65–69	0.19	0.12	0.26	11.30	9.91	12.69	0.17	0.10	0.24	11.50	10.10	12.90	−0.20	−0.36	−0.03	−1.73
	70–74	0.24	0.15	0.33	8.74	7.17	10.32	0.24	0.15	0.32	8.88	7.29	10.47	−0.14	−0.37	0.09	−1.57
	75–79	0.30	0.19	0.40	6.56	4.67	8.44	0.30	0.17	0.43	6.69	4.78	8.59	−0.13	−0.43	0.17	−1.92
	80–84	0.39	0.25	0.53	4.82	2.30	7.34	0.35	0.20	0.49	5.04	2.49	7.58	−0.21	−0.72	0.29	−4.25
	85+	0.31	0.15	0.47	3.92	−0.14	7.98	0.30	0.14	0.46	3.98	−0.12	8.08	−0.06	−1.08	0.96	−1.58
Women	50–54	0.02	−0.02	0.07	24.11	23.19	25.02	0.04	−0.03	0.10	24.16	23.23	25.08	−0.05	−0.25	0.15	−0.21
	55–59	0.13	0.09	0.18	19.62	18.72	20.53	0.13	0.08	0.18	19.73	18.83	20.63	−0.11	−0.19	−0.02	−0.55
	60–64	0.11	0.06	0.15	15.78	14.88	16.68	0.10	0.06	0.14	15.89	14.99	16.79	−0.11	−0.20	−0.03	−0.72
	65–69	0.21	0.14	0.28	11.90	10.98	12.82	0.20	0.13	0.28	11.98	11.06	12.90	−0.08	−0.19	0.03	−0.67
	70–74	0.38	0.29	0.48	8.60	7.68	9.51	0.39	0.30	0.49	8.65	7.73	9.57	−0.05	−0.17	0.06	−0.63
	75–79	0.40	0.29	0.52	6.23	5.33	7.14	0.40	0.28	0.51	6.33	5.43	7.24	−0.10	−0.24	0.04	−1.56
	80–84	0.49	0.38	0.60	4.14	3.23	5.04	0.47	0.35	0.59	4.21	3.30	5.11	−0.07	−0.21	0.07	−1.70
	85+	0.61	0.45	0.77	2.69	1.61	3.77	0.61	0.45	0.77	2.66	1.58	3.74	0.03	−0.22	0.28	0.97

Table 24 Portugal

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.05	0.00	0.10	25.92	24.74	27.09	0.05	0.00	0.09	25.92	24.75	27.08	0.00	−0.17	0.17	0.01
	55–59	0.18	0.03	0.33	21.91	20.72	23.11	0.18	0.02	0.35	21.89	20.71	23.07	0.02	−0.11	0.15	0.09
	60–64	0.04	0.01	0.07	18.67	17.46	19.87	0.04	0.01	0.08	18.66	17.46	19.85	0.01	−0.12	0.14	0.05
	65–69	0.22	0.07	0.36	14.82	13.55	16.09	0.23	0.07	0.39	14.84	13.59	16.09	−0.02	−0.15	0.12	−0.12
	70–74	0.09	0.03	0.15	12.05	10.71	13.40	0.08	0.02	0.14	12.14	10.82	13.47	−0.09	−0.26	0.08	−0.75
	75–79	0.22	0.10	0.35	8.92	7.40	10.43	0.22	0.09	0.34	8.97	7.47	10.47	−0.05	−0.28	0.18	−0.56
	80–84	0.24	0.05	0.44	6.84	4.99	8.69	0.23	0.02	0.45	6.85	5.03	8.68	−0.01	−0.39	0.36	−0.20
	85+	0.03	−0.02	0.09	5.66	2.94	8.39	0.05	−0.02	0.11	5.60	2.91	8.29	0.06	−0.79	0.92	1.11
Women	50–54	0.21	0.07	0.34	26.16	25.22	27.10	0.21	0.07	0.36	27.11	26.18	28.03	−0.95	−1.05	−0.85	−3.50
	55–59	0.09	0.03	0.14	22.47	21.57	23.38	0.09	0.03	0.14	23.47	22.58	24.35	−0.99	−1.08	−0.90	−4.23
	60–64	0.16	0.05	0.26	18.26	17.36	19.16	0.10	0.04	0.16	19.26	18.38	20.14	−1.01	−1.10	−0.91	−5.22
	65–69	0.10	0.05	0.15	14.45	13.56	15.33	0.10	0.04	0.16	15.20	14.33	16.08	−0.76	−0.84	−0.67	−4.97
	70–74	0.27	0.10	0.44	10.41	9.52	11.30	0.20	0.08	0.32	11.20	10.31	12.08	−0.78	−0.90	−0.67	−7.01
	75–79	0.21	0.08	0.34	7.37	6.51	8.22	0.19	0.07	0.31	7.83	6.96	8.69	−0.46	−0.58	−0.34	−5.88
	80–84	0.41	0.23	0.59	4.17	3.29	5.04	0.38	0.19	0.56	4.57	3.67	5.47	−0.40	−0.55	−0.26	−8.83
	85+	0.72	0.50	0.93	2.02	1.07	2.98	0.67	0.38	0.96	2.38	1.38	3.37	−0.35	−0.56	−0.14	−14.83

Table 25 Slovenia

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.07	0.03	0.11	26.01	25.15	26.87	0.07	0.03	0.11	25.91	25.04	26.78	0.10	0.00	0.20	0.38
	55–59	0.05	0.03	0.08	21.98	21.12	22.84	0.06	0.03	0.08	21.90	21.02	22.77	0.08	0.01	0.16	0.38
	60–64	0.09	0.04	0.13	18.19	17.30	19.08	0.09	0.04	0.13	18.11	17.20	19.01	0.08	−0.01	0.17	0.45
	65–69	0.08	0.04	0.12	14.72	13.79	15.66	0.09	0.04	0.13	14.64	13.69	15.59	0.08	−0.02	0.18	0.56
	70–74	0.17	0.11	0.24	11.46	10.45	12.47	0.18	0.11	0.25	11.38	10.35	12.40	0.08	−0.04	0.19	0.68
	75–79	0.20	0.12	0.28	8.95	7.80	10.09	0.20	0.13	0.28	8.89	7.73	10.04	0.06	−0.09	0.21	0.67
	80–84	0.14	0.06	0.21	7.04	5.61	8.48	0.15	0.06	0.23	6.99	5.54	8.45	0.05	−0.19	0.29	0.71
	85+	0.04	−0.01	0.09	5.54	3.34	7.74	0.04	−0.01	0.09	5.54	3.30	7.77	0.00	−0.52	0.52	0.03
Women	50–54	0.12	0.07	0.16	29.09	28.40	29.78	0.13	0.08	0.18	28.84	28.14	29.53	0.25	0.19	0.32	0.88
	55–59	0.13	0.03	0.23	25.04	24.37	25.70	0.15	0.03	0.26	24.84	24.17	25.51	0.19	0.14	0.25	0.78
	60–64	0.10	0.04	0.16	21.15	20.50	21.80	0.11	0.04	0.18	21.03	20.37	21.68	0.12	0.07	0.18	0.59
	65–69	0.16	0.07	0.25	17.24	16.60	17.88	0.17	0.08	0.26	17.15	16.51	17.79	0.09	0.02	0.15	0.52
	70–74	0.17	0.07	0.28	13.78	13.16	14.39	0.18	0.07	0.29	13.72	13.10	14.34	0.05	−0.01	0.11	0.38
	75–79	0.22	0.11	0.32	10.48	9.88	11.09	0.21	0.11	0.32	10.48	9.87	11.08	0.01	−0.06	0.08	0.08
	80–84	0.17	0.10	0.23	7.89	7.29	8.48	0.17	0.09	0.24	7.87	7.27	8.47	0.01	−0.06	0.09	0.16
	85+	0.21	0.10	0.32	5.56	4.86	6.26	0.21	0.10	0.32	5.55	4.85	6.26	0.01	−0.12	0.13	0.17

Table 26 Spain

Gender	Age	Replicated weights						Education-adjusted weights						Bias			
		π	95% CI		HEX	95% CI		π	95% CI		HEX	95% CI		ΔHEX	95% CI		Δ%
Men	50–54	0.03	0.00	0.06	29.05	28.47	29.63	0.03	0.00	0.05	29.09	28.51	29.67	−0.04	−0.11	0.02	−0.14
	55–59	0.05	0.02	0.08	24.79	24.21	25.37	0.04	0.01	0.07	24.82	24.24	25.40	−0.02	−0.08	0.03	−0.10
	60–64	0.05	0.02	0.08	20.81	20.23	21.40	0.05	0.02	0.07	20.82	20.23	21.40	0.00	−0.06	0.05	−0.02
	65–69	0.04	0.01	0.07	17.04	16.44	17.64	0.04	0.01	0.07	17.02	16.42	17.62	0.02	−0.03	0.07	0.12
	70–74	0.04	0.01	0.06	13.40	12.77	14.04	0.03	0.01	0.05	13.39	12.76	14.03	0.01	−0.05	0.07	0.08
	75–79	0.10	0.06	0.14	10.01	9.31	10.72	0.11	0.07	0.14	9.98	9.27	10.69	0.03	−0.03	0.09	0.30
	80–84	0.16	0.10	0.22	7.24	6.40	8.08	0.16	0.10	0.22	7.24	6.39	8.08	0.01	−0.09	0.10	0.08
	85+	0.19	0.11	0.26	5.30	4.14	6.47	0.19	0.11	0.26	5.30	4.14	6.47	0.00	−0.15	0.15	0.01
Women	50–54	0.01	0.00	0.02	34.02	33.58	34.45	0.01	0.00	0.03	34.01	33.58	34.45	0.00	−0.04	0.04	0.01
	55–59	0.01	0.00	0.03	29.40	28.97	29.84	0.01	0.00	0.03	29.41	28.97	29.84	0.00	−0.04	0.03	−0.01
	60–64	0.01	0.00	0.02	24.86	24.43	25.30	0.01	0.00	0.02	24.86	24.43	25.30	0.00	−0.04	0.04	0.00
	65–69	0.04	0.02	0.06	20.36	19.92	20.80	0.04	0.02	0.06	20.36	19.92	20.80	0.00	−0.04	0.04	0.01
	70–74	0.05	0.02	0.08	16.13	15.69	16.56	0.05	0.02	0.08	16.12	15.68	16.56	0.01	−0.03	0.05	0.05
	75–79	0.09	0.05	0.12	12.13	11.69	12.57	0.09	0.06	0.13	12.12	11.68	12.56	0.00	−0.03	0.04	0.02
	80–84	0.14	0.09	0.19	8.65	8.19	9.10	0.14	0.09	0.19	8.66	8.20	9.11	−0.01	−0.06	0.04	−0.10
	85+	0.25	0.19	0.32	5.93	5.43	6.43	0.25	0.18	0.32	5.95	5.46	6.45	−0.02	−0.08	0.03	−0.39

References

1. Sanders, B.S.: Measuring community health levels. Am. J. Public Health **54**, 1063–1070 (1964)
2. Sullivan, D.F.: A single index of mortality and morbidity. HSMHA Health Rep. **86**, 347–354 (1971). https://doi.org/10.2307/4594169
3. European Commission: Turning Europe into a true Innovation Union. http://europa.eu/rapid/press-release_MEMO-10-473_en.htm
4. European Commission: What is the European innovation partnership on active and healthy ageing (EIP on AHA)?. https://ec.europa.eu/eip/ageing/about-the-partnership_en
5. Bogaert, P., Van Oyen, H., Beluche, I., Cambois, E., Robine, J.M.: The use of the global activity limitation Indicator and healthy life years by member states and the European Commission. Arch. Public Health **76**, 1–7 (2018). https://doi.org/10.1186/s13690-018-0279-z
6. Reinikainen, J., Tolonen, H., Borodulin, K., Härkänen, T., Jousilahti, P., Karvanen, J., Koskinen, S., Kuulasmaa, K., Männistö, S., Rissanen, H., Vartiainen, E.: Participation rates by educational levels have diverged during 25 years in Finnish health examination surveys. Eur. J. Public Health **28**, 237–243 (2018). https://doi.org/10.1093/eurpub/ckx151
7. Demarest, S., Van Der Heyden, J., Charafeddine, R., Tafforeau, J., Van Oyen, H., Van Hal, G.: Socio-economic differences in participation of households in a Belgian national health survey. Eur. J. Public Health **23**, 981–985 (2013). https://doi.org/10.1093/eurpub/cks158
8. Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Helenius, H.: Non-response and related factors in a nation-wide health survey. Eur. J. Epidemiol. **17**, 991–999 (2001)
9. Mirowsky, J., Ross, C.E.: Education, Social Status, and Health. Routledge, New York (2003)
10. Eide, E.R., Showalter, M.H.: Estimating the relation between health and education: What do we know and what do we need to know? Econ. Educ. Rev. **30**, 778–791 (2011). https://doi.org/10.1016/j.econedurev.2011.03.009
11. Cutler, D.M., Lleras-Muney, A.: Education and health: evaluating theories and evidence. NBER Work. Pap. Ser. 12352 (2006)
12. Schurer, S., Shields, M.A., Jones, A.M.: Socio-economic inequalities in bodily pain over the life cycle: longitudinal evidence from Australia, Britain and Germany. J. R. Stat. Soc. Ser. A Stat. Soc. **177**, 783–806 (2014). https://doi.org/10.1111/rssa.12058
13. Eurostat: Life expectancy by age, sex and educational attainment level. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_mlexpecedu&lang=en
14. Eurostat: Self-perceived long-standing limitations in usual activities due to health problem by sex, age and educational attainment level (EU28). http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do
15. Cutler, D.M., Lleras-Muney, A.: Understanding differences in health behaviors by education. J. Health Econ. **29**, 1–28 (2010). https://doi.org/10.1016/j.jhealeco.2009.10.003
16. Luy, M., Di Giulio, P., Caselli, G.: Differences in life expectancy by education and occupation in Italy, 1980–94: indirect estimates from maternal and paternal orphanhood. Popul. Stud. (NY) **65**, 137–155 (2011). https://doi.org/10.1080/00324728.2011.568192
17. Matthews, R.J., Jagger, C., Hancock, R.M.: Does socio-economic advantage lead to a longer, healthier old age? Soc. Sci. Med. **62**, 2489–2499 (2006). https://doi.org/10.1016/j.socscimed.2005.11.019
18. Cambois, E., Egidi, V., Jagger, C., Jeune, B., Cambois, E., Solé-Auró, A., Brønnum-Hansen, H., Egidi, V., Jagger, C., Jeune, B., Nusselder, W.J., Oyen, H.Van, White, C., Robine, J.: Educational differentials in disability vary across an within welfare regimes: a comparison of 26 European countries in 2009. J. Epidemiol. Community Health **70**, 331–338 (2015)
19. Tubeuf, S., Jusot, F.: Social health inequalities among older Europeans: the contribution of social and family background. Eur. J. Health Econ. **12**, 61–77 (2011). https://doi.org/10.1007/s10198-010-0229-3
20. Van Der Heyden, J., De Bacquer, D., Gisle, L., Demarest, S., Charafeddine, R., Drieskens, S., Tafforeau, J., Van Oyen, H., Van Herck, K.: Additional weighting for education affects estimates

from a National Health Interview Survey. Eur. J. Public Health **27**, 892–897 (2017). https://doi.org/10.1093/eurpub/ckx005
21. Ekholm, O., Gundgaard, J., Hansen, E.H., Rasmussen, N.K.R.: The effect of health, socio-economic position, and mode of data collection on non-response in health interview surveys. Scand. J. Public Health. **38**, 699–706 (2010). https://doi.org/10.1177/1403494810382474
22. Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R.: Survey Methodology. Wiley, Hoboken (2009)
23. Christensen, A.I., Ekholm, O., Glümer, C., Juel, K.: Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys. Eur. J. Public Health **24**, 327–332 (2014). https://doi.org/10.1093/eurpub/ckt067
24. Shavers, V.L., Lynch, C.F., Burmeister, L.F.: Racial differences in factors that influence the willingness to participate in medical research studies. Ann. Epidemiol. **12**, 248–256 (2002)
25. Van Loon, A.J.M., Tijhuis, M., Picavet, H.S.J., Surtees, P.G., Ormel, J.: Survey non-response in the Netherlands: effects on prevalence estimates and associations. Ann. Epidemiol. **13**, 105–110 (2003). https://doi.org/10.1016/S1047-2797(02)00257-0
26. Jagger, C., Weston, C., Cambois, E., Van Oyen, H., Nusselder, W., Doblhammer, G., Rychtarikova, J., Robine, J.M.: Inequalities in health expectancies at older ages in the European Union: findings from the Survey of Health and Retirement in Europe (SHARE). J. Epidemiol. Community Health **65**, 1030–1035 (2011). https://doi.org/10.1136/jech.2010.117705
27. Jagger, C., McKee, M., Christensen, K., Lagiewka, K., Nusselder, W., Van Oyen, H., Cambois, E., Jeune, B., Robine, J.M.: Mind the gap—reaching the European target of a 2-year increase in healthy life years in the next decade. Eur. J. Public Health **23**, 829–833 (2013). https://doi.org/10.1093/eurpub/ckt030
28. Luy, M., Minagawa, Y.: Gender gaps—life expectancy and proportion of life in poor health. Health Rep. **25**, 12–19 (2014)
29. Pongiglione, B., De Stavola, B.L., Ploubidis, G.B.: A systematic literature review of studies analyzing inequalities in health expectancy among the older population. PLoS One **10**, 1–21 (2015). https://doi.org/10.1371/journal.pone.0130747
30. Robine, J.-M., Cambois, E.: Healthy life expectancy in Europe. Popul. Soc. (2013)
31. Rubio-Valverde, J.R., Nusselder, W.J., Mackenbach, J.P.: Educational inequalities in Global Activity Limitation Indicator disability in 28 European Countries: does the choice of survey matter? Int. J. Public Health. **64**, 461–474 (2019). https://doi.org/10.1007/s00038-018-1174-7
32. Jagger, C.: Trends in life expectancy and healthy life expectancy. Gov. Off. Sci. (2015)
33. Eurostat: Life table. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_mlifetable&lang=en
34. Saito, Y., Robine, J.M., Crimmins, E.M.: The methods and materials of health expectancy. Stat. J. IAOS **30**, 209–223 (2014). https://doi.org/10.3233/SJI-140840
35. Brønnum-Hansen, H., Baadsgaard, M., Eriksen, M.L., Andersen-Ranberg, K., Jeune, B.: Educational inequalities in health expectancy during the financial crisis in Denmark. Int. J. Public Health **60**, 927–935 (2015). https://doi.org/10.1007/s00038-015-0726-3
36. Jeune, B., Andersen Ranberg, K., Eriksen, M.L., Brønnum Hansen, H.: Improvement in health expectancy at ages 50 and 65 in Denmark during the period 2004-2011. Scand. J. Public Health **43**, 254–259 (2015)
37. De Luca, G.: Weights. In: SHARE Release Guide 6.1.0, pp. 34–41 (2018)
38. De Luca, G., Rossetti, C.: Computing Calibrated Weights. SHARE Man. (2018)
39. Malter, F., Börsch-Supan, A. (eds.): SHARE Wave 4: Innovations and Methodology. Max Planck Institute for Social Law and Social Policy, Munich (2013)
40. Börsch-Supan, A., Brandt, M., Litwin, H., Weber, G. (eds.): Active ageing and solidarity between generations in Europe: First results from SHARE after the economic crisis. De Gruyter, Berlin (2013)
41. Börsch-Supan, A.: Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 7.0.0. SHARE-ERIC. Data set., https://doi.org/10.6103/share.w4.700
42. Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S.: Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). Int. J. Epidemiol. **42**, 992–1001 (2013)
43. Lynn, P., De Luca, G., Ganninger, M.: Sample design in SHARE wave four. In: Malter, F., Börsch-Supan, A. (eds.) SHARE Wave 4: Innovations and Methodology, pp. 74–123 (2013)
44. Bergmann, M., Kneip, T., Luca, G. De, Scherpenzeel, A.: Survey participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1-6. SHARE Work. Pap. Ser. 31-2017 (2017)
45. Eurostat: International Standard Classification of Education (ISCED). http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_%28ISCED%29#Correspondence_between_ISCED_2011_and_ISCED_1997
46. Van Oyen, H., Cox, B., Jagger, C., Cambois, E., Nusselder, W., Gilles, C., Robine, J.M.: Gender gaps in life expectancy and expected years with activity limitations at age 50 in the European Union: associations with macro-level structural indicators. Eur. J. Ageing **7**, 229–237 (2010)
47. Renard, F., Devleesschauwer, B., Van Oyen, H., Gadeyne, S., Deboosere, P.: Evolution of educational inequalities in life and health expectancies at 25 years in Belgium between 2001 and 2011: a census- based study. Arch. Public Health **77**, 1–10 (2019). https://doi.org/10.1186/s13690-019-0330-8
48. Jagger, C., Gillies, C., Cambois, E., Van Oyen, H., Nusselder, W., Robine, J.M.: The Global Activity Limitation Index measured function and disability similarly across European countries. J. Clin. Epidemiol. **63**, 892–899 (2010). https://doi.org/10.1016/j.jclinepi.2009.11.002
49. Van Oyen, H., Heyden, J., Perenboom, R., Jagger, C.: Monitoring population disability: evaluation of a new Global Activity Limitation Indicator (GALI). Soz. Praventivmed. **51**, 153–161 (2006). https://doi.org/10.1007/s00038-006-0035-y
50. Srisurapanont, M., Suttajit, S., Eurviriyanukul, K., Varnado, P.: Discrepancy between objective and subjective cognition in adults with major depressive disorder. Sci. Rep. **7**, 1–7 (2017)
51. Peracchi, F., Rossetti, C.: Heterogeneity in health responses and anchoring vignettes. Empir. Econ. **42**, 513–538 (2012)
52. Schneider, U., Pfarr, C., Schneider, B.S., Ulrich, V.: I feel good! Gender differences and reporting heterogeneity in self-assessed health. Eur. J. Health Econ. **13**, 251–265 (2012)
53. Bago d'Uva, T., O'Donnell, O., Van Doorslaer, E.: Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. Int. J. Epidemiol. **37**, 1375–1383 (2008)
54. Hardy, M.A., Acciai, F., Reyes, A.M.: How health conditions translate into self-ratings: a comparative study of older adults across Europe. J. Health Soc. Behav. **55**, 320–341 (2014)
55. Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., O'Donnell, O.: Does reporting heterogeneity bias the measurement of health diesparities? Health Econ. **17**, 351–375 (2008)
56. Jürges, H.: True health vs response style: exploring cross-country differences in self-reported health. Health Econ. **16**, 163–178 (2007)
57. Spitzer, S., Weber, D.: Reporting biases in self-assessed physical and cognitive health status of older Europeans. PLoS One **14**, e0223526 (2019). https://doi.org/10.1371/journal.pone.0223526

58. Idler, E.L.: Age differences in self-assessments of health: age changes, cohort differences, or survivorship? J. Gerontol. **48**, 289–300 (1993)
59. Henchoz, K., Cavalli, S., Girardin, M.: Health perception and health status in advanced old age: a paradox of association. J. Aging Stud. **22**, 282–290 (2008). https://doi.org/10.1016/j.jaging.2007.03.002
60. Cruz-Jentoft, A.J., Baeyens, J.P., Bauer, J.M., Boirie, Y., Cederholm, T., Landi, F., Martin, F.C., Michel, J.-P., Rolland, Y., Schneider, S.M., Topinkova, E., Vandewoude, M., Zamboni, M.: Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing **39**, 412–423 (2010). https://doi.org/10.1093/ageing/afq034
61. Steiber, N.: Strong or weak handgrip? Normative reference values for the German population across the life course stratified by sex, age, and body height. PLoS One **11**, 1–14 (2016). https://doi.org/10.1371/journal.pone.0163917
62. Rijk, J.M., Roos, P.R.K.M., Deckx, L., Van den Akker, M., Buntinx, F.: Prognostic value of handgrip strength in people aged 60 years and older: a systematic review and meta-analysis. Geriatr. Gerontol. Int. **16**, 5–20 (2016). https://doi.org/10.1111/ggi.12508
63. Andersen-Ranberg, K., Petersen, I., Frederiksen, H., Mackenbach, J.P., Christensen, K.: Cross-national differences in grip strength among 50+ year-old Europeans: results from the SHARE study. Eur. J. Ageing **6**, 227–236 (2009). https://doi.org/10.1007/s10433-009-0128-6
64. Roberts, H.C., Denison, H.J., Martin, H.J., Patel, H.P., Syddall, H., Cooper, C., Sayer, A.A.: A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. Age Ageing **40**, 423–429 (2011). https://doi.org/10.1093/ageing/afr051
65. Eurostat: 2011 Census Hub. https://ec.europa.eu/eurostat/web/population-and-housing-census/census-data/2011-census
66. Battaglia, M.P., Izrael, D., Hoaglin, D.C., Frankel, M.R.: Practical considerations in raking survey data. Surv. Pract. 2, 1–37 (2009). http://surveypractice.org/2009/06/29/raking-survey-data/
67. Winter, N.: SURVWGT: Stata module to create and manipulate survey weights. Stat. Softw. Components S427503, Dep. Econ. Bost. Coll. (2018)
68. Deville, J.-C., Särndal, C.-E.: Calibration estimators in survey sampling. J. Am. Stat. Assoc. **87**, 376–382 (1992). https://doi.org/10.1093/biomet/90.4.937
69. Pacifico, D.: SREWEIGHT: a Stata command to reweight survey data to external totals. Stata J. **14**, 4–21 (2014)
70. Kolenikov, S.: Calibrating survey data using iterative proportional fitting (raking). Stata J. **14**, 22–59 (2014)
71. Preston, S.H., Heuveline, P., Guillot, M.: Demography: measuring and modeling population processes. Blackwell Publishing Ltd, Malden (2001)
72. Jagger, C., Van Oyen, H., Robine, J.-M.: Health Expectancy Calculation by the Sullivan Method : A Practical Guide, 4th edn. http://www.eurohex.eu/pdf/Sullivan_guide_prefinal_oct2014.pdf
73. Rogers, A., Rogers, R.G., Belanger, A.: Longer life but worse health? measurement and dynamics. Gerontologist **30**, 640–649 (1990). https://doi.org/10.1093/geront/30.5.640
74. Rogers, R.G., Rogers, A., Belanger, A.: Active life among the elderly in the United States: multistate life-table estimates and population projections. Milibank Q. **67**, 370–411 (1989)
75. Mathers, C.D., Robine, J.M.: How good is Sullivan's method for monitoring changes in population health expectancies. J. Epidemiol. Community Health **51**, 80–86 (1997). https://doi.org/10.1136/jech.51.1.80
76. Jousilahti, P., Salomaa, V., Kuulasmaa, K., Niemelä, M., Vartiainen, E.: Total and cause specific mortality among participants and non-participants of population based health surveys: a comprehensive follow up of 54 372 Finnish men and women. J. Epidemiol. Community Health **59**, 310–315 (2005). https://doi.org/10.1136/jech.2004.024349
77. Chinn, D.J., White, M., Howel, D., Harland, J.O.E., Drinkwater, C.K.: Factors associated with non-participation in a physical activity promotion trial. Public Health **120**, 309–319 (2006). https://doi.org/10.1016/j.puhe.2005.11.003
78. Lorant, V., Demarest, S., Miermans, P.J., Van Oyen, H.: Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census. Int. J. Epidemiol. **36**, 1292–1299 (2007). https://doi.org/10.1093/ije/dym191
79. Jones, A.M., Koolman, X., Rice, N.: Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighted estimators in non-linear models. J. R. Stat. Soc. Ser. A Stat. Soc. **169**, 543–569 (2006). https://doi.org/10.1111/j.1467-985x.2006.00399.x
80. Haapea, M., Miettunen, J., Veijola, J., Lauronen, E., Tanskanen, P., Isohanni, M.: Non-participation may bias the results of a psychiatric survey. An analysis from the survey including magnetic resonance imaging within the Northern Finland 1966 Birth Cohort. Soc. Psychiatry Psychiatr. Epidemiol. **42**, 403–409 (2007). https://doi.org/10.1007/s00127-007-0178-z
81. Bingley, P., Martinello, A.: Measurement error in the Survey of Health, Ageing and Retirement in Europe: a validation study with administrative data for education level, income and employment. SHARE Work. Pap. Ser. 16-2014 (2014)
82. Eurostat: Employment rates by sex, age and educational attainment level (%). http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do
83. Zmerli, S., Van der Meer, T.W.G.: Handbook on political trust. Edward Elgar Publishing, Cheltenham (2014)
84. Uslaner, E.: Trust and civic engagement in East and West. In: Badescu, G., Uslaner, E. (eds.) Social Capital and the Transition to Democracy. Routledge, New York (2003)
85. Tolonen, H., Kuulasmaa, K., Ahonen, S., Jentoft, S., Heldal, J.: Differences in participation rates and lessons learned about recruitment of participants—the European Health Examination Survey Pilot Project. Scand. J. Public Health. **43**, 212–219 (2015). https://doi.org/10.1177/1403494814565692
86. Samir, K.C., Potančoková, M., Bauer, R., Goujon, A.: Data and methods. In: Lutz, W., Buth, W.P., Samir, K.C. (eds.) World Population and Human Capital in the Twenty-First Century, pp. 434–518. Oxford University Press, Oxford (2014)
87. Schumacher, E.F.: Small is Beautiful: A Study of Economics as if People Mattered. Harper and Row, New York (1973)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.2 Health measures and health perception (2nd Publication)

The second publication of the dissertation answers RQ 2: "*How accurate is self-assessed health status, and how are health measures biased by individual health misperception?*". This is a joint research paper with Daniela Weber[2] and was published on October 8th 2020 as

Spitzer, S. & Weber, D. (2019). Reporting biases in self-assessed physical and cognitive health status of older Europeans. *PLoS ONE*, 14(10): e0223526. https://doi.org/10.1371/journal.pone.0223526

Abstract: This paper explores which demographic characteristics substantially bias self-reported physical and cognitive health status of older Europeans. The analysis utilises micro-data for 19 European countries from the Survey of Health, Ageing and Retirement in Europe to compare performance-tested outcomes of mobility and memory with their self-reported equivalents. Relative importance analysis based on multinomial logistic regressions shows that the bias in self-reported health is mostly due to reporting heterogeneities between countries and age groups, whereas gender contributes little to the discrepancy. Concordance of mobility and cognition measures is highly related; however, differences in reporting behaviour due to education and cultural background have a larger impact on self-assessed memory than on self-assessed mobility. Southern as well as Central and Eastern Europeans are much more likely to misreport their physical and cognitive abilities than Northern and Western Europeans. Overall, our results suggest that comparisons of self-reported health between countries and age groups are prone to significant biases, whereas comparisons between genders are credible for most European countries. These findings are crucial given that self-assessed data are often the only information available to researchers and policymakers when asking health-related questions.

The article is reprinted by permission of the publisher.

[2]Daniela Weber is affiliated with the Health Economics and Policy Division of the Vienna University of Economics and Business (WU), Vienna, Austria as well as with the World Population Program at the International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital, Laxenburg, Austria.

RESEARCH ARTICLE

Reporting biases in self-assessed physical and cognitive health status of older Europeans

Sonja Spitzer [1] *, **Daniela Weber** [1,2]

1 World Population Program at the International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital, Laxenburg, Austria, **2** Health Economics and Policy Division of the Vienna University of Economics and Business, Vienna, Austria

* sonja.spitzer@iiasa.ac.at

OPEN ACCESS

Citation: Spitzer S, Weber D (2019) Reporting biases in self-assessed physical and cognitive health status of older Europeans. PLoS ONE 14 (10): e0223526. https://doi.org/10.1371/journal.pone.0223526

Editor: Federica Maria Origo, Universitá degli Studi di Bergamo, ITALY

Received: March 20, 2019

Accepted: September 23, 2019

Published: October 8, 2019

Data Availability Statement: The data underlying the results presented in the study are available from the Survey of Health, Ageing and Retirement in Europe (http://www.share-project.org). In particular, Wave 2 (10.6103/SHARE.w2.611), Wave 4 (10.6103/SHARE.w4.611), and Wave 5 (10.6103/SHARE.w5.611) are utilised.

Funding: This project received funding from the Austrian Federal Ministry of Science, Research, and Economy in the framework of the Joint Programming Initiative "More Years, Better Lives – The Challenges and Opportunities of Demographic

Abstract

This paper explores which demographic characteristics substantially bias self-reported physical and cognitive health status of older Europeans. The analysis utilises micro-data for 19 European countries from the Survey of Health, Ageing and Retirement in Europe to compare performance-tested outcomes of mobility and memory with their self-reported equivalents. Relative importance analysis based on multinomial logistic regressions shows that the bias in self-reported health is mostly due to reporting heterogeneities between countries and age groups, whereas gender contributes little to the discrepancy. Concordance of mobility and cognition measures is highly related; however, differences in reporting behaviour due to education and cultural background have a larger impact on self-assessed memory than on self-assessed mobility. Southern as well as Central and Eastern Europeans are much more likely to misreport their physical and cognitive abilities than Northern and Western Europeans. Overall, our results suggest that comparisons of self-reported health between countries and age groups are prone to significant biases, whereas comparisons between genders are credible for most European countries. These findings are crucial given that self-assessed data are often the only information available to researchers and policymakers when asking health-related questions.

Introduction

Understanding the bias in self-reported health and its determinants is of utmost importance, because subjective data are often the only information at hand when researchers and policymakers ask health-related questions. These data are readily available as their collection takes less time and is more cost-effective than performance-based health measures. However, several studies show discrepancies between tested and self-reported health indicators [1–9]. In a meta-analysis, [1] find that correlation coefficients of tested and self-reported functional ability range from -0.72 to 0.60. Thus, subjective health measures are prone to bias. Assuming an underlying true but unobservable health status, survey respondents will report a higher or lower level of health depending on their demographic characteristics. Over- and underestimating health does not only harm the reliability of survey data, but also individuals themselves.

Change". Furthermore, the research leading to these results received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC under Grant ERC2012-AdG 323947-Re-Ageing). IIASA has made funds available for publishing of this paper. IIASA encourages and actively supports its researchers to publish their research in journal articles or books that are made available for free to all users (gold open access). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This paper uses data from SHARE Wave 2 (DOI: 10.6103/SHARE.w2.611), Wave 4 (DOI: 10.6103/SHARE.w4.611), and Wave 5 (DOI: 10.6103/SHARE.w5.611). The SHARE data collection has primarily been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

Competing interests: The authors have declared that no competing interests exist.

Overrating health, for example, is associated with riskier health behaviour. Older individuals that overestimate their physical ability are more prone to suffer fall-induced injuries [10].

Research analysing the reporting bias in subjective health is growing and can be categorised into three streams based on the methods applied. A common strategy is to analyse the determinants of and variation in general self-rated health [11–15]. A second approach is the application of vignette methods, in which it is assumed that survey participants rate vignettes similarly to their own health [16–18]. However, there is evidence that the vignette method does not capture the full scale of reporting heterogeneity in health [16,17]. Finally, reporting biases can be evaluated directly by matching survey participants' reports on their health with their actual tested health. In comparison with other techniques, the most important advantage of this method is that the response behaviour of each survey participant can be directly evaluated in view of his or her individual characteristics, while being fully flexible on the specification of the relationship between the tested and the self-reported variables. To date, however, this strategy has only been applied in small-scale studies evaluating either self-assessed physical health [1,2] or self-assessed cognitive abilities [3,19,20], but never both of them simultaneously.

Our scientific contribution is threefold. First, we quantify which demographic characteristics most relevantly contribute to the overall bias in subjective health. The demographic characteristics analysed in this study are those commonly used for health comparisons and thus collected in most surveys, namely country of residence, gender, age and education. To this end, we conduct a relative importance analysis allowing us to clearly identify which characteristics contribute the largest bias and consequently should not be compared based on self-reports only. To the best of our knowledge, no previous research has decomposed the bias in subjective health into its contributing determinants. Second, we directly match performance-based health measures with their self-reported equivalent for a large cross-country dataset that allows country comparisons of reporting behaviour. As a result, we can quantify the cultural bias in self-reports based on the direct comparison of objective and subjective measures, without using indirect methods such as vignettes. Third, we analyse and compare discrepancies in self-reported data for two health dimensions simultaneously, namely, self-reported physical and cognitive abilities. This allows us to explore whether the two health dimensions are correlated due to similarities in reporting style.

The analysis utilises data from the Survey of Health, Ageing and Retirement in Europe (SHARE), which comprises more than 200,000 observations of adults aged 50 to 94 from 19 European countries. We construct three-category outcome variables that indicate if an individual overestimates his or her health, underestimates his or her health, or achieves concordance between performance-tested and self-reported indicators. Multinomial logit regression allows a clear estimation of the effects of demographic characteristics on reporting behaviour. Then the relative importance of these characteristics for explaining the reporting biases is evaluated by decomposing the regression's fit statistics. Hence, we quantify the contribution of demographic characteristics to the bias in self-reported health based on how much of the variation in concordance these characteristics explain.

Our findings show that misreporting of physical and cognitive health differs substantially between countries and age groups. The large variation in reporting style between age groups can partly be explained by differences in employment status. Educational attainment influences reporting behaviour too, especially when individuals are asked to evaluate their cognitive ability. Men and women also evaluate their health status differently, but these differences are less important in explaining the overall reporting bias. We provide a range of robustness analyses to observe whether our results are sensitive to the definition of physical and cognitive impairment, sample composition and model specifications.

The remainder of this paper is structured as follows. The dataset is introduced in Section 2 with a description of both the self-reported and performance-based variables utilised. Next, the methods used are explained in Section 3. Sections 4 and 5 present our results, which are discussed and compared with previous work in Section 6. Additional estimations along with robustness analyses are provided in S1 Appendix.

Data and variables

The data analysed are provided by SHARE, a cross-country panel study of non-institutionalised individuals aged 50 and older who regularly live in one of the participating European countries [21–25]. The survey was launched in 2004/2005 in 11 European countries with more countries joining in the follow-up waves, resulting in 18 countries participating in 2015 in Wave 6. SHARE was reviewed and approved by the Ethics Committee of the University of Mannheim and the Ethics Council of the Max Planck Society [26].

For our analysis, we require pairs of tested and self-assessed health measures that can be matched directly. SHARE provides two such pairs, namely for mobility and cognition. Since the performance-based test for mobility is conducted in Wave 2 (2006/2007) and Wave 5 (2013) only, we pool these waves to analyse self-reports of physical health [27,28]. Wave 4 (2010–2012) and Wave 5 provide suitable data for the analysis of self-assessed cognitive health [29]. In summary, the analysis is based on pooled cross-sectional data with 88,087 observations from 17 different countries for mobility and 115,785 observations from 17 different countries for cognition.

Outcome variables

We investigate the reporting behaviour in two health dimensions, mobility and cognition, by comparing the results of a performance test and its adequate self-report. The self-reports are requested prior to the respective performance test for mobility and cognition, and thus the test results do not influence the subjective health measures.

We assume that the performance test and its self-report cover the same health dimension. Therefore, we are able to assess whether the two variables coincide, after dichotomising them where necessary (see Subsection 2.1.2). Consequently, three different combinations of objective and subjective health measures are possible for each survey participant in the study. First, respondents achieve concordance if they have the same outcome in both the performance-tested and self-reported variable. Importantly, we do not distinguish between positive agreements (i.e. no impairment according to the test as well as the self-report) and negative agreement (i.e. impairment according to the test as well as the self-report). Second, respondents are considered to be overestimating their health if they report no impairment but are actually impaired according to the performance test. Third, respondents are considered to be underestimating their health if they report impairments but show no impairment during the performance test.

Mobility indicators. Performance-based mobility is measured by a chair stand test conducted in Waves 2 and 5. While all individuals were asked to perform a chair stand test in Wave 5, only individuals aged 75 years or younger were asked to do this test within Wave 2. Because Greece, Ireland, and Poland only participated in Wave 2, concordance of mobility measures can only be observed for the population aged 50–75 in these three countries.

For the mobility performance task, survey participants were asked to stand up from a chair without using their arms. Specifically, the interviewer gave the instruction, "I would like you to fold your arms across your chest and sit so that your feet are on the floor; then stand up keeping your arms folded across your chest. Like this. . .". Following this introduction, survey

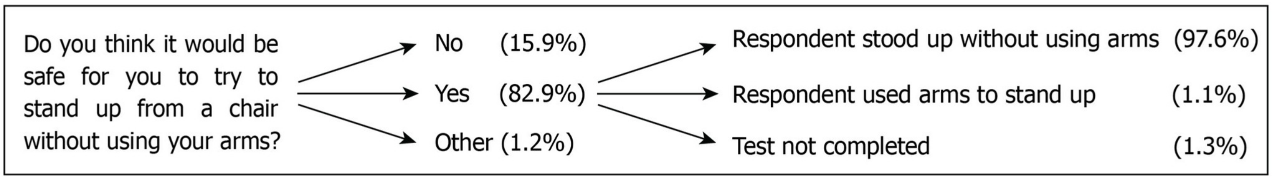


Fig 1. Sequence of questions and proportions of answers ascertaining tested mobility.

https://doi.org/10.1371/journal.pone.0223526.g001

participants were asked whether they thought it would be safe to try standing up from a chair without using their arms (Fig 1 summarises the exact sequence of questions). Everybody completing the performance test successfully is coded as unimpaired, whereas individuals are considered impaired if they did not complete the test or if they thought it was unsafe to try in the first place. Moreover, a small percentage (1.1%) of individuals used their arms to stand up from the chair; this is also considered to be unimpaired. We provide sensitivity analyses in which individuals who thought it was unsafe to perform are excluded from the analysis, and a second set of sensitivity analyses in which individuals using their arms to stand up from the chair are considered as impaired (Tables A and B in S1 Appendix).

The self-reported mobility measure is based on the survey question, "Please tell me whether you have any difficulty doing each of the everyday activities [. . .]. Exclude any difficulties that you expect to last less than three months". Among other everyday activities, survey respondents could choose difficulties in "getting up from a chair after sitting for long periods". Individuals are considered impaired if they reported having difficulties getting up from a chair.

Cognition indicators. Cognition was addressed with a memory test in Waves 4 to 6. Because the self-reported memory item has more than 80% missing values in Wave 6, this study only considers Waves 4 and 5.

Self-reported memory is evaluated with the survey question, "How would you rate your memory at the present time?", which was answered on a Likert scale with categories (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor. Every individual reporting fair or poor memory is considered impaired [30]. The memory performance task reports the ability to immediately recall as many words as possible. The interviewer reads aloud a list of 10 words and asks the survey participant to recall as many of the words as he or she can within 1 minute, in any order. In this study, individuals are considered to be cognitively impaired if they recall only three words or less [31,32]. Additionally, in robustness analyses, individuals are considered impaired if they recall only two or fewer words (Tables C and D in S1 Appendix). Since the subjective memory question might refer to immediate and delayed memory, we conduct an additional sensitivity analysis in which we operationalise objective cognition with delayed word recall (Table E in S1 Appendix).

Determinants of concordance

Scientific studies on health-related questions as well as governmental health reports usually include separate analyses for one or more subpopulations. The subpopulations that are most commonly compared are individuals from different countries, different genders, age groups and educational groups. Often, these analyses are based on self-assessed health data, which is crucial since these demographic characteristics are frequently identified in the literature as important factors of health misreporting [11,13,14,16,17,33,34]. For example, [14] showed that variations in self-assessed health between European countries would be much smaller if all countries had the same reporting behaviour. These disparities are explained by cultural differences in reporting behaviour, different perceptions of how restricting poor health is and compositional differences [11]. It was also shown that older individuals often overestimate their health [35], possibly due to peer effects [36]. Some evidence suggests that women tend to underestimate their health [9], which could be related to them reporting limitations more frequently [37–39]. However, other studies find no effect of gender on reporting behaviour [15]. Finally, evidence on educational attainment shows that highly educated older Europeans are more likely to rate their health state negatively and that consequently, health inequalities appear lower than they actually are [16]. Similar results were found for non-European countries [33].

Based on the observation that demographic characteristics are most commonly used for comparative health studies, and that the same characteristics are associated with deviations in reporting behaviour, this study focuses on the main demographic characteristics only (i.e. country of residence, gender, age and educational attainment). In accordance with the International Standard Classification of Education, education levels are combined into three groups [40]. The group of low education includes everyone with lower secondary education and less. Medium education refers to survey participants with upper secondary or post-secondary non-tertiary education, and tertiary education includes individuals with tertiary education. Age is operationalised as a categorical variable, grouping 5-year age groups. Only participants between the ages 50 and 94 are considered, resulting in a total of nine age groups.

In addition to the main demographic characteristics, other individual factors such as marital status, parenthood or employment status might contribute to or mediate the effect of demographic characteristics on reporting behaviour. For example, employment status might impact health perception since persons working in analytical jobs experience their level of cognition regularly and persons conducting manual labour are likely aware of their mobility impairments. The employment status of older Europeans is highly correlated with their age, since most individuals exit the labour force at a set retirement age. Thus, parts of the effect of age on reporting behaviour might be due to differences in the employment status. Furthermore, employment might also mediate the effect of education on health perception, since highly educated individuals are more likely to work in jobs that require strong cognitive skills. While results for such subordinate channels are not presented in the main document, supplementary analyses including additional determinants are provided in S1 Appendix.

Methods

We first investigate trends with descriptive statistics. Following this, the relationship between demographic characteristics and the probability to overestimate or underestimate health is estimated. Finally, a relative importance analysis highlights the magnitude of each explanatory variable's contribution to the overall reporting bias. The empirical strategy employed is based on a recent study by Angel et al. [41], who analysed the reporting bias in survey-based income data. All of our analyses are first applied to indicators of mobility and then to indicators of cognition.

Multinomial logistic regression

A multinomial logit model is applied to estimate the effects of demographic characteristics on the probability to overestimate or underestimate health. The characteristics of interest are gender, age, education, and country of residence. In addition, we control for the survey wave to account for potential time effects.

The outcome variables used in the regression models are three-category variables that indicate if an individual overestimates his or her health, underestimates his or her health, or achieves concordance between performance-tested and self-reported indicators. Concordance is used as the reference category; hence, the log odds of the variables explaining overestimating and underestimating have to be interpreted relative to the outcome category of concordance. More specifically, the non-linear baseline models are as follows:

$$\ln\left(\frac{P(y = \text{over} - \text{estimating})}{P(y = \text{concordance})}\right) = \beta_{1.0} + \beta_{1.1}\text{COUNTRY}_i + \beta_{1.2}\text{AGE}_i + \beta_{1.3}\text{EDUC}_i + \beta_{1.4}\text{GENDER}_i + \beta_{1.5}\text{WAVE}_i + \varepsilon_i \quad (1)$$

$$\ln\left(\frac{P(y = \text{under} - \text{estimating})}{P(y = \text{concordance})}\right) = \beta_{2.0} + \beta_{2.1}\text{COUNTRY}_i + \beta_{2.2}\text{AGE}_i + \beta_{2.3}\text{EDUC}_i + \beta_{2.4}\text{GENDER}_i + \beta_{2.5}\text{WAVE}_i + \varepsilon_i \quad (2)$$

COUNTRY_i is a dummy variable indicating the country of residence of each individual with the reference country being Slovenia. AGE_i indicates the 5-year age group of individual i with age group 60–64 as the reference category. The binary variable GENDER_i is 1 if the survey participant is female. EDUC_i is a three-category variable, and medium education serves as the reference category. WAVE_i is a dummy variable indicating the respective survey wave. When analysing mobility, the reference category is Wave 2; when analysing memory, the reference category is Wave 4. The standard errors are clustered at the individual level since respondents could participate in more than one wave. First, models 1 and 2 are estimated for the pooled sample including all countries. Then the models are estimated for each country separately to analyse how the effects vary by country. In the country-specific estimations, the wave dummies are only included if the respective country participated in both waves.

Relative importance analysis

To analyse the contribution of individual characteristics to the overall bias in self-reported mobility and cognition, relative importance analysis is conducted. More specifically, the fit statistics of the regression models are decomposed to evaluate how much of the variation in concordance, overestimating, and underestimating is explained by the regressors COUNTRY_i, AGE_i, GENDER_i, EDUC_i, and WAVE_i.

We utilise the user-written programme domin for Stata to calculate the relative contributions [42,43]. For this purpose, different models with all possible combinations of the five explanatory variables except the constant-only model are estimated. The fit statistic, in our case a Pseudo R^2, varies depending on the constellation of the regressors. Based on this variation, the relative contribution of each explanatory variable can be computed. Importantly, only explained variation can be decomposed. Hence, only the contribution of variables actually included in the model can be quantified. We calculate the relative importance of each explanatory variable in the pooled model, as well as in the country-specific models.

Robustness analyses

In addition to the main model specification described above, we provide robustness analyses in S1 Appendix to analyse if the results are sensitive to the definition of physical and cognitive impairment, sample composition and model specifications. First, we control for additional variables to analyse the robustness of the estimated coefficients. In particular, we add employment status, a dummy variable that indicates whether the survey participant has children, and a dummy variable that indicates whether the individual is married or in a registered partnership to the models (Tables J-O in S1 Appendix). Furthermore, education is interacted with gender to determine if the effects of education vary with gender (Tables P and Q in S1 Appendix). We also investigate whether learning effects influence the estimates. That is, when individuals had their mobility or memory tested in a previous wave, they might be more likely to achieve concordance in a subsequent wave. To control for a potential learning effect, dummy variables are added to the model, which indicate if an individual performed a test in any wave prior to the one investigated (Tables R and S in S1 Appendix).

We also analyse whether the results are sensitive to the definition of mobility impairment. In particular, we investigate whether the results change when individuals that have to use their arms to stand up from a chair are considered impaired (Table A in S1 Appendix) or when individuals that refuse standing up from a chair are dropped from the analysis (Tables A and B in S1 Appendix). We also investigate whether the results are robust to different thresholds defining memory impairment (Tables C and D in S1 Appendix). Furthermore, we use delayed word recall instead of immediate word recall to operationalise memory for a sensitivity analysis (Table E in S1 Appendix).

Finally, we investigate if the results are robust to different sample compositions. First, all frail individuals are excluded from the sample [44,45]. This allows us to account for the fact that frail individuals might be more likely to live in institutions in some countries than in other countries and consequently are not always included in our target population. These differences in sample compositions could alter the results, if poor health has an impact on reporting behaviour (Tables F and G in S1 Appendix). Second, we run the models on the exact same sample for both health dimensions. For the main analysis, Wave 2 and Wave 5 are utilised to estimate concordance of mobility measures, and Wave 4 and Wave 5 are utilised to estimate concordance of cognition measures. Since we want to compare the results for concordance of mobility and cognition measures, we also compute estimates based on Wave 5 only, which provides data for both health dimensions. Thus, we ensure that differences between the two samples are not mistakenly interpreted as differences in reporting behaviour (Tables H and I as well as Figs A and B in S1 Appendix).

Results on mobility

Descriptive results

When asked about their mobility, 19.2% of the survey participants report difficulties getting up from a chair after sitting for long periods. However, when tested, only 17.2% are unable to stand up from a chair or considered it unsafe to try. Overall, 80.4% of the survey participants show concordance between their reported and tested mobilities, yet the outcome varies substantially by individual characteristics. Men are more likely to report their actual level of mobility than females, mainly because women tend to more frequently underestimate their health. Interestingly, 12.0% of all women rate their mobility lower than it actually is compared to 7.9% of all men (Table 1).

Concordance strongly declines with age. In the 50–54 age group, 85.5% report their correct level of mobility, but in the 90–94 age group, only 65.6% achieve concordance. Overestimating increases from 7.1% at ages 50–54 to 24.7% at ages 90–94. Underestimating increases less steeply and not linearly from 7.4% to 9.7%. There is also a clear education gradient in reporting behaviour. Highly educated individuals are more likely to achieve concordance (86.3%) than less-educated individuals (76.4%). In addition, the less educated more often overestimate their health, whereas the highly educated more often underestimate their health.

Finally, concordance varies strongly between countries. Overall, it is much higher in Northern and Western European countries than in Southern European countries, Central and Eastern European (CEE) countries, and Ireland. Denmark has the highest average concordance of 87.7%, and Poland has the lowest with only 70.4%. The variation in concordance may stem from differences in overestimating rather than underestimating, as participants from Southern

Table 1. Summary statistics showing heterogeneities in self-reported mobility and cognition.

	Mobility						Cognition					
	Impairment		Concordance				Impairment		Concordance			
	S	T	S = T	S>T	S<T		S	T	S = T	S > T	S < T	
	%	%	%	%	%	N	%	%	%	%	%	N
Total	19.2	17.2	80.4	9.4	10.2	88,087	29.4	16.1	71.8	7.5	20.7	115,785
Gender												
Men	14.9	15.2	82.8	9.3	7.9	39,417	28.1	17	72.3	8.3	19.3	51,013
Women	22.7	18.8	78.4	9.6	12.0	48,670	30.4	15.3	71.4	6.8	21.8	64,772
Age												
50–54	10.3	10.0	85.5	7.1	7.4	11,229	17.6	6.3	80.6	4.0	15.4	13,244
55–59	12.7	11.6	83.9	7.5	8.5	16,196	20.5	7.1	77.9	4.3	17.7	19,461
60–64	14.9	12.5	82.3	7.6	10.0	16,836	22.9	8.7	75.4	5.2	19.4	21,098
65–69	16.6	14.7	80.2	9.0	10.8	15,721	26.5	11.3	72.9	6.0	21.1	19,447
70–74	20.7	19.5	78.0	10.5	11.5	12,906	33.8	17.0	66.9	8.2	24.9	16,180
75–79	26.9	25.0	75.8	11.7	12.5	7,347	42.0	27.6	62.2	11.8	26.0	12,350
80–84	34.4	36.7	71.4	15.9	12.7	4,664	48.5	39.3	61.4	14.9	23.7	8,525
85–89	42.6	49.8	69.1	19.5	11.4	2,438	52.3	50.0	63.5	17.4	19.1	4,283
90–94	46.9	60.2	65.6	24.7	9.7	750	53.2	55.0	63.9	19.5	16.5	1,197
Education												
Low	24.7	23.6	76.4	12.2	11.4	35,808	39.7	27.4	64.8	11.6	23.6	46,113
Medium	16.9	14.4	81.4	8.4	10.3	31,953	24.8	9.6	74.4	5.2	20.4	43,362
High	11.8	9.5	86.3	6.0	7.7	19,058	17.7	5.7	80.7	3.7	15.6	24,337
Country												
Austria	20.8	17.9	80.1	9.0	11.0	5,032	17.8	11.6	80.8	6.4	12.8	9,028
Belgium	19.5	14.1	80.8	7.4	11.9	7,932	24.4	13.5	73.8	7.7	18.5	10,511
Czechia	23.2	21.3	78.1	10.6	11.2	7,651	30.0	11.6	71.8	5.0	23.2	10,609
Denmark	12.7	7.6	87.7	4.2	8.1	6,014	17.3	9.0	81.3	5.2	13.5	6,171
Estonia	29.1	26.3	76.6	10.3	13.1	5,454	51.4	16.5	56.2	4.4	39.4	11,792
France	16.3	17.2	79.9	11.0	9.0	6,566	31.9	17.6	68.4	8.6	23.0	9,796
Germany	19.6	13.8	80.3	7.5	12.1	7,700	22.4	10.1	76.3	5.7	17.9	7,099
Greece	18.1	18.7	78.6	13.6	7.8	2,601	.	.	.	.	.	.
Hungary	.	.	.	.	.	.	34.2	17.2	67.8	7.6	24.6	2,938
Ireland	18.0	20.1	78.3	13.6	8.1	792	.	.	.	.	.	.
Italy	19.4	24.1	76.1	15.0	8.9	6,919	32.9	22.7	69.6	10.3	20.1	7,895
Luxembourg	21.2	16.1	78.8	8.3	12.9	1,561	18.5	15.5	77.4	9.9	12.6	1,543
Netherlands	14.7	10.1	85.8	5.1	9.1	6,258	15.7	10.8	80.7	7.2	12.1	6,770
Poland	29.5	29.3	70.4	17.0	12.6	1,969	32.8	24.4	69.0	11.1	19.9	1,678
Portugal	.	.	.	.	.	.	45.4	29.3	61.6	11.1	27.3	1,899
Slovenia	20.9	19.5	77.9	10.5	11.6	2,873	26.9	20.4	71.8	11.0	17.2	5,511
Spain	21.8	24.4	78.3	13.3	8.4	8,011	41.1	34.0	67.0	12.9	20.1	9,628
Sweden	15.4	10.9	83.7	6.5	9.8	6,611	29.3	12.2	71.0	6.2	22.9	6,346
Switzerland	11.2	9.3	85.6	6.6	7.9	4,143	16.5	8.2	81.6	5.2	13.3	6,571
Wave												
Wave 2	18.6	16.6	79.8	10.9	9.2	26,973	.	.	.	.	.	.
Wave 4	.	.	.	.	.	.	29.4	16.9	71.6	7.9	20.5	55,172
Wave 5	19.5	17.4	80.6	8.8	10.6	61,114	29.4	15.3	72.0	7.1	20.9	60,613

Note: S refers to self-reported impairment and T refers to tested impairment. S = T denotes concordance, S>T denotes overestimating, and S<T denotes underestimating. N = 100%

https://doi.org/10.1371/journal.pone.0223526.t001

and CEE countries as well as Ireland tend to strongly overestimate their mobility. Furthermore, all Southern countries are less likely to underestimate their ability to stand up from a chair.

Regression analysis

Most findings from the descriptive analysis are confirmed by regression analyses for both the pooled sample with all countries as well as the country-specific samples (Table 2). When estimating Models 1 and 2 for the pooled sample, the coefficients show a drastic decline of concordance with age. Individuals aged 80–84 are 2.7 times more likely to overestimate their mobility than 60- to 64-year-olds (log odds 0.976). Participants aged 90–94 are 4.4 times more likely to overestimate than the reference group (log odds 1.489). The tendency to underestimate mobility also increases with age, but less strongly than the tendency to overestimate. Furthermore, underestimating peaks at ages 80–84, but decreases again for the oldest individuals. For a better overview, S1 Fig provides the predicted values of concordance based on the country-specific estimations by age group. When employment is added to the model, the age gradient in concordance remains, but appears less steep. This finding indicates that parts of the strong age effect are due to difference in the employments status between age groups (Table J in S1 Appendix).

Women are 1.4 times more likely to underestimate their mobility than men (log odds 0.301); in regard to overestimating, the gender effects are small (log odds 0.054) and appear insignificant once we control for employment, marriage or an interaction effect between education and gender (Tables J, N and P in S1 Appendix) as well as once participants that felt unsafe are excluded from the sample (Table B in S1 Appendix).

Similar to the descriptive results, the regression results indicate a clear education gradient in concordance. Less-educated participants are 1.2 times more likely to overestimate their mobility (log odds 0.182) and also 1.2 times more likely to underestimate their mobility (log odds 0.163) compared to individuals in the medium education group. On the contrary, participants with a tertiary education have a lower tendency to both overestimate (log odds -0.287) and underestimate mobility (log odds -0.299). There is also an interaction between gender and education, where less-educated women in particular are prone to underestimating their ability to stand up from a chair (Table P in S1 Appendix). Similarly to age, the education gradient in concordance appears less steep once employment is controlled for, which supports the hypothesis that parts of the education effect are due to educational differences in employment (Table J in S1 Appendix).

Fig 2 presents the rates of concordance, overestimating, and underestimating by country. Overall, there is a tendency for higher concordance in Western and Northern European countries. By contrast, individuals in Southern European countries, CEE countries, and Ireland are less likely to achieve concordance, mainly because they tend to more often overestimate their mobility. The tendency to underestimate mobility is more evenly distributed among countries, yet there are still differences. For example, Southern Europeans underestimate their health less often.

Finally, the coefficient for the survey waves indicates that survey participants are less likely to overestimate their mobility in 2013 compared to 2006/2007 (log odds -0.414). The coefficient decreases after controlling for potential learning effects, but still remains significant (Table R in S1 Appendix). This could be due to cohort effects, but it is not possible to disentangle cohort effects from period effects using the present dataset. A second explanation for the significant time effects could be that some countries changed their interview procedure between the two survey waves.

Table 2. Multinomial logistic estimation for concordance of mobility measures.

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.195*	0.080	-0.050	0.076
Belgium	-0.422***	0.077	0.083	0.071
Czechia	-0.061	0.074	-0.053	0.071
Denmark	-0.966***	0.092	-0.307***	0.079
Estonia	-0.031	0.077	0.111	0.072
France	-0.085	0.075	-0.249***	0.075
Germany	-0.299***	0.076	0.159*	0.070
Greece	0.045	0.089	-0.302**	0.098
Ireland	0.164	0.125	-0.156	0.148
Italy	0.219**	0.072	-0.280***	0.075
Luxembourg	-0.195	0.112	0.150	0.097
Netherlands	-0.864***	0.087	-0.285***	0.076
Poland	0.395***	0.092	0.303**	0.095
Spain	0.034	0.072	-0.402***	0.074
Sweden	-0.636***	0.082	-0.195**	0.074
Switzerland	-0.607***	0.090	-0.432***	0.085
Age (Ref: 60–64)				
50–54	-0.134**	0.048	-0.356***	0.045
55–59	-0.048	0.042	-0.179***	0.038
65–69	0.193***	0.041	0.099**	0.036
70–74	0.334***	0.042	0.156***	0.039
75–79	0.569***	0.049	0.245***	0.045
80–84	0.976***	0.053	0.301***	0.054
85–89	1.199***	0.063	0.206**	0.072
90–94	1.489***	0.096	0.092	0.132
Women	0.054*	0.024	0.458***	0.024
Education (Ref: Medium)				
Low	0.182***	0.030	0.163***	0.028
High	-0.289***	0.038	-0.299***	0.035
Wave 5	-0.414***	0.030	0.028	0.029
Constant	-1.965***	0.075	-2.269***	0.072
N	86,819	Pseudo R^2		0.033

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level,

*p<0.05,

**p<0.01,

***p<0.001

https://doi.org/10.1371/journal.pone.0223526.t002

When estimating models 1 and 2 for the country-specific samples, the results from the pooled model are confirmed. However, standard errors are larger due to the smaller sample sizes, leading to less significant results. The output tables for the country-specific estimations can be provided upon request. Furthermore, the results are robust to different specifications of impaired mobility (Tables A and B in S1 Appendix) as well as to different sample compositions (Tables F and H as well as Figs A and B in S1 Appendix).

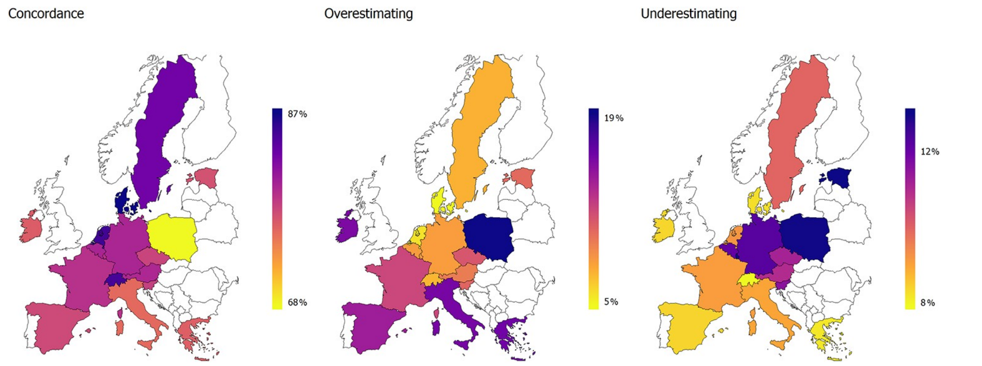


Fig 2. Concordance between tested and self-reported mobility by country (predicted shares).

https://doi.org/10.1371/journal.pone.0223526.g002

Relative importance analysis

Relative importance analysis for the pooled model shows that most of the bias in self-reported mobility stems from differences in reporting behaviour by country and age. Country differences in reporting behaviour contribute 35.0% of the explained variance in concordance, overestimating, and underestimating. Differences between age groups explain 32.1% of the bias. Together, country and age explain more than two-thirds of the variance. Reporting heterogeneity by education contribute another 17.1%, and differences by gender contribute only 11.3%. Differences by survey waves (4.6%) contribute only nominally. When employment is added to the analysis, age and education explain relatively less of the variation, which indicates again that parts of the strong age and education effects are due to differences in employment status. For additional robustness analyses, please consult S1 Appendix.

Fig 3 shows the results of the relative importance analysis for each country individually. Because Estonia, Greece, Ireland, Luxembourg, Poland, and Slovenia only participated in one

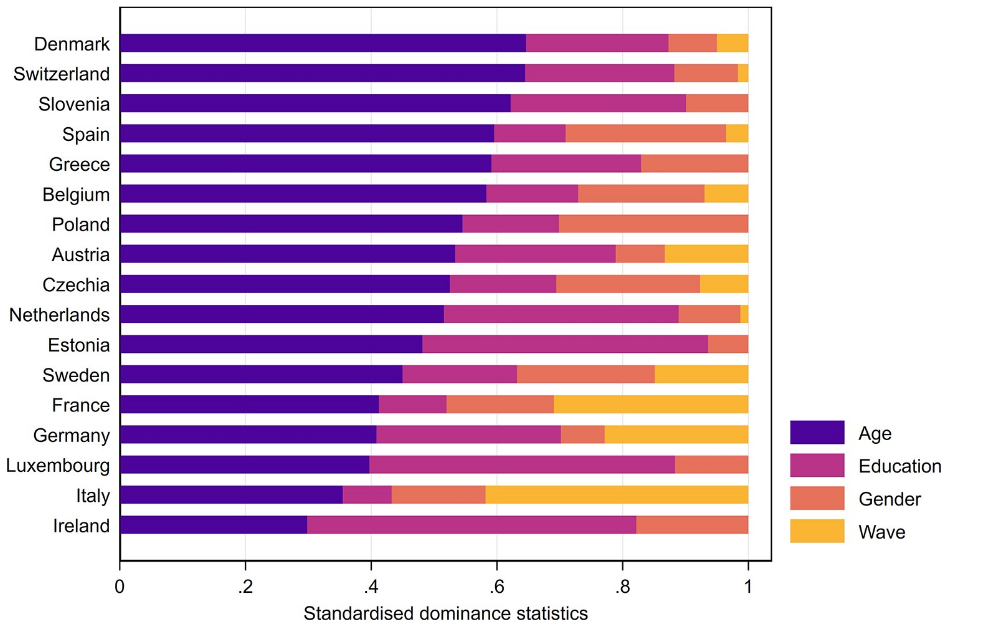


Fig 3. Decomposition of the overall bias in self-reported mobility.

https://doi.org/10.1371/journal.pone.0223526.g003

survey wave, the estimates of time effects for these countries are not provided. For the majority of the countries, age is the single most important characteristic explaining the bias of self-reported health. Depending on the country, either education or gender comes second. The contribution of time effects is negligible in most countries, except for France, Germany, and Italy. As discussed earlier, these time effects could be due to unobserved cohort effects, or because these countries changed their interview process between Wave 2 and Wave 5.

Results on cognition

Descriptive results

When asked about their memory, 29.4% of all survey participants report cognitive impairment (Table 1), yet when tested, only 16.1% recall three words or less. Overall, 71.8% of the participants show concordance between their reported and tested memories, but there is no clear difference between genders except for a slight tendency for men to overestimate and for women to underestimate their cognition. Concordance between mobility and cognition measures is highly related. According to Chi-squared tests, individuals that are prone to overestimate one dimension are also more likely overestimate the other; the same holds for underestimating and concordance.

Similar to mobility, there is a strong decline in concordance with age. While 80.6% of the 50–54 age group report their correct level of memory, only 63.9% of the 90–94 age group achieve concordance. Misreporting is even more pronounced at ages 80–84, in which 61.4% show divergence between tested and self-reported measures. Unlike mobility, it is not clear from the numbers whether the decrease in concordance with age is due to an increase in overestimating or underestimating. While the tendency to overestimate cognition increases steadily with age, under-estimating is highest at ages 75–79 (26.0%) and decreases thereafter.

There is a pronounced education gradient in the concordance between tested and self-reported cognition, where again Western and Northern countries have lower discrepancies. Switzerland has the highest rate of concordance (81.6%) and Estonia has the lowest (56.2%). However, the division is not as clear as for mobility, mainly because Sweden has a relatively low rate of concordance (71.0%), similar to that of Slovenia and Czechia.

Regression analysis

Regression analyses also show concordance decreasing strongly with age (Table 3). Individuals aged 80–84 are three times as likely to overestimate their memory than the reference group of 60- to 64-year-olds (log odds 1.095). The oldest individuals, aged 90–94, are 3.7 times as likely to overestimate their cognitive ability (log odds 1.297). Similar to mobility, the probability to underestimate memory increases up to ages 75–79 (log odds 0.386), but slightly decreases again for the oldest individuals. Based on the country specific samples, S2 Fig provides the values of concordance by country and age. Contrary to mobility, the strong age gradient in concordance does not change once employment is controlled for (Table K in S1 Appendix).

The effect of education on concordance is even stronger for cognition than it is for mobility. Less-educated participants are 1.9 times more likely to overestimate their memory (log odds 0.644) and 1.3 times more likely to underestimate their memory (log odds 0.240). Tertiary education is associated with a lower probability to both overestimate (log odds -0.445) and underestimate cognition (log odds -0.308). These results remain robust even after controlling for employment (Table K in S1 Appendix).

Contrary to mobility, women are less likely to overestimate their memory than men (log odds -0.290). However, females are slightly more likely to underestimate their cognition in the pooled model. In the country-specific estimations, this finding holds for Belgium, Estonia,

Table 3. Multinomial logistic estimation for concordance between cognition measures.

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.613***	0.066	-0.386***	0.053
Belgium	-0.392***	0.062	0.090	0.049
Czechia	-0.854***	0.066	0.251***	0.047
Denmark	-0.654***	0.076	-0.264***	0.058
Estonia	-0.690***	0.067	1.075***	0.045
France	-0.339***	0.061	0.332***	0.048
Germany	-0.473***	0.071	0.029	0.052
Hungary	-0.287***	0.086	0.495***	0.059
Italy	-0.325***	0.062	0.036	0.051
Luxembourg	-0.124	0.100	-0.429***	0.087
Netherlands	-0.622***	0.069	-0.499***	0.058
Poland	-0.072	0.098	0.201**	0.077
Portugal	-0.133	0.093	0.583***	0.068
Spain	-0.165**	0.059	0.058	0.049
Sweden	-0.686***	0.073	0.235***	0.051
Switzerland	-0.822***	0.076	-0.365***	0.058
Age (Ref: 60–64)				
50–54	-0.258***	0.056	-0.247***	0.032
55–59	-0.196***	0.049	-0.113***	0.027
65–69	0.162***	0.045	0.111***	0.026
70–74	0.526***	0.044	0.321***	0.028
75–79	0.885***	0.045	0.386***	0.030
80–84	1.095***	0.047	0.288***	0.035
85–89	1.182***	0.056	0.032	0.048
90–94	1.297***	0.085	-0.099	0.089
Women	-0.290***	0.025	0.091***	0.017
Education (Ref: Medium)				
Low	0.644***	0.031	0.240***	0.020
High	-0.445***	0.043	-0.308***	0.024
Wave 5	-0.127***	0.024	0.116***	0.015
Constant	-2.202***	0.059	-1.653***	0.046
N	113,812	Pseudo R^2		0.055

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level,
*p<0.05,
**p<0.01,
***p<0.001

https://doi.org/10.1371/journal.pone.0223526.t003

France, Italy, Portugal, and Spain. However, in Austria, Denmark, and The Netherlands, women are less likely to underestimate their memory. The gender differences increase when memory impairment is based on delayed word recall, which indicates that women and men either interpret the subjective memory question differently, or relationship between immediate and delayed word recall differs between genders (Table E in S1 Appendix).

Concordance between tested and self-reported cognition differs among the countries observed. Again, Southern European and CEE countries have lower rates of concordance than

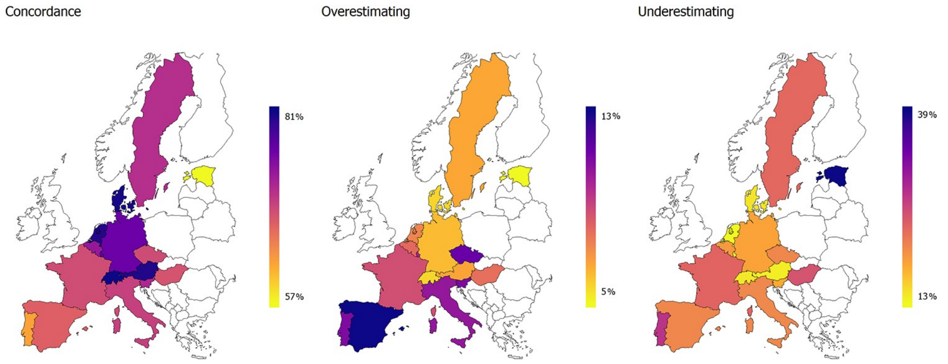


Fig 4. Concordance between tested and self-reported cognition by country (predicted shares).

https://doi.org/10.1371/journal.pone.0223526.g004

Western and Northern European countries (Fig 4). Two exceptions are Czech Republic, which achieves a relatively high rate of concordance, and Sweden, which achieves a medium level of concordance. As with mobility, the tendency to overestimate cognitive ability is much greater in Southern and CEE countries.

Interestingly, participants of Wave 5 are less likely to overestimate and instead more likely to underestimate. This finding does not change when additionally controlling for a potential learning effect (Table S in S1 Appendix). As with mobility, this could indicate a cohort and/or time effect or differences in the interview procedure over time, both of which the available data cannot account for. Finally, all results are robust to changes in the threshold of cognitive impairment (Tables C and D in S1 Appendix), to differences in the sample composition (Tables G and I in S1 Appendix) as well as to different model specifications (Tables K, M, O and Q in S1 Appendix).

Relative importance analysis

The bias in self-reported cognition is mainly due to differences in reporting behaviour by country, which explain 44.9 per cent in the pooled model. Differences by age group contribute 29.7 per cent to the explained variation. Education is much more relevant in explaining the reporting bias in self-reported cognition (22.7 per cent) than it is for measures of mobility. Variations in reporting behaviour by gender (2.1 per cent) and survey wave (0.6 per cent) are even less important for self-reported memory than they are for self-reported mobility. This finding holds also when estimates are based on Wave 5 only (Tables H and I as well as Figs A and B in S1 Appendix).

Fig 5 shows country specific decompositions of the fit statistic. Age is still very relevant for explaining the reporting bias in cognition measures, yet education is just as important in some countries. On the contrary, gender and wave are neglectable when it comes to explaining the reporting bias. Two exceptions are Estonia and Austria, where the survey wave seems to contribute to the explained variance. Similar to the results on mobility, these exceptions could either be due to cohort effects, or because interviews were conducted differently in Wave 4 and Wave5.

Discussion

In this study on older Europeans, we investigate the discrepancy between tested and self-reported health measures and explore which demographic characteristics are most important in explaining health misreporting. In particular, we focus on the demographic characteristics

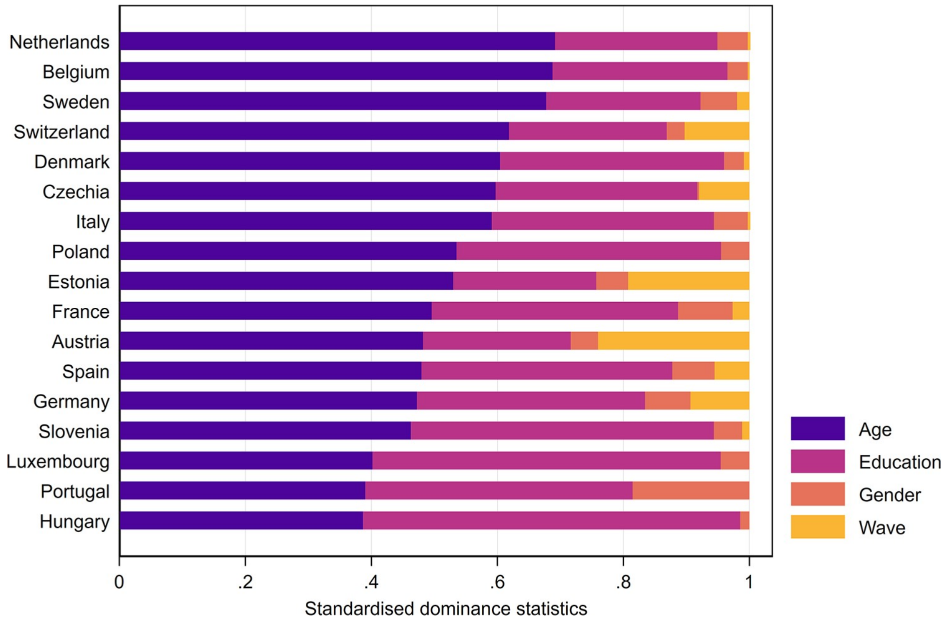


Fig 5. Decomposition of the overall bias in self-reported cognition.

https://doi.org/10.1371/journal.pone.0223526.g005

most frequently used for health comparisons, namely country of residence, gender, age and educational attainment. Furthermore, we investigate subordinate channels that might explain or mediate the effect of demographic characteristics on reporting behaviour, particularly employment status, parenthood and marital status. Conducting a relative importance analysis, we find that differences in reporting style between countries and age groups explain most of the bias in self-reported health. These findings suggest that comparisons of health between countries and age groups based on subjective data have to be treated particularly careful. In addition, for self-reported cognition specifically, misreporting varies substantially between educational groups. Parts of the strong age and education effects on reporting style can be explained by differences in employment by age and education. Parenthood and being married, however, add little to the bias. Sensitivity analyses show that the results are robust to changes in the definition of physical and cognitive impairment, sample composition and model specifications (S1 Appendix).

Concordance as well as the tendency to overestimate and underestimate health vary strongly across Europe. Results from the relative importance analyses show that 35% of the reporting bias in mobility and 45% of the bias in memory are due to differences in reporting behaviour between countries. Overall, Northern and Western European countries have fewer discrepancies than CEE or Southern European countries. Southern Europeans seem particularly prone to overestimating their health, which is contrary to the results of [14], who finds that Scandinavians overrate their health the most. Previous studies also identified country differences in reporting style for European countries [14,46,47], low- and middle-income countries [4], as well as within countries and across subpopulations [5]. It was shown that self-reports are influenced by culture-specific reporting behaviour, compositional differences between countries and differences in the perception of how restricting poor health is [11]. In addition, the strong country effects could also be due to different health care policies. For instance, the proportion of elderly persons in residential care varies across Europe, thus frail persons might be sampled differently across countries. If frailty affects response behaviour, different shares of frail individuals in the country samples could explain differences in aggregated concordance. We controlled for this possibility by excluding all frail individuals from the analysis, yet the results remained robust (Tables F and G in S1 Appendix). Speculatively, the between-country discrepancies could also be due to differences in regional development. For a subset of our country sample, early results on the relationship between a regional developmental index [48] and discrepancies in mobility suggest that countries with better living conditions show more concordance than their counterparts. However, further research with data on the whole lifecycle is needed to investigate the potential development effect properly.

In addition to the cultural bias in self-reported data, we find a strong decrease in concordance with age for both health dimensions. This result is in accordance with earlier research on several physical performance measures [6–8]. Further, previous research supports our finding that subjective health measures of older individuals are often upward biased [35]. One explanation could be that octogenarians and nonagenarians tend to compare their health status with peers suffering from worse health, which enables them to maintain a positive perception of their own health state [36]. This so called downward comparison makes older persons feel more satisfied with their lives, especially, when they are frail themselves [49]. Resilience strategies like these help individuals to flexibly adapt to changes of their physical and cognitive health while maintaining a positive self-image [50].

Overall, the age-related decline in concordance between performance based and perceived memory measures is robust to controlling for employment (Table K in S1 Appendix). However, concordance between mobility measures declines less steeply with age once the employment status is considered. This indicates that a part of the strong age effect is due to variation in the share of employed persons across age groups. The causal direction, however, remains unclear. It could either be that employed individuals are more aware of their physical ability, or that persons that are more aware of their own health status are more likely to be employed. Thus, future studies could fruitfully explore the interrelations between health perception, age and employment.

We also identify a clear education gradient in concordance for mobility and an even stronger effect for cognition. Less-educated individuals tend to misreport their mobility and memory more frequently, whereas the highly educated are less likely to misreport. Previous research does not provide conclusive results on this matter. Some studies report that higher education results in a more optimistic view on health [8], while others find the exact opposite [33,51,52] or no significant education effect at all [53,54]. Overall, our results on education can be interpreted as additional evidence for the phenomenon that higher educated individuals have higher health awareness and literacy [55,56]. For example, higher educated are more familiar with the risks of tobacco smoking [57], less likely to misjudge their weight [58] and, as shown in this study, also less likely to have a biased view on their physical and cognitive abilities. Since health literacy is an important determinant of health behaviour and consequently health itself [59–61], enhancing health literacy of low educated individuals could improve their health outcomes. It may also be hypothesised that the gender gap in the education of older Europeans contributes to differences in misreporting. On average, men at advanced age are higher educated than women within our investigated cohorts. What supports this hypothesis is our finding that less-educated women are particularly prone to underestimate their mobility (Table P in S1 Appendix). In addition, employment status at higher ages varies by gender and education with higher educated being more likely to work longer [62]. Our robustness analyses showed that the education gradient in concordance appears less pronounced for mobility once employment is accounted for, but interestingly does not change for cognition (Tables J and K in S1 Appendix). The educational differences in cognition only changed when delayed word recall is used and education is less important to explain the differences (Table O in S1 Appendix).

We also find differences in reporting behaviour between men and women, but they are less pronounced and explain very little of the overall reporting bias. In particular, women tend to underestimate their health more frequently in both health dimensions. One explanation for these gender differences might be the tendency of women to report limitations more frequently [37–39], while men tend to underreport their health status [63]. Recent research also showed that reporting morbidity was more legitimate in female-dominated work environments, indicating an association of gender norms with gender difference in reporting behaviour [39]. This might also be related to women looking for medical advice more often than men [64,65]. Interestingly, our findings on overestimating health vary by health dimension with women being less likely to overestimate their memory than men, but being more likely to overestimate their mobility. Moreover, difference between genders increases when delayed word recall is instead of immediate word recall, which indicates that women and men might interpret the subjective memory question differently. Our small and sometimes ambiguous gender effects are in line with the literature, which does not provide conclusive results either. While some studies comparing self-assessed and clinical data find clear evidence that women are more likely to overestimate their health [66], others identify women to be more likely to underestimate their health [67,68]. A recent study based on SHARE data found no clear gender-specific pattern in reporting behaviour [15].

In general, our results not only give guidance on how to carefully interpret self-reported health measures, but might also contribute to a reduction in adverse health outcomes due to mistaken self-assessments. For instance, overestimating lower body functioning might contribute to higher risks of fall-induced injuries [10]. Further, overestimating cognitive abilities might result in illusory self-awareness of everyday functioning [69]. In psychology, the consequences of wrong self-awareness of cognitive abilities are discussed as the Dunning-Kruger effect, which states that unable individuals are especially prone to overestimate their abilities [70,71]. If the tendency to overestimate ones physical and cognitive capacity has an adverse impact on health-related behaviour of older Europeans, then awareness should be created in particular among the oldest old, among men and among Southern Europeans.

A major contribution to the literature is that we are able to compare reporting behaviour of mobility and cognition simultaneously. The results show that concordance of the two health dimensions is highly related. Individuals that are prone to misreport one dimension are also more likely to misreport the other. This indicates that correlations between the two health dimensions are, to a certain degree, due to similarities in reporting behaviour. However, we also find differences in the reporting styles of subjective physical and cognitive health. For instance, concordance is slightly higher between mobility measures than between memory measures. Furthermore, the composition of the bias in self-reports differs between the two health dimensions. The cultural bias in subjective data, i.e. differences across countries, is more relevant for cognition than for mobility. Additionally, reporting heterogeneities between education groups result in larger biases in self-reported memory than in self-reported mobility. Gender, however, explains relatively little of the bias in both health dimensions.

Controlling for wave effects shows that participants in Wave 5 are less likely to overestimate their mobility as well as their cognition, even after controlling for potential learning effects. These findings indicate that cohort or time effects influence the reporting style, which is crucial since the analysis of mobility and memory are based on different waves. To ensure that the differences in reporting style of physical and cognitive health do not stem from differences in the sample composition, we conducted a robustness analysis for which we restricted our analysis to Wave 5, which is the only wave that provides relevant data for mobility and memory. Tables H and I in S1 Appendix show that the overall findings remain even after both health dimensions are analysed based on the same subsample.

The main limitations of this study are threefold. First, the population composition is likely to vary across countries. We conducted robustness analyses for different sample shares of frail individuals, but additional deviations in the sample composition could also influence the results. Second, the questionnaire is conducted in the national language, which could result in some bias when it comes to self-assessed health because the wording differs across languages. Third, it appears that some of the effects are influenced by time or cohort effects, however, disentangling these effects is not feasible with the data at hands.

In conclusion, self-reported measures of mobility and cognition have to be treated cautiously, in particular when comparing health across countries and age groups. In addition, the education gradient in concordance needs to be considered when analysing memory. Finally, men and women show different reporting behaviours, yet the impact of gender on the overall bias between tested and self-reported health is less pronounced than that of other demographic characteristics.

Supporting information

S1 Appendix. Robustness analyses.
(PDF)

S1 Fig. Predicted values of concordance between tested and self-reported mobility by country and age.
(TIF)

S2 Fig. Predicted values of concordance between tested and self-reported cognition by country and age.
(TIF)

Acknowledgments

We are grateful to Mujaheed Shaikh, Bernhard Hammer and the participants of various conferences and seminars for their helpful comments. Furthermore, we thank the editor and three anonymous reviewers whose suggestions helped improving the quality of this paper. This work uses data from SHARE Wave 2 (DOI: 10.6103/SHARE.w2.611), Wave 4 (DOI: 10.6103/SHARE.w4.611), and Wave 5 (DOI: 10.6103/SHARE.w5.611). The SHARE data collection has primarily been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04–064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

Author Contributions

Conceptualization: Sonja Spitzer, Daniela Weber.

Data curation: Sonja Spitzer.

Formal analysis: Sonja Spitzer.

Methodology: Sonja Spitzer, Daniela Weber.

Validation: Daniela Weber.

Visualization: Sonja Spitzer, Daniela Weber.

Writing – original draft: Sonja Spitzer.

Writing – review & editing: Daniela Weber.

References

1. Coman L, Richardson J. Relationship between self-report and performance measures of function: a systematic review. Can J aging. 2006; 25(3):253–70. PMID: 17001588
2. Fors S, Thorslund M, Parker MG. Do actions speak louder than words? Self-assessed and performance-based measures of physical and visual function among old people. Eur J Ageing. 2006; 3(1):15–21. https://doi.org/10.1007/s10433-006-0021-5 PMID: 28794746
3. Beaudoin M, Desrichard O. Are memory self-efficacy and memory performance related? A meta-analysis. Psychol Bull. 2011 Mar; 137(2):211–41. https://doi.org/10.1037/a0022106 PMID: 21244133
4. Capistrant BD, Glymour MM, Berkman LF. Assessing mobility difficulties for cross-national comparisons: Results from the WHO study on AGEing and Adult Health. J Am Geriatr Soc. 2014; 62(2):329–35. https://doi.org/10.1111/jgs.12633 PMID: 24438515
5. Jackson JD, Rentz DM, Aghjayan SL, Buckley RF, Meneide TF, Sperling RA, et al. Subjective cognitive concerns are associated with objective memory performance in Caucasian but not African-American persons. Age Ageing. 2017; 46(6):988–93. https://doi.org/10.1093/ageing/afx077 PMID: 29088363
6. Grant LH, Ward MM. Association of measured physical performance and demographic and health characteristics with self-reported physical function: implications for the interpretation of self-reported limitations. Health Qual Life Outcomes. 2010; 8(1):84.
7. Crossley TF, Kennedy S. The reliability of self-assessed health status. J Health Econ. 2001; 21:643–58.
8. Huang Z, Maurer J. Validity of self-rated memory among middle-aged and older chinese adults: Results from the China Health and Retirement Longitudinal Study (CHARLS). Assessment. 2017;1–12.
9. Merrill SS, Seeman TE, Kasl S V., Berkman LF. Gender differences in the comparison of self-reported disability and performance measures. Journals Gerontol—Ser A Biol Sci Med Sci. 1997; 52(1):19–26.
10. Sakurai R, Fujiwara Y, Ishihara M, Higuchi T, Uchida H, Imanaka K. Age-related self-overestimation of step-over ability in healthy older adults and its relationship to fall risk. BMC Geriatr. 2013; 13(44):1–9.
11. Hardy MA, Acciai F, Reyes AM. How health conditions translate into self-ratings: A comparative study of older adults across Europe. J Health Soc Behav. 2014; 55(3):320–41. https://doi.org/10.1177/0022146514541446 PMID: 25138200
12. Verropoulou G. Determinants of change in self-rated health among older adults in Europe: A longitudinal perspective based on SHARE data. Eur J Ageing. 2012; 9(4):305–18. https://doi.org/10.1007/s10433-012-0238-4 PMID: 28804430
13. Schneider U, Pfarr C, Schneider BS, Ulrich V. I feel good! Gender differences and reporting heterogeneity in self-assessed health. Eur J Heal Econ. 2012; 13(3):251–65.
14. Jürges H. True health vs response style: exploring cross-country differences in self-reported health. Health Econ. 2007; 16:163–78. https://doi.org/10.1002/hec.1134 PMID: 16941555
15. Oksuzyan A, Dańko MJ, Caputo J, Jasilionis D, Shkolnikov VM. Is the story about sensitive women and stoical men true? Gender differences in health after adjustment for reporting behavior. Soc Sci Med. 2019; 228(March):41–50.
16. Bago d'Uva T, Van Doorslaer E, Lindeboom M, O'Donnell O. Does reporting heterogeneity bias the measurement of health diesparities? Health Econ. 2008; 17:351–75. https://doi.org/10.1002/hec.1269 PMID: 17701960
17. Peracchi F, Rossetti C. Heterogeneity in health responses and anchoring vignettes. Empir Econ. 2012; 42(2):513–38.
18. Voňková H, Hullegie P. Is the anchoring vignette method sensitive to the domain and choice of the vignette? J R Stat Soc Ser A Stat Soc. 2011; 174(3):597–620.
19. Furnham A. Self-estimates of intelligence : culture and gender difference in self and other estimates of both general (g) and multiple intelligences. Pers Individ Dif. 2001; 31:1381–405.
20. Mabe PA, West SG. Validity of self-evaluation of ability: A review and meta-analysis. J Appl Psychol. 1982; 67(3):280–96.

21. Börsch-Supan A, Brandt M, Litwin H, Weber G, editors. Active ageing and solidarity between generations in Europe: First results from SHARE after the economic crisis. Berlin: De Gruyter; 2013.
22. Malter F, Börsch-Supan A, editors. SHARE Wave 4: Innovations and Methodology. Munich: Max Planck Institute for Social Law and Social Policy; 2013.
23. Börsch-Supan A, Brandt M, Hunkler C, Kneip T, Korbmacher J, Malter F, et al. Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). Int J Epidemiol. 2013; 42(4):992–1001. https://doi.org/10.1093/ije/dyt088 PMID: 23778574
24. Malter F, Börsch-Supan A. SHARE Wave 5: Innovations & Methodology. Munich: MEA, Max Planck Institute for Social Law and Social Policy; 2015.
25. Börsch-Supan A, Brugiavini A, Jürges H, Kapteyn A, Mackenbach J, Siegrist J, et al. First results from the Survey of Health, Ageing and Retirement in Europe (2004–2007). Starting the longitudinal dimension. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA); 2008.
26. Ethics Council of the Max Planck Society. Opinion of the Ethics Council of the Max Planck Society on the "SHARE" project [Internet]. 2018. www.share-project.org/fileadmin/pdf_documentation/MPG_Ethics_Council_SHARE_overall_approval_14.06.2018__en_.pdf
27. Börsch-Supan A. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 6.1.1. SHARE-ERIC. Data set. [Internet]. 2018.
28. Börsch-Supan A. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 6.1.1. SHARE-ERIC. Data set. 2018
29. Börsch-Supan A. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 6.1.1. SHARE-ERIC. Data set. [Internet]. 2018.
30. Gardner RC, Langa KM, Yaffe K. Subjective and objective cognitive function among older adults with a history of traumatic brain injury: A population-based cohort study. PLoS Med. 2017; 14(3):1–16.
31. Grodstein F, Chen J, Wilson RS, Manson JE. Type 2 diabetes and cognitive function in community-dwelling elderly women. Diabetes Care. 2001; 24(6):1060–5. https://doi.org/10.2337/diacare.24.6.1060 PMID: 11375371
32. Purser JL, Fillenbaum GG, Pieper CF, Wallace RB. Mild cognitive impairment and 10-year trajectories of disability in the Iowa established populations for epidemiologic studies of the elderly cohort. J Am Geriatr Soc. 2005; 53(11):1966–72. https://doi.org/10.1111/j.1532-5415.2005.53566.x PMID: 16274380
33. Bago d'Uva T, O'Donnell O, Van Doorslaer E. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. Int J Epidemiol. 2008; 37 (6):1375–83. https://doi.org/10.1093/ije/dyn146 PMID: 18676985
34. Srisurapanont M, Suttajit S, Eurviriyanukul K, Varnado P. Discrepancy between objective and subjective cognition in adults with major depressive disorder. Sci Rep. 2017; 7(1):1–7.
35. Idler EL. Age differences in self-assessments of health: Age changes, cohort differences, or survivorship? J Gerontol. 1993; 48(6):289–300.
36. Henchoz K, Cavalli S, Girardin M. Health perception and health status in advanced old age: A paradox of association. J Aging Stud. 2008; 22(3):282–90.
37. Verbrugge LM, Wingard DL. Sex differentials in health and mortality. Women Health. 1987 Nov 4; 12 (2):103–45. https://doi.org/10.1300/J013v12n02_07 PMID: 3424846
38. Luy M, Minagawa Y. Gender gaps—life expectancy and proportion of life in poor health. Heal Reports. 2014; 25(12):12–9.
39. Caroli E, Weber-Baghdiguian L. Self-reported health and gender: The role of social norms. Soc Sci Med. 2016 Mar 1; 153:220–9. https://doi.org/10.1016/j.socscimed.2016.02.023 PMID: 26921837
40. Eurostat. International Standard Classification of Education (ISCED) [Internet]. 2018 [cited 2018 Jul 20]. http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard _Classification_of_Education_%28ISCED%29#Correspondence_between_ISCED_2011 _and_ISCED_1997
41. Angel S, Disslbacher F, Humer S, Schnetzer M. What did you really earn last year?: explaining measurement error in survey income data. J R Stat Soc Ser A Stat Soc. 2019;
42. Luchman JN. DOMIN: Stata module to conduct dominance analysis [Internet]. 2013 [cited 2018 Dec 17]. https://ideas.repec.org/c/boc/bocode/s457629.html
43. Luchman JN. Relative importance analysis with multicategory dependent variables: An extension and review of best practices. Organ Res Methods. 2014; 17(4):452–71.
44. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001; 56A(3):M146–56.
45. Santos-Eggimann B, Cuénoud P, Spagnoli J, Junod J. Prevalence of frailty in middle-aged and older community-dwelling Europeans living in 10 countries. Journals Gerontol—Ser A Biol Sci Med Sci. 2009; 64A(6):675–81.

46. Pfarr C, Schmid A, Schneider U. Reporting heterogeneity in self-assessed health among elderly Europeans: The impact of mental and physical health status. Health Econ Rev. 2012; 2(21):1–14.
47. Meijer E, Kapteyn A, Andreyeva T. Internationally comparable health indices. Health Econ. 2011 May 1; 20(5):600–19. https://doi.org/10.1002/hec.1620 PMID: 20572201
48. Weber D, Skirbekk V, Freund I, Herlitz A. The changing face of cognitive gender differences in Europe. Proc Natl Acad Sci. 2014; 111(32):11673–8. https://doi.org/10.1073/pnas.1319538111 PMID: 25071201
49. Frieswijk N, Buunk BP, Steverink N, Slaets JPJ. The effect of social comparison information on the life satisfaction of frail older persons. Psychol Aging. 2004; 19(1):183–90. https://doi.org/10.1037/0882-7974.19.1.183 PMID: 15065941
50. Brandtstädter J, Greve W. The aging self: Stabilizing and protecitve processes. Dev Rev. 1994; (14):52–80.
51. Layes A, Asada Y, Kepart G. Whiners and deniers—What does self-rated health measure? Soc Sci Med. 2012; 75(1):1–9. https://doi.org/10.1016/j.socscimed.2011.10.030 PMID: 22265085
52. Delpierre C, Lauwers-Cances V, Datta GD, Lang T, Berkman L. Using self-rated health for analysing social inequalities in health: A risk for underestimating the gap between socioeconomic groups? J Epidemiol Community Health. 2009; 63(6):426–32. https://doi.org/10.1136/jech.2008.080085 PMID: 19218251
53. Guralnik JM, Branch LG, Cummings SR, Curb JD. Physical performance measures in aging research. J Gerontol. 1989 Sep; 44(5):M141–6. https://doi.org/10.1093/geronj/44.5.m141 PMID: 2768767
54. Kempen GIJM, Steverink N, Ormel J, Deeg DJH. The assessment of ADL among frail elderly in an interview survey: Self- report versus performance-based tests and determinants of discrepancies. Journals Gerontol—Ser B Psychol Sci Soc Sci. 1996; 51(5):254–60.
55. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: An updated systematic review. Ann Intern Med. 2011; 155(2).
56. Jansen T, Rademakers J, Waverijn G, Verheij R, Osborne R, Heijmans M. The role of health literacy in explaining the association between educational attainment and the use of out-of-hours primary care services in chronically ill people: a survey study. BMC Health Serv Res. 2018; 18(394):1–13.
57. Siahpush M, McNeill A, Hammond D, Fong GT. Socioeconomic and country variations in knowledge of health risks of tobacco smoking and toxic constituents of smoke: Results from the 2002 International Tobacco Control (ITC) Four Country Survey. Tob Control. 2006; 15(3):65–70.
58. Muttarak R. Normalization of Plus Size and the Danger of Unseen Overweight and Obesity in England. Obesity. 2018; 26(7):1125–9. https://doi.org/10.1002/oby.22204 PMID: 29932517
59. Von Wagner C, Steptoe A, Wolf MS, Wardle J. Health literacy and health actions: A review and a framework from health psychology. Heal Educ Behav. 2009; 36(5):860–77.
60. Smith SG, Curtis LM, Wardle J, von Wagner C, Wolf MS. Skill Set or Mind Set? Associations between Health Literacy, Patient Activation and Health. PLoS One. 2013; 8(9):1–7.
61. Von Wagner C, Knight K, Steptoe A, Wardle J. Functional health literacy and health-promoting behaviour in a national sample of British adults. J Epidemiol Community Health. 2007; 61(12):1086–90. https://doi.org/10.1136/jech.2006.053967 PMID: 18000132
62. Loichinger E. Labor force projections up to 2053 for 26 EU countries, by age, sex, and highest level of educational attainment. Demogr Res. 2015; 32(15).
63. Oksuzyan A, Juel K, Vaupel JW, Christensen K. Men: good health and high mortality. Sex differences in health and aging. Aging Clin Exp Res. 2008 Apr 25; 20(2):91–102. PMID: 18431075
64. Benyamini Y, Leventhal EA, Leventhal H. Gender differences in processing information for making self-assessments of health. Psychosom Med. 2000; 62:354–64. https://doi.org/10.1097/00006842-200005000-00009 PMID: 10845349
65. Courtenay WH. Constructions of masculinty and their influence on men's well being: a theory of gender and health. Soc Sci Med. 2000; 50:1385–401. https://doi.org/10.1016/s0277-9536(99)00390-1 PMID: 10741575
66. Dave GJ, Bibeau DL, Schulz MR, Aronson RE, Ivanov LL, Black A, et al. Predictors of congruency between self-reported hypertension status and measured blood pressure in the stroke belt. J Am Soc Hypertens. 2013; 7(5):370–8. https://doi.org/10.1016/j.jash.2013.04.007 PMID: 23706250
67. Dey AK, Alyass A, Muir RT, Black SE, Swartz RH, Murray BJ, et al. Validity of self-report of cardiovascular risk factors in a population at high risk for stroke. J Stroke Cerebrovasc Dis. 2015; 24(12):2860–5. https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.08.022 PMID: 26411693
68. Short ME, Goetzel RZ, Pei X, Tabrizi MJ, Ozminkowski RJ, Gibson TB, et al. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with

administrative data. J Occup Environ Med. 2009; 51(7):786–96. https://doi.org/10.1097/JOM.0b013e3181a86671 PMID: 19528832

69. Tucker-Drob EM. Neurocognitive functions and everyday functions change together in old age. Neuropsychology. 2011 May; 25(3):368–77. https://doi.org/10.1037/a0022348 PMID: 21417532

70. Dunning D. The Dunning–Kruger Effect: On being ignorant of one's own ignorance. Adv Exp Soc Psychol. 2011 Jan 1; 44:247–96.

71. Dunning D, Heath C, Suls JM. Flawed self-assessment: implications for health, education, and the workplace. Psychol Sci Public Interes. 2004 Dec 24; 5(3):69–106.

S1 Appendix. Robustness analyses

We conduct a range of robustness analyses to investigate whether the results presented in the main document are sensitive to changes in the definition of physical and cognitive impairment, sample composition and model specifications. Furthermore, we explore additional channels that may help to explain the effects of demographic characteristics on concordance, overestimating and underestimating.

Additional definitions of physical impairment

As a robustness analysis, we apply a stricter scenario where individuals are considered physically impaired when they have to use their arms to stand up from the chair, which is considered unimpaired in the main analysis. All trends described in the main text hold (Table A). Most of the average values are very similar to those when individuals are allowed to use their arms. However, there is slightly less concordance and a small increase in overestimating when individuals are not allowed to use their arms. This shift is plausible, since the question on mobility does not ask whether or not individuals use their arms when standing up from a chair. Respondents simply might not interpret having to use their arms as an impairment.

An additional specification of impairment is also applied, for which individuals who think it is unsafe to try the chair stand test are excluded from the analysis rather than considering them as impaired (Table A). The reduced sample includes 73,912 observations instead of 88,087. As expected, this specification alters the results. Concordance increases in each subgroup, mainly because overestimating drops to only 0.9%. This indicates that individuals that are unable to stand up from a chair avoid the test in the first place rather than failing the test. Individuals who report having no problem getting up from a chair might prefer not to be tested if they expect to perform badly at the test. Even though the level of overestimating is much lower with the new specification, most observed trends still hold. Concordance is still higher for men and highly educated individuals and decreases with age. Yet, the results by country vary from those in the first specifications. All Central and Eastern European (CEE) countries are still in the bottom half of concordance, but Southern European countries have higher relative rates of concordance in the new specification since large numbers of overestimating respondents are dropped in that specification. Still, most Southern and CEE European countries as well as Ireland have above-average rates of overestimation. While Northern European countries still have above-average concordance, Western European countries have a scattered distribution of results using this new specification.

Table B displays results of applying Models 1 and 2 on the reduced sample (i.e. where everyone refusing to do the chair stand test is dropped). We find that Southern European countries have much higher concordance rates for mobility measures. Furthermore, Austria, Germany, and Luxembourg have relatively low concordance, as their tendency to underestimate mobility is relatively more important. Concordance still decreases with age, mainly due to an increase in underestimating opposed to an increase in overestimating. In summary, overestimating may mainly be due to not taking part in the test, which is especially relevant for Southern European and CEE countries. Once these observations are dropped, underestimating is more prevalent, especially in Western European countries. Similar to the descriptive evidence, these results indicate self-selection of individuals in performing the test.

Additional definitions of cognitive impairment

Table C provides summary statistics for an additional specification of cognitive impairment. In our main analysis, individuals are considered to be cognitively impaired if they recall three words or less in the memory test. For this sensitivity analysis, a more lenient threshold is applied in which participants are considered to be impaired when they recall two words or less. Applying this specification results in a much lower proportion of impaired individuals (7.6% versus 16.1% using the original specification). While the overall rate of concordance hardly changes, the tendency to overestimate is much lower and the tendency to underestimate is much higher with the new specification. This is to be expected as the new specification considers fewer individuals to be impaired.

Although the overall levels of overestimating and underestimating change with the new specification, the trends observed in the main analysis hold. Men are still more likely to achieve concordance than women. While men tend to overestimate their cognition, women tend to underestimate theirs. The result still shows a clear decrease in concordance with age and both overestimating and underestimating show the same patterns with age as with the original specification of impairment. We still observe a strong education gradient in concordance and the country ranking is almost identical to that of the original specification. Switzerland has still the highest rate of concordance (83.2%), while Estonia has the lowest (53.1%).

Table D displays the regression results for Models 1 and 2 when using the new specification of cognitive impairment. The magnitude of the coefficients changes, yet the findings remain the same as within the main analysis. The pattern of age effects and between countries are almost identical to the main findings. The only difference is that the level of overestimating is lower and the level of underestimating is higher with the new specification. In conclusion, the threshold of impairment impacts the level of overestimating and underestimating, but not the overall trends in concordance between tested and self-reported cognition.

In our main analysis, objective cognition was based on immediate word recall. However, the self-assessment of memory might also refer to delayed word recall. Thus, we also provide an additional analysis of objective cognitive impairment based on delayed word recall. During the interview, survey participants are first asked to repeat a list of ten words, which is the basis for the immediate word recall measure. Following that, the participants perform some additional tests, for example on numeracy. After these additional tests, which take approximately 5 minutes, the interviewer asks "A little while ago, I read you a list of words and you repeated the ones you could remember. Please tell me any of the words that you can remember now?", which is the basis for a delayed word recall measure. While survey participants recall on average 5.2 words immediately, they only recall 3.9 words in the delayed test. As a consequence, concordance is lower when objective cognition is based on delayed word recall, because, by default, more individuals overestimate their cognition when the new definition is applied.

Table E presents regression results for when objective cognition is based on delayed word recall. While the trend in age is similar to that of immediate word recall, the decrease in concordance with age appears less steep. Furthermore, differences between educational attainment groups are smaller when the new specification is applied. On the contrary, the difference between the genders increases. In line with these findings, the results based on the relative importance analysis show that age and education appear slightly less important in explaining the variance in response behaviour, whereas gender appears more relevant. The main conclusions and the relative ranking of determinants remain. Specifically, in the model with immediate (delayed) word recall, country differences contribute 45% (45%) to the explained variance, age differences 30% (24%), educational differences 23% (21%), gender 2% (8%) and time effects 1% (2%).

Additional sample compositions

We also analyse whether the results are sensitive to different sample compositions. For example, frail individuals might be more likely to live in institutions in some countries than in other countries and consequently are not always included in our target population of non-institutionalised population. This could be relevant for the results since the survey respondent's overall level of health might affect concordance, especially when they suffer from very poor health. Thus, we exclude frail individuals from the sample and analyse if they influence the outcomes. To measure frailty, we rely on a well-established indicator introduced by [1], for which individuals are considered frail if they show three or more of the following components: exhaustion, weakness, slowness, shrinking and low activity levels. We follow exactly the operationalisation by [2], who adapted the indicator for SHARE data. According to the frailty measure, 8% of the survey participants are considered frail in our mobility sample (Waves 2 and 5), and 9% in our cognition sample (Waves 4 and 5). Consequently, 6,335 observations are dropped for the robustness analysis of mobility, and 9,996 observations for cognition.

The results for mobility are presented in Table F. Country coefficients change marginally in magnitude when frail individuals are excluded, while all other coefficients remain almost identical. Similarly, results based on relative importance analysis hardly change when frail survey participants are dropped. In the model with (without) frail individuals, country differences contribute 35% (39%) to the explained variance, age differences contribute 29% (32%), education differences contribute 17% (15%), gender differences contribute 11% (11%) and time effects contribute 5% (6%). Thus, the only difference is that age and education contribute marginally less to the explained variance in concordance, which appears plausible since frailty is highly correlated with age and education. Consequently, all other determinants explain relatively more of the variation once frailty is accounted for. The results for cognition hardly change when frail individuals are dropped from the sample (Table G). Country coefficients change slightly in magnitude, but not in sign. All other coefficients are virtually identical to those of the main regression analysis. Similarly, results based on relative importance analysis remain unaffected. In summary, the results appear robust to different compositions of frail individuals and their reporting behaviour.

In the main analysis, we describe differences in reporting behaviour between physical and cognitive impairment. Physical impairment is taken from Wave 2 and Wave 5, cognitive impairment from Wave 4 and Wave 5. Since the results for the two health dimensions are not based on the same sample, these differences could stem from differences in the sample rather than differences in reporting behaviour. Thus, we run additional analyses based on Wave 5 only, in which information on concordance of physical as well as cognitive health care measures is provided, i.e. we can estimate the relationship between demographic characteristics and the probability to overestimate or underestimate physical and cognitive health based on the exact same group of individuals. The regression results are provided in Table H and Table I. Since wave dummies are not needed for this specification, they are excluded from the model. Although some of the coefficients slightly change in magnitude and significance, the main results appear robust. Results from the relative importance analysis cannot be directly compared with the main model, since the wave dummy is now missing. In Wave 5, the explained variation in concordance of mobility measures can be decomposed as follows: country differences 29%, age differences 43%, educational differences 19%, gender differences 10%; thus, the main difference to the estimations based on both waves is that age appears more relevant now than when both waves are combined. The variation in concordance of cognition measures can be decomposed as follows: country differences 50%, age differences 27%, educational differences 21% and gender 1%. Thus, the results are very similar to the main computations. Results for each country individually can be found in Figs A and B.

Additional model specifications

In addition to demographic characteristics, other factors might have an impact on concordance and/or further explain the effect of demographic characteristics on reporting behaviour. In particular, we analyse whether the results change after we account for employment status, marital status and whether a person has children (Tables J-O). Furthermore, Tables P to S provide regression results including learning effects and an interaction term between gender and education.

Whether an individual works or not is likely to influence health perception. First, persons working regularly might be more aware of their mobility impairments. Further, during their working tasks they might face limitations of their memory abilities, which might be particularly relevant for individuals working in analytical jobs. Since age is highly correlated with an individual's employment status, parts of the strong effect of age on concordance might be explained by younger survey participants that are still in employment. Furthermore, employment might be an important mediator for the effect of educational attainment on concordance between measures of cognitive health, since highly educated individuals are more likely to work in jobs that demand strong cognitive skills. To test the employment channel, we add a dummy variable to the models that indicates if an individual is employed, as opposed to retired, unemployed, permanently sick or a homemaker.

In the mobility sample, 27% of the survey participants are employed and in the cognition sample, it is 26%. In both samples, employment has a strong negative correlation with age and a strong positive correlation with educational attainment. Furthermore, summary statistics show that employed individuals are more likely to achieve concordance. Tables J and K present regression results for mobility and cognition respectively. As expected, employed individuals are less likely to overestimate or underestimate their physical and cognitive health. Furthermore, the age gradient in concordance appears less pronounced. In addition, the education gradient in concordance appears less pronounced for mobility once employment is accounted for but does not change for cognition.

In addition to employment, having children or being in a relationship might influence health perception. For example, if family members comment on the survey participant's health status or if the health of other family members serves as a reference point. Thus, we provide results for two more models, in which we control for whether the survey participant has children (Tables L and M) and for whether the survey participant is married or in a registered partnership (Tables N and O). The coefficients for children and marriage either have the expected sign or are insignificant. What is more relevant for the work at hands, however, is that the inclusion of these variables has almost no impact on all other coefficients.

Relative importance analysis confirms that the employment channel explains part of the strong age effect, at least for reporting behaviour related to mobility. When employment status, marital status and a dummy for children are added to the model for mobility, country differences still contribute 32% percent to the explained variation, but age differences drop to 20%, probably, because differences in employment status explain 17%. Likely, for the same reason, the contribution of educational differences slightly decreases to 13%. Gender remains at 9% and wave at 4%. Being married (3%) and having children (1%) explains only little of the variation. Similar results are found for cognition, although employment seems relatively less important in explaining concordance. Country differences contribute 44% to the explained variation, age differences 22%, differences in employment status 11%, educational differences 20%, gender 2% and wave less than one per cent. Again, the contribution of having children and being married is negligible.

Including additional mediators in the model identified potential pathways, but more detailed analyses are required to draw concrete conclusions. For instance, the effect of labour market participation should be investigated more thoroughly considering factors such as the number of working hours, part-time retirement and type of occupation; however, this goes beyond the scope of this study.

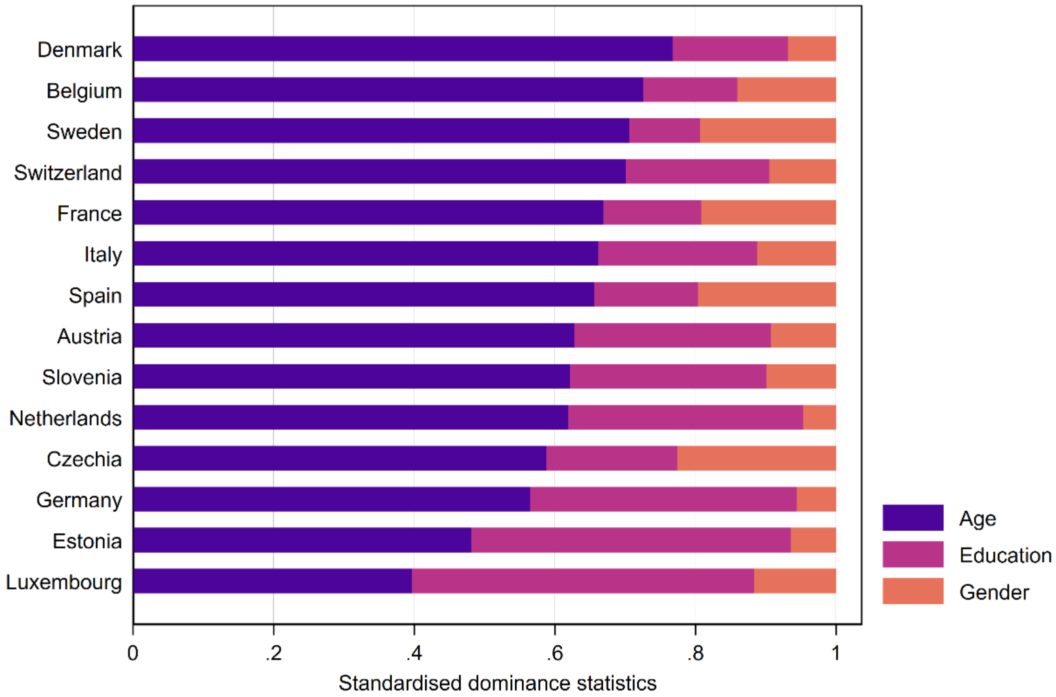


Fig A. Decomposition of the overall bias in self-reported mobility (based on Wave 5 only)

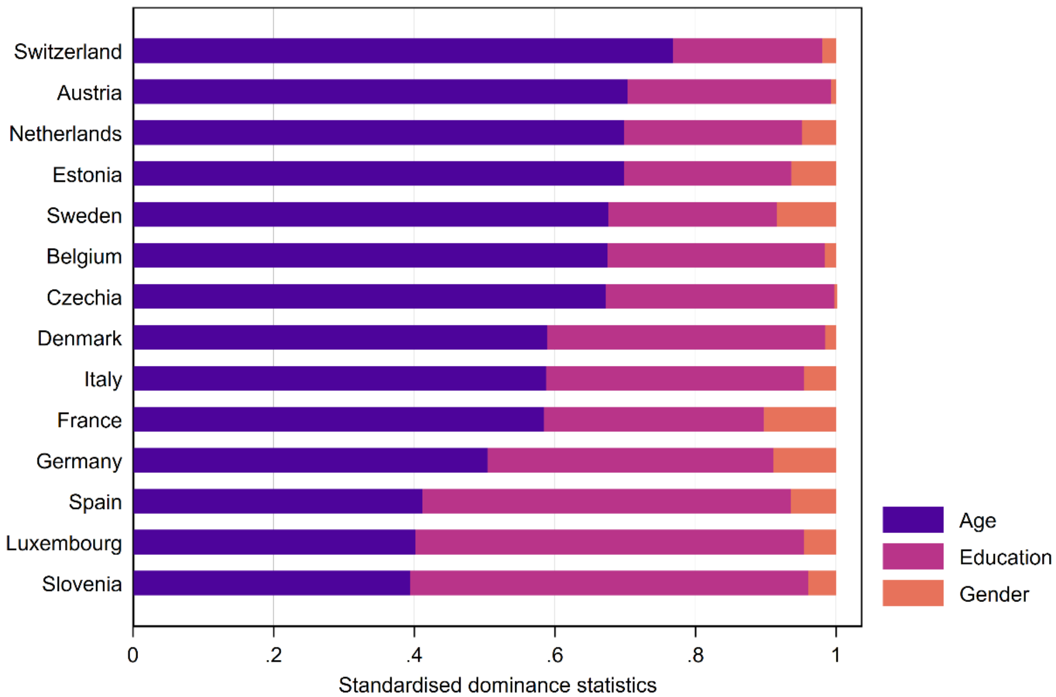


Fig B. Decomposition of the overall bias in self-reported cognition (based on Wave 5 only)

Table A. Summary statistics showing different specifications of impaired mobility

	Chair stand without using arms						Chair stand without participants that felt unsafe					
	Impairment		Concordance				Impairment		Concordance			
	S	T	S=T	S>T	S<T		S	T	S=T	S > T	S < T	
	%	%	%	%	%	N	%	%	%	%	%	N
Total	19.2	18.0	80.0	10.0	9.9	88,087	19.2	1.3	86.9	0.9	12.1	73,912
Gender												
Men	14.9	16.0	82.5	9.8	7.7	39,417	14.9	1.2	89.8	1.0	9.2	33,832
Women	22.7	19.7	78.1	10.2	11.7	48,670	22.7	1.4	84.5	0.9	14.6	40,080
Age	10.3	10.5	85.2	7.5	7.3	11,229	10.3	1.1	90.9	1.0	8.1	10,219
50–54	12.7	12.2	83.7	8.0	8.4	16,196	12.7	1.3	89.4	1.0	9.5	14,501
55–59	14.9	13.1	82.0	8.1	9.9	16,836	14.9	1.0	87.9	0.8	11.3	14,886
60–64	16.6	15.5	80.0	9.5	10.5	15,721	16.6	1.1	86.6	0.9	12.6	13,569
65–69	20.7	20.6	77.5	11.3	11.2	12,906	20.7	1.5	84.9	1.0	14.1	10,553
70–74	26.9	26.1	75.4	12.4	12.2	7,347	26.9	1.2	82.8	0.7	16.5	5,579
75–79	34.4	38.2	71.0	16.8	12.2	4,664	34.4	1.9	79.1	1.2	19.7	3,012
80–84	42.6	52.1	68.3	21.1	10.6	2,438	42.6	4.4	76.1	2.2	21.7	1,281
85–89	46.9	62.2	65.2	25.9	8.9	750	46.9	4.2	73.4	3.2	23.4	312
90–94	10.3	10.5	85.2	7.5	7.3	11,229	10.3	1.1	90.9	1.0	8.1	10,219
Education												
Low	24.7	25.0	75.9	13.1	11.0	35,808	24.7	1.8	84.1	1.2	14.7	27,858
Medium	16.9	15.0	81.2	8.8	10.1	31,953	16.9	1.1	87.3	0.8	11.9	27,644
High	11.8	10.0	86.0	6.3	7.6	19,058	11.8	0.7	90.9	0.6	8.5	17,374
Country												
Austria	20.8	18.3	79.9	9.2	10.8	5,032	20.8	1.2	86.0	0.8	13.2	4,182
Belgium	19.5	14.6	80.7	7.6	11.7	7,932	19.5	0.5	85.9	0.4	13.7	6,845
Czechia	23.2	22.7	77.8	11.4	10.7	7,651	23.2	1.3	84.9	1.0	14.1	6,102
Denmark	12.7	7.7	87.6	4.3	8.1	6,014	12.7	0.3	91.1	0.2	8.7	5,578
Estonia	29.1	26.9	76.5	10.7	12.8	5,454	29.1	1.4	81.6	1.0	17.5	4,079
France	16.3	17.6	79.8	11.3	8.9	6,566	16.3	2.3	87.8	1.6	10.6	5,563
Germany	19.6	14.4	80.1	7.9	12.0	7,700	19.6	1.1	85.2	0.8	13.9	6,712
Greece	18.1	21.5	77.5	15.5	7.0	2,601	18.1	0.8	89.8	0.7	9.5	2,133
Ireland	18.0	20.6	77.8	14.1	8.1	792	18.0	2.8	88.0	2.2	9.8	651
Italy	19.4	25.8	75.6	16.0	8.4	6,919	19.4	2.5	86.8	1.7	11.5	5,383
Luxembourg	21.2	16.5	78.5	8.6	12.9	1,561	21.2	0.7	84.2	0.5	15.3	1,318
Netherlands	14.7	10.4	85.6	5.4	9.0	6,258	14.7	0.6	89.7	0.3	10.0	5,663
Poland	29.5	29.7	70.5	17.1	12.3	1,969	29.5	3.7	79.9	3.0	17.2	1,445
Slovenia	20.9	20.1	78.0	10.8	11.2	2,873	20.9	0.5	85.3	0.4	14.3	2,325
Spain	21.8	27.0	76.7	15.3	7.9	8,011	21.8	2.4	87.1	2.0	10.9	6,207
Sweden	15.4	11.3	83.6	6.7	9.6	6,611	15.4	0.7	88.6	0.5	10.9	5,932
Switzerland	11.2	9.9	85.3	7.0	7.7	4,143	11.2	1.0	90.6	0.8	8.6	3,794
Wave												
Wave 2	18.6	17.7	79.4	11.7	8.9	26,973	18.6	1.6	87.9	1.2	10.9	22,867
Wave 5	19.5	18.2	80.3	9.3	10.3	61,114	19.5	1.1	86.5	0.8	12.7	51,045

Note: S refers to self-reported impairment and T refers to tested impairment. S=T denotes concordance, S>T denotes overestimating, and S<T denotes underestimating. N = 100%

Table B. Multinomial logistic estimation for concordance between mobility measures (excl. participants that felt unsafe)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	0.745*	0.375	-0.052	0.077
Belgium	-0.141	0.391	0.046	0.072
Czechia	0.840*	0.363	-0.022	0.072
Denmark	-0.790	0.452	-0.379***	0.080
Estonia	0.958**	0.370	0.222**	0.073
France	1.210***	0.354	-0.333***	0.076
Germany	0.743*	0.360	0.141*	0.071
Greece	0.131	0.440	-0.328***	0.099
Ireland	1.424**	0.441	-0.155	0.149
Italy	1.254***	0.356	-0.296***	0.076
Luxembourg	0.142	0.531	0.139	0.099
Netherlands	-0.366	0.413	-0.364***	0.077
Poland	1.785***	0.379	0.413***	0.097
Spain	1.410***	0.356	-0.396***	0.075
Sweden	0.10	0.385	-0.290***	0.075
Switzerland	0.602	0.380	-0.543***	0.085
Age (Ref: 60–64)				
50–54	0.177	0.141	-0.383***	0.045
55–59	0.298*	0.126	-0.190***	0.038
65–69	0.121	0.134	0.116**	0.037
70–74	0.230	0.138	0.220***	0.039
75–79	-0.008	0.194	0.382***	0.046
80–84	0.668***	0.196	0.604***	0.055
85–89	1.244***	0.225	0.728***	0.075
90–94	1.733***	0.344	0.853***	0.145
Women	0.020	0.078	0.516***	0.025
Education (Ref: Medium)				
Low	0.234*	0.096	0.229***	0.029
High	-0.141	0.119	-0.325***	0.035
Wave 5	-0.351***	0.093	0.018	0.029
Constant	-5.336***	0.355	-2.234***	0.073
N	72,876	Pseudo R^2		0.036

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table C. Summary statistics showing a different specification of impaired cognition

Cognition, impaired if able to recall two words or less						
	Impairment		**Concordance**			
	S	T	S=T	S>T	S<T	
	%	%	%	%	%	N
Total	29.4	7.6	72.4	3.0	24.7	115,785
Gender						
Men	28.1	7.8	73.2	3.3	23.5	51,013
Women	30.4	7.4	71.7	2.7	25.6	64,772
Age						
50–54	17.6	2.7	81.6	1.7	16.7	13,244
55–59	20.5	2.8	79.0	1.6	19.4	19,461
60–64	22.9	3.4	76.8	1.8	21.4	21,098
65–69	26.5	4.3	73.7	2.1	24.2	19,447
70–74	33.8	7.0	67.3	3.0	29.7	16,180
75–79	42.0	12.7	61.8	4.6	33.7	12,350
80–84	48.5	21.3	60.3	6.4	33.3	8,525
85–89	52.3	30.6	62.2	8.3	29.5	4,283
90–94	53.2	37.7	64.4	10.5	25.1	1,197
Education						
Low	39.7	13.4	64.5	4.7	30.8	46,113
Medium	24.8	4.0	75.5	1.9	22.6	43,362
High	17.7	2.7	81.8	1.6	16.6	24,337
Country						
Austria	17.8	5.7	81.9	2.9	15.2	9,028
Belgium	24.4	6.5	75.1	3.5	21.3	10,511
Czechia	30.0	4.9	72.1	1.5	26.4	10,609
Denmark	17.3	3.8	82.9	1.8	15.3	6,171
Estonia	51.4	8.2	53.1	1.8	45.0	11,792
France	31.9	8.4	69.3	3.6	27.2	9,796
Germany	22.4	4.8	77.7	2.4	19.8	7,099
Hungary	34.2	7.9	67.4	3.2	29.4	2,938
Italy	32.9	11.1	70.7	3.9	25.3	7,895
Luxembourg	18.5	7.4	79.6	4.8	15.6	1,543
Netherlands	15.7	4.4	83.1	2.8	14.1	6,770
Poland	32.8	12.1	70.0	4.5	25.5	1,678
Portugal	45.4	13.9	59.1	4.6	36.3	1,899
Slovenia	26.9	8.7	74.2	3.9	21.9	5,511
Spain	41.1	17.4	65.8	5.1	29.0	9,628
Sweden	29.3	4.9	71.8	2.1	26.1	6,346
Switzerland	16.5	3.0	83.2	1.8	15.1	6,571
Wave						
Wave 4	29.4	7.9	72.2	3.1	24.7	55,172
Wave 5	29.4	7.2	72.6	2.8	24.6	60,613

Note: S refers to self-reported impairment and T refers to tested impairment. S=T denotes concordance, S>T denotes overestimating, and S<T denotes underestimating. N = 100%

Table D. Multinomial logistic estimation for concordance between cognition measures (impaired if able to recall two words or less)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.337***	0.098	-0.419***	0.049
Belgium	-0.150	0.094	0.004	0.046
Czechia	-0.983***	0.110	0.169***	0.043
Denmark	-0.676***	0.122	-0.326***	0.054
Estonia	-0.400***	0.101	1.079***	0.042
France	-0.192*	0.093	0.267***	0.045
Germany	-0.267*	0.107	-0.048	0.049
Hungary	-0.047	0.130	0.496***	0.055
Italy	-0.249**	0.096	0.000	0.046
Luxembourg	0.173	0.142	-0.445***	0.080
Netherlands	-0.560***	0.107	-0.597***	0.055
Poland	0.088	0.147	0.209**	0.071
Portugal	0.130	0.138	0.662***	0.063
Spain	-0.037	0.090	0.169***	0.045
Sweden	-0.747***	0.118	0.142**	0.048
Switzerland	-0.832***	0.120	-0.448***	0.055
Age (Ref: 60–64)				
50–54	-0.065	0.087	-0.247***	0.031
55–59	-0.146	0.079	-0.114***	0.027
65–69	0.151*	0.074	0.158***	0.025
70–74	0.569***	0.071	0.411***	0.026
75–79	0.985***	0.070	0.562***	0.028
80–84	1.307***	0.071	0.554***	0.032
85–89	1.502***	0.080	0.399***	0.042
90–94	1.703***	0.113	0.211**	0.079
Women	-0.295***	0.037	0.052**	0.016
Education (Ref: Medium)				
Low	0.747***	0.047	0.361***	0.019
High	-0.273***	0.065	-0.359***	0.024
Wave 5	-0.107**	0.037	0.099***	0.014
Constant	-3.463***	0.092	-1.546***	0.042
N	113,812	Pseudo R-squared		0.063

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table E. Multinomial logistic estimation for concordance between cognition measures (delayed word recall)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.653***	0.042	-0.343***	0.070
Belgium	-0.453***	0.040	0.304***	0.063
Czechia	-0.463***	0.039	0.279***	0.062
Denmark	-0.731***	0.047	-0.117	0.073
Estonia	-0.890***	0.042	0.974***	0.058
France	-0.583***	0.041	0.532***	0.062
Germany	-0.491***	0.043	0.059	0.068
Hungary	-0.510***	0.055	0.332***	0.080
Italy	-0.342***	0.042	0.061	0.068
Luxembourg	-0.465***	0.067	-0.235*	0.109
Netherlands	-0.562***	0.044	-0.226**	0.074
Poland	-0.151*	0.065	-0.172	0.118
Portugal	-0.532***	0.064	0.571***	0.088
Spain	-0.363***	0.041	-0.109	0.068
Sweden	-0.755***	0.046	0.425***	0.066
Switzerland	-0.863***	0.047	-0.146*	0.072
Age (Ref: 60–64)				
50–54	-0.257***	0.030	-0.159***	0.038
55–59	-0.151***	0.026	-0.022	0.033
65–69	0.176***	0.025	0.076*	0.032
70–74	0.339***	0.026	0.159***	0.034
75–79	0.414***	0.028	0.018	0.039
80–84	0.484***	0.032	-0.223***	0.048
85–89	0.472***	0.040	-0.621***	0.070
90–94	0.609***	0.069	-0.847***	0.152
Women	-0.307***	0.015	0.178***	0.021
Education (Ref: Medium)				
Low	0.259***	0.018	0.007	0.025
High	-0.429***	0.023	-0.205***	0.028
Wave 5	-0.125***	0.014	0.134***	0.019
Constant	-0.364***	-0.037	-2.058***	-0.060
N	113,721	Pseudo R-squared		0.036

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table F. Multinomial logistic estimation for concordance between mobility measures (frail individuals are excluded from the sample)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.266**	0.086	-0.041	0.079
Belgium	-0.542***	0.084	0.055	0.075
Czechia	-0.074	0.079	-0.045	0.074
Denmark	-1.112***	0.100	-0.341***	0.083
Estonia	0.004	0.082	0.102	0.076
France	-0.136	0.080	-0.340***	0.081
Germany	-0.341***	0.081	0.166*	0.074
Greece	-0.013	0.094	-0.356***	0.103
Ireland	0.152	0.130	-0.237	0.158
Italy	0.228**	0.077	-0.391***	0.081
Luxembourg	-0.242*	0.121	0.113	0.103
Netherlands	-0.974***	0.093	-0.305***	0.080
Poland	0.352***	0.098	0.304**	0.101
Spain	0.018	0.078	-0.476***	0.079
Sweden	-0.677***	0.086	-0.235**	0.078
Switzerland	-0.682***	0.096	-0.452***	0.088
Age (Ref: 60–64)				
50–54	-0.133**	0.050	-0.374***	0.047
55–59	-0.043	0.044	-0.171***	0.039
65–69	0.195***	0.043	0.107**	0.038
70–74	0.310***	0.045	0.156***	0.041
75–79	0.586***	0.053	0.251***	0.049
80–84	1.009***	0.059	0.334***	0.060
85–89	1.313***	0.074	0.364***	0.084
90–94	1.735***	0.123	0.290	0.170
Women	0.067**	0.026	0.458***	0.026
Education (Ref: Medium)				
Low	0.175***	0.031	0.163***	0.030
High	-0.274***	0.039	-0.297***	0.036
Wave 5	-0.460***	0.032	0.012	0.030
Constant	-1.952***	-0.079	-2.271***	-0.075
N	80,484	Pseudo R^2		0.034

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table G. Multinomial logistic estimation for concordance between cognition measures (frail individuals are excluded from the sample)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.628***	0.070	-0.435***	0.056
Belgium	-0.446***	0.067	0.062	0.052
Czechia	-0.881***	0.071	0.241***	0.049
Denmark	-0.683***	0.081	-0.292***	0.061
Estonia	-0.647***	0.072	1.095***	0.047
France	-0.367***	0.066	0.327***	0.051
Germany	-0.499***	0.076	0.026	0.055
Hungary	-0.260**	0.097	0.426***	0.065
Italy	-0.293***	0.067	0.044	0.053
Luxembourg	-0.165	0.107	-0.496***	0.093
Netherlands	-0.674***	0.073	-0.531***	0.061
Poland	-0.005	0.107	0.239**	0.082
Portugal	-0.058	0.101	0.613***	0.073
Spain	-0.096	0.064	0.073	0.052
Sweden	-0.723***	0.078	0.227***	0.053
Switzerland	-0.828***	0.079	-0.389***	0.060
Age (Ref: 60–64)				
50–54	-0.244***	0.058	-0.254***	0.033
55–59	-0.198***	0.050	-0.107***	0.028
65–69	0.166***	0.047	0.134***	0.027
70–74	0.516***	0.047	0.343***	0.029
75–79	0.933***	0.047	0.443***	0.032
80–84	1.186***	0.051	0.364***	0.038
85–89	1.276***	0.064	0.133*	0.055
90–94	1.343***	0.112	-0.019	0.118
Women	-0.305***	0.026	0.078***	0.018
Education (Ref: Medium)				
Low	0.668***	0.032	0.264***	0.021
High	-0.429***	0.044	-0.314***	0.025
Wave 5	-0.114***	0.025	0.119***	0.016
Constant	-2.237***	-0.063	-1.690***	-0.048
N	103,816	Pseudo R-squared		0.058

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table H. Multinomial logistic estimation for concordance between mobility measures (based on Wave 5 only)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.225**	0.085	-0.088	0.078
Belgium	-0.339***	0.082	0.141	0.073
Czechia	0.123	0.076	-0.041	0.074
Denmark	-0.848***	0.100	-0.246**	0.082
Estonia	-0.027	0.077	0.112	0.072
France	-0.267**	0.083	-0.157*	0.078
Germany	-0.386***	0.084	0.161*	0.072
Italy	-0.001	0.079	-0.345***	0.080
Luxembourg	-0.205	0.112	0.150	0.097
Netherlands	-0.757***	0.094	-0.297***	0.082
Spain	0.077	0.074	-0.465***	0.077
Sweden	-0.738***	0.091	-0.249**	0.079
Switzerland	-0.514***	0.098	-0.409***	0.090
Age (Ref: 60–64)				
50–54	-0.168*	0.068	-0.347***	0.057
55–59	-0.071	0.058	-0.148**	0.048
65–69	0.150**	0.055	0.098*	0.045
70–74	0.276***	0.056	0.133**	0.047
75–79	0.552***	0.056	0.265***	0.050
80–84	0.934***	0.058	0.313***	0.056
85–89	1.158***	0.067	0.219**	0.074
90–94	1.444***	0.099	0.105	0.133
Women	0.065*	0.030	0.412***	0.028
Education (Ref: Medium)				
Low	0.260***	0.037	0.133***	0.033
High	-0.280***	0.046	-0.308***	0.039
Constant	-2.389***	-0.075	-2.203***	-0.069
N	60,233	Pseudo R^2		0.033

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table I. Multinomial logistic estimation for concordance between cognition measures (based on Wave 5 only)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.954***	-0.102	-0.198**	-0.066
Belgium	-0.323***	-0.083	0.078	-0.062
Czechia	-0.870***	-0.094	0.336***	-0.059
Denmark	-0.660***	-0.099	-0.276***	-0.069
Estonia	-0.694***	-0.098	1.242***	-0.057
France	-0.350***	-0.087	0.367***	-0.062
Germany	-0.394***	-0.087	-0.056	-0.062
Italy	-0.225**	-0.083	0.032	-0.063
Luxembourg	-0.103	-0.109	-0.397***	-0.092
Netherlands	-0.525***	-0.09	-0.496***	-0.07
Spain	-0.11	-0.078	-0.05	-0.061
Sweden	-0.554***	-0.091	0.187**	-0.062
Switzerland	-0.948***	-0.113	-0.214**	-0.072
Age (Ref: 60–64)				
50–54	-0.240**	-0.079	-0.232***	-0.044
55–59	-0.201**	-0.069	-0.105**	-0.037
65–69	0.176**	-0.063	0.116**	-0.035
70–74	0.438***	-0.063	0.312***	-0.036
75–79	0.859***	-0.062	0.444***	-0.039
80–84	1.023***	-0.066	0.336***	-0.045
85–89	1.132***	-0.077	0.114	-0.061
90–94	1.312***	-0.113	-0.144	-0.116
Women	-0.260***	-0.033	0.058**	-0.021
Education (Ref: Medium)				
Low	0.628***	-0.042	0.231***	-0.026
High	-0.510***	-0.058	-0.281***	-0.03
Constant	-2.326***	-0.08	-1.563***	-0.057
N	59,742	Pseudo R^2		0.059

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table J. Multinomial logistic estimation for concordance between mobility measures (incl. indicator for employment)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.187*	0.08	-0.028	0.076
Belgium	-0.449***	0.078	0.124	0.071
Czechia	-0.043	0.074	-0.003	0.071
Denmark	-0.956***	0.094	-0.210**	0.080
Estonia	0.021	0.077	0.196**	0.073
France	-0.059	0.075	-0.194*	0.076
Germany	-0.268***	0.077	0.225**	0.071
Greece	0.066	0.089	-0.259**	0.099
Ireland	0.196	0.126	-0.082	0.148
Italy	0.236**	0.072	-0.238**	0.075
Luxembourg	-0.189	0.114	0.175	0.098
Netherlands	-0.897***	0.089	-0.224**	0.076
Poland	0.358***	0.092	0.288**	0.095
Spain	0.059	0.072	-0.351***	0.074
Sweden	-0.581***	0.082	-0.088	0.075
Switzerland	-0.569***	0.091	-0.331***	0.085
Age (Ref: 60–64)				
50–54	0.015	0.052	-0.151**	0.048
55–59	0.051	0.045	-0.036	0.040
65–69	0.114**	0.042	-0.009	0.037
70–74	0.242***	0.043	0.035	0.040
75–79	0.473***	0.050	0.120**	0.046
80–84	0.871***	0.055	0.180**	0.055
85–89	1.094***	0.065	0.113	0.073
90–94	1.338***	0.104	-0.008	0.139
Women	0.036	0.025	0.434***	0.025
Education (Ref: Medium)				
Low	0.158***	0.030	0.142***	0.028
High	-0.265***	0.038	-0.260***	0.035
Wave 5	-0.416***	0.030	0.042	0.029
Employment	-0.343***	0.040	-0.486***	0.037
Constant	-1.872***	0.075	-2.186***	-0.072
N	86,157	Pseudo R^2		0.035

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table K. Multinomial logistic estimation for concordance between cognition measures (incl. indicator for employment)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.614***	0.066	-0.371***	0.053
Belgium	-0.405***	0.063	0.124*	0.049
Czechia	-0.836***	0.066	0.288***	0.047
Denmark	-0.623***	0.077	-0.192**	0.059
Estonia	-0.641***	0.067	1.153***	0.045
France	-0.322***	0.062	0.376***	0.049
Germany	-0.460***	0.072	0.077	0.053
Greece	-0.284**	0.087	0.502***	0.059
Ireland	-0.300***	0.063	0.076	0.051
Italy	-0.101	0.101	-0.426***	0.088
Luxembourg	-0.604***	0.070	-0.450***	0.058
Netherlands	-0.067	0.098	0.202**	0.077
Poland	-0.111	0.093	0.618***	0.068
Spain	-0.134*	0.059	0.108*	0.049
Sweden	-0.648***	0.074	0.329***	0.052
Switzerland	-0.794***	0.077	-0.284***	0.059
Age (Ref: 60–64)				
50–54	-0.131*	0.060	-0.072*	0.034
55–59	-0.107*	0.051	0.014	0.029
65–69	0.097*	0.046	0.021	0.027
70–74	0.450***	0.046	0.216***	0.028
75–79	0.809***	0.046	0.279***	0.031
80–84	1.013***	0.049	0.183***	0.035
85–89	1.101***	0.058	-0.084	0.049
90–94	1.163***	0.092	-0.174	0.094
Women	-0.302***	0.025	0.074***	0.017
Education (Ref: Medium)				
Low	0.633***	0.031	0.221***	0.020
High	-0.421***	0.043	-0.279***	0.024
Wave 5	-0.120***	0.024	0.122***	0.015
Employment	-0.282***	0.045	-0.410***	0.026
Constant	-2.142***	0.060	-1.574***	0.046
N	112,906	Pseudo R^2		0.057

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table L. Multinomial logistic estimation for concordance between mobility measures (incl. indicator for children)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.220**	0.081	-0.046	0.076
Belgium	-0.438***	0.077	0.087	0.071
Czechia	-0.066	0.074	-0.056	0.071
Denmark	-0.973***	0.092	-0.303***	0.079
Estonia	-0.042	0.077	0.115	0.072
France	-0.101	0.075	-0.248**	0.076
Germany	-0.318***	0.077	0.161*	0.070
Greece	0.040	0.090	-0.282**	0.098
Ireland	0.138	0.125	-0.144	0.148
Italy	0.204**	0.073	-0.274***	0.075
Luxembourg	-0.210	0.112	0.156	0.097
Netherlands	-0.877***	0.087	-0.286***	0.076
Poland	0.388***	0.092	0.304**	0.095
Spain	0.020	0.073	-0.385***	0.074
Sweden	-0.637***	0.082	-0.193**	0.074
Switzerland	-0.631***	0.091	-0.423***	0.085
Age (Ref: 60–64)				
50–54	-0.143**	0.049	-0.349***	0.045
55–59	-0.052	0.043	-0.179***	0.038
65–69	0.192***	0.041	0.097**	0.037
70–74	0.330***	0.042	0.153***	0.039
75–79	0.574***	0.049	0.245***	0.046
80–84	0.971***	0.053	0.303***	0.054
85–89	1.177***	0.064	0.206**	0.072
90–94	1.449***	0.098	0.100	0.132
Women	0.057*	0.024	0.457***	0.025
Education (Ref: Medium)				
Low	0.181***	0.030	0.163***	0.028
High	-0.293***	0.038	-0.298***	0.035
Wave 5	-0.415***	0.030	0.028	0.029
Children	-0.247***	0.039	0.080	0.042
Constant	-1.728***	0.083	-2.344***	0.082
N	86,173	Pseudo R^2		0.033

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table M. Multinomial logistic estimation for concordance between cognition measures (incl. indicator for children)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.627***	0.066	-0.387***	0.053
Belgium	-0.406***	0.062	0.087	0.049
Czechia	-0.844***	0.066	0.250***	0.047
Denmark	-0.657***	0.076	-0.263***	0.058
Estonia	-0.697***	0.067	1.077***	0.045
France	-0.352***	0.062	0.336***	0.049
Germany	-0.479***	0.071	0.03	0.053
Greece	-0.298***	0.086	0.497***	0.059
Ireland	-0.334***	0.063	0.038	0.051
Italy	-0.135	0.100	-0.426***	0.087
Luxembourg	-0.625***	0.069	-0.504***	0.058
Netherlands	-0.079	0.098	0.201**	0.077
Poland	-0.138	0.093	0.579***	0.068
Spain	-0.171**	0.059	0.061	0.049
Sweden	-0.689***	0.073	0.235***	0.051
Switzerland	-0.841***	0.076	-0.362***	0.058
Age (Ref: 60–64)				
50–54	-0.268***	0.057	-0.252***	0.033
55–59	-0.202***	0.049	-0.116***	0.027
65–69	0.162***	0.045	0.110***	0.026
70–74	0.527***	0.045	0.321***	0.028
75–79	0.888***	0.045	0.387***	0.030
80–84	1.092***	0.047	0.287***	0.035
85–89	1.172***	0.056	0.037	0.048
90–94	1.288***	0.086	-0.096	0.090
Women	-0.290***	0.025	0.091***	0.017
Education (Ref: Medium)				
Low	0.644***	0.031	0.240***	0.020
High	-0.446***	0.043	-0.306***	0.024
Wave 5	-0.130***	0.024	0.116***	0.015
Children	-0.230***	0.039	0.021	0.029
Constant	-1.984***	0.070	-1.672***	0.053
N	113,081	Pseudo R^2		0.056

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table N. Multinomial logistic estimation for concordance between mobility measures (incl. indicator for marriage or registered partnership)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.215**	0.081	-0.046	0.076
Belgium	-0.421***	0.077	0.08	0.071
Czechia	-0.076	0.074	-0.061	0.071
Denmark	-0.982***	0.093	-0.310***	0.080
Estonia	-0.059	0.077	0.11	0.072
France	-0.105	0.075	-0.250***	0.076
Germany	-0.297***	0.077	0.157*	0.071
Greece	0.067	0.090	-0.315**	0.100
Ireland	0.157	0.125	-0.163	0.148
Italy	0.239**	0.073	-0.272***	0.075
Luxembourg	-0.195	0.112	0.157	0.097
Netherlands	-0.854***	0.087	-0.274***	0.076
Poland	0.400***	0.092	0.300**	0.096
Spain	0.053	0.073	-0.390***	0.074
Sweden	-0.635***	0.082	-0.188*	0.075
Switzerland	-0.621***	0.091	-0.446***	0.085
Age (Ref: 60–64)				
50–54	-0.138**	0.049	-0.358***	0.046
55–59	-0.044	0.042	-0.183***	0.038
65–69	0.190***	0.041	0.098**	0.036
70–74	0.322***	0.042	0.152***	0.039
75–79	0.545***	0.049	0.239***	0.046
80–84	0.936***	0.053	0.289***	0.054
85–89	1.129***	0.064	0.185*	0.073
90–94	1.391***	0.097	0.059	0.133
Women	0.020	0.025	0.444***	0.025
Education (Ref: Medium)				
Low	0.173***	0.030	0.157***	0.028
High	-0.293***	0.038	-0.300***	0.035
Wave 5	-0.411***	0.030	0.017	0.029
Married	-0.212***	0.027	-0.077**	0.027
Constant	-1.782***	0.079	-2.192***	0.076
N	85,781	Pseudo R^2		0.033

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table O. Multinomial logistic estimation for concordance between cognition measures (incl. indicator for marriage or registered partnership)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.642***	0.066	-0.388***	0.053
Belgium	-0.410***	0.062	0.093	0.049
Czechia	-0.875***	0.066	0.251***	0.047
Denmark	-0.668***	0.076	-0.273***	0.058
Estonia	-0.719***	0.067	1.075***	0.045
France	-0.355***	0.062	0.337***	0.049
Germany	-0.472***	0.071	0.024	0.053
Greece	-0.303***	0.086	0.499***	0.059
Ireland	-0.315***	0.063	0.034	0.051
Italy	-0.125	0.101	-0.426***	0.087
Luxembourg	-0.619***	0.069	-0.507***	0.058
Netherlands	-0.070	0.098	0.200*	0.078
Poland	-0.124	0.093	0.588***	0.068
Spain	-0.153**	0.059	0.058	0.049
Sweden	-0.687***	0.073	0.238***	0.052
Switzerland	-0.832***	0.077	-0.360***	0.058
Age (Ref: 60–64)				
50–54	-0.280***	0.058	-0.259***	0.033
55–59	-0.197***	0.049	-0.119***	0.028
65–69	0.161***	0.045	0.112***	0.026
70–74	0.520***	0.044	0.322***	0.028
75–79	0.864***	0.045	0.387***	0.030
80–84	1.056***	0.047	0.288***	0.035
85–89	1.117***	0.057	0.033	0.048
90–94	1.200***	0.086	-0.099	0.090
Women	-0.337***	0.025	0.089***	0.017
Education (Ref: Medium)				
Low	0.639***	0.031	0.237***	0.020
High	-0.442***	0.043	-0.308***	0.024
Wave 5	-0.129***	0.024	0.117***	0.015
Married	-0.216***	0.027	-0.003	0.019
Constant	-2.007***	0.064	-1.649***	0.049
N	112,713	Pseudo R^2		0.056

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table P. Multinomial logistic estimation for concordance between mobility measures (incl. interaction effect)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.195*	0.080	-0.049	0.075
Belgium	-0.420***	0.077	0.087	0.071
Czechia	-0.059	0.074	-0.049	0.071
Denmark	-0.965***	0.092	-0.309***	0.079
Estonia	-0.027	0.077	0.116	0.072
France	-0.084	0.075	-0.247**	0.075
Germany	-0.300***	0.076	0.162*	0.070
Greece	0.046	0.089	-0.299**	0.098
Ireland	0.168	0.125	-0.150	0.148
Italy	0.222**	0.072	-0.276***	0.075
Luxembourg	-0.195	0.112	0.151	0.097
Netherlands	-0.863***	0.087	-0.283***	0.076
Poland	0.395***	0.092	0.305**	0.095
Spain	0.037	0.072	-0.398***	0.074
Sweden	-0.632***	0.082	-0.192**	0.074
Switzerland	-0.607***	0.090	-0.430***	0.085
Age (Ref: 60–64)				
50–54	-0.132**	0.048	-0.356***	0.045
55–59	-0.048	0.042	-0.179***	0.038
65–69	0.193***	0.041	0.099**	0.036
70–74	0.333***	0.042	0.156***	0.039
75–79	0.568***	0.049	0.245***	0.045
80–84	0.975***	0.053	0.300***	0.054
85–89	1.197***	0.063	0.206**	0.072
90–94	1.485***	0.096	0.088	0.132
Women	0.029	0.041	0.388***	0.039
Education (Ref: Medium)				
Low	0.147***	0.042	0.094*	0.045
High	-0.272***	0.053	-0.378***	0.054
Interaction Effects				
Low x Women	0.061	0.054	0.109*	0.054
High x Women	-0.040	0.075	0.130	0.069
Wave 5	-0.414***	0.030	0.028	0.029
Constant	-1.953***	0.077	-2.228***	0.074
N	86,819	Pseudo R^2		0.033

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table Q. Multinomial logistic estimation for concordance between cognition measures (incl. interaction effect)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.613***	0.066	-0.385***	0.053
Belgium	-0.380***	0.062	0.096	0.049
Czechia	-0.844***	0.066	0.256***	0.047
Denmark	-0.653***	0.076	-0.264***	0.058
Estonia	-0.672***	0.067	1.082***	0.045
France	-0.333***	0.061	0.334***	0.048
Germany	-0.473***	0.071	0.032	0.052
Hungary	-0.288***	0.086	0.495***	0.059
Italy	-0.312***	0.062	0.041	0.051
Luxembourg	-0.124	0.100	-0.427***	0.087
Netherlands	-0.616***	0.069	-0.496***	0.058
Poland	-0.068	0.098	0.204**	0.077
Portugal	-0.120	0.093	0.588***	0.068
Spain	-0.151*	0.059	0.064	0.049
Sweden	-0.670***	0.073	0.241***	0.051
Switzerland	-0.821***	0.076	-0.363***	0.058
Age (Ref: 60–64)				
50–54	-0.255***	0.056	-0.247***	0.032
55–59	-0.195***	0.049	-0.114***	0.027
65–69	0.160***	0.045	0.111***	0.026
70–74	0.524***	0.044	0.320***	0.028
75–79	0.882***	0.045	0.385***	0.030
80–84	1.090***	0.047	0.286***	0.035
85–89	1.175***	0.056	0.030	0.048
90–94	1.285***	0.085	-0.104	0.089
Women	-0.465***	0.046	0.020	0.027
Education (Ref: Medium)				
Low	0.501***	0.041	0.165***	0.030
High	-0.483***	0.055	-0.362***	0.035
Interaction Effects				
Low x Women	0.285***	0.056	0.128***	0.037
High x Women	0.071	0.087	0.098*	0.048
Wave 5	-0.126***	0.024	0.116***	0.015
Constant	-2.128***	0.061	-1.616***	0.047
N	113,812	Pseudo R^2		0.056

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table R. Multinomial logistic estimation for concordance between mobility measures (incl. learning effect)

	Overestimating	SE	Underestimating	SE
Country (Ref: Slovenia)				
Austria	-0.153	0.080	-0.069	0.076
Belgium	-0.342***	0.078	0.046	0.071
Czechia	-0.004	0.074	-0.079	0.071
Denmark	-0.877***	0.093	-0.349***	0.080
Estonia	-0.032	0.077	0.112	0.072
France	-0.009	0.075	-0.284***	0.076
Germany	-0.245**	0.077	0.134	0.071
Greece	0.117	0.090	-0.335***	0.099
Ireland	0.237	0.126	-0.189	0.148
Italy	0.306***	0.073	-0.321***	0.076
Luxembourg	-0.193	0.112	0.149	0.097
Netherlands	-0.783***	0.088	-0.323***	0.077
Poland	0.469***	0.092	0.269**	0.096
Spain	0.092	0.073	-0.429***	0.074
Sweden	-0.560***	0.082	-0.231**	0.075
Switzerland	-0.537***	0.091	-0.465***	0.085
Age (Ref: 60–64)				
50–54	-0.169***	0.048	-0.337***	0.045
55–59	-0.063	0.042	-0.171***	0.038
65–69	0.197***	0.041	0.098**	0.036
70–74	0.342***	0.042	0.153***	0.039
75–79	0.585***	0.049	0.239***	0.045
80–84	0.947***	0.053	0.314***	0.054
85–89	1.132***	0.064	0.236**	0.073
90–94	1.418***	0.097	0.123	0.132
Women	0.057*	0.024	0.457***	0.024
Education (Ref: Medium)				
Low	0.183***	0.030	0.163***	0.028
High	-0.290***	0.038	-0.299***	0.035
Wave 5	-0.337***	0.032	-0.006	0.031
Learning effect	-0.311***	0.043	0.115***	0.035
Constant	-2.033***	0.075	-2.238***	0.073
N	86,819	Pseudo R^2		0.033

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

Table S. Multinomial logistic estimation for concordance between cognition measures (incl. learning effect)

	Overestimating	**SE**	**Underestimating**	**SE**
Country (Ref: Slovenia)				
Austria	-0.575***	0.066	-0.452***	0.053
Belgium	-0.332***	0.063	0.000	0.049
Czechia	-0.820***	0.066	0.203***	0.047
Denmark	-0.593***	0.076	-0.349***	0.058
Estonia	-0.683***	0.067	1.047***	0.045
France	-0.281***	0.062	0.231***	0.049
Germany	-0.467***	0.072	0.043	0.052
Hungary	-0.326***	0.086	0.560***	0.060
Italy	-0.268***	0.063	-0.046	0.051
Luxembourg	-0.203*	0.101	-0.254**	0.088
Netherlands	-0.562***	0.070	-0.585***	0.058
Poland	0.074	0.100	-0.050	0.078
Portugal	-0.169	0.093	0.644***	0.068
Spain	-0.137*	0.059	0.027	0.049
Sweden	-0.636***	0.074	0.186***	0.051
Switzerland	-0.764***	0.077	-0.460***	0.058
Age (Ref: 60–64)	-0.319***	0.057	-0.141***	0.033
50–54	-0.205***	0.049	-0.098***	0.027
55–59	0.165***	0.045	0.107***	0.026
65–69	0.534***	0.044	0.311***	0.028
70–74	0.893***	0.045	0.376***	0.030
75–79	1.105***	0.047	0.274***	0.035
80–84	1.193***	0.056	0.015	0.048
85–89	1.307***	0.086	-0.114	0.090
90–94	-0.319***	0.057	-0.141***	0.033
Women	-0.287***	0.025	0.085***	0.017
Education (Ref: Medium)				
Low	0.643***	0.031	0.238***	0.020
High	-0.447***	0.043	-0.308***	0.024
Wave 5	-0.084***	0.025	0.003	0.016
Learning effect	-0.193***	0.027	0.337***	0.018
Constant	-2.165***	0.059	-1.722***	0.046
N	113,812	Pseudo R^2		0.058

Note: The dependent variable is a three-category variable that indicates if an individual achieved concordance (reference category), overestimated or underestimated his or her health. Coefficients are given in log odds, standard errors are clustered at the individual level, *p<0.05, **p<0.01, ***p<0.001

References Appendix

1. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56A(3):M146-56.
2. Santos-Eggimann B, Cuénoud P, Spagnoli J, Junod J. Prevalence of frailty in middle-aged and older community-dwelling Europeans living in 10 countries. Journals Gerontol - Ser A Biol Sci Med Sci. 2009;64A(6):675–81.

Figure S1 Fig.: Predicted values of concordance between tested and self-reported mobility by country and age

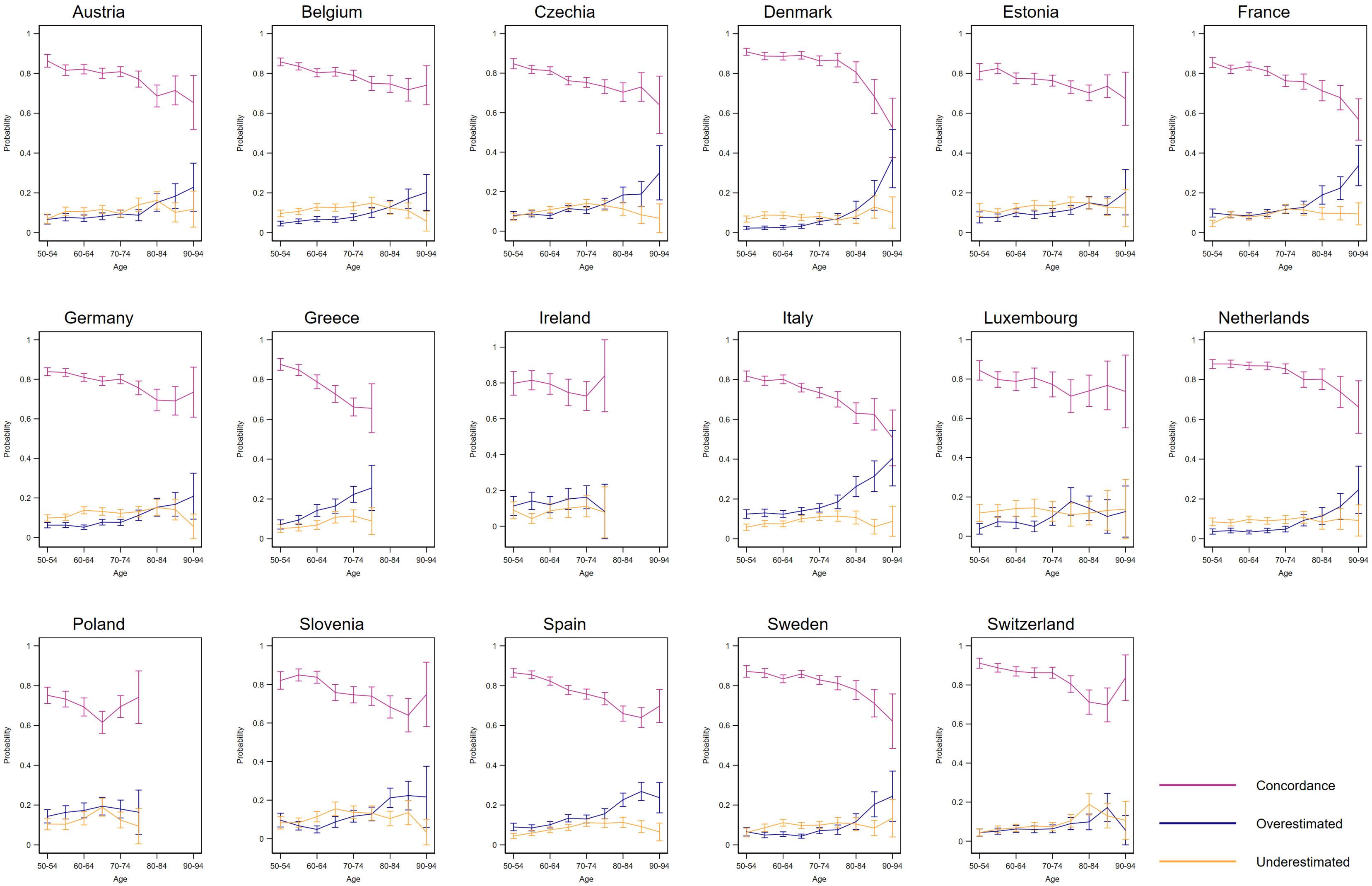
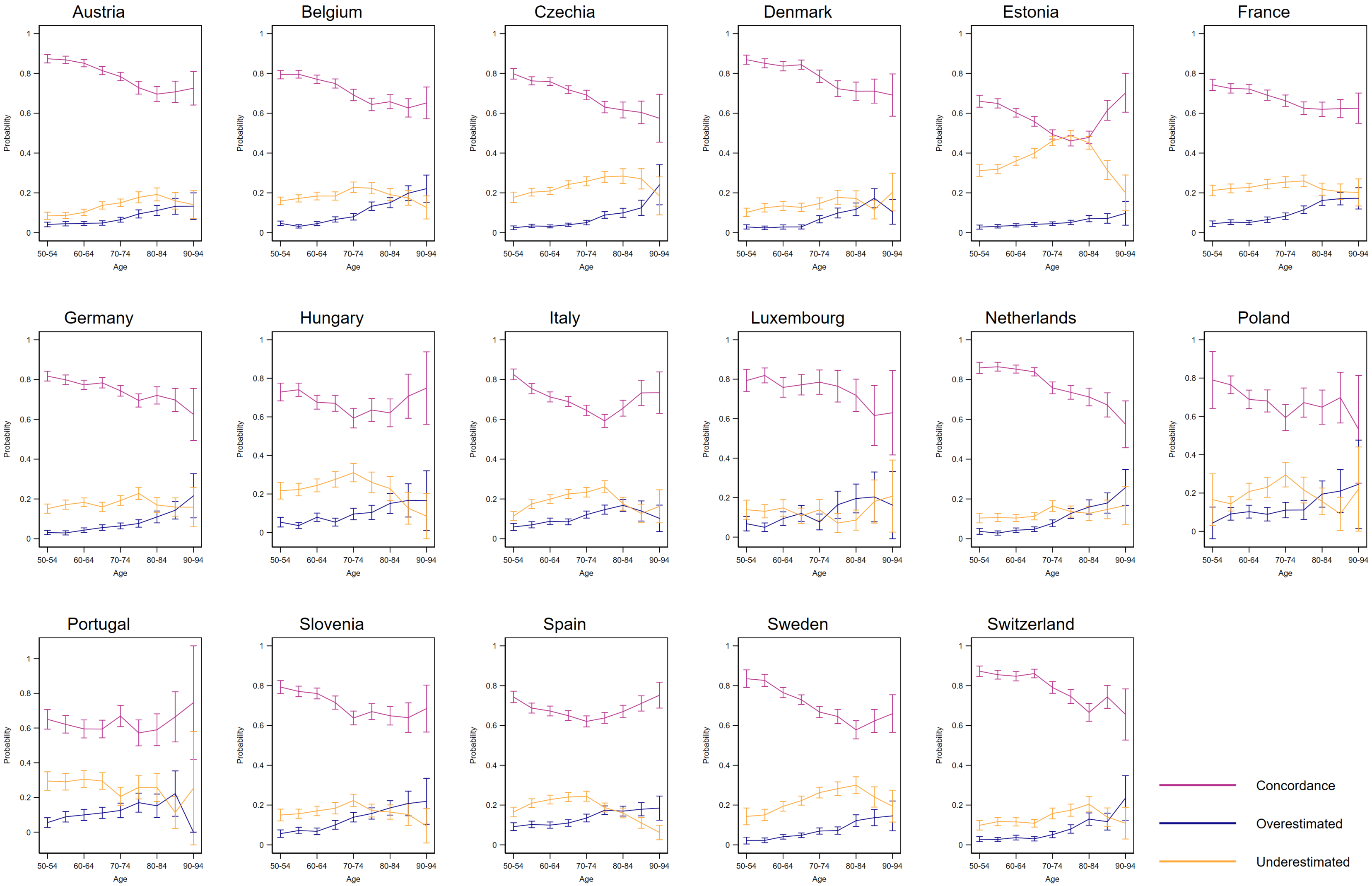


Figure S2 Fig.: Predicted values of concordance between tested and self-reported cognition by country and age



3.3 Health perception and healthcare utilisation (3rd Publication)

The third publication of the dissertation answers RQ 3: "*How does individual health misperception affect healthcare utilisation?*". This is a joint research paper with Mujaheed Shaikh[3] and was published in January 2020 as

Spitzer, S. & Shaikh, M. (2020). Health misperception and healthcare utilisation among older Europeans. *Vienna Institute of Demography Working Papers*, 01/2020.

Abstract: Health perception biases can have serious consequences on health. Despite their relevance, the role of such biases in determining healthcare utilisation is severely underexplored. Here we study the relationship between health misperception, doctor visits, and concomitant out-of-pocket expenditures for the population 50+ in Europe. We conceptualise health misperception as arising from either overconfidence or underconfidence, where overconfidence is measured as overestimation of health and underconfidence is measured as underestimation of health. Comparing objective performance measures and their self-reported equivalents from the Survey of Health, Ageing and Retirement in Europe, we find that individuals who overestimate their health visit the doctor 14% less often than individuals who correctly assess their health, which is crucial for preventive care such as screenings. Lower healthcare utilisation is accompanied by lower out-of-pocket spending (38% less). In contrast, individuals who underestimate their health visit the doctor more often (28% more) and have higher out-of-pocket spending (17% more). We project that underestimating health of the population 50+ will cost the average European country Intl$ 71 million in 2020 and Intl$ 81 million by 2060. Country-specific estimates based on population and demographic projections show that countries such as Germany, Denmark and The Netherlands will experience significantly large costs of such misperception. The results are robust to several sensitivity tests and, more important, to various conceptualisations of the misperception measure.

The article is reprinted by permission of the publisher.

[3]Mujaheed Shaikh is affiliated with the Hertie School, Berlin, Germany.

WWW.OEAW.AC.AT

ÖAW AUSTRIAN ACADEMY OF SCIENCES

VIENNA INSTITUTE OF DEMOGRAPHY

WORKING PAPERS

01/2020

HEALTH MISPERCEPTION AND HEALTHCARE UTILISATION AMONG OLDER EUROPEANS

SONJA SPITZER AND MUJAHEED SHAIKH

Vienna Institute of Demography
Austrian Academy of Sciences
Vordere Zollamtsstraße 3 | 1030 Vienna, Austria
vid@oeaw.ac.at | www.oeaw.ac.at/vid

Vienna Institute of Demography

VID – VIENNA INSTITUTE OF DEMOGRAPHY

Abstract
Health perception biases can have serious consequences on health. Despite their relevance, the role of such biases in determining healthcare utilisation is severely underexplored. Here we study the relationship between health misperception, doctor visits, and concomitant out-of-pocket expenditures for the population 50+ in Europe. We conceptualise health misperception as arising from either overconfidence or underconfidence, where overconfidence is measured as overestimation of health and underconfidence is measured as underestimation of health. Comparing objective performance measures and their self-reported equivalents from the Survey of Health, Ageing and Retirement in Europe, we find that individuals who overestimate their health visit the doctor 14% less often than individuals who correctly assess their health, which is crucial for preventive care such as screenings. Lower healthcare utilisation is accompanied by lower out-of-pocket spending (38% less). In contrast, individuals who underestimate their health visit the doctor more often (28% more) and have higher out-of-pocket spending (17% more). We project that underestimating health of the population 50+ will cost the average European country Intl$ 71 million in 2020 and Intl$ 81 million by 2060. Country-specific estimates based on population and demographic projections show that countries such as Germany, Denmark and The Netherlands will experience significantly large costs of such misperception. The results are robust to several sensitivity tests and, more important, to various conceptualisations of the misperception measure.

Keywords
Healthcare utilisation, health perception, overconfidence and underconfidence, doctor visits, out-of-pocket expenditures, SHARE data.

Authors
Sonja Spitzer (corresponding author), Wittgenstein Centre for Demography and Global Human Capital (Univ. Vienna, IIASA, VID/ÖAW), Austria. Email: sonja.spitzer@iiasa.ac.at

Mujaheed Shaikh, Hertie School, Germany. Email: shaikh@hertie-school.org

Acknowledgements
We are very grateful to Daniela Weber, Miguel Sanchez-Romero, Michael Kuhn, Raf Van Gestel, Anne Goujon, Marcel Bilger, Elsa Fontainha, and Monika Oczkowska for their valuable input. Also, we thank the participants of the 16th National Conference on Health Economics in Lisbon, the participants of the 6th International Workshop on the Socio-Economics of Ageing in Lisbon, the participants of the Wittgenstein Centre Conference 2019 in Vienna, and the participants of the Annual Meeting of the Austrian Economic Association 2020 in Vienna for their comments.

Health Misperception and Healthcare Utilisation among Older Europeans

Sonja Spitzer and Mujaheed Shaikh

1 Introduction

Biased perception of one's own ability is a hallmark of human nature. The literature in psychology, economics, and evolutionary biology has repeatedly demonstrated this phenomenon. Zell & Krizan (2014) conducted a meta-synthesis across different scientific areas and concluded that people have only moderate knowledge of their ability. Johnson & Fowler (2011) presented an evolutionary model of one such bias, namely, overconfidence, and the conditions under which it prevails. Such biases have significant implications for education, labour market outcomes, savings, investment choices, and political decisions (Anderson et al. 2017, Ortoleva & Snowberg 2015, Reuben et al. 2017). They are particularly relevant for health, as they can directly affect risk for accident and injury (Preston & Harris 1965, Sakurai et al. 2013) and have serious long-lasting effects on wellbeing and mortality. Recent work in this domain shows that overconfidence is related to engagement in risky health behaviours (Arni et al. 2019).

Despite the relevance of biased perception for health, its role in healthcare seeking is largely unexplored. Here we study the relationship between misperception of one's own health and future healthcare utilisation and medical expenditures. We categorise misperception as arising from either overconfidence or underconfidence in one's own health. Following the literature in psychology, we measure overconfidence as the overestimation of one's actual health and measure underconfidence as the underestimation of one's actual health (Moore & Healy 2008). It is a priori ambiguous how over- or underconfidence might relate to healthcare use. On the one hand, individuals who overestimate their health may be less likely to visit the doctor when necessary, seek medical attention, or receive timely screenings because they believe their health is perfect. These individuals might also engage in more physical activity, which decreases healthcare utilisation (Rocca et al. 2015). On the other hand, the same individuals might engage in activity or behaviour detrimental to health and thus end up in the hospital more often. For example, older individuals who overestimate their mobility are more prone to fall-induced injuries (Sakurai et al. 2013). Similarly, individuals who underestimate their health may overutilise healthcare services by seeking care and purchasing relatively more medication when it is not necessary—at least in the short run. In the long run, however, they might need less care and use fewer services because of their frequent doctor visits and timely diagnoses. Assessing the relationship between health perception and healthcare utilisation thus remains an empirical task that we undertake in this study.

Measuring over- or underconfidence bias in health is anything but trivial. It requires a subjective health measure and its objective equivalent, the lack of which often dissuades researchers from engaging in such research. We use a novel indicator to measure over- and underconfidence that is derived from the objective performance measures in the Survey of Health, Ageing and Retirement (SHARE). We analyse differences between subjective and objective health based on individuals' self-reported and tested ability to stand up from a chair. Individuals who subjectively report being able to stand but objectively are unable to do so are classified as overconfident, whereas those who subjectively report being unable to stand but objectively are able to are classified as underconfident. Individuals who do not differ in their subjective report and objective assessment are classified as concordant. Prior research has shown the chair stand test to be a good predictor of overall health (Ferrer et al. 1999, Sainio et al. 2006, Pinheiro et al. 2016, Spitzer & Weber 2019). Our approach distinguishes our measure of overconfidence from overplacement and overprecision because we focus only on individual judgements of completing a task rather than on relative comparisons with others or the estimated accuracy of such judgements (Moore & Healy 2008).

To assess utilisation, we use self-reported data on the annual number of doctor visits, which includes emergency room visits and outpatient clinic visits. Using count models, a rich set of controls, and longitudinal data, we find that relative to individuals who achieve concordance (i.e., those who estimate their health accurately), individuals who underestimate their health visit the doctor more often (approximately two more visits per year). In contrast, individuals who overestimate their health visit the doctor less often. We also analyse concomitant out-of-pocket (OOP) expenditures via log-Gamma models and find that individuals who underestimate their health have higher expenses, whereas individuals who overestimate their health have lower expenses. Our results are not biased by other individual characteristics, such as education, age, employment, or marital status, nor are they a manifestation of the inverse relationship between healthcare utilisation and the estimation of one's health as already stated. The results are robust to different model specifications, estimation methods, and measures of health perception.

We use data from 15 European countries from the SHARE survey, which provides other advantages besides a measure of confidence. First, the longitudinal nature of the survey allows us to assess the relationship between confidence today and healthcare utilisation in the next wave of the survey. Thus, an important source of bias in our estimates—reverse causality—is not a first-order issue in our analyses. Second, utilising health services is conditional on having access to such services; a fair comparison of utilisation requires no significant difference in accessibility among the entities being compared. Universal coverage in European countries ensures that everyone has a certain level of access to the health system, unlike in the United States (OECD & European Commission 2018). Finally, Europe is a policy-relevant setting because of its rapidly ageing population (Lutz et al. 2003, Eurostat 2019) and fiscal pressures to reduce expenditures and unnecessary care (Christensen et al. 2009, European Commission 2018).

To quantify the public expenditure associated with health misperception, we perform a back-of-the-envelope calculation of the costs of health misperception. We project that underestimating health will cost the average European country Intl$ 71 million in 2020 and Intl$ 81 million by 2060. Although overestimating health results in negative costs due to lower numbers of doctor visits, these are in the short run only. In the long run, overestimation may result in individuals skipping timely screening and preventive care and lead to worse health, resulting in higher healthcare expenditures.

The contribution of this study is twofold. First, we introduce and advance a measure of health misperception in the health economics literature. Our measure of over- and under-confidence is simple and easy to calculate and an accurate indicator of health status. The medical literature has shown the chair stand test to be strongly correlated with physical health (Ferrer et al. 1999, Sainio et al. 2006, Pinheiro et al. 2016). Moreover, it is regularly performed in other surveys, such as the English Longitudinal Study of Ageing, which provides the opportunity to study different settings and make subsequent comparative analyses between countries.

Second, we contribute to at least two strands of the literature in health economics. The literature has repeatedly shown that individuals frequently over- or underestimate their own health status (Bago d'Uva et al. 2008, Beaudoin & Desrichard 2011, Coman & Richardson 2006, Furnham 2001, Jürges 2007). In addition, health perception differs by sociodemographic characteristics such as age (Srisurapanont et al. 2017, Crossley & Kennedy 2001), gender (Schneider et al. 2012, Merrill et al. 1997), country of residence (Spitzer & Weber 2019, Capistrant et al. 2014, Jürges 2007), education (Bago d'Uva et al. 2008, Choi & Cawley 2017), and race (Jackson et al. 2017). The difference between subjective and predicted survival probability affects healthcare utilisation (Bíró 2016*a*), and individuals with higher expected longevity are more likely to go for cancer screening (Picone et al. 2004), suggesting that health perception affects healthcare utilisation. Our paper contributes to this strand by directly studying over- and underconfidence in one's own health.

It also contributes to the literature on the determinants of healthcare use. In explaining variation in health expenditures and healthcare utilisation, this literature focuses on either the supply side (i.e., provider confidence and precision) (Baumann et al. 1991, Berner & Graber 2008, Cutler et al. 2013, Meyer et al. 2013) or easily observable demand characteristics (e.g., age, gender, income, social class, employment and education) (Bíró 2013, Cameron et al. 2010, Tavares & Zantomio 2017, Vallejo-Torres & Morris 2013, Van Doorslaer et al. 2004, Zhang et al. 2018). Our paper makes a novel contribution by extending this literature to assess a difficult-to-observe demand variable that has consistently been shown to affect health.

The remainder of this paper is structured as follows. In Section 2, we describe the data and variables. In Section 3, we introduce our methodology. Section 4 presents and discusses the results, Section 5 describes a range of robustness analyses, Section 6 provides estimates for the total public cost of health misperception, and Section 7 concludes the paper.

2 Data and Descriptive Statistics

We analyse the relationship between health misperception and healthcare utilisation based on SHARE, a representative cross-country panel study of noninstitutionalised individuals ages 50 and older as well as their younger spouses (Börsch-Supan et al. 2013).[1] The survey provides rich information on health, socioeconomic background, employment, and social networks based on about 380,000 interviews with around 140,000 individuals. It is particularly well suited for studying European countries, as the data are ex-ante harmonised. Also, because it focuses on older individuals, who generally have higher healthcare needs than the young, it is the ideal data source for our analyses. SHARE was previously used to anal-

[1]This study uses data from SHARE Waves 1, 2, 4, 5, and 6 (DOIs: 10.6103/SHARE.w1.700, 10.6103/SHARE.w2.700, 10.6103/SHARE.w4.700, 10.6103/SHARE.w5.700, 10.6103/SHARE.w6.700). SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N211909, SHARE-LEAP: GA N227822, SHARE M4: GA N261982), and Horizon 2020 (SHARE-DEV3: GA N676536, SERISS: GA N654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C), and various national funding sources is gratefully acknowledged (see www.share-project.org).

yse healthcare utilisation by, among others, Bíró (2014), Bolin et al. (2009), Paccagnella et al. (2013), and Tavares & Zantomio (2017).

2.1 Sample Construction

The chair stand test, which we use to determine our measure of over- and underconfidence, is used only in SHARE Wave 2 (2006/2007) and Wave 5 (2013). Because we are assessing the relationship between this measure of confidence and healthcare utilisation in the next wave, our dependent variables, namely, annual number of doctor visits and concomitant OOP expenditures, are taken from the next waves, that is, Wave 4 (2010–2012)[2] and Wave 6 (2015) (Börsch-Supan 2019*b*,*c*). Hence, we treat the data as pooled cross-sections by matching individuals' misperception at Waves 2 and 5 (w) with their utilisation at Waves 4 and 6 ($w+1$), respectively.

Our dependent variables are taken from wave $w+1$, which is why we drop all observations that do not provide information on doctor visits at wave $w+1$. This affects mostly respondents who participated in Wave 2 but not in the subsequent Wave 4 or respondents who participated in Wave 5 but not in the subsequent Wave 6. We also exclude all respondents younger than 50 years and all observations based on proxy respondents. Overall, this results in 58,897 observations from 15 European countries, namely, Austria, Belgium, Czechia, Denmark, Estonia, France, Germany, Italy, Luxembourg, The Netherlands, Poland, Slovenia, Spain, Sweden, and Switzerland. The sample for OOP payments is smaller (41,868 observations), as OOP payments were not captured in Wave 4 (Section 2.2.2).

Based on their results on the chair stand test, we categorise individuals into three groups: those who achieve concordance (i.e. subjectively report having no problem standing up from the chair and objectively are able to or subjectively report having problems standing up from the chair and objectively are not able to), those who are overconfident (i.e., overestimate their health; subjectively report being able to stand up but objectively are unable to), and those who are underconfident (i.e., underestimate their health; subjectively report being unable to stand up but objectively are able to). With concordance as the reference category, the sample is split into two groups: those who are overconfident and those who are underconfident. Further details are provided in Section 2.3.

For the main analysis, health misperception is based on the chair stand variables, because they are binary and therefore clearly indicate whether an individual is unimpaired or im-

[2]SHARE Wave 3 focuses on people's life histories and thus is not utilised in our analyses.

paired. For robustness, we use additional measures of health perception based on subjective cognition and walking ability and their objective counterparts. We therefore add more waves to the analyses for robustness (Section 5.5).

2.2 Outcome Variables

In line with the literature, we use the annual number of doctor visits as a proxy for healthcare utilisation (see Bago d'Uva & Jones 2009, Bíró 2016*b*, Bolin et al. 2009, Lugo-Palacios & Gannon 2017, Tavares & Zantomio 2017, Zhang et al. 2018, among others). By analysing this number, we are able to capture the effects of health perception on public expenditures, as doctor visits are frequently subsidised by the public. In addition, doctor visits are good indicators of healthcare seeking in general and preventive healthcare and screenings in particular. In addition to doctor visits, we analyse annual OOP payments for doctor visits, which allows us to analyse the effects of health perception on private healthcare expenses.

2.2.1 Annual Doctor Visits

The annual number of doctor visits, emergency room visits, and outpatient clinic visits is ascertained by answers to the following question: "Now please think about the last 12 months. About how many times in total have you seen or talked to a medical doctor or qualified/ registered nurse about your health? Please exclude dentist visits and hospital stays, but include emergency room or outpatient clinic visits." The survey question is phrased almost identically in Waves 4 and 6; however, the words "or qualified/registered nurse" are excluded in Wave 4. For this and other reasons, we run separate estimations for each wave as a sensitivity analysis (Section 5.4).

The number of doctor visits is top-coded at 98 visits per year. On average, individuals in our sample visit the doctor seven times per year. The median, however, is lower (five times), which demonstrates the variable's strong right-skewness (Table 2). Naturally, individuals who suffer from chronic diseases or activity limitations visit the doctor more frequently than healthy individuals; thus, the number of doctor visits also increases with age. Gender differences in doctor visits are clear: Women have more annual doctor visits than men. A socioeconomic gradient is also observed with respect to education: The number of doctor visits decreases as education increases. It is interesting that individuals with supplementary insurance have fewer doctor visits than those without, a finding quite contrary to the literature, which predicts moral hazard with supplemental insurance (Coulson et al. 1995, Buchmueller et al. 2004) (see Table A.1 in the Appendix).

2.2.2 OOP expenditures for doctor visits

If participants report that they have seen or talked to a doctor, they are asked, "Did you pay anything yourself for your doctor visits (in the last twelve months)? Please also include expenses for diagnostic exams, such as imaging or laboratory diagnostics." If they answer "yes", they are then asked, "Overall, how much did you pay yourself for your doctor visits (in the last twelve months), that is how much did you pay without getting reimbursed by (a health insurance/ your national health system/ a third party payer)?" The amount of OOP payments is based on the latter question; it is set to zero if the respondent did not visit a doctor at all or if he or she claims zero payments for doctor visits. All values are presented in Euros. Implausibly large values are set to missing, as suggested by SHARE (Jürges 2015). This affects 3,006 observations.

OOP payments are available in Wave 6 but not in Wave 4; thus, we assess the association between health perception at Wave 5 (w) and OOP expenditures at Wave 6 ($w+1$) only. Consequently, the sample is smaller for analyses of OOP payments than those of doctor visits. Because potential deductibles include expenditures for not only doctor visits but also other healthcare services, such as dentist visits and hospital stays, we do not consider deductibles when calculating the OOP expenditures variable.

The mean OOP expenditure is 73 Euros per year. However, 61% of the participants have zero OOP payments at Wave 6; thus, the median is zero (Table 2). It is interesting that OOP payments do not increase with the number of chronic diseases or activity limitations, but educational attainment has a strong positive correlation with OOP expenditures. Furthermore, mean OOP payments vary substantially between countries: They are highest in Luxembourg, Switzerland, Italy, and Austria, reflecting differences in utilisation and/or cost-sharing mechanisms (Paccagnella et al. 2013) (see Table A.2 in the Appendix).

2.3 Explanatory Variable: Health Perception

Following the literature in psychology, our measure of misperception relates to the most common interpretation of over- and underconfidence, namely, over- and underestimating one's performance, actual ability, chance of success, or level of control (Moore & Healy 2008). Assuming an underlying true level of health, we group individuals according to their perception of their health status. More specifically, we differentiate among individuals who perceive their health status correctly (concordance), those who believe that they are healthier than they really are (overestimation), and those who believe that they are unhealthier than they really are (underestimation). The true level of health is proxied by objective perfor-

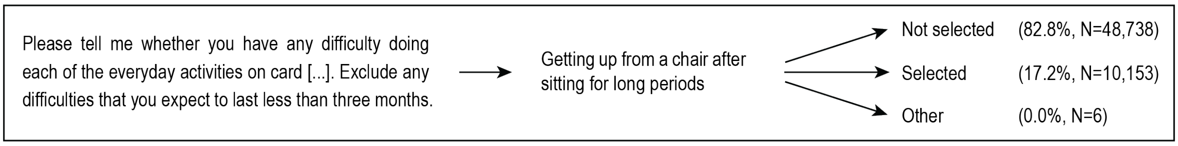


Figure 1: Survey question ascertaining subjective impairment (response category proportions in brackets)

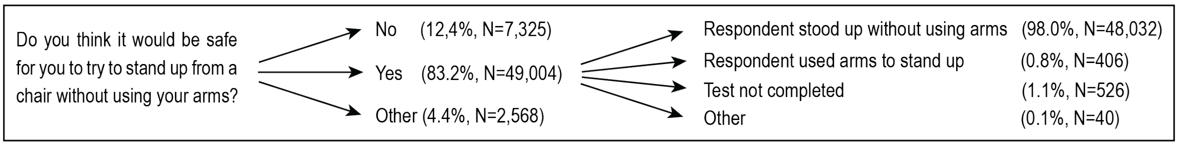


Figure 2: Sequence of questions ascertaining objective impairment (response category proportions in brackets)

mance measures data based on physical performance measures. This objective information about the respondent's health is matched with the respondent's subjective assessment of his or her health, thus revealing whether that individual's beliefs are correct or not.

SHARE provides several objective performance measures that can be utilised as proxies for true health. The measure most suited to analysing differences between objective and subjective health is the ability to stand up from a chair, as this self-assessed variable relates directly to its tested equivalent. This measure has been used previously in Spitzer & Weber (2019). In additional analyses, we also observe the differences between subjective and objective cognition as well as subjective and objective walking ability (Section 5.5).

To evaluate subjective ability to get up from a chair, survey participants are asked whether they have difficulties getting up from a chair. Figure 1 provides the detailed survey question. Individuals are considered subjectively impaired if they report difficulties getting up from a chair and subjectively unimpaired if they do not. Overall, 17.0% of the survey participants in our sample are considered subjectively impaired. Both the impaired and unimpaired groups are then subjected to the objective assessment.

In the objective assessment, individuals are asked to physically stand up from a chair.[3] The chair stand test is introduced with the interviewer saying, "The next test measures the strength and endurance in your legs. I would like you to fold your arms across your chest and sit so that your feet are on the floor; then stand up keeping your arms folded across your chest. Like this ..." The exact sequence of questions leading to the chair stand test is shown in Figure 2. Individuals are considered objectively unimpaired if they stand up

[3] It is important to note that the chair stand test in Wave 2 was only conducted among those younger than 76 years. Thus, the sample is younger than 76 for any country that participated only in Wave 2.

Table 1: Overview health perception categories

	Objectively			
Subjectively	Unimpaired		Impaired	
Unimpaired	Pos. concordance:	87.6%	Overestimating:	56.9%
Impaired	Underestimating:	12.4%	Neg. concordance:	43.1%
Total		**100.0%**		**100.0%**

Note: No weights applied

without using their arms and objectively impaired if they are not able to stand up from the chair, if they have to use their arms to stand up, or if they think it is unsafe to try to stand up from the chair.

Following the subjective report of impairment (i.e., unimpaired or impaired) and the subsequent objective test, individuals can either achieve concordance, overestimate their own health, or underestimate their own health. If they subjectively report being unimpaired but are objectively impaired, they overestimate their health. Likewise, if they subjectively report being impaired but are objectively unimpaired, they underestimate their health. Although the categorisation of over- and underestimation is straightforward, the categorisation of concordance (i.e., accurate beliefs about their health status) requires further consideration. Given true (objective) health, it is important to distinguish between two types of concordance. Individuals with a poor health status (i.e., objectively impaired) are classified as "negative concordance" if they also subjectively report being impaired. Likewise, individuals with a good health status(i.e., objectively unimpaired) are classified as "positive concordance" if they also subjectively report being unimpaired. The four health perception outcomes are shown in Table 1.

Distinguishing between the two types of concordance ensures that we use the appropriate reference category for over- and underestimation in regression analyses. Overestimation can only be measured in the group whose objective health is impaired yet who subjectively report being unimpaired. Therefore, an appropriate group of individuals to compare to are those who are also objectively impaired (i.e., negative concordance). Underestimation can only be measured in the group whose objective health is unimpaired yet subjectively report impaired. The appropriate comparator for these individuals is the group that is also objectively unimpaired. This separation of the concordance group also provides an important empirical advantage; it ensures that we compare like with like in terms of true initial health thereby ridding ourselves of an important source of endogeneity, namely variation in health that can determine utilisation.

Table 2: Summary statistics

	N	Mean	Std. Dev.	Min.	Max.	Median
Healthcare utilisation						
Annual number of doctor visits at w+1	58,764	7.332	9.423	0	98	5
Annual out-of-pocket expenditure for doctor visits at w+1	39,988	73.349	298.196	0	47,500	0
Health perception						
Positive concordance (1 = yes)	56,152	0.743	0.437	0	1	1
Underestimating (1 = yes)	56,152	0.101	0.302	0	1	0
Negative concordance (1 = yes)	56,152	0.060	0.237	0	1	0
Overestimating (1 = yes)	56,152	0.096	0.295	0	1	0
Impairment						
Subjective impairment (1 = impaired)	58,758	0.170	0.376	0	1	0
Objective impairment (1 = impaired)	56,157	0.156	0.363	0	1	0
Health variables						
Number of chronic diseases at w	58,702	1.145	1.217	0	10	1
Number of chronic diseases at w+1	58,754	1.207	1.231	0	9	1
Number of activity limitations at w	58,755	0.357	1.177	0	13	0
Number of activity limitations at w+1	58,752	0.490	1.451	0	13	0
Control variables						
Age (in number of years)	58,764	64.521	9.765	50	100	63
Gender (1 = female)	58,764	0.545	0.498	0	1	1
Low education (1 = yes)	57,979	0.430	0.495	0	1	0
Medium education (1 = yes)	57,979	0.369	0.483	0	1	0
High education (1 = yes)	57,979	0.201	0.401	0	1	0
Is retired (1 = yes)	58,471	0.509	0.500	0	1	1
Is married (1 = yes)	56,883	0.680	0.466	0	1	1
Household income (in Euros per year)	58,764	46,569.89	76,244.77	0	1,200,000	24,000
Health access (1 = difficult)	39,120	0.163	0.370	0	1	0

Note: Calibrated cross-sectional individual weights are applied. For more detailed cross-tabulations see Tables A.1 to A.3 in the Appendix.

As shown in Table 1, in the objectively impaired group, 57% overestimate their health status; in the unimpaired group, only 12% underestimate. The large number of people reporting overconfidence is not surprising, as it has been documented in psychology and evolutionary theory as being favoured by natural selection and providing adaptive gains. Individuals tend to be overconfident because it increases morale and ambition and may thus improve potential (Johnson & Fowler 2011). Furthermore, our sample consists of older people, among whom overconfidence is particularly prevalent (Idler 1993, Spitzer & Weber 2019) and is seen as a resilience strategy to maintain a positive self-image (Brandtstädter & Greve 1994).

2.4 Additional Control Variables

We control for a range of variables that might otherwise confound our results. Summary statistics for these control variables are provided in Table 2, and cross-tabulations of control variables, doctor visits, health expenditures, and health perception are provided in Tables A.1 to A.3 in the Appendix. Most important, we control for other health factors at wave

w. In particular, we include the number of chronic diseases and the number of limitations in instrumental activities of daily living (IADLs) in our model. Chronic conditions that we consider are heart problems, high blood pressure or hypertension, high blood cholesterol, stroke or cerebral vascular disease, diabetes, chronic lung diseases, cancer, stomach or duodenal ulcer, Parkinson's disease, cataracts, hip fractures, other fractures, and Alzheimer's disease. A total of 35% of the sample have no chronic diseases at wave w; the weighted mean is 1.2 diseases. IADLs that we consider are difficulties dressing, walking across a room, bathing or showering, eating and cutting up food, getting in or out of bed, using the toilet, using a map, preparing a hot meal, shopping for groceries, making a telephone call, taking medications, doing work around the house or garden, and managing money. A total of 81% of the sample have no IADLs at wave w; the weighted mean is 0.5 IADLs. We only consider chronic diseases and IADLs that are included in both Wave 2 and Wave 5.

We also control for sociodemographic characteristics, as they are expected to influence health perception as well as healthcare utilisation (Avitabile et al. 2011, Lange 2011). In particular, we include age and age squared, gender, and educational attainment according to the International Standard Classification of Education (Eurostat 2018). Because pensioners appear to have higher healthcare utilisation (Bíró 2016*b*, Zhang et al. 2018), we also consider whether an individual is retired as opposed to all other employment options (employed, self-employed, unemployed, permanently sick or disabled, homemaker, other). Also, we control for whether the survey participant is married or in a registered partnership as opposed to never married, divorced, or widowed.

The effects of economic resources on healthcare utilisation are considered via equivalised household income. Because there are many missing values for household income in SHARE, the data set comes with two additional imputed variables. We use one of these imputed variables in our model and conduct a robustness analysis with the other (Section 5.4). We equivalise household income by using the square root scale, in which household income is divided by the square root of household size. Using the Organisation for Economic Co-operation and Development equivalence scale is not feasible, as children cannot be identified unambiguously. Furthermore, we use a cube root transformation to normalise the skewed income distribution (Cox 2011). Standard log normalisation is not feasible because of the substantial number of zero values.[4] We run a robustness analysis in which we use equivalised household income that was not normalised (Section 5.4).

[4]Results are robust to dropping observations with zero values in household income.

3 Method

Ideally, we would randomly assign health perception to individuals to elicit causal effects of (mis)perception on healthcare utilisation and expenditures. In the absence of such random assignment, we rely on the panel dimension of the SHARE survey and control for a rich set of variables to account for confounding effects and bias due to reverse causation. Health perception is expected to affect healthcare utilisation, but the opposite mechanism, that healthcare utilisation precedes health perception, appears plausible too. For example, individuals who frequently visit the doctor are more likely to achieve concordance, as they receive more information about their health status. To overcome potential endogeneity, we analyse the effects of current health perception (wave w) on future healthcare utilisation (wave $w+1$).

The main outcome variable—annual doctor visits—is strongly skewed to the right, yet without severe mass at zero. To accommodate this, we use a negative binomial model with mean dispersion, which is used frequently in the healthcare literature. We refrain from using a simple Poisson model, as the variance in the outcome variable is much larger than its mean. However, we perform robustness analyses using different models (Section 5.4). Thus, the number of doctor visits of individual i at wave $w+1$ ($\text{DOCTOR}_{i,w+1}$) is assumed to follow a Poisson distribution but with a negative binomial specification for which each individual unit has a separate, Gamma-distributed mean. More specifically,

$$\text{DOCTOR}_{i,w+1} \sim \text{Poisson}(\mu_{i,w+1}), \tag{1}$$

where

$$\mu_{i,w+1} = \exp(\beta \times \text{HEALTH PERCEPTION}_{i,w} + \gamma \times \text{HEALTH}_{i,w} + \delta \times \text{X}_{i,w} + \nu_i), \tag{2}$$

and

$$\exp(\nu_i) \sim \text{Gamma}(1/\alpha, \alpha) \tag{3}$$

HEALTH PERCEPTION is a binary variable that indicates whether individual i achieves concordance or misperceives his or her health at wave w. The vector HEALTH includes two variables, namely the number of chronic diseases in period w as well as the number of

IADLs in period w (thus, in the same period as health perception). The vector of control variables $X_{i,w}$ includes age and age squared, the individual's gender, educational attainment, household income, and control dummies for the survey wave as well as for the country of residence. The terms β, γ, and δ represent coefficients.

As discussed earlier, the sample is split into individuals who are overconfident (i.e., overestimate their health status) and individuals who are underconfident (i.e., underestimate it). The regression coefficients are therefore interpreted relative to those who estimate their health correctly (i.e., achieve concordance). For heterogeneity analyses, we further split the sample by gender, country, and number of chronic diseases.

When analysing the effects of health perception on OOP expenditures, we use a nonlinear model with a log link and Gamma family instead of the negative binomial model to account for the continuous nature of the outcome variable as well as for the excess zeros (Deb & Norton 2018). The specification of the variables included, however, remains identical to that described in Equation 2.

A total of 32% of the survey respondents participate in Waves 2, 4, 5, and 6, which allows us to analyse how health perception varies between Wave 2 and Wave 5 for these observations. For the majority (75%), health perception does not vary with age. If health perception changes, the most common changes are from underestimating to concordance (7.6%), from concordance to overestimating (6.5%), and from concordance to underestimating (5.6%). Because there is not enough variation in health perception within individuals, we refrain from using individual fixed effects in our analyses.

In Section 5.5, we explore whether our results are robust to different specifications of health perception. In particular, we estimate Equation 2 using cognition and the ability to walk as bases for the health perception variable.

4 Results

We first present the main results for the link between health misperception and healthcare utilisation and expenditures. We then examine heterogeneity in the relationships and present the results of important robustness analyses. Finally, we provide results for alternative measures of health perception bias.

4.1 Main Results

Table 3: Annual number of doctor visits and OOP expenditures for doctor visits at w+1

	(1) Objectively Unimpaired Doctor visits	(2) Objectively Unimpaired OOP	(3) Objectively Impaired Doctor visits	(4) Objectively Impaired OOP
Health perception (ref.: concordance)				
Underestimating	0.244***	0.166*		
	(0.018)	(0.077)		
Overestimating			-0.146***	-0.378**
			(0.029)	(0.138)
Chronic diseases	0.181***	0.118***	0.149***	0.118*
	(0.005)	(0.031)	(0.009)	(0.055)
Activity limitations	0.096***	0.088	0.048***	0.031
	(0.010)	(0.045)	(0.007)	(0.030)
Age	-0.001	0.082	0.021	0.165*
	(0.011)	(0.058)	(0.018)	(0.082)
Age squared	0.000	-0.001	-0.000	-0.001*
	(0.000)	(0.000)	(0.000)	(0.001)
Woman	0.042***	0.125*	0.014	0.422***
	(0.013)	(0.060)	(0.028)	(0.117)
Educ. group (ref.: low)				
Medium	0.006	0.419***	-0.006	0.112
	(0.016)	(0.072)	(0.033)	(0.126)
High	-0.003	0.881***	-0.087*	0.544***
	(0.018)	(0.093)	(0.042)	(0.152)
Retired	0.029	-0.031	0.014	0.398
	(0.017)	(0.104)	(0.033)	(0.228)
Married	-0.034*	0.029	0.019	0.405***
	(0.015)	(0.072)	(0.030)	(0.115)
Equiv. hh income (cube root)	-0.001	0.014***	-0.001	0.006
	(0.001)	(0.003)	(0.002)	(0.007)
Wave 5	-0.089***		-0.046	
	(0.015)		(0.038)	
Constant	1.507***	0.626	1.434*	-1.232
	(0.356)	(1.963)	(0.646)	(2.828)
Country dummies	Yes	Yes	Yes	Yes
N	46,067	32,564	7,801	5,603
AIC	260,957	297,483	50,545	50,035
BIC	261,202	297,684	50,740	50,194

Note: "Unimpaired" refers to the sample that is objectively unimpaired, i.e. able to stand up from the chair and "Impaired" refers to the sample that is objectively impaired, i.e. unable to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave w+1, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. The dependent variable "OOP" is based on annual out-of-pocket payments for doctor visits at wave w+1, i.e. Wave 6. All explanatory variables are taken from wave w, i.e. Wave 5. The coefficients are estimated based on a generalised linear model model with log link and a Gamma family. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

4.1.1 Healthcare Utilisation

Healthcare utilisation is measured by the annual number of doctor visits. Table A.1 of the Appendix shows that overall, individuals who overestimate their health have fewer doctor visits (8.6 visits) compared to their reference group (i.e., negative concordance = 11.9 visits). Similarly, those who underestimate their health have significantly more doctor visits in a year (10.1 visits) compared to their relevant reference group (i.e., positive concordance = 6.2 visits).

The table also shows cross-tabulations by other characteristics of the sample. Using number of chronic conditions and activity limitations as proxies for doctor visits provides two important insights. First, we find that as illness increases, so does the number of doctor visits irrespective of the category of perception bias. Second, at every level of illness, individuals who underestimate their health visit the doctor more often than those who are concordant. Similarly, overall, at almost each level of health (barring a few exceptions), individuals who overestimate their health visit the doctor less often. This shows that despite the same underlying health status, there is variation in doctor visits by health misperception category. Starkly similar results are observed for increasing age.

Although the picture is somewhat mixed across education categories, we observe fewer doctor visits for overestimators relative to their concordant counterparts at every level of education. Similarly, underestimators have higher healthcare utilisation than their concordant counterparts at each level of education. Accessibility to health professionals strongly determines health access; the pattern of utilisation across this variable by our misperception category remains the same as before: Overestimation shows fewer doctor visits, and underestimation shows more visits. Similar results are observed by supplementary insurance status.

These descriptive findings show that despite conditioning on individual characteristics, there is clear variation in healthcare utilisation in the form of doctor visits among the different health perception categories. In the regression analyses, we control for these and other variables such as country dummies. Table 3 shows the regression results. Columns 1 and 3 show the results for the two groups (i.e., overestimators and underestimators categorised based on the objective health status as impaired or unimpaired). All coefficients are to be interpreted relative to the concordance category.

We find a strong and significant association between health misperception and healthcare utilisation. Individuals who underestimate their health visit the doctor 27.6% more often in the subsequent period than individuals who achieve concordance. Computing marginal effects at means shows that this results in approximately two additional doctor visits per year.

We also find a strong and significant link between overestimation and the annual number of doctor visits. Individuals who overestimate their health go to the doctor less often than those who achieve concordance. Overestimating health at wave w results in 13.6% fewer doctor visits at wave $w + 1$ compared to perceiving one's health correctly. The marginal effect at means of overestimating health on healthcare utilisation is approximately 1.3 fewer doctor visits per year.

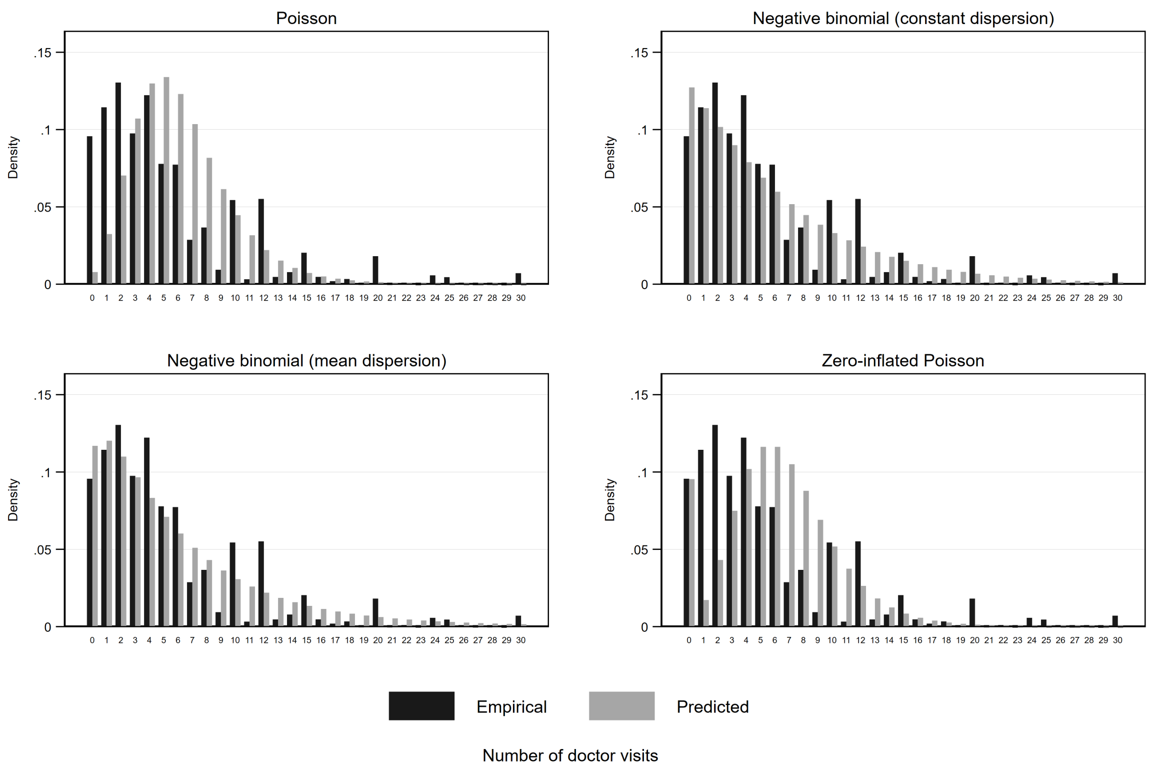
The results for doctor visits in Table 3 are based on a negative binomial model with mean dispersion. Figures 3 and 4 show that this model has the best fit among a simple Poisson model, a negative binomial model with constant dispersion, and a zero-inflated Poisson model.

4.1.2 OOP Expenditures

Individuals who visit a doctor also report OOP expenses, if any, measured in Euros. Table A.2 shows descriptive cross-tabulations of OOP expenditures incurred by other individual characteristics and by misperception category. Although no consistent pattern in OOP spending emerges by the number of chronic conditions, number of activity limitations, or increasing age, women show slightly higher expenditures than men. A clear education gradient is also observed in OOP spending, with higher education relating positively to spending. Similarly, retired individuals spend more than those who are not retired, as do married individuals compared to single ones. It is not surprising that in general individuals with supplementary insurance spend slightly less than those without it, and those with difficulty accessing healthcare spend less than those with no difficulty. Certain countries, such as Switzerland, Luxembourg, and Italy, show exceptionally high OOP expenditures, whereas others, such as France, Denmark, and Eastern European countries, show much lower OOP expenditures, which partly reflects institutional differences in user charges and the coverage of certain services.

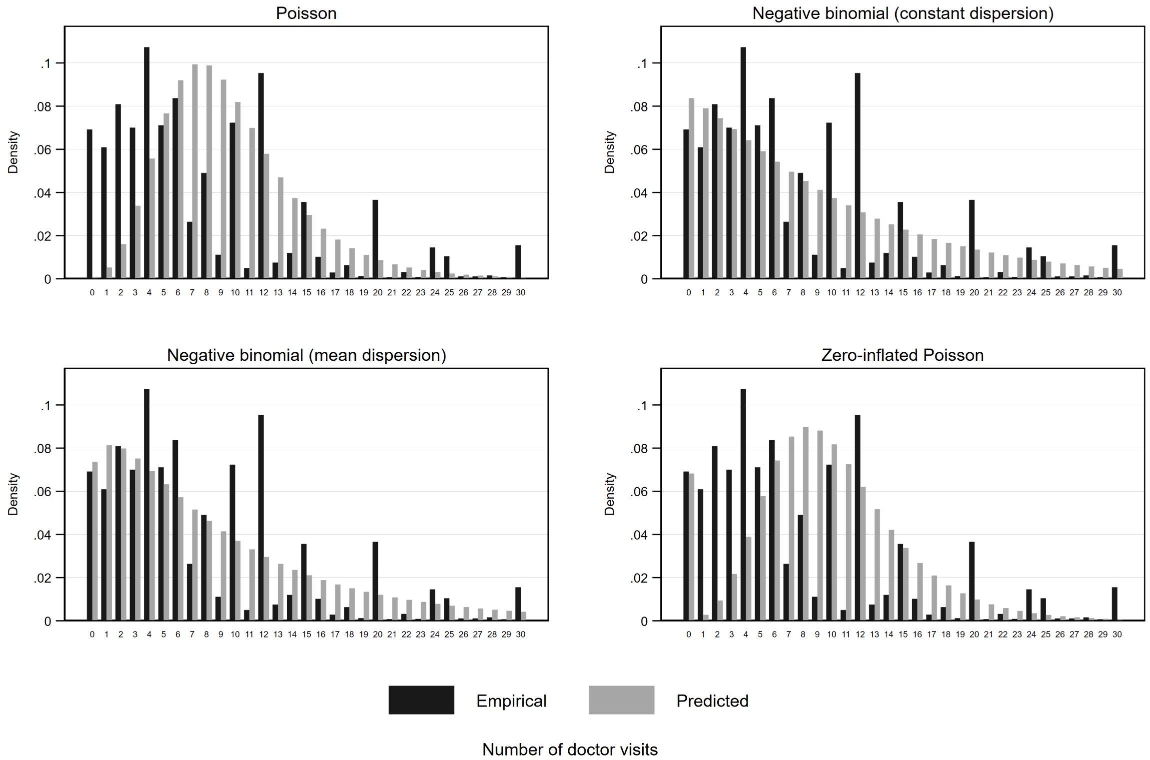
It is interesting that we observe similar patterns as for doctor visits across the misperception categories. At almost every level of chronic conditions, activity limitations, and increasing age, we find that those who underestimate their health have higher OOP spending than those who achieve concordance. The findings for overestimation are somewhat mixed. Underestimating men (women) have slightly lower (much higher) OOP spending than their concordant counterparts. Overestimating men (women) also have slightly lower (higher) OOP spending than their concordant counterparts. Although an education gradient can be seen for underestimators with higher OOP spending compared to the concordant group at

Figure 3: Count model comparison for the annual number of doctor visits in the unimpaired sample, i.e. able to stand up from the chair



Note: Doctor visits are top coded at 98 visits by year. This figure shows only the first 30 doctor visits for better visualisation. Dark bars represent the empirically observed numbers of doctor visits and light bars represent the predicted values based on the respective count model.

Figure 4: Count model comparison for the annual number of doctor visits in the impaired sample, i.e. unable to stand up from the chair



Note: Doctor visits are top coded at 98 visits by year. This figure shows only the first 30 doctor visits for better visualisation. Dark bars represent the empirically observed numbers of doctor visits and light bars represent the predicted values based on the respective count model.

each level of education, the same is not observed consistently for overestimators in the case of lower spending. Although underestimators consistently spend more OOP across the other individual characteristics, mixed findings are seen for overestimators. Regression results controlling for these characteristics in Columns 2 and 4 of Table 3 show that individuals who underestimate their health have significantly higher OOP expenses. On average, expenditures are 16.6% higher for those who underestimate their health compared to those who achieve concordance. In contrast, individuals who overestimate their health spend 37.8% less in OOP expenditures per year relative to their concordant group.

The results for OOP payments are based on a log-Gamma model. According to Akaike's information criterion and the Bayesian information criterion, the log-Gamma model has a better fit than either a log-Gaussian model or a log-Poisson model.

4.2 Heterogeneity of effects

We assess the heterogeneity of our main results in several ways. In particular, we consider gender differences, country specificities, and differences by health status.

4.2.1 Gender Differences

The literature has shown differences in health perception by individual characteristics, most important by gender (Merrill et al. 1997, Schneider et al. 2012). Gender differences in effects of health beliefs on healthcare utilisation may partly explain the well-documented differences in healthcare seeking between men and women, as men tend to have lower healthcare use (Galdas et al. 2005, Mansfield et al. 2003, Schlichthorst et al. 2016). Thus, we assess whether the relationship between health (mis)perception and utilisation also differs between men and women. As noted earlier, Table A.1 shows that, overall, women have slightly more doctor visits annually compared to men; this is true also within the misperception category, but the difference is not large. Furthermore, both under- and overestimating men and women have more doctor visits relative to their respective concordant comparators. In the case of OOP expenditures, however, whereas both under- and overestimating women have higher spending relative to their concordant group, both under- and overestimating men have lower spending relative to their concordant group. However, under- and overestimating women tend to spend more than under- or overestimating men.

Regression analyses by gender reveal that the association between health misperception and the annual number of doctor visits is slightly larger in magnitude for men than for women (Table 4). Marginal effects at means show that men who underestimate their health visit the doctor an additional 1.8 times compared to men who achieve concordance. For women, the difference is an additional 1.5 doctor visits. Men who overestimate their health have 1.5 fewer annual doctor visits compared to men who achieve concordance. For women, it is 1.3 fewer visits. A Wald test, however, reveals that the coefficients for women and men are not statistically different from each other.

4.2.2 Country Specificity

Differences in reporting behaviour by country are well documented (Capistrant et al. 2014, Jürges 2007, Spitzer & Weber 2019). To ensure that our findings are not driven by differential reporting due to cultural biases in reporting health or the oversampling of certain countries in the SHARE survey, we rerun our analyses for each country separately. By and large, we find similar results for all countries, for both under- and overestimation, with the exception of a few countries for which we do not find statistically significant results because of small sample sizes (see Tables A.4 and A.5 in the Appendix). However, it is worth noting that the magnitude of the coefficient for underestimation is much larger in certain countries, such as Denmark, Germany, and The Netherlands, than in others, perhaps reflecting differences in accessibility or other institutional differences in terms of, for example, user charges.

4.2.3 Differences by Health Status

Separating the underestimators and overestimators by objective health status allows us to overcome an important endogeneity concern related to initial health status that affects both health perception and healthcare utilisation. However, because we assess healthcare utilisation at $w+1$, we also assess heterogeneity by health status at $w+1$, which allows us to understand whether current health status drives differential utilisation in any way. Note that current utilisation at $w+1$ will not drive health misperception because we assess misperception at w.

The descriptive statistics in Table A.3 in the Appendix indicate a slight decrease in concordance as the number of chronic diseases increases; however, this trend is far from obvious and might also be due to the correlation between health and age. To disentangle these effects, we run separate regressions for those individuals who do not have any chronic diseases at wave $w+1$ (healthy) and those who report one or more chronic diseases at wave $w+1$ (unhealthy). The results are reported in Table 5. Although health perception affects the doctor visits of impaired individuals with and without chronic diseases similarly, underestimation has a bigger effect on those without chronic diseases than on those with chronic diseases—this is confirmed by a Wald test. However, marginal effects reveal no substantial difference between the healthy and the unhealthy subsamples with respect to the relationship between over- or underestimation and doctor visits. Because we categorise based on

Table 4: Annual number of doctor visits at $w + 1$ by gender

	(1) Objectively Unimpaired Men Doctor visits	(2) Objectively Impaired Men Doctor visits	(3) Objectively Unimpaired Women Doctor visits	(4) Objectively Impaired Women Doctor visits
Health perception (ref.: concordance)				
Underestimating	0.267***		0.229***	
	(0.035)		(0.021)	
Overestimating		-0.139**		-0.136***
		(0.048)		(0.034)
Chronic diseases	0.191***	0.168***	0.176***	0.143***
	(0.008)	(0.014)	(0.007)	(0.012)
Activity limitations	0.096***	0.045***	0.096***	0.050***
	(0.018)	(0.012)	(0.011)	(0.008)
Age	0.018	0.038	-0.008	0.014
	(0.017)	(0.031)	(0.013)	(0.022)
Age squared	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Educ. group (ref.: low)				
Medium	0.042	-0.074	-0.027	0.044
	(0.026)	(0.052)	(0.020)	(0.041)
High	0.017	-0.151*	-0.024	-0.055
	(0.029)	(0.061)	(0.024)	(0.057)
Retired	0.031	0.075	0.026	-0.026
	(0.031)	(0.067)	(0.021)	(0.040)
Married	-0.045	0.032	-0.047*	-0.007
	(0.025)	(0.055)	(0.018)	(0.036)
Equiv. hh income (cube root)	-0.001	0.000	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.002)
Wave 5	-0.116***	0.090	-0.066***	-0.124**
	(0.022)	(0.058)	(0.019)	(0.047)
Constant	0.654	0.604	1.983***	1.858*
	(0.583)	(1.059)	(0.441)	(0.788)
Country dummies	Yes	Yes	Yes	Yes
N	20,693	2,864	25,374	4,937
AIC	116,397	18,426	144,362	32,072
BIC	116,611	18,587	144,582	32,248

Note: "Unimpaired" refers to the sample that is objectively unimpaired, i.e. able to stand up from the chair and "Impaired" refers to the sample that is objectively impaired, i.e. unable to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w + 1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

health (in other words, fix health at the same level) we can conclude that the results are not driven by health differences: Both the healthy group's and the unhealthy group's healthcare utilisation is affected by their health perception in the same direction and to a similar magnitude.

Table 5: Annual number of doctor visits at $w + 1$ by chronic diseases at $w + 1$

	(1) Objectively Unimpaired No chronic dis. at w+1	(2) Objectively Unimpaired Chronic dis. at w+1	(3) Objectively Impaired No chronic dis. at w+1	(4) Objectively Impaired Chronic dis. at w+1
Health perception (ref.: concordance)				
Underestimating	0.351***	0.197***		
	(0.042)	(0.020)		
Overestimating			-0.259***	-0.116***
			(0.066)	(0.031)
Chronic diseases	0.185***	0.093***	0.108***	0.106***
	(0.015)	(0.006)	(0.029)	(0.010)
Activity limitations	0.142***	0.087***	0.073***	0.045***
	(0.025)	(0.010)	(0.015)	(0.007)
Age	-0.027	-0.028*	-0.019	0.006
	(0.018)	(0.012)	(0.039)	(0.020)
Age squared	0.000*	0.000**	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman	0.155***	0.008	0.116	-0.002
	(0.024)	(0.014)	(0.063)	(0.031)
Educ. group (ref.: low)				
Medium	0.022	0.018	0.023	-0.002
	(0.030)	(0.018)	(0.075)	(0.035)
High	0.066	-0.000	0.014	-0.071
	(0.034)	(0.021)	(0.083)	(0.047)
Retired	0.023	0.020	0.094	-0.005
	(0.034)	(0.019)	(0.076)	(0.035)
Married	-0.033	-0.034*	-0.003	0.036
	(0.027)	(0.017)	(0.075)	(0.031)
Equiv. hh income (cube root)	0.002	-0.002*	0.004	-0.001
	(0.001)	(0.001)	(0.003)	(0.002)
Wave 5	-0.046	-0.067***	-0.145	-0.001
	(0.027)	(0.017)	(0.074)	(0.042)
Constant	1.835**	2.821***	1.590	2.320***
	(0.614)	(0.419)	(1.348)	(0.704)
Country dummies	Yes	Yes	Yes	Yes
N	17,362	28,696	1,827	5,973
AIC	84,631	173,306	10,120	40,020
BIC	84,849	173,537	10,274	40,207

Note: "Unimpaired" refers to the sample that is objectively unimpaired, i.e. able to stand up from the chair and "Impaired" refers to the sample that is objectively impaired, i.e. unable to stand up from the chair. The sample "No chronic dis. at $w + 1$" includes those that have zero chronic diseases at wave $w + 1$, whereas "Chronic dis. at $w + 1$" refers to those that have one ore more chronic diseases at wave $w + 1$. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w + 1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

5 Robustness Tests

We conduct a range of robustness analyses to observe whether our results are sensitive to model specifications and sample composition. These results are presented in Tables A.6 and A.7 in the Appendix along with the original model specification (Column 1).

5.1 Income

First, we utilise different income variables. We exchange the first imputed income variable provided by SHARE with the second imputed income variable (Column 2), and we use income that is not normalised with the cube root method but only equivalised (Column 3). These adjustments have no effects on the results. We also replace income with wealth (Column 4), and the results remain robust.

5.2 Wave Specific Analyses

Second, we separate the sample by survey wave to explore whether the slight change in the phrasing of the survey question about doctor visits in Wave 6 (Section 2.2.1), the restriction of the chair stand test to those younger than 76 years in Wave 2 (Section 2.3), or the different time gaps between w and $w+1$ affect the results. The estimates in Table A.9 in the Appendix reveal that the effect of health misperception on healthcare utilisation is slightly stronger at Wave 5 than at Wave 2; however, the difference is not statistically significant according to a Wald test.

5.3 Response Reliability

Third, we exclude anyone diagnosed with Alzheimer's disease, dementia, or another serious memory impairment, as their survey answers might not be reliable (Column 5). The results remain robust, perhaps because the number of individuals observations with severe cognitive impairments in the survey is small.

5.4 Robustness to Further Controls

Finally, we conduct robustness analyses that are only possible for Wave 5, as they include variables only collected in this wave (see Table A.8 in the Appendix). First, we analyse whether differences in access to healthcare affect the number of doctor visits. For this, the household respondent is asked "How easy is it to get to your general practitioner or the nearest health center? Would you say it is very easy, easy, difficult or very difficult?" We dichotomise the variable by comparing the first two and the last two possible answers and add it to the model (Columns 2 and 5). The coefficients show, however, that the results do not depend on access to healthcare. Second, we investigate whether the results are robust to individuals purchasing supplementary health insurance. Although supplementary insurance increases healthcare utilisation (Moreira & Barros Pita 2010, Paccagnella et al. 2013), we find no significant changes in our results (Column 6) when controlling for this variable.

5.5 Additional Measures of Health Perception

For the main analyses, health perception is operationalised based on tested and self-reported ability to stand up from a chair. We also analyse whether the results hold for other health dimensions, in particular health perception concerning cognition and walking ability.

5.5.1 Cognition

Similar to previous work, we use the difference between subjective and objective cognition as an additional measure of health perception (Spitzer & Weber 2019). Objective cognition is operationalised based on a memory test, which is conducted in Waves 4 to 6. In particular, individuals are asked to recall a list of 10 words in any order within a minute.

Subjective cognition is based on the question "How would you rate your memory at the present time?" which is answered on a Likert scale with the categories excellent, very good, good, fair, and poor. Because the subjective cognition variable has more than 80% missing values in Wave 6, we only utilise data from Waves 4 and 5. Hence, the estimates for cognition are based on a different sample. For the main results presented in Section 4, health perception from Waves 2 and 5 is matched with healthcare utilisation from Waves 4 and 6. For the results for cognition, health perception from Waves 4 and 5 is matched with healthcare utilisation from Waves 5 and 6.

Defining cognitive impairment is not as straightforward as defining the ability to stand up from a chair. Whereas the chair stand variables are binary and therefore clearly indicate whether an individual is impaired, both the subjective and objective cognition variables are categorical. Thus, we rely on previous literature to define the threshold marking cognitive impairment. Participants are considered objectively impaired if they recall three words or fewer (Grodstein et al. 2001, Purser et al. 2005). In addition, in robustness analyses, individuals are considered impaired if they recall two words or fewer. Individuals are considered subjectively impaired if they report having a fair or poor memory (Gardner et al. 2017).

Tables 6 provides regression results for this new specification of health perception. The results confirm our earlier findings. Individuals who underestimate their cognitive ability at wave w are more likely to visit the doctor at wave $w+1$ than individuals who achieve concordance between objective and subjective measures of memory. By contrast, survey participants who overestimate their health have fewer annual doctor visits than those who achieve concordance. Modifying the threshold for objective impairment from three to two words changes the magnitude of the coefficient for overestimation but not its sign. The magnitude of the coefficient for overestimation remains virtually identical.

5.5.2 Walking Ability

We also operationalise health perception based on walking ability. Objective walking ability is based on a walking speed test in which participants have to walk a distance of 2.5 m. Individuals are considered objectively impaired if their walking speed is 0.4 m per second or slower. This threshold is in line with the previous literature (Jürges 2007, Steel et al. 2003). Because the test is only conducted in Waves 1 and 2, the analysis is restricted to those waves (Börsch-Supan 2019*a*). The walking speed test is supposed to be conducted only for individuals older than 75 years. However, the data set includes information for those 75 and younger too. The variable has many missing values (∼90%) and thus needs to be handled with caution.

Subjective walking impairment is based on the following question: "Please look at card [. . .]. We need to understand difficulties people may have with various activities because of a health or physical problem. Please tell me whether you have any difficulty doing each of the everyday activities on card [. . .]. Exclude any difficulties that you expect to last less than three months." Participants are coded as having subjectively impaired walking ability if they report difficulty walking 100 m.

When analysing health perception based on walking ability, we do not control for IADLs, as the ability to walk across a room is itself considered an IADL. Also, the second imputed income variable is used for this analysis, as the first one is not available in Wave 1. The robustness analysis in Section 5.4 shows, however, that both variables produce the same results.

Results for the effects of health perception on the annual number of doctor visits based on walking ability are provided in Table 7. The coefficients in Table 7 confirm once again that individuals who underestimate their health have more annual doctor visits than those who assess their health correctly. The results also show that those who overestimate their health have fewer doctor visits. Thus, our results are robust to different specifications of health perception.

Table 6: Health perception based on cognition

	(1) Objectively Unimpaired 3 words	(2) Objectively Impaired 3 words	(3) Objectively Unimpaired 2 words	(4) Objectively Impaired 2 words
Health perception (ref.: concordance)				
Underestimating	0.079***			
	(0.013)			
Overestimating		-0.067*		
		(0.027)		
Underestimating			0.077***	
			(0.012)	
Overestimating				-0.145***
				(0.042)
Chronic diseases	0.194***	0.143***	0.189***	0.147***
	(0.004)	(0.009)	(0.004)	(0.014)
Activity limitations	0.114***	0.057***	0.106***	0.045***
	(0.005)	(0.007)	(0.005)	(0.009)
Age	0.013	0.042*	0.017*	0.055
	(0.009)	(0.021)	(0.008)	(0.031)
Age squared	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman	0.078***	-0.003	0.072***	-0.043
	(0.011)	(0.029)	(0.011)	(0.045)
Educ. group (ref.: low)				
Medium	0.006	0.074*	0.010	0.013
	(0.014)	(0.037)	(0.013)	(0.056)
High	0.013	0.016	0.008	0.026
	(0.016)	(0.059)	(0.016)	(0.100)
Retired	0.033*	-0.046	0.022	0.010
	(0.016)	(0.035)	(0.015)	(0.052)
Married	-0.025	-0.035	-0.028*	-0.030
	(0.013)	(0.031)	(0.012)	(0.045)
Equiv. hh income (cube root)	-0.001*	-0.000	-0.001*	-0.003
	(0.001)	(0.002)	(0.001)	(0.002)
Wave 5	-0.037***	-0.054*	-0.036***	-0.100*
	(0.010)	(0.026)	(0.010)	(0.040)
Constant	1.086***	0.406	0.937**	0.245
	(0.302)	(0.750)	(0.287)	(1.114)
Control variables country	Yes	Yes	Yes	Yes
N	64,609	9,091	70,122	3,578
AIC	373,563	55,873	407,761	21,793
BIC	373,808	56,065	408,008	21,960

Note: In columns 1 and 2, individuals are considered objectively impaired if they recall 3 words or less ("3 words"), while in columns 3 and 4 the cutoff is at 2 words or less ("2 words"). The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 5 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 4 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table 7: Health perception based on walking ability

	(1) Objectively Unimpaired	(2) Objectively Impaired
Health perception (ref.: concordance)		
Underestimating	0.249**	
	(0.090)	
Overestimating		-0.328*
		(0.150)
Chronic diseases	0.129***	0.012
	(0.020)	(0.061)
Age	0.152*	0.337**
	(0.061)	(0.128)
Age squared	-0.001*	-0.002**
	(0.000)	(0.001)
Woman	-0.186**	-0.185
	(0.066)	(0.149)
Educ. group (ref.: low)		
Medium	0.189*	-0.041
	(0.075)	(0.203)
High	0.097	-0.235
	(0.088)	(0.239)
Retired	-0.040	-0.202
	(0.085)	(0.178)
Married	-0.109	0.127
	(0.066)	(0.134)
Equiv. hh income (cube root) 2	-0.007	-0.023
	(0.005)	(0.013)
Wave 2	-0.315*	-0.126
	(0.137)	(0.204)
Constant	-3.134	-9.194
	(2.197)	(4.759)
Control variables country	Yes	Yes
N	1,545	233
AIC	9,447	1,548
BIC	9,580	1,634

Note: The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 2 or Wave 4. All explanatory variables are taken from wave w, i.e. Wave 1 or Wave 2 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

6 The Total Public Cost of Health Misperception

We perform a back-of-the-envelope calculation to estimate additional health expenditures due to health misperception. The total cost of over- or underestimating health for the population 50+ in the respective country C_t in year t is calculated as follows:

$$C_{c,t} = d_c \times m_c \times f_c \times p_{c,t} \tag{4}$$

where d_c is the predicted cost per outpatient visit in 2010 Intl$ according to the World Health Organisation for the respective country (World Health Organization 2011) and m_c denotes the marginal effect at means of over- or underestimating health on doctor visits (i.e., the difference in doctor visits between concordance and over- or underestimation according to our estimates). The term f_c denotes the fraction of individuals in the SHARE sample who over- or underestimate their health, and $p_{c,t}$ is the population older than age 50 in the respective country and year according to predictions from the Wittgenstein Centre for Demography and Global Human Capital (2018).

We project that on average across the European Union, underestimation of the population 50+ will cost Intl$ 71 million in 2020 per country and increase to Intl$ 81 million by 2060. Overestimation will result in negative healthcare costs of approximately Intl$ 37 million per country in 2020 and Intl$ 45 million in 2060. Altogether, we project a net cost of Intl$ 34 million per country in 2020. Note that although overestimation results in negative costs, these are in the short run only. In the long run, overestimation may result in individuals skipping timely screening and preventive care and lead to worse health, resulting in higher healthcare expenditures. Longer panel data will aid in evaluating the full cost of overestimation in future work.

The costs of misperception are also projected separately for each country in our sample. Figure 5 shows the total costs of underestimation in 2020, 2040, and 2060 – it is highest in Germany, Italy, and France. When dividing the total cost of underestimation by the population size, Germany, Denmark, and The Netherlands have the highest cost. Germany has a high marginal effect at means of underestimating health on doctor visits m_c along with a relatively high fraction of individuals that underestimate their health f_c. As a result, additional outpatient visits due to underestimation are predicted to cost Germany Intl$ 503 million in 2020 and Intl$ 538 million in 2060. Denmark and The Netherlands have much lower f_c, but their high cost per outpatient visit d_c along with large marginal effects m_c result in high public cost of underestimating health. Countries such as Poland and Czechia have much lower

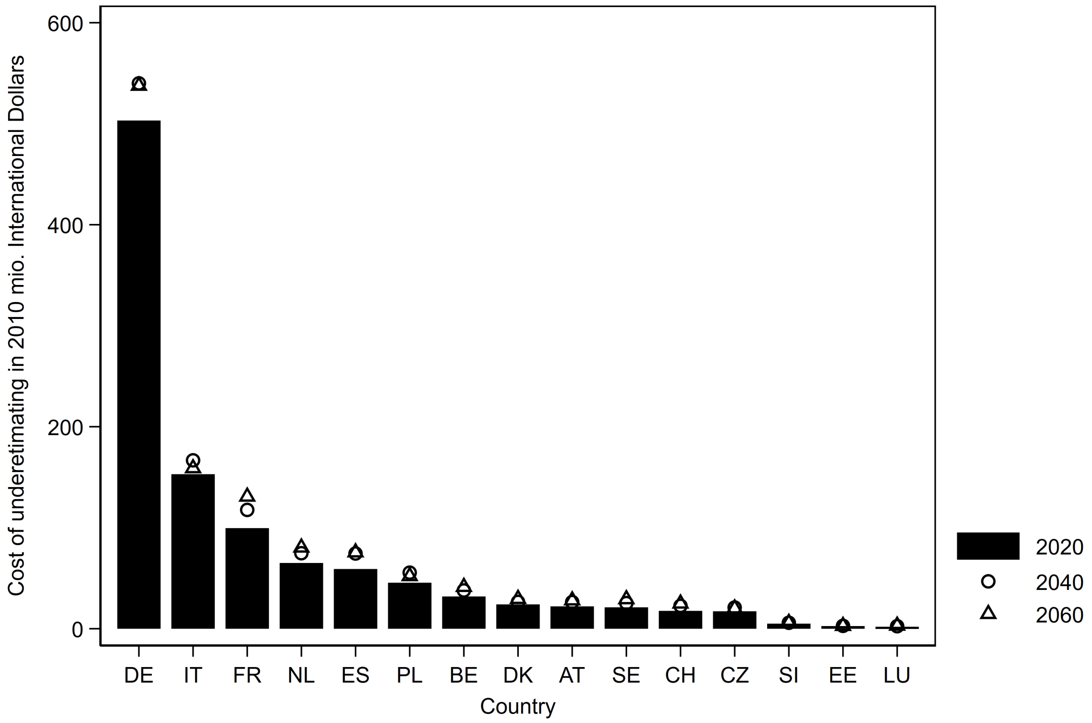


Figure 5: Projected total public cost of underestimating health for the population 50+

costs because approximately similar percentages report over- and underestimation and also much lower consultation costs. Spain has the lowest predicted cost of underestimation per capita. In total, it is Intl$ 59 million in 2020 and Intl$ 76 million in 2060. Table 8 shows projected misperception costs for all countries in 2020, 2040, and 2060.

7 Conclusion

We utilise rich longitudinal data from 15 European countries from SHARE to explore the effects of health (mis)perception on healthcare utilisation. We categorise misperception as arising due to overconfidence or underconfidence. Following the literature in psychology, overconfidence is measured as overestimation of one's health, whereas underconfidence is defined analogously as underestimation of one's own health. Healthcare utilisation is measured as the annual number of doctor visits. In addition, we assess the relationship between misperception and OOP expenditures incurred by those who visit the doctor. Our results based on count models and log-Gamma models suggest that individuals who underestimate their health visit the doctor more often and have higher OOP expenditures than those who assess their health correctly. By contrast, survey participants who overestimate their health visit the doctor less often and have lower OOP payments.

Table 8: Projected total cost of health misperception by country

Country	Year	Population 50+	Cost	Cost underestimating	Cost overestimating	Balance
		in 1,000	per visit	in mio. 2010 Intl$	in mio. 2010 Intl$	in mio. 2010 Intl$
Austria	2020	3,725.3	46.8	22.1	-30.9	-8.8
	2040	4,444.5	46.8	26.4	-36.9	-10.5
	2060	4,726.9	46.8	28.1	-39.2	-11.2
Belgium	2020	4,554.0	44.4	31.8	-16.3	15.6
	2040	5,416.9	44.4	37.9	-19.3	18.5
	2060	5,922.0	44.4	41.4	-21.1	20.3
Czechia	2020	4,137.6	31.3	17.2	-24.1	-6.8
	2040	5,038.8	31.3	21.0	-29.3	-8.3
	2060	4,829.6	31.3	20.1	-28.1	-8.0
Denmark	2020	2,324.7	47.0	24.0	0.1	24.2
	2040	2,569.4	47.0	26.5	0.2	26.7
	2060	2,859.2	47.0	29.5	0.2	29.7
Estonia	2020	519.2	25.7	2.50	-1.7	0.9
	2040	583.6	25.7	2.80	-1.9	1.0
	2060	568.9	25.7	2.80	-1.8	0.9
France	2020	26,121.9	40.8	99.5	-50.1	49.4
	2040	30,875.5	40.8	117.6	-59.2	58.4
	2060	34,332.8	40.8	130.8	-65.8	65.0
Germany	2020	37,597.6	44.0	503.1	-83.5	419.6
	2040	40,352.9	44.0	540.0	-89.6	450.3
	2060	40,176.2	44.0	537.6	-89.2	448.4
Italy	2020	27,580.7	38.4	152.9	-35.8	117.1
	2040	30,070.5	38.4	166.7	-39.0	127.7
	2060	28,662.7	38.4	158.9	-37.2	121.7
Luxembourg	2020	214.5	89.5	1.7	-2.6	-0.9
	2040	313.5	89.5	2.5	-3.7	-1.3
	2060	386.3	89.5	3.0	-4.6	-1.6
Netherlands	2020	7,082.6	48.1	65.0	-47.0	18.0
	2040	8,138.9	48.1	74.7	-54.0	20.7
	2060	8,750.5	48.1	80.4	-58.1	22.3
Poland	2020	14,398.8	25.5	45.5	-70.2	-24.7
	2040	17,587.8	25.5	55.5	-85.7	-30.2
	2060	16,498.9	25.5	52.1	-80.4	-28.3
Slovenia	2020	880.7	32.6	4.9	-4.4	0.6
	2040	1,048.1	32.6	5.9	-5.2	0.7
	2060	1,026.1	32.6	5.8	-5.1	0.7
Spain	2020	19,709.0	37.8	59.0	-186.8	-127.7
	2040	24,851.7	37.8	74.5	-235.5	-161.1
	2060	25,235.6	37.8	75.6	-239.2	-163.6
Sweden	2020	3914.8	45.8	21.1	-9.2	11.9
	2040	4703.5	45.8	25.4	-11.1	14.3
	2060	5419.9	45.8	29.2	-12.7	16.5
Switzerland	2020	3,513.4	55.2	17.6	8.7	26.2
	2040	4,516.4	55.2	22.6	11.1	33.7
	2060	4,986.1	55.2	24.9	12.3	37.2

Our results are robust to a range of sensitivity analyses with different model specifications, sample compositions, estimation methods, and health dimensions. In addition, we account for potential endogeneity by exploiting the panel structure of our data. Specifically, arguments concerning individuals' health perception improving as a result of frequent doctor visits do not apply because we focus on current misperception and future doctor visits. Descriptive cross-tabulations show that individual characteristics such as education, illnesses, age, retirement, supplementary insurance, and others do not matter for the relationship between health misperception and healthcare utilisation; regressions controlling for these variables confirm the stability of the results.

The main limitation of this study is related to panel attrition. Individuals who suffer from diseases are less likely to participate in consecutive survey waves and thus are less likely to be included in our sample. However, we address this limitation by running our analyses separately by the number of diseases that a participant is suffering from and find no difference in the results between healthy and unhealthy participants, which suggests that panel attrition is not a concern in our study. Future work could fruitfully explore the long-term effects of health misperception on healthcare utilisation, for example, exploiting national panel data collected over a longer period of time than SHARE data. Longer panels would also allow for panel regressions and thus enable researchers to control for unobserved heterogeneity between observations.

The policy implications of our results are straightforward. First, addressing rising health expenditures has been a top priority on policymakers' agenda in many countries. Excessive hospital admissions use more than 37 million bed days across the European Union every year, significantly increasing public expenditures (OECD & European Commission 2018). Containing sources of waste and inefficiency in healthcare on either the demand or supply side is important in this regard. Our paper provides new insights, highlighting demand-side misperception as a possible source of wasteful spending. Given our results, we perform a back-of-the-envelope calculation of the costs of health misperception. (see Appendix 6 for detail). We project that on average across the European Union, underestimation will cost Intl$ 71 million in 2020 per country and increase to Intl$ 81 million by 2060. While overestimating health reduces public healthcare expenditure in the short run, it is likely that in the long run it will increase cost due to forgone preventive care.

Second, if individuals' own perceptions of health are what drive healthcare demand beyond actual health and other socioeconomic characteristics, then equipping them through personalised or public health campaigns with the necessary tools and information to accurately assess their health and determine when to seek healthcare is perhaps a valuable long-term strategy for reducing unnecessary healthcare use. This is a particularly relevant measure in countries with ageing populations that suffer from cognitive dissonance and thereby increased health misperception (Brandtstädter & Greve 1994, Frieswijk et al. 2004, Henchoz et al. 2008, Idler 1993, Spitzer & Weber 2019). Reaching out to those who overestimate their health by providing information about the benefits of screening and preventive care might also improve their health and thus prevent suffering and costs in the long run. Initiatives to increase health literacy, such as the National Action Plan on Health Literacy, are already in place in Germany (Vogt et al. 2018). Other countries can follow similar approaches to evaluate health literacy levels and take strategic action to educate people.

Finally, wait time is often used as a non-price rationing measure in healthcare by policymakers (Barzel 1974, Iversen & Siciliani 2011). Identifying patients with health misperception and reducing unnecessary visits to the doctor can have important implications for the effectiveness of such rationing mechanisms. Not only will they free up physician capacity, but they can also directly ensure timely care for other patients who are in need of urgent intervention. Moreover, with the advent of artificial intelligence and technology, providing individuals with the option to use online physician chatbots and telephone consultations will further reduce the burden of unnecessary doctor visits due to misperception rather than true health need.

References

Anderson, A., Baker, F. & Robinson, D. T. (2017), 'Precautionary savings, retirement planning and misperceptions of financial literacy', *Journal of Financial Economics* **126**(2), 383–398.

Arni, P., Dragone, D., Goette, L. & Ziebarth, N. R. (2019), 'Biased Health Perceptions and Risky Health Behaviors — Theory and Evidence', *Preliminary Draft* .

Avitabile, C., Jappelli, T. & Padula, M. (2011), 'Cognitive abilities, healthcare and screening tests', *Journal of Population Ageing* **4**, 251–269.

Bago d'Uva, T. & Jones, A. M. (2009), 'Health care utilisation in Europe: New evidence from the ECHP', *Journal of Health Economics* **28**(2), 265–279.

Bago d'Uva, T., O'Donnell, O. & Van Doorslaer, E. (2008), 'Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans', *International Journal of Epidemiology* **37**(6), 1375–1383.

Barzel, Y. (1974), 'A Theory of Rationing by Waitin', *The Journal of Law & Economics* **17**(1), 73–95.

Baumann, A. O., Deber, R. B. & Thompson, G. G. (1991), 'Overconfidence among physicians and nurses: The 'micro-certainty, macro-uncertainty' phenomenon', *Social Science and Medicine* **32**(2), 167–174.

Beaudoin, M. & Desrichard, O. (2011), 'Are memory self-efficacy and memory performance related? A meta-analysis.', *Psychological Bulletin* **137**(2), 211–241.

Berner, E. S. & Graber, M. L. (2008), 'Overconfidence as a Cause of Diagnostic Error in Medicine', *American Journal of Medicine* **121**(5 SUPPL.), S2–S23.

Bíró, A. (2013), 'Discount rates and the education gradient in mammography in the UK', *Health economics* (22), 1021–1036.

Bíró, A. (2014), 'Supplementary private health insurance and health care utilization of people aged 50+', *Empirical Economics* **46**(2), 501–524.

Bíró, A. (2016*a*), 'Differences between subjective and predicted survival probabilities and their relation to preventive care use', *B.E. Journal of Economic Analysis and Policy* **16**(2), 807–835.

Bíró, A. (2016*b*), 'Outpatient visits after retirement in Europe and the US', *International Journal of Health Economics and Management* **16**(4), 363–385.

Bolin, K., Lindgren, A., Lindgren, B. & Lundborg, P. (2009), 'Utilisation of physician services in the 50+ population: The relative importance of individual versus institutional factors in 10 European countries', *International Journal of Health Care Finance and Economics* **9**(1), 83–112.

Börsch-Supan, A. (2019*a*), 'Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 7.0.0. SHARE-ERIC. Data set.'.
URL: *10.6103/SHARE.w1.700*

Börsch-Supan, A. (2019*b*), 'Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 7.0.0. SHARE-ERIC. Data set.'.
URL: *10.6103/SHARE.w4.700*

Börsch-Supan, A. (2019*c*), 'Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 7.0.0. SHARE-ERIC. Data set.'.
URL: *10.6103/SHARE.w6.700*

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S. & Zuber, S. (2013), 'Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)', *International Journal of Epidemiology* **42**(4), 992–1001.

Brandtstädter, J. & Greve, W. (1994), 'The aging self: Stabilizing and protecitve processes', *Developmental Review* (14), 52–80.

Buchmueller, T. C., Couffinhal, A., Grignon, M. & Perronnin, M. (2004), 'Access to physician services: Does supplemental insurance matter? Evidence from France', *Health Economics* **13**, 669–687.

Cameron, K. A., Song, J., Manheim, L. M. & Dunlop, D. D. (2010), 'Gender disparities in health and healthcare use among older adults', *Journal of Women's Health* **19**(9), 1643–1650.

Capistrant, B. D., Glymour, M. M. & Berkman, L. F. (2014), 'Assessing mobility difficulties for cross-national comparisons: Results from the WHO study on AGEing and Adult Health', *Journal of the American Geriatrics Society* **62**(2), 329–335.

Choi, A. & Cawley, J. (2017), 'Health disparities across education: The role of differential reporting error', *Health Economics* **27**(3), e1–e29.

Christensen, K., Doblhammer, G., Rau, R. & Vaupel, J. W. (2009), 'Ageing populations: the challenges ahead', *Lancet* **374**(9696), 1196–1208.

Coman, L. & Richardson, J. (2006), 'Relationship between self-report and performance measures of function: a systematic review.', *Canadian journal on aging* **25**(3), 253–270.

Coulson, E., Terza, J. V., Neslusan, C. A. & Stuart, B. C. (1995), 'Estimating the Moral-Hazard Effect of Supplemental Medical Insurance in the Demand for Prescription Drugs by the Elderly', *American Economic Review* **85**, 122–126.

Cox, N. (2011), 'Stata tip 96: Cube roots', *Stata Journal* **11**(1), 149–154.

Crossley, T. F. & Kennedy, S. (2001), 'The reliability of self-assessed health status.', *Journal of Health Economics* **21**, 643–658.

Cutler, D., Skinner, J., Stern, A. D. & Wennberg, D. (2013), 'Physician beliefs and patient preferences : A new look at regional variation in health care spending', *NBER Working Paper Series* (19320).

Deb, P. & Norton, E. C. (2018), 'Modeling Health Care Expenditures and Use', *Annual Review of Public Health Modeling* **39**, 489–505.

European Commission (2018), The 2018 Ageing Report: Economic & budgetary projections for the 28 EU member states (2016-2070), Technical report.

Eurostat (2018), 'International Standard Classification of Education (ISCED)'.
URL: *http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_%28ISCED%29#Correspondence_between_ISCED_2011_and_ISCED_1997*

Eurostat (2019), *Ageing Europe: Looking at the lives of older People in the EU*, Publications Office of the European Union, Luxembourg.

Ferrer, M., Lamarca, R., Orfila, F. & Alonso, J. (1999), 'Comparison of performance-based and self-rated functional capacity in Spanish elderly', *American Journal of Epidemiology* **149**(3), 228–235.

Frieswijk, N., Buunk, B. P., Steverink, N. & Slaets, J. P. (2004), 'The effect of social comparison information on the life satisfaction of frail older persons', *Psychology and Aging* **19**(1), 183–190.

Furnham, A. (2001), 'Self-estimates of intelligence : culture and gender difference in self and other estimates of both general (g) and multiple intelligences', *Personality and Individual Differences* **31**, 1381–1405.

Galdas, P. M., Cheater, F. & Marshall, P. (2005), 'Men and health help-seeking behaviour: Literature review', *Journal of Advanced Nursing* **49**(6), 616–623.

Gardner, R. C., Langa, K. M. & Yaffe, K. (2017), 'Subjective and objective cognitive function among older adults with a history of traumatic brain injury: A population-based cohort study', *PLoS Medicine* **14**(3), 1–16.

Grodstein, F., Chen, J., Wilson, R. S. & Manson, J. E. (2001), 'Type 2 diabetes and cognitive function in community-dwelling elderly women', *Diabetes Care* **24**(6), 1060–1065.

Henchoz, K., Cavalli, S. & Girardin, M. (2008), 'Health perception and health status in advanced old age: A paradox of association', *Journal of Aging Studies* **22**(3), 282–290.

Idler, E. L. (1993), 'Age differences in self-assessments of health: Age changes, cohort differences, or survivorship?', *Journal of Gerontology* **48**(6), 289–300.

Iversen, T. & Siciliani, L. (2011), Non-Price Rationing and Waiting Times, *in* S. Glied & P. C. Smith, eds, 'The Oxford Handbook of Health Economics', Oxford University Press, Oxford New York.

Jackson, J. D., Rentz, D. M., Aghjayan, S. L., Buckley, R. F., Meneide, T. F., Sperling, R. A. & Amariglio, R. E. (2017), 'Subjective cognitive concerns are associated with objective memory performance in Caucasian but not African-American persons', *Age and Ageing* **46**(6), 988–993.

Johnson, D. D. P. & Fowler, J. H. (2011), 'The evolution of overconfidence', *Nature* **477**, 317–320.

Jürges, H. (2007), 'True health vs response style: exploring cross-country differences in self-reported health', *Health economics* **16**, 163–178.

Jürges, H. (2015), Health care utilization and out-of-pocket expenses, *in* F. Malter & A. Börsch-Supan, eds, 'SHARE Wave 5 : Innovations & Methodology', Munich Center for the Economics of Ageing (MEA) at the Max Planck Institute for Social Law and Socia, Munich, pp. 37–42.

Lange, F. (2011), 'The role of education in complex health decisions: Evidence from cancer screening', *Journal of Health Economics* **30**(1), 43–54.

Lugo-Palacios, D. G. & Gannon, B. (2017), 'Health care utilisation amongst older adults with sensory and cognitive impairments in Europe', *Health Economics Review* **7**(1), 1–15.

Lutz, W., O'Neill, B. C. & Scherbov, S. (2003), 'Europe's population at a turning point', *Science* **299**, 1991–1992.

Mansfield, A., Addis, M. & Mahalik, J. (2003), '"Why Won't He Go to the Doctor?": The Psychology of Men's Help Seeking', *International Journal of Men's Health* **2**(2), 93–109.

Merrill, S. S., Seeman, T. E., Kasl, S. V. & Berkman, L. F. (1997), 'Gender differences in the comparison of self-reported disability and performance measures', *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* **52**(1), 19–26.

Meyer, A. N., Payne, V. L., Meeks, D. W., Rao, R. & Singh, H. (2013), 'Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study', *JAMA Internal Medicine* **173**(21), 1952–1961.

Moore, D. A. & Healy, P. J. (2008), 'The Trouble With Overconfidence', *Psychological Review* **115**(2), 502–517.

Moreira, S. & Barros Pita, P. (2010), 'Double health insurance coverage and health care utilisation: Evidence from quantile regression', *Health economics* **19**, 1075–1092.

OECD & European Commission (2018), *Health at a Glance: Europe 2018: State of health in the EU cycle*, OECD Publishing, Paris.

Ortoleva, P. & Snowberg, E. (2015), 'Overconfidence in political behavior', *American Economic Review* **105**(2), 504–535.

Paccagnella, O., Rebba, V. & Weber, G. (2013), 'Voluntary private health insurance among the over 50s in Europe', *Health economics* **22**, 289–315.

Picone, G., Sloan, F. & Taylor, D. (2004), 'Effects of risk and time preference and expected longevity on demand for medical tests', *Journal of Risk and Uncertainty* **28**(1), 39–53.

Pinheiro, P. A., Carneiro, J. A., Coqueiro, R. S., Pereira, R. & Fernandes, M. H. (2016), 'Chair stand test as simple tool for sarcopenia screening in elderly women', *Journal of Nutrition, Health and Aging* **20**(1), 56–59.

Preston, C. E. & Harris, S. (1965), 'Psychology of drivers in traffic accidents', *Journal of Applied Psychology* **49**(4), 284–288.

Purser, J. L., Fillenbaum, G. G., Pieper, C. F. & Wallace, R. B. (2005), 'Mild cognitive impairment and 10-year trajectories of disability in the Iowa established populations for epidemiologic studies of the elderly cohort', *Journal of the American Geriatrics Society* **53**(11), 1966–1972.

Reuben, E., Wiswall, M. & Zafar, B. (2017), 'Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender', *Economic Journal* **127**(604), 2153–2186.

Rocca, P., Beckman, A., Ekvall Hansson, E. & Ohlsson, H. (2015), 'Is the association between physical activity and healthcare utilization affected by self-rated health and socioeconomic factors?', *BMC Public Health* **15**(1), 1–8.

Sainio, P., Koskinen, S., Heliövaara, M., Martelin, T., Härkänen, T., Hurri, H., Miilunpalo, S. & Aromaa, A. (2006), 'Self-reported and test-based mobility limitations in a representative sample of Finns aged 30+', *Scandinavian Journal of Public Health* **34**(4), 378–386.

Sakurai, R., Fujiwara, Y., Ishihara, M., Higuchi, T., Uchida, H. & Imanaka, K. (2013), 'Age-related self-overestimation of step-over ability in healthy older adults and its relationship to fall risk', *BMC Geriatrics* **13**(44), 1–9.

Schlichthorst, M., Sanci, L. A., Pirkis, J., Spittal, M. J. & Hocking, J. S. (2016), 'Why do men go to the doctor? Socio-demographic and lifestyle factors associated with healthcare utilisation among a cohort of Australian men', *BMC Public Health* **16**(Suppl 3), 81–90.

Schneider, U., Pfarr, C., Schneider, B. S. & Ulrich, V. (2012), 'I feel good! Gender differences and reporting heterogeneity in self-assessed health', *European Journal of Health Economics* **13**(3), 251–265.

Spitzer, S. & Weber, D. (2019), 'Reporting biases in self-assessed physical and cognitive health status of older Europeans', *Plos One* **14**(10), e0223526.

Srisurapanont, M., Suttajit, S., Eurviriyanukul, K. & Varnado, P. (2017), 'Discrepancy between objective and subjective cognition in adults with major depressive disorder', *Scientific Reports* **7**(1), 1–7.

Steel, N., Huppert, F. & Melzer, D. (2003), Physical and cognitive function, *in* M. Marmot, J. Banks, R. Blundell, C. Lessof & J. Nazroo, eds, 'Health, Wealth and Lifestyles of the Older Population in England: The 2002 English Longitudinal Study of Ageing', IFS, London, chapter Physical a.

Tavares, L. P. & Zantomio, F. (2017), 'Inequity in healthcare use among older people after 2008: The case of southern European countries', *Health Policy* **121**(10), 1063–1071.

Vallejo-Torres, L. & Morris, S. (2013), 'Income-related inequity in healthcare utilisation among individuals with cardiovascular disease in England - account for vertical inquity', *Health economics* (22), 533–553.

Van Doorslaer, E., Koolman, X. & Jones, A. M. (2004), 'Explaining income-related inequalities in doctor utilisation in Europe', *Health Economics* **13**(7), 629–647.

Vogt, D., Schaeffer, D., Bauer, U., Koplatzik, K. & Hurrelmann, K. (2018), 'The German National Action Plan on health literacy: Empirical base, development and implementation', *European Journal of Public Health* **28**.

Wittgenstein Centre for Demography and Global Human Capital (2018), 'Wittgenstein Centre Data Explorer Version 2.0 (Beta)'.
URL: *http://www.wittgensteincentre.org/dataexplorer*

World Health Organization (2011), 'Country-specific inpatient and outpatient estimates in 2010 currency'.
URL: *https://www.who.int/choice/cost-effectiveness/inputs/ country_inpatient_outpatient_2010.pdf?ua=1*

Zell, E. & Krizan, Z. (2014), 'Do People Have Insight Into Their Abilities? A Metasynthesis', *Perspectives on Psychological Science* **9**, 111–125.

Zhang, Y., Salm, M. & Soest, A. V. (2018), 'The effect of retirement on healthcare utilization : Evidence from China', *Journal of Health Economics* **62**, 165–177.

A Appendix

Table A.1: Crosstable mean doctor visits at $w+1$ (weighted)

	Health perception				
	Pos. concordance	**Underestimating**	**Neg. concordance**	**Overestimating**	**Total**
	Mean doctor visits				
No. chronic diseases at w					
0	4.6	7.2	10.5	6.1	5.1
1	6.3	9.5	10.5	8.0	7.0
2	8.1	10.6	12.1	11.0	9.0
3	8.8	11.3	11.7	12.4	10.0
4	11.0	16.2	13.8	15.5	13.1
5	13.5	13.5	16.2	11.3	13.9
6	11.7	18.9	14.8	17.2	14.6
7	11.0	14.1	26.9	32.1	21.1
8	11.3	19.5	15.0	14.9	13.2
9	20.0				20.0
10	11.7				11.7
Total	6.2	10.1	11.9	8.6	7.2
No. activity limitations at w					
0	6.0	9.0	11.1	7.8	6.5
1	8.3	10.3	10.8	9.8	9.3
2	9.8	14.0	11.6	13.9	12.0
3	11.7	11.5	11.9	15.1	12.2
4	10.2	31.1	11.6	12.3	15.5
5	7.8	11.7	12.3	18.6	12.8
6	5.2	11.4	14.7	12.4	13.2
7	8.8	15.8	13.4	15.6	13.8
8	4.7	8.6	15.3	9.8	13.6
9	7.6	9.6	20.2	11.5	18.9
10		30.0	20.5	6.7	19.4
11		9.1	9.8	6.0	9.7
12		6.2	16.6	30.6	16.4
13	5.5	8.6	14.5	98.0	10.3
Total	6.2	10.1	11.8	8.6	7.2
5-year age groups					
50-54	5.2	10.6	11.7	6.8	5.9
55-59	5.4	10.9	13.8	7.2	6.4
60-64	6.0	9.1	10.5	8.9	6.8
65-69	6.8	9.5	13.5	9.4	7.7
70-74	7.7	10.8	10.9	9.1	8.5
75-79	8.1	9.0	12.0	12.1	9.1
80-84	8.0	11.4	10.4	9.0	9.0
85-89	6.4	9.4	12.8	9.3	8.7
90+	5.5	7.8	9.2	13.9	9.0
Total	6.2	10.1	11.8	8.6	7.2
Gender					
Men	6.0	10.0	11.8	8.1	6.7
Women	6.4	10.1	11.9	9.1	7.5
Total	6.2	10.1	11.8	8.6	7.2
Education					
Low	6.6	10.1	11.8	9.2	7.7
Medium	6.2	9.7	12.1	8.2	7.0
High	5.7	10.8	11.3	7.5	6.4
Total	6.2	10.1	11.8	8.6	7.2

Note: Calibrated cross-sectional individual weights are applied.

Table A.1, continued: Crosstable mean doctor visits at $w+1$ (weighted)

	Health perception				
	Pos. concordance	**Underestimating**	**Neg. concordance**	**Overestimating**	**Total**
	Mean doctor visits				
Is retired					
No	5.5	10.2	11.6	7.9	6.4
Yes	7.0	10.0	11.9	9.3	7.9
Total	6.2	10.1	11.8	8.6	7.2
Is married					
No	6.4	9.9	11.5	9.2	7.5
Yes	6.1	10.2	12.0	8.5	7.0
Total	6.2	10.1	11.8	8.7	7.2
Healt access					
Not difficult	5.9	10.2	11.0	8.4	6.7
Difficult	6.5	10.0	12.0	10.6	8.3
Total	6.0	10.1	11.3	8.9	7.0
Supplementary insurance					
No	6.1	10.4	11.7	9.7	7.3
Yes	5.7	9.4	10.6	7.1	6.4
Total	6.0	10.0	11.4	8.9	7.0
Has children					
No	6.1	10.8	12.0	10.1	7.3
Yes	6.2	10.0	11.8	8.5	7.2
Total	6.2	10.1	11.8	8.7	7.2
Country					
Austria	6.1	8.8	11.4	8.9	7.0
Germany	6.9	11.2	13.4	10.1	8.0
Sweden	3.8	5.7	10.6	5.4	4.3
Netherlands	5.0	7.4	9.8	8.6	5.5
Spain	5.2	8.6	10.2	7.5	6.2
Italy	7.3	13.1	13.8	10.2	8.6
France	5.6	7.7	9.3	6.7	6.1
Denmark	4.3	8.2	10.6	6.4	4.9
Switzerland	4.4	7.2	8.6	8.0	5.0
Belgium	7.0	10.3	16.4	9.5	8.0
Czechia	6.5	9.7	11.2	9.1	7.5
Poland	6.7	9.9	10.2	6.8	7.5
Luxembourg	7.7	10.5	17.3	12.1	9.0
Slovenia	4.5	7.4	9.0	7.5	5.5
Estonia	4.9	7.6	8.4	6.2	5.9
Total	6.2	10.1	11.8	8.6	7.2
Survey wave					
Wave 2	6.5	10.1	12.6	8.3	7.4
Wave 5	6.0	10.0	11.4	8.9	7.0
Total	6.2	10.1	11.8	8.6	7.2

Note: Calibrated cross-sectional individual weights are applied.

Table A.2: Crosstable mean OOP expenditures in Euros at $w+1$ (weighted)

	Health perception				
	Pos. concordance Mean OOP	**Underestimating** Mean OOP	**Neg. concordance** Mean OOP	**Overestimating** Mean OOP	**Total** Mean OOP
No. chronic diseases at w					
0	64.1	98.9	65.0	73.5	66.7
1	73.9	100.4	56.3	66.5	75.2
2	75.2	79.6	65.8	79.0	75.3
3	80.4	76.5	94.4	102.0	83.6
4	79.4	76.9	58.8	83.3	74.8
5	81.1	158.1	95.0	41.3	98.1
6	71.7	181.6	352.4	6.4	156.3
7	38.3	95.1	9.0	67.7	40.2
8	23.2	334.6	50.7	8.8	40.6
10	0.0				0.0
Total	70.8	92.4	74.6	74.6	73.6
No. activity limitations at w					
0	71.8	90.5	75.7	73.7	73.5
1	62.5	96.5	72.9	61.7	71.3
2	53.6	50.9	52.3	108.2	60.7
3	44.8	140.5	160.1	103.2	121.3
4	35.6	86.4	34.0	34.6	46.9
5	38.1	57.9	75.4	57.3	64.6
6	9.8	54.0	90.8	42.6	73.3
7	0.0	123.1	18.7	527.8	60.6
8	34.2	0.0	47.2	59.3	39.9
9	163.4	19.0	15.6	0.0	16.7
10		0.0	219.9	0.0	197.9
11		352.6	28.2	0.0	60.5
12		289.9	46.6	8.6	100.2
13	2.2	298.1	141.1		240.2
Total	70.8	92.4	74.5	74.4	73.5
5-year age groups					
50-54	60.8	147.3	57.4	46.8	65.8
55-59	67.4	80.9	72.5	103.9	71.0
60-64	70.6	78.8	86.0	68.4	71.8
65-69	73.9	80.6	110.8	89.8	77.6
70-74	89.7	99.7	86.3	58.6	88.1
75-79	76.2	96.6	102.2	81.5	82.7
80-84	67.5	74.3	42.5	86.9	67.1
85-89	48.3	70.0	45.3	52.9	51.5
90+	49.0	15.4	10.8	47.2	35.1
Total	70.8	92.4	74.5	74.4	73.5
Gender					
Men	74.5	70.4	62.2	60.9	72.7
Women	67.1	104.2	79.6	84.6	74.2
Total	70.8	92.4	74.5	74.4	73.5
Education					
Low	62.0	89.3	78.5	64.3	66.8
Medium	71.1	92.0	71.0	70.6	73.3
High	85.4	108.5	55.7	120.8	88.4
Total	71.1	93.7	74.3	74.1	73.9

Note: Calibrated cross-sectional individual weights are applied.

Table A.2, continued: Crosstable mean OOP expenditures in Euros at $w+1$ (weighted)

	Health perception				
	Pos. concordance Mean OOP	**Underestimating** Mean OOP	**Neg. concordance** Mean OOP	**Overestimating** Mean OOP	**Total** Mean OOP
Is retired					
No	64.8	101.1	54.4	81.8	68.5
Yes	77.0	87.4	88.7	70.0	78.5
Total	70.7	92.8	75.4	75.0	73.6
Is married					
No	69.5	72.9	42.0	68.5	67.2
Yes	72.5	100.8	103.1	81.5	77.5
Total	71.6	90.8	74.8	76.2	74.1
Healt access					
Not difficult	71.5	98.6	78.2	84.0	75.4
Difficult	67.7	81.4	66.7	43.8	66.6
Total	71.0	95.0	74.1	74.5	73.9
Supplementary insurance					
No	70.5	101.1	73.0	75.8	74.3
Yes	71.2	76.6	80.2	71.3	72.2
Total	70.7	92.5	74.6	74.5	73.5
Has children					
No	89.3	131.7	92.9	73.9	91.6
Yes	68.3	88.9	72.7	74.8	71.3
Total	70.8	92.8	75.0	74.7	73.6
Country					
Austria	123.9	141.1	124.0	58.3	121.0
Germany	55.8	45.6	30.7	75.9	54.2
Sweden	67.9	75.2	75.6	66.9	68.9
Spain	10.1	33.4	64.2	16.5	18.0
Italy	139.7	270.5	135.8	132.5	149.2
France	28.6	37.3	27.2	29.8	29.5
Denmark	5.5	2.9	4.9	1.5	5.1
Switzerland	386.5	442.0	636.0	380.6	397.0
Belgium	90.6	183.0	155.2	90.9	105.1
Czechia	7.3	11.2	9.6	10.9	8.3
Luxembourg	171.3	209.6	292.1	206.8	185.3
Slovenia	13.9	20.2	9.9	2.8	13.2
Estonia	14.8	12.0	14.6	13.9	14.3
Total	70.8	92.4	74.5	74.4	73.5
Survey wave					
Wave 5	70.8	92.4	74.5	74.4	73.5
Total	70.8	92.4	74.5	74.4	73.5

Note: Calibrated cross-sectional individual weights are applied.

Table A.3: Crosstable health perception (weighted)

	Health perception								
	Pos. concordance		**Underestimating**		**Neg. concordance**		**Overestimating**		**Total**
	Row %	95% CI	Row %	95% CI	Row %	95% CI	Row %	95% CI	Row %
Objective impairment									
Unimpaired (n=47,913)	88.0	[87.5,88.5]	12.0	[11.5,12.5]	0.0		0.0		100.0
Impaired (n=8,239)	0.0		0.0		38.3	[36.6,40.0]	61.7	[60.0,63.4]	100.0
Total (n=56,152)	74.3	[73.6,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(3) =	5.62e+04								
Design-based F(2.77, 155429.77) =	7984.2277	Pr =	0.000						
No. chronic diseases at w									
0 (n=20,630)	82.1	[81.2,82.9]	5.7	[5.2,6.2]	3.1	[2.7,3.5]	9.1	[8.5,9.9]	100.0
1 (n=17,715)	76.2	[75.1,77.3]	10.2	[9.5,11.0]	4.4	[3.9,4.9]	9.2	[8.5,10.0]	100.0
2 (n=10,246)	67.8	[66.2,69.3]	13.9	[12.8,15.1]	8.2	[7.4,9.1]	10.1	[9.1,11.2]	100.0
3 (n=4,712)	60.4	[58.0,62.7]	16.2	[14.6,18.0]	11.7	[10.3,13.4]	11.6	[10.1,13.3]	100.0
4 (n=1,889)	48.2	[44.4,51.9]	20.4	[17.4,23.8]	20.1	[17.2,23.3]	11.4	[9.1,14.1]	100.0
5 (n=650)	39.5	[32.7,46.8]	25.7	[19.9,32.5]	24.1	[18.8,30.4]	10.7	[7.3,15.3]	100.0
6 (n=180)	39.9	[29.2,51.8]	16.3	[10.0,25.4]	28.5	[19.2,40.2]	15.2	[8.2,26.5]	100.0
7 (n=51)	22.1	[10.0,41.8]	19.9	[7.7,42.6]	53.0	[30.4,74.5]	5.0	[1.7,13.8]	100.0
8 (n=18)	50.9	[17.5,83.5]	3.7	[0.5,22.2]	36.6	[9.0,77.1]	8.8	[1.5,38.2]	100.0
9 (n=1)	100.0		0.0		0.0		0.0		100.0
10 (n=2)	100.0		0.0		0.0		0.0		100.0
Total (n=56,094)	74.3	[73.6,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(30) =	3717.8201								
Design-based F(24.19, 1.36e+06) =	56.8893	Pr =	0.000						
No. activity limitations at w									
0 (n=47,381)	80.6	[80.0,81.2]	7.9	[7.5,8.3]	2.6	[2.3,2.8]	8.9	[8.5,9.4]	100.0
1 (n=4,717)	50.3	[47.9,52.7]	22.7	[20.8,24.9]	12.8	[11.3,14.5]	14.2	[12.5,16.0]	100.0
2 (n=1,683)	30.2	[26.6,34.0]	26.5	[23.2,30.0]	29.9	[26.3,33.8]	13.5	[11.1,16.4]	100.0
3 (n=851)	21.3	[16.7,26.6]	25.7	[21.2,30.7]	38.9	[33.5,44.6]	14.2	[10.4,19.0]	100.0
4 (n=485)	15.4	[10.4,22.2]	20.9	[14.4,29.3]	48.6	[40.4,56.8]	15.2	[10.9,20.7]	100.0
5 (n=340)	11.4	[6.9,18.2]	19.3	[13.3,27.2]	52.0	[43.0,60.9]	17.3	[11.0,26.1]	100.0
6 (n=224)	7.0	[2.5,18.1]	13.9	[8.4,22.1]	63.1	[51.7,73.2]	16.0	[9.3,26.1]	100.0
7 (n=159)	4.6	[1.6,12.6]	20.0	[10.5,34.7]	69.9	[55.8,81.1]	5.4	[2.2,12.9]	100.0
8 (n=93)	3.5	[0.8,13.2]	17.0	[7.8,33.0]	75.0	[59.6,85.9]	4.5	[1.8,10.8]	100.0
9 (n=71)	0.7	[0.2,3.4]	9.3	[3.0,25.7]	87.6	[71.8,95.1]	2.4	[0.4,12.7]	100.0
10 (n=42)	0.0		6.5	[0.9,34.1]	81.1	[58.6,92.9]	12.4	[4.0,32.4]	100.0
11 (n=27)	0.0		7.4	[1.7,26.5]	91.6	[73.2,97.7]	1.1	[0.1,7.6]	100.0
12 (n=26)	0.0		22.8	[8.5,48.3]	61.8	[36.9,81.7]	15.4	[4.2,43.0]	100.0
13 (n=49)	4.4	[0.9,19.7]	66.2	[49.3,79.8]	29.2	[16.8,45.7]	0.1	[0.0,0.8]	100.0
Total (n=56,148)	74.3	[73.6,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(39) =	1.49e+04								
Design-based F(33.61, 1.89e+06) =	177.1679	Pr =	0.000						
5-year age groups									
50-54 (n=7,593)	81.7	[80.1,83.2]	7.0	[6.1,8.0]	2.8	[2.3,3.5]	8.5	[7.4,9.7]	100.0
55-59 (n=10,672)	78.6	[77.3,80.0]	8.7	[7.9,9.7]	4.0	[3.4,4.7]	8.6	[7.7,9.6]	100.0
60-64 (n=11,137)	77.0	[75.7,78.2]	10.6	[9.7,11.6]	4.2	[3.7,4.8]	8.2	[7.4,9.0]	100.0
65-69 (n=10,290)	73.6	[72.2,74.9]	10.7	[9.8,11.6]	5.4	[4.8,6.2]	10.3	[9.4,11.3]	100.0
70-74 (n=8,143)	68.7	[67.0,70.3]	12.5	[11.3,13.8]	8.3	[7.4,9.3]	10.5	[9.5,11.7]	100.0
75-79 (n=4,390)	63.7	[61.2,66.1]	13.9	[12.3,15.8]	11.8	[10.2,13.5]	10.6	[9.2,12.3]	100.0
80-84 (n=2,645)	55.7	[52.4,59.0]	13.2	[11.1,15.6]	16.5	[14.1,19.2]	14.6	[12.3,17.2]	100.0
85-89 (n=1,047)	47.4	[42.3,52.5]	13.8	[10.7,17.7]	20.6	[16.8,25.1]	18.1	[14.6,22.2]	100.0
90+ (n=235)	38.2	[28.3,49.1]	7.3	[4.1,12.6]	27.5	[18.9,38.2]	27.0	[18.4,37.8]	100.0
Total (n=56,152)	74.3	[73.6,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(24) =	2476.8374								
Design-based F(22.74, 1.28e+06) =	37.9231	Pr =	0.000						

Note: Calibrated cross-sectional individual weights are applied.

Table A.3, continued: Crosstable health perception (weighted)

	Health perception								
	Pos. concordance		**Underestimating**		**Neg. concordance**		**Overestimating**		**Total**
	Row %	95% CI	Row %	95% CI	Row %	95% CI	Row %	95% CI	Row %
Gender									
Men (n=24,503)	79.4	[78.5,80.2]	7.3	[6.8,7.8]	3.8	[3.5,4.2]	9.5	[8.9,10.2]	100.0
Women (n=31,649)	69.9	[69.0,70.7]	12.5	[11.9,13.2]	7.8	[7.3,8.3]	9.8	[9.2,10.4]	100.0
Total (n=56,152)	74.3	[73.6,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(3) =	922.0913								
Design-based F(2.99, 167664.12) =	107.2160	Pr =	0.000						
Education									
Low (n=21,346)	68.2	[67.2,69.2]	11.0	[10.4,11.7]	8.8	[8.3,9.5]	12.0	[11.3,12.7]	100.0
Medium (n=21,228)	76.1	[75.0,77.1]	10.3	[9.6,11.0]	4.8	[4.3,5.3]	8.8	[8.1,9.6]	100.0
High (n=12,833)	83.2	[82.0,84.3]	7.8	[7.1,8.6]	2.5	[2.1,3.0]	6.5	[5.7,7.3]	100.0
Total (n=55,407)	74.3	[73.7,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(6) =	1177.9308								
Design-based F(5.94, 329109.56) =	68.3745	Pr =	0.000						
Is retired									
No (n=25,298)	77.6	[76.7,78.5]	8.7	[8.2,9.3]	4.8	[4.4,5.3]	8.8	[8.2,9.5]	100.0
Yes (n=30,601)	71.0	[70.1,71.8]	11.6	[11.0,12.2]	7.0	[6.6,7.5]	10.5	[9.9,11.1]	100.0
Total (n=55,899)	74.4	[73.8,75.0]	10.1	[9.7,10.5]	5.9	[5.6,6.2]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(3) =	344.7396								
Design-based F(2.99, 167064.47) =	39.5269	Pr =	0.000						
Is married									
No (n=14,874)	69.8	[68.5,71.1]	11.4	[10.6,12.4]	8.1	[7.4,8.9]	10.7	[9.8,11.6]	100.0
Yes (n=39,474)	76.1	[75.4,76.7]	9.6	[9.2,10.1]	5.1	[4.8,5.5]	9.2	[8.8,9.7]	100.0
Total (n=54,348)	74.2	[73.5,74.8]	10.2	[9.7,10.6]	6.0	[5.7,6.3]	9.7	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(3) =	290.4041								
Design-based F(2.98, 162037.55) =	30.4726	Pr =	0.000						
Healt access									
Not difficult (n=32,586)	78.2	[77.4,78.9]	9.7	[9.2,10.3]	5.0	[4.6,5.4]	7.2	[6.7,7.7]	100.0
Difficult (n=6,341)	60.8	[58.5,63.0]	13.1	[11.7,14.7]	14.6	[13.1,16.4]	11.5	[10.1,13.0]	100.0
Total (n=38,927)	75.3	[74.6,76.1]	10.3	[9.7,10.8]	6.5	[6.1,7.0]	7.9	[7.4,8.4]	100.0
Pearson: Uncorrected chi2(3) =	1158.8844								
Design-based F(2.99, 116441.54) =	116.8392	Pr =	0.000						

Note: Calibrated cross-sectional individual weights are applied.

Table A.3, continued: Crosstable health perception (weighted)

	Health perception								
	Pos. concordance		**Underestimating**		**Neg. concordance**		**Overestimating**		**Total**
	Row %	95% CI	Row %	95% CI	Row %	95% CI	Row %	95% CI	Row %
Supplementary insurance									
No (n=26,149)	73.0	[72.1,74.0]	10.3	[9.6,10.9]	7.8	[7.3,8.4]	8.9	[8.3,9.5]	100.0
Yes (n=15,280)	78.8	[77.6,79.9]	10.4	[9.6,11.2]	4.4	[3.8,4.9]	6.5	[5.9,7.2]	100.0
Total (n=41,429)	75.1	[74.4,75.8]	10.3	[9.8,10.8]	6.6	[6.2,7.0]	8.0	[7.6,8.5]	100.0
Pearson: Uncorrected chi2(3) =	282.3894								
Design-based F(2.99, 123937.16) =	31.5801	Pr =	0.000						
Has children									
No (n=5,121)	74.2	[72.1,76.1]	9.3	[8.0,10.7]	6.0	[5.0,7.2]	10.5	[9.2,12.1]	100.0
Yes (n=50,336)	74.2	[73.6,74.9]	10.2	[9.8,10.7]	6.0	[5.7,6.3]	9.5	[9.1,10.0]	100.0
Total (n=55,457)	74.2	[73.6,74.8]	10.1	[9.7,10.6]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(3) =	10.8682								
Design-based F(2.99, 165995.64) =	1.0866	Pr =	0.353						
Country									
Austria (n=3,241)	72.7	[70.7,74.6]	11.5	[10.1,12.9]	7.1	[6.1,8.3]	8.7	[7.5,10.0]	100.0
Germany (n=5,222)	75.8	[74.4,77.2]	12.7	[11.6,13.8]	5.1	[4.4,5.8]	6.4	[5.7,7.2]	100.0
Sweden (n=4,722)	80.3	[78.9,81.6]	9.6	[8.6,10.5]	3.9	[3.3,4.7]	6.2	[5.4,7.1]	100.0
Netherlands (n=1,376)	84.2	[82.1,86.2]	8.5	[7.1,10.1]	2.9	[2.1,4.0]	4.4	[3.4,5.7]	100.0
Spain (n=5,384)	71.7	[69.8,73.5]	8.6	[7.5,9.7]	8.7	[7.7,9.8]	11.0	[9.8,12.4]	100.0
Italy (n=4,868)	70.3	[68.7,71.8]	8.3	[7.4,9.2]	6.9	[6.1,7.8]	14.5	[13.4,15.8]	100.0
France (n=4,311)	76.6	[75.1,78.1]	9.0	[8.1,10.0]	4.5	[3.8,5.2]	9.9	[8.8,11.0]	100.0
Denmark (n=4,475)	86.1	[85.0,87.1]	7.9	[7.1,8.8]	2.6	[2.2,3.2]	3.4	[2.8,4.0]	100.0
Switzerland (n=3,344)	83.9	[82.5,85.2]	7.4	[6.6,8.4]	2.6	[2.1,3.3]	6.1	[5.2,7.0]	100.0
Belgium (n=5,599)	77.0	[75.7,78.2]	11.5	[10.6,12.5]	5.2	[4.6,5.9]	6.4	[5.6,7.1]	100.0
Czechia (n=5,147)	71.6	[69.6,73.5]	10.2	[9.1,11.5]	8.3	[7.3,9.5]	9.9	[8.5,11.4]	100.0
Poland (n=1,222)	61.4	[58.5,64.3]	14.1	[12.2,16.3]	9.7	[8.1,11.6]	14.7	[12.7,17.0]	100.0
Luxembourg (n=1,013)	73.4	[70.3,76.2]	12.5	[10.5,14.8]	6.3	[4.9,8.2]	7.8	[6.1,9.9]	100.0
Slovenia (n=2,222)	72.0	[69.8,74.2]	10.8	[9.5,12.4]	7.3	[6.2,8.6]	9.9	[8.4,11.5]	100.0
Estonia (n=4,006)	64.9	[63.3,66.5]	13.8	[12.7,15.0]	12.6	[11.5,13.7]	8.7	[7.8,9.7]	100.0
Total (n=56,152)	74.3	[73.6,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(42) =	1357.8340								
Design-based F(20.44, 1.15e+06) =	26.8724	Pr =	0.000						
Survey wave									
Wave 2 (n=14,623)	73.1	[72.0,74.1]	9.9	[9.2,10.6]	5.1	[4.7,5.7]	11.9	[11.2,12.7]	100.0
Wave 5 (n=41,529)	75.1	[74.3,75.8]	10.3	[9.8,10.8]	6.6	[6.2,7.0]	8.0	[7.6,8.5]	100.0
Total (n=56,152)	74.3	[73.6,74.9]	10.1	[9.7,10.5]	6.0	[5.7,6.3]	9.6	[9.2,10.1]	100.0
Pearson: Uncorrected chi2(3) =	272.0734								
Design-based F(3.00, 168272.56) =	30.3672	Pr =	0.000						

Note: Calibrated cross-sectional individual weights are applied.

Table A.4: Underestimating health and annual number of doctor visits at $w+1$ by country

	(1) Austria	(2) Belgium	(3) Czechia	(4) Denmark	(5) Estonia	(6) France	(7) Germany	(8) Italy
Health perception (ref.: concordance)								
Underestimating	0.157*	0.180***	0.171***	0.510***	0.241***	0.173***	0.306***	0.211***
	(0.063)	(0.051)	(0.045)	(0.090)	(0.072)	(0.048)	(0.053)	(0.063)
Chronic diseases	0.166***	0.171***	0.181***	0.206***	0.226***	0.177***	0.188***	0.218***
	(0.021)	(0.016)	(0.014)	(0.019)	(0.021)	(0.014)	(0.016)	(0.018)
Activity limitations	0.072*	0.163***	0.060*	0.075*	0.128***	0.045**	0.132***	0.030
	(0.032)	(0.029)	(0.028)	(0.036)	(0.026)	(0.017)	(0.036)	(0.017)
Age	0.006	-0.035	0.036	0.004	0.006	-0.048	-0.045	0.081*
	(0.041)	(0.028)	(0.035)	(0.036)	(0.040)	(0.029)	(0.028)	(0.037)
Age squared	0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman	0.094	0.134***	0.054	0.003	-0.061	0.069	0.058	0.161***
	(0.053)	(0.038)	(0.035)	(0.047)	(0.049)	(0.036)	(0.036)	(0.042)
Educ. group (ref.: low)								
Medium	0.035	-0.051	0.012	-0.007	0.033	-0.035	0.040	-0.047
	(0.063)	(0.047)	(0.037)	(0.071)	(0.058)	(0.041)	(0.060)	(0.057)
High	0.050	-0.039	0.014	-0.016	0.073	-0.001	0.015	-0.178*
	(0.070)	(0.045)	(0.052)	(0.073)	(0.074)	(0.052)	(0.065)	(0.084)
Retired	0.074	0.056	-0.025	0.045	0.038	0.001	0.097	-0.003
	(0.065)	(0.046)	(0.067)	(0.065)	(0.067)	(0.049)	(0.054)	(0.060)
Married	-0.053	-0.082	0.065	-0.120*	-0.042	-0.016	-0.013	0.007
	(0.053)	(0.042)	(0.038)	(0.050)	(0.050)	(0.039)	(0.044)	(0.059)
Equiv. hh income (cube root)	-0.000	-0.001	-0.002	-0.004	0.011	-0.002	-0.001	-0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.007)	(0.002)	(0.002)	(0.002)
Constant	0.867	2.702**	0.172	1.307	0.893	3.183**	3.161***	-1.246
	(1.392)	(0.925)	(1.161)	(1.249)	(1.334)	(0.969)	(0.908)	(1.214)
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2,593	4,666	3,991	4,082	3,048	3,530	4,587	3,757
AIC	15,395	28,098	23,833	21,083	16,761	19,215	27,479	23,323
BIC	15,478	28,188	23,921	21,171	16,839	19,301	27,570	23,410

Note: The coefficients are based on the sample that is objectively unimpaired, i.e. able to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table A.4, continued: Underestimating health and annual number of doctor visits at $w+1$ by country

	(1) Luxembourg	(2) Netherlands	(3) Poland	(4) Slovenia	(5) Spain	(6) Sweden	(7) Switzerland
Health perception (ref.: concordance)							
Underestimating	0.088	0.370*	0.120	0.302***	0.172**	0.277***	0.240**
	(0.105)	(0.174)	(0.079)	(0.079)	(0.057)	(0.066)	(0.079)
Chronic diseases	0.107***	0.162***	0.255***	0.222***	0.181***	0.123***	0.204***
	(0.031)	(0.043)	(0.026)	(0.028)	(0.018)	(0.022)	(0.025)
Activity limitations	0.266***	0.028	0.045	-0.002	0.038	0.159***	0.222**
	(0.063)	(0.071)	(0.029)	(0.051)	(0.023)	(0.040)	(0.072)
Age	-0.098	-0.003	0.149	0.010	0.007	0.026	0.026
	(0.073)	(0.129)	(0.103)	(0.058)	(0.034)	(0.040)	(0.046)
Age squared	0.001	0.000	-0.001	0.000	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Woman	0.091	-0.054	0.038	-0.076	0.085	-0.087	-0.011
	(0.093)	(0.086)	(0.071)	(0.062)	(0.046)	(0.050)	(0.057)
Educ. group (ref.: low)							
Medium	0.228*	-0.253*	0.035	0.009	-0.043	0.103	0.007
	(0.098)	(0.099)	(0.074)	(0.069)	(0.075)	(0.068)	(0.076)
High	-0.143	0.004	-0.045	-0.053	0.049	0.080	0.031
	(0.113)	(0.112)	(0.105)	(0.090)	(0.079)	(0.069)	(0.094)
Retired	0.044	-0.236*	0.086	0.019	0.040	0.082	0.153*
	(0.104)	(0.102)	(0.086)	(0.088)	(0.050)	(0.090)	(0.077)
Married	-0.161	-0.003	0.131	0.054	0.075	-0.065	-0.147*
	(0.111)	(0.108)	(0.094)	(0.072)	(0.051)	(0.061)	(0.070)
Equiv. hh income (cube root)	0.006	-0.001	0.005	0.004	-0.001	-0.008	-0.003
	(0.003)	(0.005)	(0.007)	(0.005)	(0.003)	(0.005)	(0.002)
Constant	5.093*	1.545	-3.495	0.431	0.966	-0.038	0.578
	(2.391)	(4.046)	(3.186)	(1.906)	(1.160)	(1.388)	(1.575)
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	878	1,196	911	1,810	3,948	4,130	2,940
AIC	5,424	6,581	5,525	9,590	21,252	20,692	15,472
BIC	5,486	6,647	5,588	9,661	21,340	20,780	15,556

Note: The coefficients are based on the sample that is objectively unimpaired, i.e. able to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table A.5: Overestimating health and annual number of doctor visits at $w+1$ by country

	(1) Austria	(2) Belgium	(3) Czechia	(4) Denmark	(5) Estonia	(6) France	(7) Germany	(8) Italy
Health perception (ref.: concordance)								
Overestimating	-0.205	-0.112	-0.190**	0.006	-0.192**	-0.062	-0.068	-0.022
	(0.122)	(0.084)	(0.060)	(0.157)	(0.074)	(0.085)	(0.106)	(0.090)
Chronic diseases	0.183***	0.142***	0.138***	0.203***	0.162***	0.058	0.123***	0.237***
	(0.040)	(0.027)	(0.022)	(0.038)	(0.025)	(0.029)	(0.034)	(0.031)
Activity limitations	0.044	0.110***	0.022	0.099**	0.022	0.073**	0.049*	0.048**
	(0.024)	(0.021)	(0.016)	(0.034)	(0.021)	(0.027)	(0.022)	(0.015)
Age	0.034	0.013	0.087*	-0.185*	0.146*	-0.010	-0.110	0.065
	(0.074)	(0.055)	(0.042)	(0.081)	(0.060)	(0.056)	(0.073)	(0.051)
Age squared	-0.000	-0.000	-0.001*	0.001*	-0.001**	0.000	0.001	-0.000
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Woman	-0.085	0.151	-0.077	0.176	-0.043	0.016	0.014	0.210**
	(0.135)	(0.087)	(0.064)	(0.154)	(0.073)	(0.083)	(0.109)	(0.073)
Educ. group (ref.: low)								
Medium	-0.185	-0.128	0.013	0.093	0.018	0.013	0.107	-0.149
	(0.121)	(0.100)	(0.060)	(0.194)	(0.082)	(0.093)	(0.154)	(0.109)
High	0.139	-0.110	0.066	-0.210	-0.091	0.083	-0.047	0.014
	(0.171)	(0.094)	(0.120)	(0.181)	(0.104)	(0.126)	(0.175)	(0.146)
Retired	0.061	-0.017	-0.098	-0.014	-0.109	0.100	0.306*	-0.008
	(0.141)	(0.097)	(0.101)	(0.193)	(0.121)	(0.100)	(0.132)	(0.082)
Married	-0.004	0.150	0.055	0.171	0.185**	0.001	-0.014	-0.065
	(0.115)	(0.083)	(0.064)	(0.158)	(0.072)	(0.094)	(0.123)	(0.084)
Equiv. hh income (cube root)	-0.004	-0.013**	-0.004	-0.015	0.004	-0.001	0.005	-0.001
	(0.005)	(0.004)	(0.004)	(0.009)	(0.006)	(0.005)	(0.005)	(0.003)
Constant	1.124	1.677	-0.796	8.084**	-3.010	2.221	5.848*	-0.460
	(2.636)	(1.944)	(1.450)	(2.863)	(2.078)	(1.986)	(2.437)	(1.746)
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	423	591	964	236	834	554	532	950
AIC	2,857	4,062	6,325	1,471	5,144	3,355	3,734	6,501
BIC	2,913	4,124	6,393	1,520	5,205	3,416	3,793	6,569

Note: The coefficients are based on the sample that is objectively impaired, i.e. unable to stand up from the chair. The dependent variable “doctor visits” is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table A.5, continued: Overestimating health and annual number of doctor visits at $w+1$ by country

	(1) Luxembourg	(2) Netherlands	(3) Poland	(4) Slovenia	(5) Spain	(6) Sweden	(7) Switzerland
Health perception (ref.: concordance)							
Overestimating	-0.134	-0.346	-0.160	-0.207	-0.298***	-0.135	0.096
	(0.166)	(0.271)	(0.112)	(0.113)	(0.071)	(0.166)	(0.149)
Chronic diseases	0.112*	0.274**	0.136***	0.073*	0.142***	0.107**	0.150*
	(0.055)	(0.102)	(0.035)	(0.035)	(0.024)	(0.040)	(0.072)
Activity limitations	0.122***	-0.065	0.054	-0.019	0.017	0.066*	0.118*
	(0.032)	(0.058)	(0.034)	(0.021)	(0.018)	(0.031)	(0.055)
Age	0.373**	0.603	0.072	0.070	-0.006	-0.155	0.166
	(0.130)	(0.345)	(0.130)	(0.076)	(0.043)	(0.115)	(0.104)
Age squared	-0.003**	-0.005	-0.001	-0.001	0.000	0.001	-0.001
	(0.001)	(0.003)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Woman	-0.057	-0.059	-0.138	0.221*	-0.125	-0.003	-0.001
	(0.180)	(0.242)	(0.117)	(0.094)	(0.087)	(0.154)	(0.157)
Educ. group (ref.: low)							
Medium	0.164	0.476	0.005	-0.083	-0.130	0.290	-0.018
	(0.168)	(0.324)	(0.110)	(0.105)	(0.149)	(0.217)	(0.192)
High	-0.428	-0.294	0.238	-0.409*	-0.179	-0.197	-0.146
	(0.409)	(0.274)	(0.240)	(0.184)	(0.184)	(0.161)	(0.227)
Retired	-0.630**	-0.329	0.183	0.134	0.049	0.496*	0.122
	(0.214)	(0.331)	(0.122)	(0.121)	(0.079)	(0.238)	(0.225)
Married	-0.083	-0.179	0.044	-0.044	0.056	-0.032	-0.302
	(0.173)	(0.274)	(0.118)	(0.115)	(0.087)	(0.177)	(0.175)
Equiv. hh income (cube root)	-0.011**	0.030*	-0.003	0.002	-0.001	0.005	-0.004
	(0.004)	(0.012)	(0.012)	(0.008)	(0.005)	(0.010)	(0.005)
Constant	-9.187*	-17.679	-0.059	0.432	2.510	6.492	-4.023
	(4.345)	(10.819)	(4.064)	(2.540)	(1.525)	(4.365)	(3.622)
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	120	87	305	377	1,151	417	260
AIC	868	587	1,928	2,315	7,028	2,430	1,652
BIC	904	619	1,976	2,366	7,099	2,486	1,702

Note: The coefficients are based on the sample that is objectively impaired, i.e. unable to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table A.6: Robustness analyses for annual doctor visits of the unimpaired sample

	(1) Main	(2) Income 1	(3) Income 2	(4) Wealth	(5) Alzheimer dropped
Health perception (ref.: concordance)					
Underestimating	0.244***	0.244***	0.244***	0.240***	0.243***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)
Chronic diseases	0.181***	0.181***	0.181***	0.179***	0.182***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Activity limitations	0.096***	0.096***	0.096***	0.095***	0.097***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Age	-0.001	-0.000	-0.001	0.004	-0.001
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Age squared	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman	0.042***	0.042***	0.042***	0.042***	0.042***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Educ. group (ref.: low)					
Medium	0.006	0.008	0.005	0.018	0.007
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
High	-0.003	0.002	-0.004	0.020	-0.003
	(0.018)	(0.019)	(0.018)	(0.019)	(0.019)
Retired	0.029	0.031	0.029	0.030	0.028
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Married	-0.034*	-0.031*	-0.035*	-0.018	-0.035*
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Equiv. hh income (cube root)	-0.001				-0.001
	(0.001)				(0.001)
Wave 5	-0.089***	-0.085***	-0.089***	-0.092***	-0.089***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Equiv. hh income (cube root) 2		-0.002			
		(0.001)			
Equiv. hh income not normalised			-0.000		
			(0.000)		
Wealth quintile (ref.: 1st)					
2nd				-0.058**	
				(0.022)	
3rd				-0.103***	
				(0.021)	
4th				-0.110***	
				(0.022)	
5th				-0.130***	
				(0.022)	
Constant	1.507***	1.499***	1.485***	1.381***	1.491***
	(0.356)	(0.357)	(0.356)	(0.356)	(0.358)
Country dummies	Yes	Yes	Yes	Yes	Yes
N	46,067	46,067	46,067	46,067	45,917
AIC	260,957	260,954	260,958	260,875	259,975
BIC	261,202	261,199	261,203	261,146	260,219

Note: The coefficients are based on the sample that is objectively unimpaired, i.e. able to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w + 1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table A.7: Robustness analyses for annual doctor visits of the impaired sample

	(1) Main	(2) Income 1	(3) Income 2	(4) Wealth	(5) Alzheimer dropped
Health perception (ref.: concordance)					
Overestimating	-0.146***	-0.146***	-0.146***	-0.145***	-0.144***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Chronic diseases	0.149***	0.149***	0.149***	0.146***	0.152***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Activity limitations	0.048***	0.048***	0.048***	0.047***	0.049***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Age	0.021	0.022	0.021	0.023	0.024
	(0.018)	(0.018)	(0.018)	(0.019)	(0.018)
Age squared	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman	0.014	0.014	0.014	0.015	0.009
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Educ. group (ref.: low)					
Medium	-0.006	-0.005	-0.007	0.006	-0.007
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
High	-0.087*	-0.084*	-0.088*	-0.069	-0.084*
	(0.042)	(0.043)	(0.042)	(0.043)	(0.042)
Retired	0.014	0.016	0.014	0.019	0.011
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Married	0.019	0.023	0.019	0.034	0.012
	(0.030)	(0.031)	(0.030)	(0.030)	(0.030)
Equiv. hh income (cube root)	-0.001				-0.001
	(0.002)				(0.002)
Wave 5	-0.046	-0.042	-0.046	-0.047	-0.045
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Equiv. hh income (cube root) 2		-0.001			
		(0.002)			
Equiv. hh income not normalised			-0.000		
			(0.000)		
Wealth quintile (ref.: 1st)					
2nd				-0.050	
				(0.038)	
3rd				-0.110**	
				(0.039)	
4th				-0.119**	
				(0.040)	
5th				-0.086	
				(0.045)	
Constant	1.434*	1.434*	1.421*	1.382*	1.338*
	(0.646)	(0.645)	(0.643)	(0.645)	(0.645)
Country dummies	Yes	Yes	Yes	Yes	Yes
N	7,801	7,801	7,801	7,801	7,709
AIC	50,545	50,544	50,545	50,534	49,893
BIC	50,740	50,739	50,740	50,749	50,088

Note: The coefficients are based on the sample that is objectively impaired, i.e. unable to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table A.8: Robustness analyses for Wave 5

	(1) Objectively Unimpaired Main	(2) Objectively Unimpaired Access	(3) Objectively Unimpaired Insurance	(4) Objectively Impaired Main	(5) Objectively Impaired Access	(6) Objectively Impaired Insurance
Health perception (ref.: concordance)						
Underestimating	0.259***	0.261***	0.259***			
	(0.021)	(0.022)	(0.021)			
Chronic diseases	0.176***	0.173***	0.176***	0.143***	0.140***	0.143***
	(0.006)	(0.006)	(0.006)	(0.010)	(0.011)	(0.010)
Activity limitations	0.093***	0.095***	0.094***	0.048***	0.052***	0.049***
	(0.011)	(0.011)	(0.011)	(0.007)	(0.007)	(0.007)
Age	0.004	0.006	0.003	0.028	0.026	0.028
	(0.012)	(0.012)	(0.012)	(0.021)	(0.022)	(0.021)
Age squared	0.000	0.000	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman	0.038**	0.038**	0.039**	-0.036	-0.034	-0.036
	(0.014)	(0.015)	(0.014)	(0.032)	(0.033)	(0.032)
Educ. group (ref.: low)						
Medium	0.022	0.026	0.022	-0.005	-0.003	-0.006
	(0.018)	(0.018)	(0.018)	(0.036)	(0.038)	(0.036)
High	0.007	0.009	0.004	-0.091	-0.080	-0.089
	(0.021)	(0.022)	(0.021)	(0.048)	(0.049)	(0.048)
Retired	0.026	0.025	0.026	-0.012	-0.023	-0.013
	(0.021)	(0.021)	(0.021)	(0.038)	(0.040)	(0.038)
Married	-0.037*	-0.040*	-0.038*	0.046	0.055	0.046
	(0.016)	(0.017)	(0.016)	(0.031)	(0.032)	(0.031)
Equiv. hh income (cube root)	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Health access		0.020			0.021	
		(0.023)			(0.033)	
Supplementary insurance			0.019			-0.046
			(0.021)			(0.050)
Overestimating				-0.147***	-0.142***	-0.149***
				(0.032)	(0.033)	(0.032)
Constant	1.306**	1.226**	1.324***	1.168	1.183	1.171
	(0.398)	(0.407)	(0.398)	(0.714)	(0.753)	(0.715)
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	33,984	32,023	33,921	5,840	5,417	5,812
AIC	192,107	180,867	191,768	37,858	35,136	37,681
BIC	192,317	181,084	191,987	38,025	35,308	37,855

Note: "Unimpaired" refers to the sample that is objectively unimpaired, i.e. able to stand up from the chair and "Impaired" refers to the sample that is objectively impaired, i.e. unable to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 6. All explanatory variables are taken from Wave 5. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table A.9: Robustness analysis: Annual number of doctor visits at $w+1$ by survey wave

	(1) Objectively Unimpaired Wave 2	(2) Objectively Unimpaired Wave 5	(3) Objectively Impaired Wave 2	(4) Objectively Impaired Wave 5
Health perception (ref.: concordance)				
Underestimating	0.189***	0.259***		
	(0.037)	(0.021)		
Overestimating			-0.129*	-0.147***
			(0.062)	(0.032)
Chronic diseases	0.201***	0.176***	0.166***	0.143***
	(0.011)	(0.006)	(0.022)	(0.010)
Activity limitations	0.114***	0.093***	0.045*	0.048***
	(0.021)	(0.011)	(0.019)	(0.007)
Age	-0.077*	0.004	0.011	0.028
	(0.036)	(0.012)	(0.076)	(0.021)
Age squared	0.001*	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.001)	(0.000)
Woman	0.050*	0.038**	0.173***	-0.036
	(0.023)	(0.014)	(0.052)	(0.032)
Educ. group (ref.: low)				
Medium	-0.048	0.022	-0.014	-0.005
	(0.029)	(0.018)	(0.067)	(0.036)
High	-0.032	0.007	-0.065	-0.091
	(0.033)	(0.021)	(0.079)	(0.048)
Retired	0.023	0.026	0.067	-0.012
	(0.029)	(0.021)	(0.058)	(0.038)
Married	-0.012	-0.037*	-0.103	0.046
	(0.028)	(0.016)	(0.071)	(0.031)
Equiv. hh income (cube root)	-0.002	-0.001	-0.002	-0.001
	(0.001)	(0.001)	(0.003)	(0.002)
Constant	3.622**	1.306**	1.876	1.168
	(1.118)	(0.398)	(2.440)	(0.714)
Country dummies	Yes	Yes	Yes	Yes
N	12,083	33,984	1,961	5,840
AIC	68,559	192,107	12,603	37,858
BIC	68,736	192,317	12,737	38,025

Note: "Unimpaired" refers to the sample that is objectively unimpaired, i.e. able to stand up from the chair and "Impaired" refers to the sample that is objectively impaired, i.e. unable to stand up from the chair. The dependent variable "doctor visits" is based on the annual number of doctor visits, visits to emergency rooms and outpatient clinic visits at wave $w+1$, i.e. Wave 4 or Wave 6. All explanatory variables are taken from wave w, i.e. Wave 2 or Wave 5 respectively. The estimated coefficients are based on a negative binomial regression model with mean dispersion. Standard errors are clustered at the household level and presented in parentheses. * p<0.05, ** p<0.01, *** p<0.001

VIENNA INSTITUTE OF DEMOGRAPHY

Working Papers

Kebede, Endale, *The Relative Importance of Female Education on Fertility Desires in Sub-Saharan Africa: A Multi-Level Analysis*, VID Working Paper 09/2019.

Winkler-Dworak, Maria, Eva Beaujouan, Paola Di Giulio and Martin Spielauer, *Simulating Family Life Courses: An Application for Italy, Great Britain, and Scandinavia*, VID Working Paper 08/2019.

Hammer, Bernhard, Sonja Spitzer, Lili Vargha and Tanja Istenič, *The Gender Dimension of Intergenerational Transfers in Europe*, VID Working Paper 07/2019.

Wazir, Muhammad Asif and Anne Goujon, *Assessing the 2017 Census of Pakistan Using Demographic Analysis: A Sub-National Perspective*, VID Working Paper 06/2019.

Yegorov, Yuri, Dieter Grass, Magda Mirescu, Gustav Feichtinger and Franz Wirl, *Growth and Collapse of Empires: A dynamic Optimization Model*, VID Working Paper 05/2019.

Chisumpa, Vesper H., Clifford O. Odimegwu and Nandita Saikia, *Mortality in Sub-Saharan Africa: What is Killing Adults Aged 15-59 Years in Zambia?*, VID Working Paper 04/2019.

Striessnig, Erich and Jayanta Kumar Bora, *Under-Five Child Growth and Nutrition Status: Spatial Clustering of Indian Districts*, VID Working Paper 03/2019.

Speringer, Markus, Anne Goujon, Samir K.C., Michaela Potančoková, Claudia Reiter, Sandra Jurasszovich and Jakob Eder, *Global Reconstruction of Educational Attainment, 1950 to 2015: Methodology and Assessment*, VID Working Paper 02/2019.

Testa, Maria Rita and Danilo Bolano, *Intentions and Childbearing in a Cross-Domain Life Course Approach: The Case of Australia*, VID Working Paper 01/2019.

Goujon, Anne, Claudia Reiter and Michaela Potančoková, *Religious Affiliations in Austria at the Provincial Level: Estimates for Vorarlberg, 2001-2018*, VID Working Paper 13/2018.

Spitzer, Sonja, Angela Greulich and Bernhard Hammer, *The Subjective Cost of Young Children: A European Comparison*, VID Working Paper 12/2018.

Jatrana, Santosh, Ken Richardson and Samba Siva Rao Pasupuleti, *The Effect of Nativity, Duration of Residence, and Age at Arrival on Obesity: Evidence from an Australian Longitudinal Study*, VID Working Paper 11/2018.

---

4 Conclusion

4.1 Main findings

This thesis aims for a better understanding of how reliable health-related information from surveys is, as well as the drivers of healthcare utilisation; this is of paramount importance in the context of population ageing. The previous sections have provided clear answers to the research questions stated in the introduction. The thesis first investigated the effect of non-observation bias on the reliability of health measures. To this end, biases in health expectancies due to educational differences in the survey representation of older Europeans were quantified based on SHARE and census data. The results show that health expectancies at the age of 50 are substantially biased because the education structure in the survey does not resemble that of the general population. For most countries observed, health expectancies are upwardly biased because unhealthier, less-educated individuals are underrepresented in the data. Remarkably, many Central and Eastern European (CEE) countries analysed show the opposite pattern, namely, that high-educated individuals are less likely to participate in surveys. Health expectancies are thus downwardly biased in these countries. Overall, the findings of the first dissertation publication highlight the need to account for distortions in the education structure of survey data, for example, by utilising auxiliary information about the true education structure from censuses.

The second dissertation publication analysed biases in health measures due to distortions in health perception. In particular, it employed performance measures and their self-reported equivalents to assess which population groups over- or underestimate their physical and cognitive health. Results show that the tendency to misreport health varies strongly across Europe. Overall, northern and western European countries have fewer discrepancies than CEE or southern European countries. Southern Europeans seem particularly prone to overestimating their health. In addition to the cultural bias in self-reported data, the results show a strong decrease in correct health assessment with age for both health dimensions, some of which are explained by differences in employment status between the young and the old. Educational attainment also influences health perception, especially when individuals are asked to evaluate their own cognitive abilities. Overall, these results suggest that comparisons of self-assessed health among countries, age groups and educational groups are prone to significant biases, whereas comparisons between genders are credible for most European countries. These findings are crucial given that self-reported data are often the only information available when health-related questions are being asked.

The final publication of the dissertation extended the analysis on health misperception and analysed the link between biased health beliefs and healthcare utilisation. The results show that individuals who overestimate their health visit the doctor less frequently than those who assess their health correctly. The lower number of doctor visits is accompanied by smaller out-of-pocket expenditure. In contrast, individuals who underestimate their health visit the doctor more often and have higher out-of-pocket expenditure. Overall, excess doctor visits due to underestimating health cost the average European country 71 million International Dollars in 2020 and these costs are projected to increase to 81 million per year by 2060. Germany, Denmark and the Netherlands have especially high costs of underestimation per capita.

A common finding from all three articles of the dissertation is that socio-economic differences, in particular, educational attainment, affect every observed aspect of the analysis. Educational attainment influences survey participation, with the less-educated not being accurately represented in survey data. Socio-economic status also shapes beliefs related to health—the better educated individuals are, the better informed they are about their health status. The latter finding is also relevant for the assessment of healthcare utilisation, as health perception was shown to affect healthcare utilisation. In the context of persistent inequalities in healthcare access among socio-economic groups, focusing on health perception could be a possible starting point for increasing the healthcare utilisation of vulnerable population groups.

4.2 Contributions

The dissertation made important contributions to the topics of health measures and healthcare utilisation; these contributions have crucial implications not only for scholars working with survey data on health and ageing but also for health authorities concerned with the future organisation and sustainability of public and private healthcare provision.

The first key contribution to the existing literature is the extensive analysis of the two most important sources of biased survey statistics—non-observation error and measurement error—in the context of health measures that are regularly used to analyse the health of older adults in Europe. While previous work has considered individual domains concerned with the limitations with respect to representation and measurement in survey data, this thesis provided a more comprehensive understanding of the relationship between distortions in survey data and popular health indicators in the context of ageing societies at a micro and macro level.

In particular, the thesis explored the magnitude and direction of the bias in health expectancies due to educational differences in survey participation in Europe. Although there has been an enormous amount of research on health expectancies, past work has not addressed whether discrepancies in the education structure of surveys distort measurement. Given the widespread use of health expectancies among scholars and health authorities, it is essential to know how reliable the measurement is. Moreover, the thesis has contributed to the literature by illustrating how bias can be adjusted for with calibrated weights. Given the evidence that the gap in survey participation between socio-economic groups is increasing year on year, survey methods that adjust for flawed survey data will become even more important in the future.

The thesis also provided an assessment of credibility for self-reported data by matching survey participants' reports on their health with their actual tested health. One of the primary advantages of this approach is that the response behaviour of each survey participant can be directly evaluated in the light of his or her individual characteristics, while being fully flexible on the specification of the relationship between the tested and the self-reported variables. In the past, the strategy has been applied only to small-scale studies that evaluate either self-reported physical or cognitive health. The thesis contributed to the literature by applying this approach to a large cross-country dataset that allowed country comparisons of health perception biases for physical and cognitive health simultaneously. It also applied a previously neglected empirical strategy that enabled the bias in subjective health to be decomposed into its contributing determinants.

Finally, the thesis demonstrated the effect of health misperception on healthcare utilisation. By exploring the effect of biased beliefs on doctor visits and out-of-pocket expenditure, it shed light on a hitherto ignored driver of healthcare utilisation and contributed to the analysis of health behaviour. It also introduced a new measure of health misperception into the health economics literature and provided cost estimates of health misperception for individuals as well as the public.

4.3 Final remarks

Population ageing substantially reshapes the demand for, and configuration of, social and policy institutions. When responding to population ageing, it is crucial to base potential policy reforms on reliable information about population health. This thesis has shed light on two important topics in this context, namely, the accuracy of health measures from a macro and micro perspectives and the drivers of healthcare utilisation. More precisely, the thesis has analysed (i) the reliability of health indicators against the backdrop of survey errors, (ii) the accuracy of perceived and reported health by individuals and (iii) the effects of individuals' health perception on their health behaviour, specifically, their healthcare utilisation.

Successful efforts to enhance and monitor the health of older populations and to adapt healthcare systems so that they can provide healthcare sustainably requires reliable measures of health. This dissertation has not only revealed flaws in survey data underlying popular health measures, but has also taken a constructive stance by illustrating that surveys can still be a fruitful source of information as long as appropriate statistical methods are utilised. This is particularly true for health expectancies, a measure with immense scientific and political influence that is frequently used to assess policy targets. Above all, it is important to ensure accurate representation of all relevant population groups in surveys—in particular, vulnerable subpopulations such as less-educated individuals. In the context of ever decreasing survey response rates, however, it is becoming increasingly important to adjust for flawed survey data with post-survey adjustment such as calibrated weights. This is an important starting point for achieving the United Nations Sustainable Development Goal of "ensuring healthy lives and promote well-being for all at all ages" (United Nations, 2019).

Governments are, and will be, confronted with limited resources for public healthcare. It is thus of utmost importance to understand who is in need of healthcare (i) by capturing all relevant individuals in health surveys or by making everyone count with statistical weights, (ii) by realising how well individuals know their own health status and (iii) by understanding how these health beliefs translate into healthcare utilisation—encouraging those in need to seek the healthcare appropriate for their needs.

References

United Nations. (2019). Sustainable Development Goal 3. Retrieved December 11, 2019, from https://sustainabledevelopment.un.org/sdg3