International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria
Email: sonja.spitzer@iiasa.ac.at **Department:** World Population Program

Working paper

Who is telling the truth?
Biases in self-reported physical and cognitive health status of older Europeans

Sonja Spitzer, sonja.spitzer@iiasa.ac.at
Daniela Weber, weberd@iiasa.ac.at

WP-19-002

Approved by:

Wolfgang Lutz
Program: World Population Program
Date: 12 March 2019

Table of contents

ZVR **524808900**

This research was funded by IIASA and its National Member Organizations in Africa, the Americas, Asia, and Europe.

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 Change", and the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC under Grant ERC2012-AdG 323947-Re-Ageing and under the European Union's Horizon 2020 Research and Innovation programme (Grant Agreement No. 741105)

Abstract

This paper quantifies the contribution of individual characteristics to the bias in self-reported physical and cognitive health status of the 50-plus population in 19 European countries. The analysis utilises micro-data 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 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. For self-reported cognition specifically, education is an important factor explaining the misreporting. 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.

About the authors

Sonja Spitzer is a Research Assistant at the Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Laxenburg, Austria, and is affiliated with the Vienna Institute of Demography at the Austrian Academy of Sciences, Vienna, Austria. (ontact: sonja.spitzer@iiasa.ac.at).

Daniela Weber is a Research Scholar at the Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Laxenburg, Austria, and an Assistant Professor at the Health Economics and Policy Division of the Vienna University of Economics and Business, Vienna, Austria. (Contact: weberd@iiasa.ac.at).

Acknowledgments

We are grateful to Mujaheed Shaikh, Bernhard Hammer, Wolfgang Lutz and the participants of various conferences and seminars for their helpful comments. 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 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 and under the European Union's Horizon 2020 Research and Innovation programme (Grant Agreement No. 741105). The funder 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.

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).

1. Introduction

Understanding the bias in self-reported health and its determinants is of utmost importance, because subjective data are often the only information available to researchers and policymakers when asking 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. In a meta-analysis, Coman and Richardson (2006) 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 level of health, survey respondents will report a health level that differs from the true status depending on their individual characteristics.

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 (Hardy et al. 2014; Verropoulou 2012; Schneider et al. 2012; Jürges 2007). A second approach is the application of vignette methods, in which it is assumed that survey participants rate vignettes similarly to their own health (Bago d'Uva, Van Doorslaer, et al. 2008; Peracchi & Rossetti 2012; Voňková & Hullegie 2011). However, there is evidence that the vignette method does not capture the full scale of reporting heterogeneity in health (Bago d'Uva, Van Doorslaer, et al. 2008; Peracchi & Rossetti 2012). Finally, reporting biases can be evaluated directly by comparing survey participants' reports on their health and their actual tested health. The latter approach allows the direct evaluation of individual response behaviour. To date, however, this strategy has only been applied in small-scale studies evaluating physical health measures (Fors et al. 2006; Coman & Richardson 2006) and in small-scale studies validating self-assessed cognitive abilities (Furnham 2001; Beaudoin & Desrichard 2011; Mabe & West 1982).

Our scientific contribution is two-fold. First, we analyse discrepancies in data on self-reported physical and cognitive abilities for a large cross-country dataset that allows country comparisons. To this end, we directly match performance-based measures of mobility and memory with their self-reported equivalent. Second, we quantify which individual characteristics most relevantly contribute to the overall bias in subjective health. 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. Multinomial logit regression allows a clear estimation of the effects of individual 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 individual characteristics to the bias in self-reported health.

The remainder of this paper is structured as follows. The dataset is introduced in Section 2 with a detailed 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.

2. 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 (Börsch-Supan, Brandt, Litwin, et al. 2013; Malter & Börsch-Supan 2013; Börsch-Supan, Brandt, Hunkler, et al. 2013; Malter & Börsch-Supan 2015; Börsch-Supan et al. 2008). 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. For this paper, Wave 2 (2006/2007), Wave 4 (2010–2012), and Wave 5 (2013) are pooled (Börsch-Supan 2018a; Börsch-Supan 2018b; Börsch-Supan 2018c). The analysis of mobility is based on 88,087 observations from 17 different countries; the analysis of cognition is based on 115,785 observations from 17 different countries.

2.1 Outcome variables

We investigate the reporting behaviour of 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.

Based on the outcomes of the tested and self-reported indicators, three combinations are possible for each survey participant. First, respondents achieve concordance if they have the same outcome in both the performance-tested and self-reported variable. 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.

2.1.1 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 participants were asked whether they thought it would be safe to try standing up from a chair without using their arms (Figure 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. Furthermore, 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 (see Appendix 8.1.2).

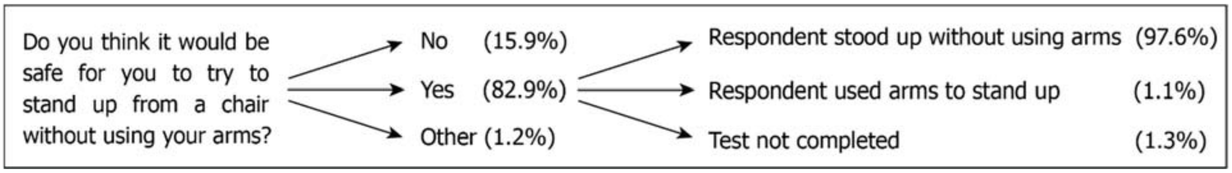

Figure 1: Sequence of questions and proportions of answers ascertaining tested mobility

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.

2.1.2 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.

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 (Grodstein et al. 2001; Purser et al. 2005). Additionally, in robustness analyses, individuals are considered impaired if they recall only two or fewer words (see Appendix 8.2.2).

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 (Gardner et al. 2017).

2.2 Determinants of concordance

The four main determinants of concordance considered in this study are are those identified in the literature as important factors of health misreporting, namely, country of residence (Hardy et al. 2014; Bago d'Uva, Van Doorslaer, et al. 2008; Jürges 2007), age (Bago d'Uva, Van Doorslaer, et al. 2008; Srisurapanont et al. 2017), gender (Bago d'Uva, Van Doorslaer, et al. 2008; Peracchi & Rossetti 2012; Schneider et al. 2012), and education (Bago d'Uva, Van Doorslaer, et al. 2008; Bago d'Uva, O'Donnell, et al. 2008). We analyse their effects on reporting behaviour and furthermore explore which of the characteristics explain most of the bias in self-reported health.

In accordance with the International Standard Classification of Education, education levels are combined into three groups (Eurostat 2018). 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.

3. Methods

We first investigate trends in the descriptive statistics. Following this, the relationship between individual 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 well established in explaining measurement errors in income survey data (Angel et al. 2018). All of our analyses are first applied to indicators of mobility and then to indicators of cognition.

3.1 Multinomial logistic regression

A multinomial logit model is applied to estimate the effects of individual 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 overestimated his or her health, underestimated his or her health, or achieved 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 baseline models are as follows:

$$\ln\left(\frac{P(y=\text{over-estimating})}{P(y=\text{concordance})}\right) = \beta_{1.0} + \beta_{1.1}COUNTRY_i + \beta_{1.2}AGE_i + \beta_{1.3}EDUC_i + \beta_{1.4}GENDER_i + \beta_{1.5}WAVE_i + \varepsilon_i \quad (1)$$

$$\ln\left(\frac{P(y=\text{under-estimating})}{P(y=\text{concordance})}\right) = \beta_{2.0} + \beta_{2.1}COUNTRY_i + \beta_{2.2}AGE_i + \beta_{2.3}EDUC_i + \beta_{2.4}GENDER_i + \beta_{2.5}WAVE_i + \varepsilon_i \quad (2)$$

$COUNTRY_i$ 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. Two more model specifications are estimated as robustness checks (see Appendices 8.1.2 and 8.2.2). First, education is interacted with gender to determine if the effects of education vary with gender. Second, we determine whether learning effects influence our estimates. Thus, if 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 before the one investigated.

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.

3.2 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 (Luchman 2013; Luchman 2014). 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.

4. Results on mobility

4.1 Descriptive results

When asked about their mobility, 19.2% of the survey participants reported difficulties getting up from a chair after sitting for long periods. However, when tested, only 17.2% were unable to stand up from a chair or considered it unsafe to try. Overall, 80.4% of the survey participants showed concordance between their reported and tested mobilities, yet the outcome varied substantially by individual characteristics. Men were 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 rated their mobility lower than it actually was compared to 7.9% of all men (see Table 1).

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

Finally, concordance varied strongly between countries. Overall, it was much higher in Northern and Western European countries than in Southern European countries, Central and Eastern European (CEE) countries, and Ireland. Denmark had the highest average concordance of 87.7%, and Poland had the lowest with only 70.4%. The variation in concordance may stem from differences in overestimating rather than underestimating, as participants from Southern and CEE countries as well as Ireland tended to strongly overestimate their mobility. Furthermore, all Southern countries were less likely to underestimate their ability to stand up from a chair.

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%

4.2 Regression analysis

Most findings were confirmed by regression analyses for both the pooled sample with all countries as well as the country-specific samples. When estimating Models 1 and 2 for the pooled sample, the coefficients showed a drastic decline of concordance with age. Individuals aged 80–84 were 2.7 times more likely to overestimate their mobility than 60- to 64-year-olds (log odds 0.976; for all coefficients see Table 2). Participants aged 90–94 were 4.4 times more likely to overestimate than the reference group (log odds 1.489). The tendency to underestimate mobility also increased with age, but less strongly than the tendency to overestimate. Furthermore, underestimating peaked at ages 80–84, but decreased again for the oldest individuals. For a better overview, Figure 6 in Appendix 8.1.1 provides the predicted values of concordance based on the country-specific estimations by age group.

Women were 1.4 times more likely to underestimate their mobility than men (log odds 0.301); in regard to overestimating, the gender effects were small (log odds 0.054). Similar to the descriptive results, the regression results indicated a clear education gradient in concordance. Less-educated participants were 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 had a lower tendency to both overestimate (log odds -0.287) and underestimate mobility (log odds -0.299). There was also an interaction between gender and education, where less-educated women in particular were prone to underestimating their ability to stand up from a chair.

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

Finally, the coefficient for the survey waves indicated that survey participants were less likely to overestimate their mobility in 2013 compared to 2006/2007 (log odds -0.414). The coefficient decreased after controlling for potential learning effects, but still remained significant. This could have been due to cohort effects, but it is not possible to disentangle cohort effects from period effects using the present dataset. When estimating Models 1 and 2 for the country-specific samples, the results from the pooled model were confirmed. However, standard errors were 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.

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

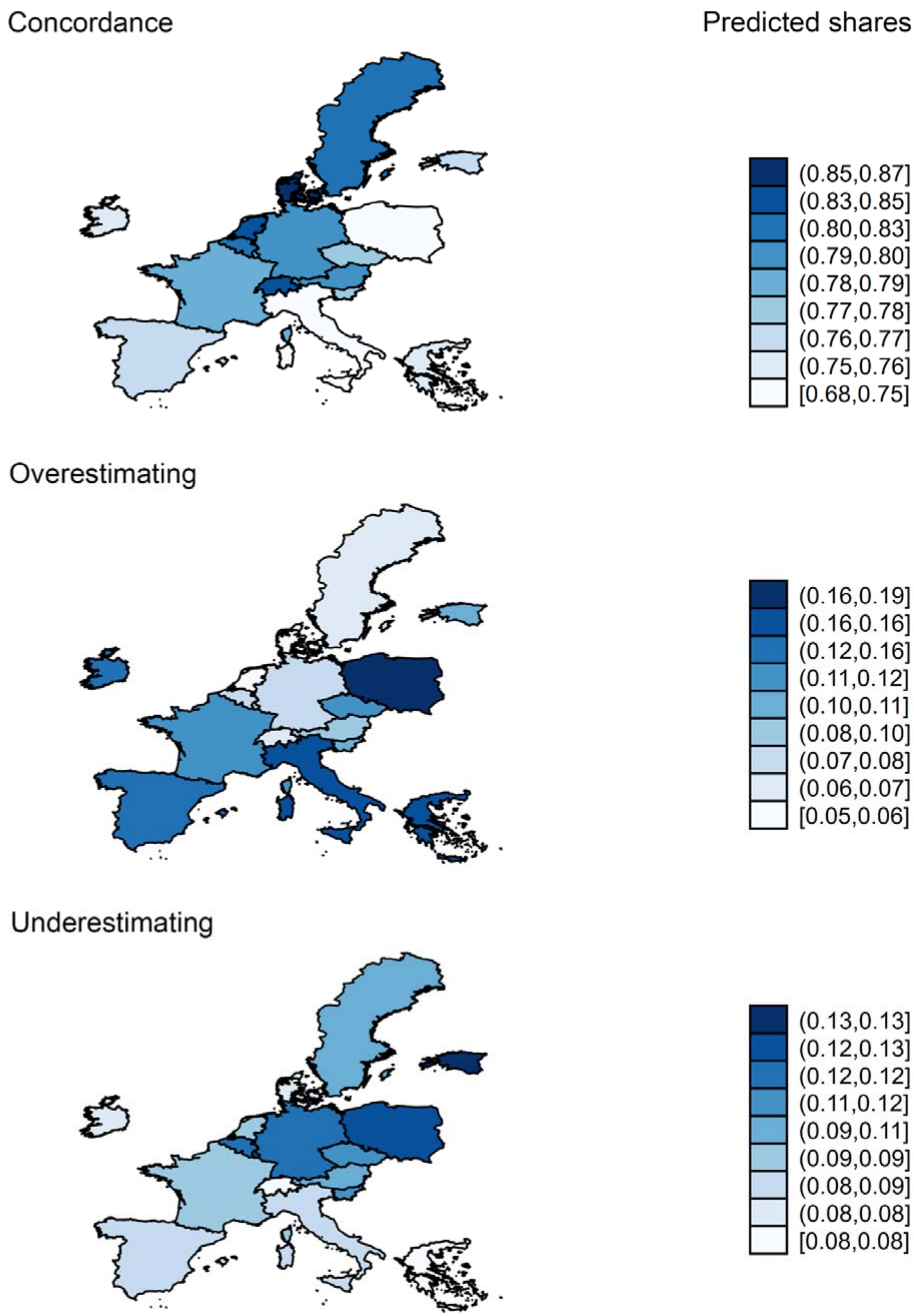

Figure 2: Concordance between tested and self-reported mobility by country (predicted shares)

4.3 Relative importance analysis

Relative importance analysis for the pooled model showed that most of the bias in self-reported mobility stemmed from differences in reporting behaviour by country and age. Country differences in reporting behaviour contributed 35.0% of the explained variance in concordance, overestimating, and underestimating. Differences between age groups explained 32.1% of the bias. Together, country and age explained more than two-thirds of the variance. Reporting heterogeneity by education contributed another 17.1%, and differences by gender contributed only 11.3%. Differences by survey waves (4.6%) contributed only nominally.

Figure 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 survey wave, the estimates of time effects for these countries are not provided. For the majority of the countries, age was the single most important characteristic explaining the bias of self-reported health. Depending on the country, either education or gender was second. The contribution of time effects was negligible in most countries, except for France, Germany, and Italy.

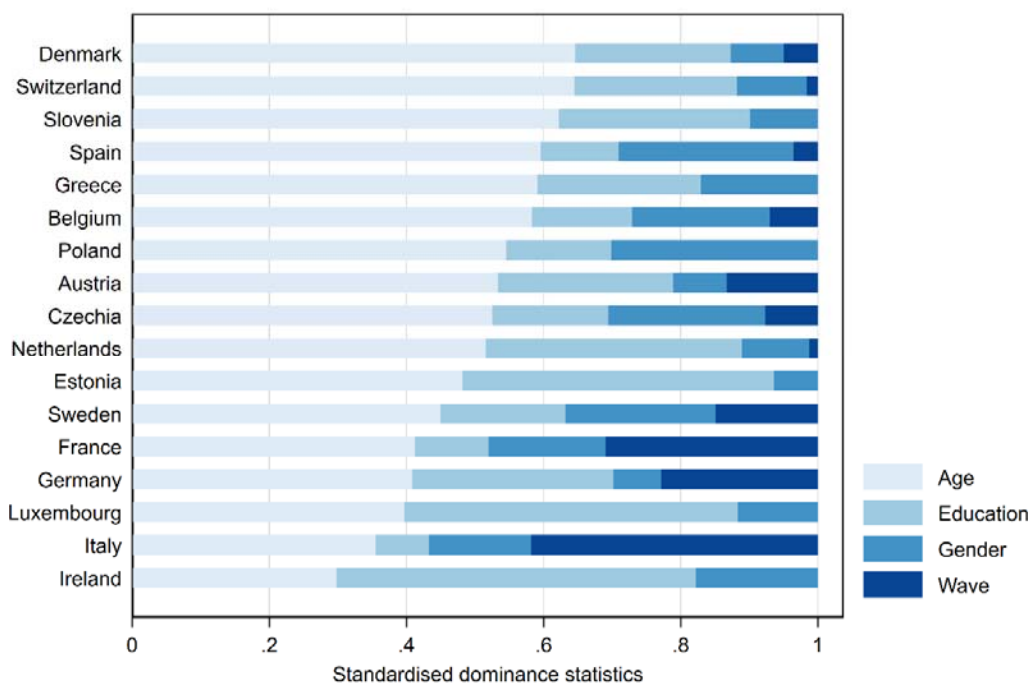

Figure 3: Decomposition of the overall bias in self-reported mobility

5. Results on cognition

5.1 Descriptive results

When asked about their memory, 29.4% of all survey participants reported cognitive impairment (see Table 1), yet when tested, only 16.1% recalled three words or less. "Overall, 71.8% of the participants showed concordance between their reported and tested memories, but there was no clear difference between genders except for a slight tendency for men to overestimate and for women to underestimate their cognition.

Similar to mobility, there was a strong decline in concordance with age. While 80.6% of the 50–54 age group reported their correct level of memory, only 63.9% of the 90–94 age group achieved concordance. Misreporting was even more pronounced at ages 80–84, in which 61.4% showed 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 increased steadily with age, under-estimating was highest at ages 75–79 (26.0%) and decreased thereafter.

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

5.2 Regression analysis

In regression analyses, concordance decreased strongly with age. Individuals aged 80–84 were 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, were 3.7 times as likely to overestimate their cognitive ability (log odds 1.297). Similar to mobility, the probability to underestimate memory increased up to ages 75–79 (log odds 0.386), but slightly decreased again for the oldest individuals. Based on the country specific samples, Figure 7 in Appendix 8.2.2 provides the values of concordance by country and age.

The effect of education on concordance was even stronger for cognition than it was for mobility. Less-educated participants were 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 was associated with a lower probability to both overestimate (log odds -0.445) and underestimate cognition (log odds -0.308).

Contrary to mobility, women were less likely to overestimate their memory than men (log odds -0.290). However, females were slightly more likely to underestimate their cognition in the pooled model. In the country-specific estimations, this finding held for Belgium, Estonia, France, Italy, Portugal, and Spain. However, in Austria, Denmark, and The Netherlands, women were less likely to underestimate their memory.

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

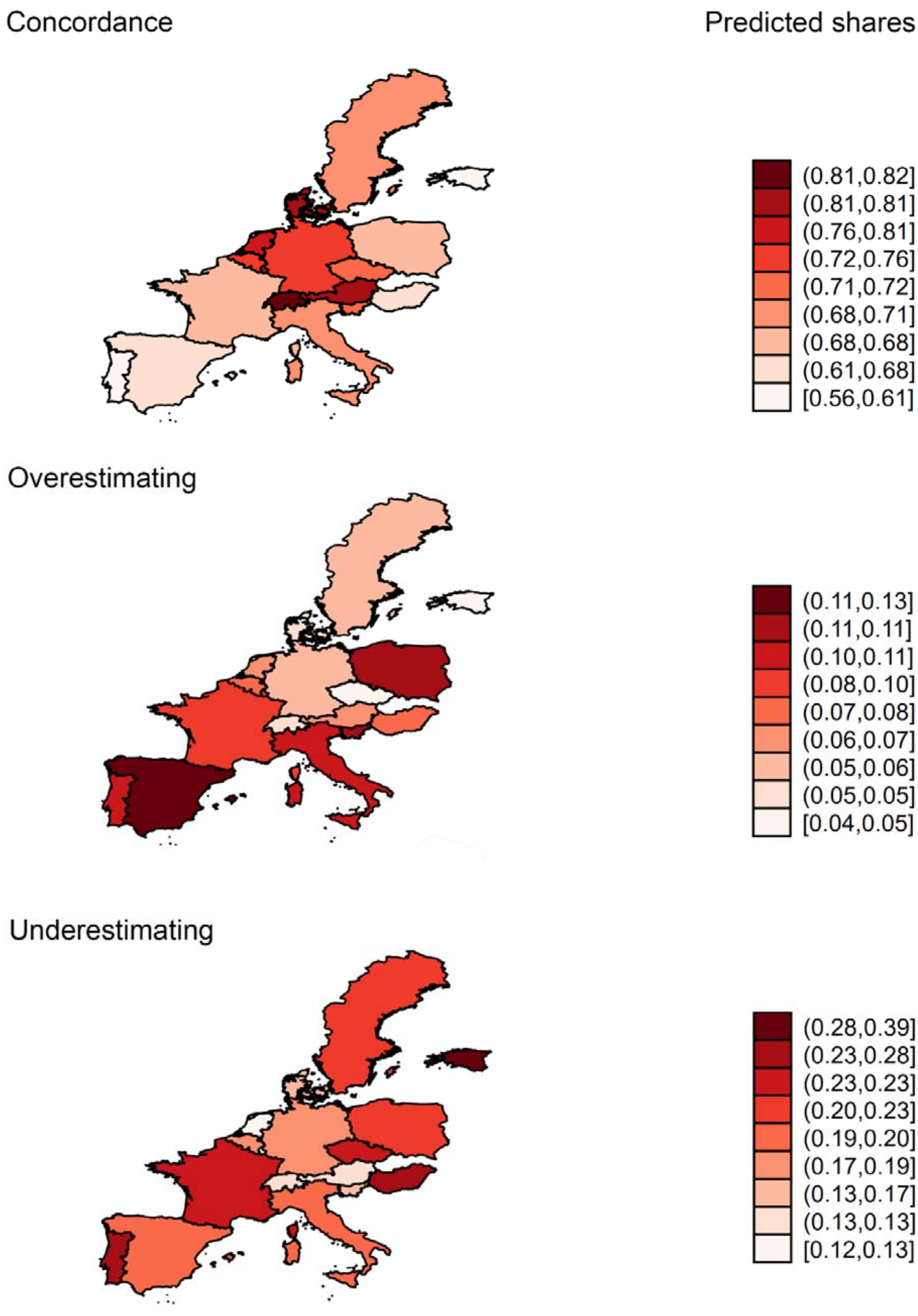

Figure 4: Concordance between tested and self-reported cognition by country (predicted shares)

Concordance between tested and self-reported cognition differed among the countries observed. Again, Southern European and CEE countries had lower rates of concordance than Western and Northern European countries (see Figure 4). Two exceptions were Czechia, which achieved a relatively high rate of concordance, and Sweden, which achieved a medium level of concordance. As with mobility, the tendency to overestimate cognitive ability was much greater in Southern and CEE countries.

Interestingly, participants of Wave 5 were less likely to overestimate and instead more likely to underestimate. This finding did not change when additionally controlling for a potential learning effect (see Table 11). As with mobility, this could indicate a cohort and/or time effect, which the available data cannot account for.

5.3 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. 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

Figures 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.

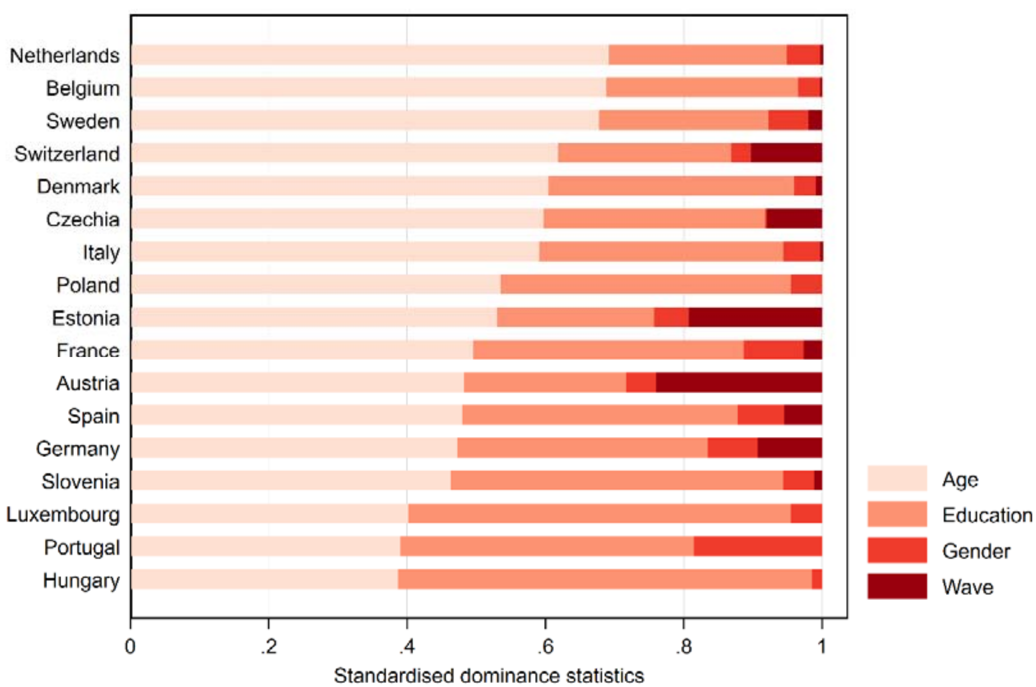

Figure 5: Decomposition of the overall bias in self-reported cognition

6. Discussion

In this study on older Europeans, we investigate the discrepancy between tested and self-reported health measures and explore the individual characteristics associated with concordance and misreporting. Overall, we find more concordance in regard to mobility than cognition. Concordance as well as the tendency to overestimate and underestimate health varies strongly across Europe. Northern and Western European countries show fewer discrepancies than CEE or Southern European countries. Southern Europeans seem particularly prone to overestimating their health. Furthermore, we find a strong decline in concordance with higher age and lower education. Reporting differences between men and women exist, but are less pronounced. We also conduct sensitivity analyses, which show that our results are robust to changes in the definition of impairment and model specifications (see Appendices 8.1.2 and 8.2.2).

Our findings show a strong decrease in concordance with age for both health dimensions, in accordance with earlier research on several physical performance measures (Grant & Ward 2010; Crossley & Kennedy 2001; Huang & Maurer 2017). Older adults that overestimate or underestimate their health might have cognitive impairments, which increase the risk of discrepancies (Fors et al. 2006).

Furthermore, we 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. Earlier research does not provide conclusive results on this matter. Some studies report that higher education results in a more optimistic view on health (Huang & Maurer 2017), while others do not find significant education effects (Guralnik et al. 1989; Kempen et al. 1996).

We also find differences in reporting discrepancies between the genders. Interestingly, women tend to underestimate their health more frequently in both health dimensions. One explanation could be that women generally tend to report limitations more frequently (Case & Paxson 2005; Verbrugge & Wingard 1987; Luy & Minagawa 2014), while men tend to underreport their health status (Oksuzyan et al. 2008). Moreover, it may be hypothesised that the gender gap in the education of older Europeans contributes to gender differences in misreporting. On average, men are more highly educated than women in the age group observed. What supports this hypothesis is that we find that less-educated women are particularly prone to underestimate their mobility (see Table 6). Regarding overestimating of health, we find that women are less likely than men to overestimate their memory, but are more likely to overestimate their mobility. The latter finding is in line with previous research (Merrill et al. 1997).

Relative importance analyses indicate that most of the bias in self-reported health stems from differences in reporting behaviour between countries, which contributes 35% to the bias in mobility and 45% in memory. Several previous studies also find that self-reports are influenced by culture, language, and education (see Coman and Richardson 2006 for an overview). Cultural biases in self-reported health have been identified in Europe (Pfarr et al. 2012; Meijer et al. 2011; Jürges 2007), in low- and middle-income countries (Capistrant et al. 2014), as well as within countries and across subpopulations (Jackson et al. 2017). Speculatively, this discrepancy might be due to differences in regional development. For a subset of our country sample, early results on the relationship between a regional developmental index (Weber et al. 2014) and discrepancies in mobility suggests 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 country effect properly.

The main limitations of this study are data-driven. First, the population composition is likely to vary across countries. For example, frail individuals might be more likely to live in institutions in certain countries than in other countries and consequently are not always included in SHARE's target population. 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 countries.

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 the other individual characteristics.

7. References

Angel, S. et al. 2018. What did you really earn last year? Explaining measurement error in survey income data. *Proceedings of the 35th IARIW General Conference*.

Bago d'Uva, T., E. Van Doorslaer, et al. 2008. Does Reporting Heterogeneity Bias the Measurement of Health Diesparities? *Health economics* 17: 351–375.

Bago d'Uva, T., O. O'Donnell, and E. Van Doorslaer. 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. and O. Desrichard. 2011. Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin* 137(2): 211–241.

Börsch-Supan, A., M. Brandt, H. Litwin, et al. eds. 2013. *Active Ageing and Solidarity between Generations in Europe: First Results from SHARE after the Economic Crisis*. Berlin: De Gruyter.

Börsch-Supan, A., M. Brandt, C. Hunkler, et al. 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. et al. 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).

Börsch-Supan, A. 2018a. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 6.1.1. SHARE-ERIC. Data set.

Börsch-Supan, A. 2018b. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 6.1.1. SHARE-ERIC. Data set.

Börsch-Supan, A. 2018c. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 6.1.1. SHARE-ERIC. Data set.

Capistrant, B.D., M.M. Glymour, and L.F. Berkman. 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.

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

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

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

Eurostat. 2018. International Standard Classification of Education (ISCED).

Fors, S., M. Thorslund, and M.G. Parker. 2006. Do actions speak louder than words? Self-assessed and performance-based measures of physical and visual function among old people. *European Journal of Ageing* 3(1): 15–21.

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.

Gardner, R.C., K.M. Langa, and K. Yaffe. 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.

Grant, L.H. and M.M. Ward. 2010. Association of measured physical performance and demographic and health characteristics with self-reported physical function: implications for the interpretation of self-reported limitations. *Health and Quality of Life Outcomes* 8(1): 84.

Grodstein, F. et al. 2001. Type 2 diabetes and cognitive function in community-dwelling elderly women. *Diabetes Care* 24(6): 1060–1065.

Guralnik, J.M. et al. 1989. Physical performance measures in aging research. *Journal of gerontology* 44(5): M141-6.

Hardy, M.A., F. Acciai, and A.M. Reyes. 2014. How health conditions translate into self-ratings: A comparative study of older adults across Europe. *Journal of Health and Social Behavior* 55(3): 320–341.

Huang, Z. and J. Maurer. 2017. Validity of Self-Rated Memory Among Middle-Aged and Older Chinese Adults: Results From the China Health and Retirement Longitudinal Study (CHARLS). *Assessment*: 1–12.

Jackson, J.D. et al. 2017. Subjective cognitive concerns are associated with objective memory performance in Caucasian but not African-American persons. *Age and Ageing* 46(6): 988–993.

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

Kempen, G.I.J.M. et al. 1996. The assessment of ADL among frail elderly in an interview survey: Self- report versus performance-based tests and determinants of discrepancies. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 51(5): 254–260.

Luchman, J.N. 2013. DOMIN: Stata module to conduct dominance analysis.

Luchman, J.N. 2014. Relative Importance Analysis With Multicategory Dependent Variables:: An Extension and Review of Best Practices. *Organizational Research Methods* 17(4): 452–471.

Luy, M. and Y. Minagawa. 2014. Gender gaps - life expectancy and proportion of life in poor health. *Health Reports* 25(12): 12–19.

Mabe, P.A. and S.G. West. 1982. Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology* 67(3): 280–296.

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

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

Meijer, E., A. Kapteyn, and T. Andreyeva. 2011. Internationally comparable health indices. *Health Economics* 20(5): 600–619.

Merrill, S.S. et al. 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.

Oksuzyan, A. et al. 2008. Men: good health and high mortality. Sex differences in health and aging. *Aging Clinical and Experimental Research* 20(2): 91–102.

Peracchi, F. and C. Rossetti. 2012. Heterogeneity in health responses and anchoring vignettes. *Empirical Economics* 42(2): 513–538.

Pfarr, C., A. Schmid, and U. Schneider. 2012. Reporting Heterogeneity in Self-Assessed Health Among Elderly Europeans: The Impact of Mental and Physical Health Status. *Health Economics Review* 2(21): 1–14.

Purser, J.L. et al. 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.

Schneider, U. et al. 2012. I feel good! Gender differences and reporting heterogeneity in self-assessed health. *European Journal of Health Economics* 13(3): 251–265.

Srisurapanont, M. et al. 2017. Discrepancy between objective and subjective cognition in adults with major depressive disorder. *Scientific Reports* 7(1): 1–7.

Verbrugge, L.M. and D.L. Wingard. 1987. Sex Differentials in Health and Mortality. *Women & Health* 12(2): 103–145.

Verropoulou, G. 2012. Determinants of change in self-rated health among older adults in Europe: A longitudinal perspective based on SHARE data. *European Journal of Ageing* 9(4): 305–318.

Voňková, H. and P. Hullegie. 2011. Is the anchoring vignette method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society. Series A: Statistics in Society* 174(3): 597–620.

Weber, D. et al. 2014. The changing face of cognitive gender differences in Europe. *Proceedings of the National Academy of Sciences* 111(32): 11673–11678.

8. Appendix

8.1 Appendix mobility

8.1.1 Predicted values by age and country

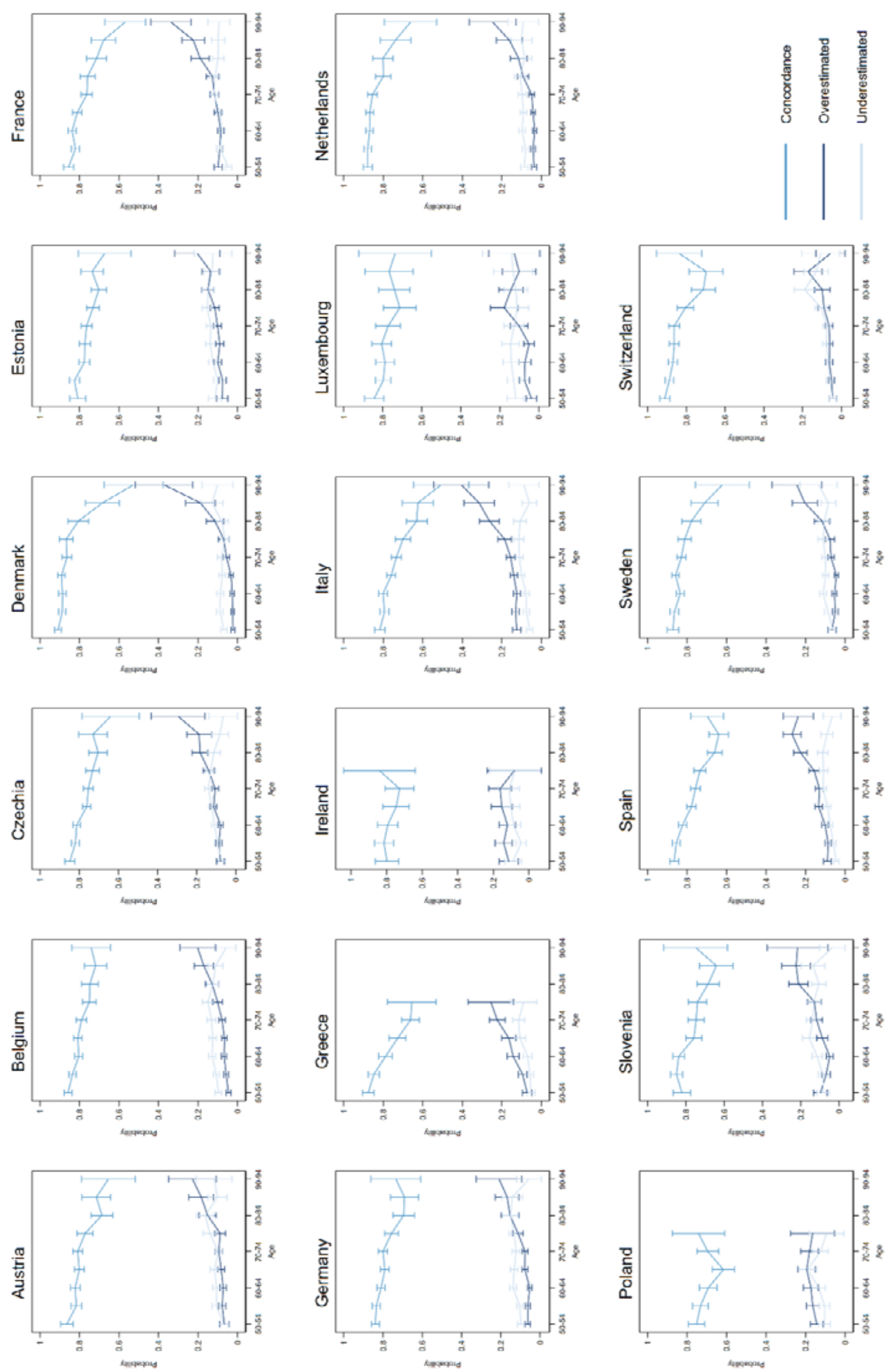

Figure 6: Predicted values of Concordance between tested and self-reported mobility by age and country

8.1.2 Robustness analyses

As a robustness check, a stricter scenario was considered, where individuals were considered impaired when they had to use their arms to stand up from the chair (Table 4). All trends described in the main text held. Most of the average values were very similar to those when individuals were allowed to use their arms. However, there was slightly less concordance and a small increase in overestimating when individuals were 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. Respondents simply might not interpret having to use their arms as a problem.

A third specification of impairment was also applied, for which individuals who thought it was unsafe to try the chair stand test were excluded from the analysis. The reduced sample included 73,912 observations instead of 88,087. As expected, this specification led to different results. Concordance increased in each subgroup, mainly because overestimating dropped to an overall of only 0.9%. This indicates that individuals that are unable to stand up from a chair avoid the test in the first place. Individuals who report having no problem getting up from a chair might prefer not to get tested if they expect to perform badly at the test. Even though the level of overestimating was much lower with the new specification, most observed trends still held. Concordance was still higher for men and highly educated individuals and decreased with age. Yet, the results by country varied from those in the first specifications. All CEE countries were still in the bottom half of concordance, but Southern European countries had higher relative rates of concordance in the new specification since large numbers of overestimating respondents were dropped in that specification. Still, most Southern and CEE European countries as well as Ireland had above-average rates of overestimation. While Northern European countries still had above-average concordance, Western European countries had a scattered distribution of results using this new specification.

Table 5 displays results for when Models 1 and 2 were estimated with the reduced sample, where everyone refusing to do the chair stand test is dropped. We found that Southern European countries had much higher concordance rates for mobility measures. Furthermore, Austria, Germany, and Luxembourg had relatively low concordance, as their tendency to underestimate mobility was relatively more important. Concordance still decreased with age, mainly due to an increase in underestimating opposed to an increase in overestimating. In summary, overestimating may have mainly been due to not taking part in the test, which is especially relevant for Southern European and CEE countries. Once these observations were dropped, underestimating was more prevalent, especially in Western European countries. As in the descriptive evidence, these results indicate self-selection of individuals in performing the test.

Table 4: 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 5: 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 6: 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 7: 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

8.2 Appendix cognition

8.2.1 Predicted values by age and country

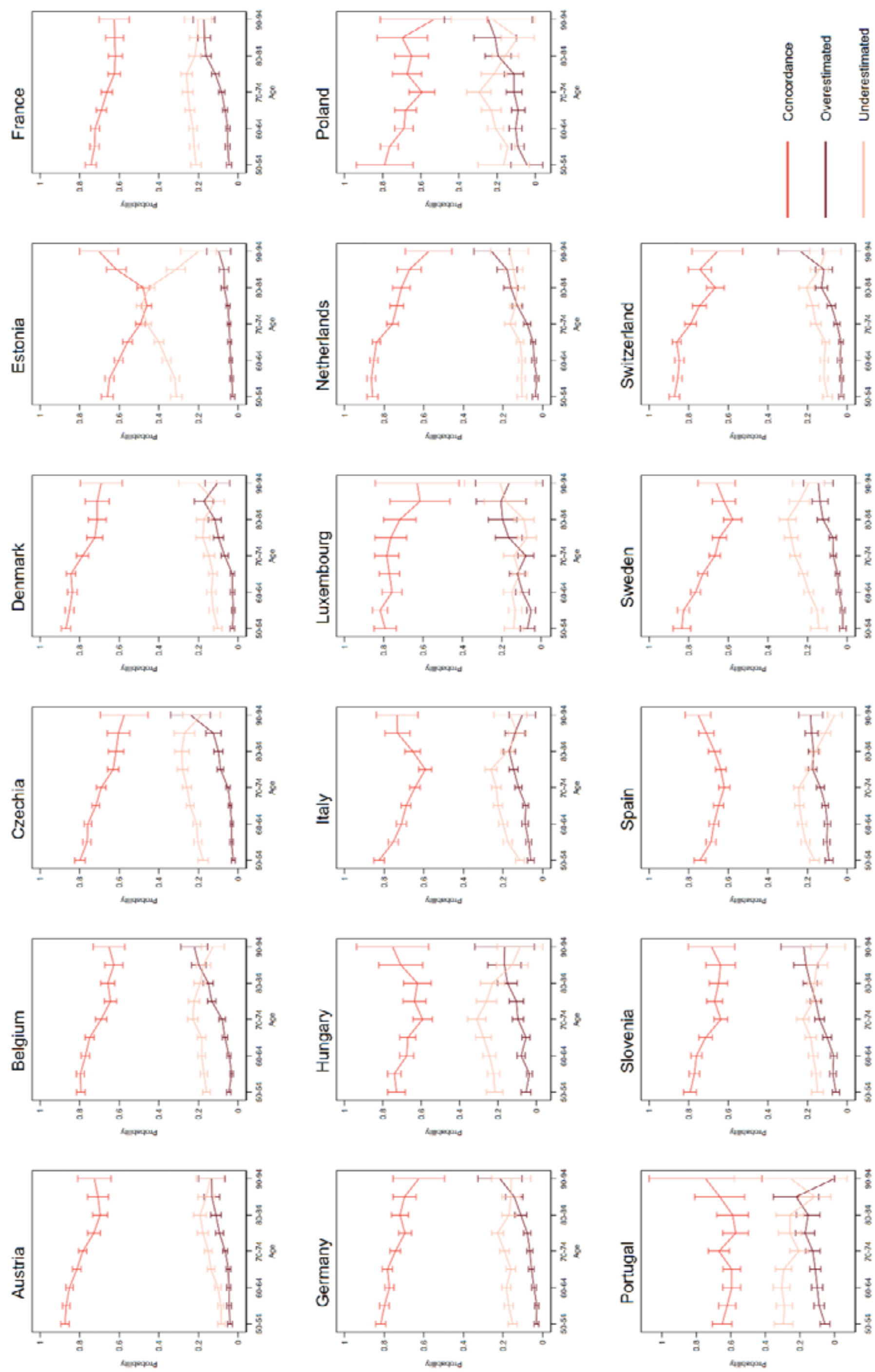

Figure 7: Predicted values of Concordance between tested and self-reported cognition by age and country

8.2.2 Robustness analyses

Table 8 provides summary statistics for an additional specification for cognitive impairment. Originally, individuals were considered to be cognitively impaired if they recalled three words or less in the memory test. For this sensitivity analysis, a more lenient threshold was applied in which participants were considered to be impaired when they recalled two words or less. Applying this specification resulted in a much lower proportion of impaired individuals (7.6% vs. 16.1% using the original specification). While the overall rate of concordance hardly changed, the tendency to overestimate was much lower and the tendency to underestimate was much higher with the new specification. This was to be expected as the new specification considered fewer individuals to be impaired.

Although the overall levels of overestimating and underestimating changed with the new specification, the trends observed in the main analysis held. Men were still more likely to achieve concordance than women. While men tended to overestimate their cognition, women tended to under-estimate theirs. The results still showed a clear decrease in concordance with age and both overestimating and underestimating showed the same patterns with age as with the original specification of impairment. We still observed a strong education gradient in concordance and the country ranking was almost identical to that of the original specification: Switzerland still had the highest rate of concordance (83.2%), while Estonia had the lowest (53.1%).

Table 9 displays the regression results for Models 1 and 2 when using the new specification of cognitive impairment. The magnitude of the coefficients changed, yet the findings remained the same as with the original specification. The patterns with age and between countries were almost identical to the main findings. The only difference was that the level of overestimating was lower and the level of underestimating was 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.

Table 8: Summary statistics after applying 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 9: Multinomial logistic estimation for concordance between cognition measures (new specification)

	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 10: 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 11: 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