Universität Bayreuth
Rechts- und Wirtschaftswissenschaftliche Fakultät
Wirtschaftswissenschaftliche Diskussionspapiere

Reporting Heterogeneity in Self-Assessed Health among Elderly Europeans: The Impact of Mental and Physical Health Status

Christian Pfarr[*], *Andreas Schmid*[†] *and Udo Schneider*[‡]

Discussion Paper 02-11

March 2011

ISSN 1611-3837

Abstract
Self-assessed health (SAH) is a frequently used measure of individuals' health status. It is also prone to reporting heterogeneity. To control for reporting heterogeneity valid measures of the objective health status are needed. The topic becomes even more complex for cross-country comparisons, as many key variables tend to vary strongly across countries, influenced by cultural and institutional differences. This study aims at exploring the key drivers for reporting heterogeneity in SAH in an international context. To this end, country specific effects are accounted for and the objective health measure is concretized, separating out effects of mental and physical health conditions. We use panel data from the Survey of Health, Ageing and Retirement in Europe (SHARE) which provides a rich dataset on the elderly European population. To obtain distinct indicators for physical and mental health conditions two indices were constructed. Finally, to identify potential reporting heterogeneity in SAH a generalized ordered probit model is estimated. We find evidence that health behaviour as well as health care utilization, mental and physical health condition as well as country characteristics affect reporting behaviour. We conclude that observed and unobserved heterogeneity play an important role when analysing SAH and have to be taken into account.

Keywords: reporting heterogeneity, SHARE, generalized ordered probit
JEL: *JEL*: C23, I10, I12

---

[*] Christian Pfarr, University of Bayreuth, Department of Law and Economics, Institute of Public Finance, D-95447 Bayreuth, e-mail: christian.pfarr@uni-bayreuth.de, phone: +49-921-554324.
[†] Andreas Schmid, University of Bayreuth, Department of Law and Economics, Institute of Public Finance, D-95447 Bayreuth, e-mail: Andreas.schmid@uni-bayreuth.de, phonw: +49-921-554324.
[‡] Udo Schneider, University of Bayreuth, Department of Law and Economics, Institute of Public Finance, D-95447 Bayreuth, e-mail: udo.schneider@uni-bayreuth.de, phone: +49-921-552929.

1 Introduction

Knowledge about the health status of individuals is paramount when health interventions are to be evaluated. Often, self-assessed health (SAH) is used as a key measure to this end. However, SAH is prone to inaccuracies due to reporting heterogeneity. Given an identical understanding of health-related questions, self-assessed health would reflect (unobservable) true health which would make it a valid indicator. However, varying reporting behaviour leads to discrepancies between self-assessed health and the underlying true health. This may result in systematic differences in the stated health across population subgroups, even if the underlying true health status is identical. This gains importance when cross country comparisons are considered. The respective institutional or cultural setting can influence asymmetries between true and self-assessed health. Objective health measures as well as SAH presented in this paper show considerable differences between countries. However, they do not reveal any sort of common pattern, which again directs the attention to potential causes for this finding.

This study aims at exploring the key drivers for reporting heterogeneity in SAH in an international context. To this end, country specific effects are accounted for and the objective health measure is further concretized, separating out effects of mental and physical health conditions.

For our analysis, we use panel data from the Survey of Health, Ageing and Retirement in Europe (SHARE) which covers countries from all parts of Europe. The survey covers individuals over the age of 50 including their partners, and provides a broad set of socio-demographic and health related variables. The focus on the elderly is beneficial to the purpose of our study, as they typically face the highest level of morbidity and usually have a long history of dealing with their health issues. This results in a comparatively rich dataset, which also allows an in-depth analysis of very specific issues like reporting heterogeneity in SAH.

The methodological approach is a consequent extension of state-of-the-art literature. Using panel data, we estimate a generalized ordered probit model to identify potential reporting heterogeneity in self-assessed health. To differentiate between true and self-assessed health, we construct indices for mental and physical conditions. Together with measures for health care utilization, socio-demographic variables and country effects these are used to evaluate their relevance for reporting heterogeneity.

Our results indicate that the reporting of self-assessed health differs across countries. For example, in countries like Germany and Spain, individuals systematically report a lower health status whereas Dutch respondents show a higher probability to opt for the best category. For both health indices, we find evidence of reporting heterogeneity. This means that a worse objective health status not only leads to a lower perception of own health, but also that the impact of the effect varies between the categories of SAH. Moreover, the magnitude of mental health problems exceeds the effect of the physical health index. We find further evidence for reporting heterogeneity when looking at aspects like health care utilization and health relevant behavior.

Our results thus allow to qualify the cross-country differences in self-assessed health. The findings imply that a straight forward comparison of health interventions on the basis of self-reported measures is highly complex and needs to be adjusted. Therefore any analysis has to control for the key factors driving reporting heterogeneity. By taking into account these factors, i.e. objective health measures, health care utilization and country-specific characteristics, the robustness of the results would be improved significantly.

The paper is organized as follows: Section 2 gives a systematic overview over the related literature. Based on the previous studies, we sketch out how our analysis contributes to this field of research. In the following section, the dataset is presented and the methodological approach is elaborated. We construct the objective health measures and provide a first descriptive cross-country comparison which illustrates the differences between reported and true health. Section 4 presents the results of the estimation of a generalized ordered probit model. The last section reflects the findings and offers suggestions for future research.

2 Literature review

There exists a broad empirical literature dealing with the application and the interpretation of health indicators such as self-assessed health. For the purpose of this paper, this literature can be classified into four fields.

The first concentrates on problems of self-reported health measures related to labour supply and retirement. The main focus of these papers is on a possible endogeneity of health that may be driven by different valuations of individual health. Butler (1987) and Bound (1991) were among the first to analyse potential measurement errors in self-reported health. Both make use of clinical measures and therefore try to objectify self-reported health. Whereas Butler et al. find that SAH is a valid indicator and measurement errors occur due to socioeconomic characteristics, Bound shows that if current health and an objective measure are imperfectly correlated this may result in problems of endogeneity. In order to obtain an objective measure of individual health stock, Disney (2006) instruments self-assessed health with explanatory variables such as personal characteristics and health indicator variables. By doing so, they find that this approach is superior to other approaches and that the retirement decision is affected by the lagged and the current health stock. Kerkhofs and Lindeboom (1995) assume that the endogeneity is driven by systematic misreporting in subjective health questions (state-dependent reporting bias). Using an objective health measure by taking various health problems into account, their results suggest that misreporting is mainly caused by labour market status and that subjective health measures lead to biased estimates. In an extension of this work, Lindeboom and Kerkhofs (2009) estimate a simultaneous model of labour supply together with a health production model. They present evidence that the reporting of health problems is characterized by a great deal of heterogeneity. Therefore, the authors suggest including more specific and therefore more objective health indicators. Coe and Zamarro (2011) study the impact of retirement on health by using a multi-country setting. In a Europe wide survey, they instrument retirement and estimate the impact on various health measures. They construct an objective health index using information of reported diseases, utilization and outcomes. The results indicate that retired individuals show a tendency to report a better health status. The impact on self-reported health is only temporary, while for the health index, long-lasting differences in health can be found.

The second field of studies deals with income inequality and reporting heterogeneity in self-assessed health. Van Doorslaer and Jones (2003) analyse differences in reporting that may be influenced by socioeconomic characteristics such as age, gender, education, individual experience with illness and the health care system. In detail, sub-groups of the population systematically use different thresholds in classifying their health into a categorical measure. Also exploring the income-health nexus, Etilé and Milcent (2006) separate the income effect on self-assessed health into a health production effect and into an effect reflecting heterogeneity, and provide evidence of a convex relationship between reporting heterogeneity and income. In a recent study, Ziebarth (2010) investigates the impact of income inequality using a concentration index as inequality measure. The widespread self-assessed health measure goes along with the highest income inequality. Concentration is significantly lower if other measures of health, e. g. the SF12 or grip strength, are used, what results in a lower degree of heterogeneity in reporting health.

The third strand of the literature focuses on methodological aspects. An interpretation of heterogeneous reporting behaviour could be that different population sub-groups use different reference points when answering health related questions. This kind of heterogeneity may express itself either in a shift of the mean or in influencing the shape of the distribution. In the literature, these effects are denoted as index and cut-point shifts (Lindeboom and van Doorslaer, 2004). When the distribution of the health measure shifts completely to the right or left, whereas the shape itself remains unchanged, an index shift can be detected. In presence of a cut-point shift, the reference points depend on the individual response behaviour and characteristics, which leads to a change in the shape of the distribution and thus to a non-parallel shift of cut-points.

Lindeboom and van Doorslaer (2004) study the relation between self-reported and objective health, using the health utility index mark III (HUI3)). They present a framework to identify the presence of cut-point and index shifts. They run estimations for population sub-groups and find evidence for both kinds of shift depending on age and gender but not on income, education or language skills. To control for both types, Hernández-Quevedo et al. (2005) study the effects of different wordings in one wave of the British Household Panel Survey. There, an index shift caused by the modified self-assessed health question can be identified. On the contrary, a cut-point shift is not present in the data. In contrast, Bago d'Uva et al. (2008) use anchoring vignettes to objectify health measures.[1] For three Asian countries (Indonesia, India and China) homogeneous reporting as well as a parallel shift of the reporting thresholds could be ruled out. According to this, the assumption of identical reference points is rejected. Therefore, answering self-assessed health questions differs between countries. Bago d'Uva et al. (2010) apply a vignettes approach to identify reporting heterogeneity, too. They analyse education-related inequity in the utilisation of health care (doctor visits). Generally, self-assessed health is often used as a proxy of medical need. Using data on elderly Europeans, they find a bias when self-assessed health is used in the analysis of the distribution of doctor visits. When correcting for this heterogeneity, depending on education levels, the inequality in health care utilisation increases.

Finally, there are two more studies related to our approach. From an international perspective, Jürges (2007) explores the differences in true vs. reported health using data from the first wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) for the year 2004/2005. To compare health across countries, he uses 15 diagnosed physical conditions together with a variable indicating depression, BMI, and two quasi-objective indicators (grip strength and walking speed) to compute a disease index as a proxy for true health. The dependent variable is self-assessed health, ranging from excellent to poor. The index is first estimated by using an ordered probit model including country dummies. Second, generalized ordered probit regression

[1] Here, respondents are asked to rate hypothetical descriptions of a fixed level of a latent construct, such as responsiveness (King et al., 2004).

allows for varying thresholds dependent on fixed country effects. From the latter regression, he constructs disability weights by dividing the estimated coefficients according to their range of the linear prediction. Jürges starts with a specification where these weights are assumed to be identical across countries. For the ten countries in the sample, he finds that self-reported health shows large cross-country heterogeneity.[2] Relaxing the assumption of identical disability weights using the Oaxaca-Blinder Decomposition of SAH allows separating prevalence and severity effects from other influences. Heterogeneity can then be reduced if SAH distributions are assumed to possess an identical underlying response style. This means that cross-country variation to a certain degree depends on differences in reporting styles. However, in contrast to previous studies, Jürges concentrates only on the construction of an objective health measure, ignoring socio-economic factors influencing the reporting of individual health. Moreover, he uses self-assessed health as a dependent variable in the index regressions, whereas the common procedure focuses on the inclusion of an objective health measure, when potential heterogeneity of self-assessed health is analysed.

Taking these two points into account, Schneider et al. (2011) analyse how both socioeconomic factors and disease experiences influence the individual valuation of health. Applying a generalized ordered probit model to German panel data, they control for observed heterogeneity in the categorical health variable allowing the thresholds to depend on ex-ante identified explanatory variables. Specifically, they concentrate on the varying answering behaviour of female and male respondents, pointing to a gender specific perception and assessment of health. The results suggest strong evidence for cut-point shifts. Especially experience with different kinds of illnesses may be one source of reporting heterogeneity. Moreover, income as a possible source of heterogeneity seems to be more central for men than for women. Other gender differences exist with respect to the influence of education on the reporting behaviour of health.

One major finding of the presented studies is that self-reporting of health questions is affected by heterogeneity. More specifically, the studies show differences between self-reported and the latent “true” health. These findings apply to retirement-related research as well as to studies concerning the heterogeneity regarding income inequality. Furthermore, more methodological papers confirm the heterogeneity in a more general context. One approach to deal with this heterogeneity is to include more objective health measures as proxy for true health. Such objective measures can be based on illness diagnoses and aspects of daily limitations. If medical

[2] In details, the sample consists of the following countries: Austria, Germany, Sweden, the Netherlands, Spain, Italy, France, Denmark, Greece and Switzerland.

diagnoses or objective diagnoses are not available from the questionnaire, one way out of this problem could be to use individuals' stated diagnoses.

Except for the study by Schneider et al. (2011), all others rely on cross-section or pooled data and omit the possibility to account for unobserved heterogeneity through panel data methods. For the purpose of this paper, especially the work of Jürges (2007) is highly relevant. First of all, to our best knowledge, this is among the first applying the question of reporting heterogeneity in self-assessed health to the Survey of Health, Ageing and Retirement in Europe (SHARE). Second, the author presents a method to compute a disease related health index. However, Jürges only concentrates on the development of the health index. In this setting, aspects of physical as well as mental health enter the index regressions. Hence, it is not possible to distinguish between the various factors. In addition, using self-assessed health as dependent variable, he neglects the impact of socioeconomic factors driving the heterogeneity. Furthermore, while he applies a generalized ordered probit model implying variation in the cut points (thresholds), the crucial assumption that this variation is only due to country effects is restrictive. Without any underlying theory, it seems to be difficult to identify factors influencing the threshold variation. In contrast to this, in a German context, Schneider et al. (2011) use a procedure proposed by Williams (2006), Boes (2007) and Pfarr et al. (2010) to identify variables that drive the heterogeneity and variables that have no impact on the regression thresholds.

Our paper aims at applying the procedure used by Schneider et al. to the SHARE data with a distinct view on the influence of mental and physical health regarding the heterogeneity of health. According to this, we use two separate indices: one related to mental conditions and one covering special diagnosed physical diseases. In contrast to Jürges, the construction of our indices does not rest on the variable self-assessed health, but is instead related to a general question capturing limited activities. Based on this approach, we include both indices to evaluate possible heterogeneous reporting behaviour. For the analysis at hand, we apply a random-effects generalized ordered probit model to European panel data, and allow the individual cut-points to vary across a broad range of explanatory variables, and test for variables causing the heterogeneity.

3 Data and estimation method

3.1 Data description

In this study, we use data from the Survey of Health, Ageing and Retirement in Europe (SHARE). The full dataset contains information on more than 45.000 elderly Europeans which was collected in two survey periods. The first wave was surveyed in 2004 and 2005, the second wave in 2006 and 2007. Each individual, aged 50 years or older, was eligible to participate, including spouses and partners irrespectively of their age. Eleven out of a total of 16 countries were surveyed in both periods, providing a panel that covers all parts of Europe. A broad set of socioeconomics variables as well as in depth surveys of special topics make SHARE a valuable tool for research. In our case, health related questions are of particular interest. This field forms one of the main three thematic pillars of the survey. It embraces hard and soft health variables as well as psychological variables, information on health care utilisation and similar related topics (Börsch-Supan and Jürges, 2005).

Health and income related variables are prone to item non-response which could make the sample less representative. In the SHARE dataset, however, a large set of important variables is imputed which again strengthens the usability and applicability of this dataset. A missing value is imputed five times, resulting in five complete datasets including all imputed and not-imputed variables. Rather than presenting a single guess for an imputed missing observation this approach aims at re-creating the distribution of the missing value in respect to a particular variable (Christelis, 2010).[3]

For the analysis of reporting heterogeneity, we use the five-point categorical variable self-assessed health. This variable ranges from excellent to poor. As explanatory factors, we account for sociodemographic characteristics, health related variables as well as country indicator variables, using an unbalanced panel structure. The complete list of variables is presented in table 1. The first group covers age and gender effects, the influence of education, and income as well as family status and nationality. Possible nonlinearity in calendar age is captured by including a linear as well as a quadratic age term. To incorporate possible impacts of income, we refer to the relative income position of a household member (Statistisches Bundesamt; Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen; Wissenschaftszentrum Berlin für Sozialforschung, 2008). This is based on the net household equivalent income. The relative position depends on the median separately computed for each country and period. To compare education across countries, the International Standard Classification of Education (ISCED 1997) is used. In six categories, school degrees as well as degrees of vocational training is included. The group of health-related variables consists of health behaviour, health condition and health

[3] The estimation with different imputations requires some caution with respect to the 'averaging' of the results (see StataCorp., 2009). For the total results, it follows that the coefficient vector of the multiple imputation analysis is given by the mean of the single estimations while for the variance-covariance estimate one has to distinguish between the within- and the between-imputation variance-covariance matrix.

care utilization. Health behaviour encompasses smoking and drinking behaviour as well as physical activities, such as sport or heavy housework. The physical and mental condition variables indicate the multimorbidity and the mental state of the respondent. Both variables are indices ranging from 0 to 100, with higher values indicating a worse condition (see chapter 3.2). Moreover, doctor visits and the number of nights in hospital are proxies for the utilization of health care. For each, the reference categories represent no doctor visits or no night in hospital respectively. To account for cross-country variation not captured by the other variables, we include country fixed effects with France as reference. In detail, the other included countries are: Austria, Germany, Sweden, Netherlands, Spain, Italy, Denmark, Greece, Switzerland and Belgium.

Table 1: Variable description

variable name	variable description
SAH	Self-assessed health, 1=excellent, 5=poor
Survey Period	1 if survey period 2006/2007
Gender	1 if female
Age	Age in years
Age^2	Age squared divided by 100
Marital status	1 if living with a partner or a spouse
Foreign	1 if foreign
Grandchildren	1 if respondent has got one or more grandchildren
Children	1 if respondent has got one or more children
Very low income	1 if income ≤ 50 % of the country's median equivalent net household income
Low income	1 if income > 50 % but ≤ 75 % of the country's median equivalent net household income
High income	1 if income > 125 % but ≤ 150 % of the country's median equivalent net household income
Very high income	1 if income > 150 % of the country's median equivalent net household income
Education	Level of education according to the ISCED scale (0 to 6)
Smoking	1 if respondent has ever been a daily smoker for at least one year
Drinking	1 if respondent has been drinking alcoholic beverages at least once or twice a week over the past six months
Physical activity	1 if respondent is engaged in vigorous physical activity like sports or heavy housework at least once a week
Physical condition	Index of respondents physical health status
Mental condition	Index of respondents mental health status
Doctor visits 1-3	1 if 1 to 3 doctor visits in the last 12 months
Doctor visits 4-11	1 if 4 to 11 doctor visits in the last 12 months
Doctor visits >11	1 if more than 11 doctor visits in the last 12 months
Hospital nights 1-6	1 if 1-6 nights in hospital in the last 12 months
Hospital nights 7-14	1 if 7-14 nights in hospital in the last 12 months
Hospital nights >14	1 if more than 14 nights in hospital in the last 12 months

The total number of observations from the two periods and eleven countries amounts to 53,931. As can be seen from table 2, the mean of self-assessed health is 2.95, indicating a slight tendency to report a poor health status. In addition to individuals aged 50 and above, the SHARE data contains information on younger partners and spouses. The mean of the age variables is 64.45 years and 56 % of the respondents are female. Almost 50 % of the respondents stated to have been a daily smoker for at least one year at some point in their life. Only 33 % report frequent drinking of alcoholic beverages during the past six months. Concerning health care utilization, 86 % visited a doctor at least once in the last twelve months, and 13 % had to stay in hospital for at least one night.

Table 2: Summary statistics

	N = 53,931	
	Mean	*SD*
Dependent variable		
SAH	2.95	1.06
Explanatory variables		
Survey Period	0.49	0.50
Gender	0.56	0.50
Age	64.45	10.35
Age^2	42.61	13.83
Marital status	0.76	0.43
Foreign	0.02	0.15
Grandchildren	0.63	0.48
Children	0.89	0.31
Very low income	0.15	0.35
Low income	0.18	0.38
High income	0.10	0.30
Very high income	0.28	0.45
Education	2.55	1.52
Smoking	0.48	0.50
Drinking	0.33	0.47
Physical activity	0.50	0.50
Physical condition	49.87	9.91
Mental condition	49.93	9.95
Doctor visits 1-3	0.33	0.47
Doctor visits 4-11	0.36	0.48
Doctor visits >11	0.17	0.38
Hospital nights 1-6	0.07	0.25
Hospital nights 7-14	0.03	0.18
Hospital nights >14	0.03	0.16
Austria	0.06	0.23
Germany	0.1	0.3
Sweden	0.1	0.3
Netherlands	0.1	0.3
Spain	0.08	0.27
Italy	0.1	0.3
Denmark	0.08	0.27
Greece	0.11	0.31
Switzerland	0.04	0.2
Belgium	0.13	0.33

3.2 Computation of physical and mental condition indices

For the purpose of this paper, it is essential to include objective health measures in the analysis of reporting behaviour of SAH. As it is clear from the literature, the identification of cut-point and index shift is only possible with an objective measure of true health. Therefore, we use a wide range of physical disabilities and mental states included in both waves of the SHARE dataset. Concerning the physical disabilities, we rely on questions regarding specific illnesses which were diagnosed by a physician. Our assessment of the individual's mental condition is closely linked to emotional health or well-being which is captured through self-reported feelings and valuations of the personal life situation. The included aspects constitute core criteria for the EURO-D scale, a depression symptom scale. For a detailed list of variables in use see tables table 3 and table 4+) Variable dropped for some countries due to collinearity.

Table 4.[4]

Referring to the work of Kerkhofs and Lindeboom (1995) or Jürges (2007), we construct indices on physical and mental conditions to objectify the reporting of illnesses or emotional distress. In a first step, we regress the binary indicator "limited activities" separately on the sets of physical and mental variables.[5] The regressions for the physical and mental conditions index are run separately by country, gender and survey period, using standard probit models. By doing so, we account for different prevalence rates of specific physical and mental conditions, gender differences and time effects. Hence, our indices will include cross-country variation and will therefore capture health related differences besides the country fixed effects. The results of the index regression for the period 2006/2007 are presented in table 3 and table 4

[4] In contrast to Jürges (2007), we refrain from using the variables walking speed and grip strength. These variables show a large number of missing values (about 10 % for grip strength) or are not available for respondents younger than 75. Jürges assumes that all individuals for which walking speed is not measured to have a normal walking speed. Further, the BMI is not included in our specification because first, it may influence both, mental and physical conditions. Second, the BMI can be seen as the result of individual behaviour rather than a diagnosed disease. Moreover, especially obesity is closely related to diseases like diabetes, cholesterol, arthritis or heart problems and influences the utilization of health care resources (Andreyeva, 2006).

[5] The wording of the corresponding question is: „For the past six months at least, to what extent have you been limited because of a health problem in activities people usually do?"

Table 3: Physical condition index

	AUT	GER	SWE	NED	ESP	ITA	FRA	DEN	GRE	SUI	BEL
Male											
heart attack	0.83***	0.59***	0.34***	0.30**	0.78***	0.84***	0.32***	0.44***	0.35***	0.46**	0.54***
high blood pressure	-0.23**	-0.22***	-0.22***	-0.31***	-0.45***	-0.45***	-0.37***	-0.37***	-0.38***	-0.53***	-0.34***
high blood cholesterol	-0.15	-0.18*	-0.31***	-0.13	-0.33***	-0.25***	-0.54***	-0.46***	-0.44***	-0.31**	-0.51***
stroke	0.95**	0.61***	0.60***	0.96***	1.18***	1.12***	0.22	0.73***	0.68***	0.69**	0.69***
diabetes	0.54***	0.08	-0.00	0.18	-0.14	0.11	0.07	0.19	-0.04	-0.28	0.27**
chronic lung disease	1.51***	0.51***	0.51**	0.77***	0.64***	0.58***	0.62***	0.51***	0.36*	0.94***	0.61***
asthma	0.41	0.33	0.08	0.37*	-0.35	0.11	0.07	-0.10	-0.13	-0.06	0.31
arthritis	0.49**	0.78***	0.53***	0.94***	0.44***	0.10	0.32***	0.35***	0.16	-0.16	0.30***
osteoporosis	0.78**	0.40	0.18	0.97***	0.17	0.63**	0.01	1.28**	0.08	0.51	0.12
cancer	0.73*	0.19	-0.16	-0.06	0.23	0.74***	0.40**	0.23	0.09	0.16	0.63***
stomach/duodenal ulcer	0.83**	0.26	0.10	-0.05	0.09	-0.29*	0.06	0.09	-0.09	0.44	-0.13
parkinson+)	.	.	.	1.05*	.	0.99**	1.00*	.	1.27**	.	.
cataracts	-0.25	-0.02	0.03	-0.16	0.24	0.17	0.35*	0.10	0.22	-0.01	-0.02
hip fracture	0.18	0.28	0.54**	1.08*	0.61	-0.34	-0.08	0.19	0.43	0.25	0.59*
other	0.31**	0.55***	0.12	0.48***	0.29***	0.24**	0.28***	0.06	0.22*	0.12	0.42***
N	540	1170	1258	1204	985	1339	1242	1166	1380	632	1421
Female											
heart attack	0.48**	0.31**	0.22**	0.34**	0.61***	0.80***	0.61***	0.67***	0.77***	0.33	0.93***
high blood pressure	-0.13	-0.17**	-0.21***	-0.13*	-0.38***	-0.17***	-0.19***	-0.38***	-0.28***	-0.29***	-0.41***
high blood cholesterol	-0.02	-0.31***	-0.31***	-0.11	-0.28***	-0.31***	-0.30***	-0.30***	-0.28***	-0.51***	-0.37***
stroke	0.77*	0.56**	0.47**	0.62**	0.59*	1.21***	0.18	0.94***	0.90***	0.53	0.50*
diabetes	0.78***	0.55***	0.10	0.18	0.29**	0.40***	0.12	0.07	-0.14	-0.10	0.16
chronic lung disease	0.63**	0.39**	1.14***	0.66***	0.38*	0.49***	0.38**	0.53***	0.45**	0.07	0.57***
asthma	0.69**	0.37*	0.14	0.59***	0.11	-0.03	-0.13	0.04	0.09	0.02	0.13
arthritis	0.66***	0.72***	0.42***	0.81***	0.48***	0.21***	0.21***	0.28***	0.20***	0.10	0.53***
osteoporosis	0.22*	0.74***	0.10	0.34***	0.22*	0.26***	-0.04	0.21	-0.15**	0.35*	0.08
cancer	0.74*	0.52***	-0.03	0.31*	0.76**	0.44**	0.19	-0.07	0.13	0.08	0.73***
stomach/duodenal ulcer	0.78**	0.30	0.21	0.41	0.21	-0.09	0.63***	0.18	-0.01	0.23	0.04
parkinson+)	.	.	.	.	0.99	0.82*	0.69	0.93**	1.33**	0.54	1.35***
cataracts	-0.10	-0.02	0.18*	0.22	0.29*	0.45***	0.29**	0.29**	0.30**	-0.10	0.20
hip fracture	1.40***	0.75	0.19	-0.15	0.78***	0.30	0.60**	0.22	0.20	0.66	1.18***
other	0.52***	0.50***	0.36***	0.45***	0.25***	0.19**	0.09	0.01	-0.03	0.21**	0.21**
N	785	1372	1470	1432	1212	1629	1660	1436	1822	806	1730

+) Variable dropped for some countries due to collinearity.

Table 4: Mental condition index

Male	AUT	GER	SWE	NED	ESP	ITA	FRA	DEN	GRE	SUI	BEL
sad or depressed last month	0.06	0.18*	0.03	0.25**	0.19	0.32***	0.10	0.13	0.34***	0.24	0.04
felt would rather be dead	0.61	-0.06	0.54**	0.25	0.66**	0.37**	0.41***	0.73**	0.58*	0.17	0.35**
feels guilty	0.62 **	0.04	-0.05	0.18	-0.39**	-0.05	-0.08	0.08	-0.07	-0.07	-0.06
trouble sleeping	0.66 ***	0.45***	0.39***	0.37***	0.28**	0.31***	0.25***	0.21**	0.26**	0.32**	0.28***
less or same interest in things	0.29	0.39**	0.49***	0.16	0.12	0.14	0.01	0.05	0.32***	-0.08	0.25*
irritability	-0.01	0.15	0.01	-0.06	0.24**	-0.00	-0.14	0.04	-0.09	-0.25*	0.06
no appetite	-0.50	-0.27	-0.61***	-0.85***	-0.32**	-0.32**	-0.46***	-0.42**	-0.28	-0.82***	-0.48***
fatigue	0.78 ***	0.55***	0.58***	0.73***	0.31***	0.62***	0.70***	0.54***	0.30***	0.53***	0.94***
difficulties concentrating											
on entertainment	0.09	-0.15	0.27*	0.33**	0.19	0.12	0.25*	0.28	-0.04	0.39**	0.03
on reading	0.59 **	0.23	0.11	0.11	0.50***	0.35***	0.12	0.41***	0.32**	0.17	0.38***
no enjoyment	-0.04	0.25**	-0.07	0.21	0.12	0.13	0.18	0.28**	0.16	0.27	-0.06
tearfulness	-0.07	0.17	-0.05	0.08	0.07	-0.13	-0.06	0.12	-0.29*	0.22	0.07
N	542	1162	1223	1178	941	1326	1175	1152	1348	629	1413
Female											
sad or depressed last month	0.46***	0.17**	0.11	-0.03	0.16*	0.23***	0.01	0.24***	0.27***	0.07	0.01
felt would rather be dead	0.32	0.33	0.22	0.28	0.16	0.64***	0.23**	0.43**	0.14	0.27	0.39***
feels guilty	-0.01	-0.06	-0.07	-0.07	-0.06	-0.09	-0.14*	-0.06	-0.15	-0.17	-0.08
trouble sleeping	0.48***	0.30***	0.26***	0.39***	0.49***	0.20***	0.28***	0.24***	0.33***	0.26**	0.25***
less or same interest in things	0.23	-0.08	0.32**	0.01	0.21*	0.10	0.08	0.45***	-0.01	0.26	0.07
irritability	-0.13	-0.13	-0.02	0.21*	-0.04	-0.24***	-0.17**	0.02	-0.34***	-0.11	-0.08
no appetite	0.12	-0.39***	-0.36**	-0.32**	-0.30**	-0.02	-0.35***	-0.32**	-0.44***	-0.66***	-0.17
fatigue	0.69***	0.72***	0.63***	0.74***	0.32***	0.67***	0.73***	0.43***	0.37***	0.54***	0.68***
difficulties concentrating											
on entertainment	-0.06	0.01	-0.19	0.44***	0.27**	0.14	0.24**	0.13	0.39***	-0.26	0.13
on reading	0.47**	0.46***	0.42***	0.04	0.16	0.33***	0.17*	0.45***	0.30***	0.56***	0.32***
no enjoyment	0.17	0.09	0.12	-0.00	0.37***	0.10	0.18	0.29*	0.16	0.62***	0.23**
tearfulness	-0.25**	0.19**	0.08	0.06	0.14	0.15*	0.08	0.08	0.11	-0.14	0.07
N	785	1359	1416	1419	1153	1597	1578	1407	1771	799	1704

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

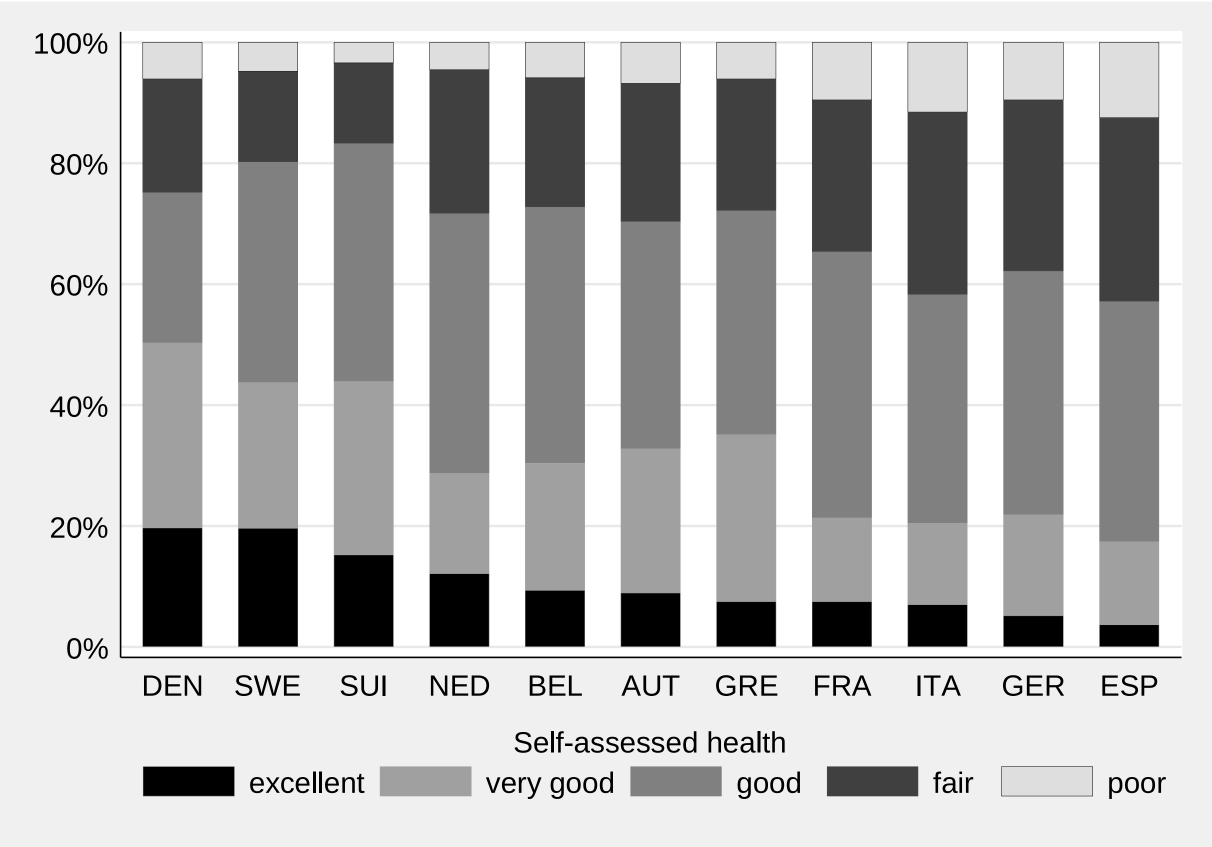
The results are reported separately for males and females and for all countries. As one can see, there is large variation between the countries. For both indices, we find gender differences regarding the magnitude, the sign and the significance of the coefficients. This means that the limiting impact of the different diagnosed illnesses on the individual's activities varies strongly. For males, the magnitude of the heart attack coefficient in the physical index regression ranges from 0.84 in Italy to 0.30 in the Netherlands. The highest impact for stroke is found in Spain (1.18), while for France we find no significance at all. Interestingly, some forms of diseases only show an impact in a few countries, e.g. hip fracture, stomach ulcer or cancer. For women, osteoporosis reveals changing signs. While the influence is highly significant and positive (0.74) for German women, it is negative for Greece (-0.15). A possible interpretation would be that German women tend to report a higher probability of being limited when suffering from osteoporosis. In contrast to this, Greek women perceive themselves as less restricted by this form of disease, perhaps taking limitations resulting from this illness at a certain age as granted and kind of "normal". Considering the mental condition index, a similar pattern is found for men and the attitude "feels guilty". While Austrians are affected negatively (higher probability for limited activities) the picture is reverse for Spain. Further items, like difficulties to concentrate on entertainment, no enjoyment and tearfulness are only partly significant. For women, the only significant coefficient for the variable "feels guilty" is found for France (-0.14). If we consider that Austria and Germany are fairly comparable, the diverging impact of tearfulness on female respondents is somewhat surprising: Estimation results indicate a positive effect for Austrian women whereas German women are negatively affected.

Subsequently, the predicted values for each sub-regression are transformed by using an inverse log transformation resulting in positive values. Last, we compute the final indices by combining the results of the country sub-regressions, i.e. we standardize the results across countries, but separately for gender and year. The final physical and mental index ranges from 0 to 100 with mean 50 and a standard deviation of 10. This can be interpreted as follows: The indices have a mean of 50 if all countries are considered, but the country-specific mean can deviate from this value. A higher index value indicates a higher degree of multimorbidity or poor mental state respectively. A country mean above the overall mean suggests that – at least in this specific year – this country faces relatively higher health problems.

3.3 Cross-country comparison

For the further analysis of reporting heterogeneity across European countries, it is important to take a closer look at the distribution of self-assessed health. To make a cross-country comparison meaningful, the figures have to be adjusted for country specific population characteristics. For this reason, we compute age-gender-standardized distributions of SAH. This standardization gives us the opportunity to compare countries with respect to different reporting patterns, while controlling for differences in the age and gender structure of the population. Figure 1 shows the standardized distribution of SAH across countries pooled for both observations periods.

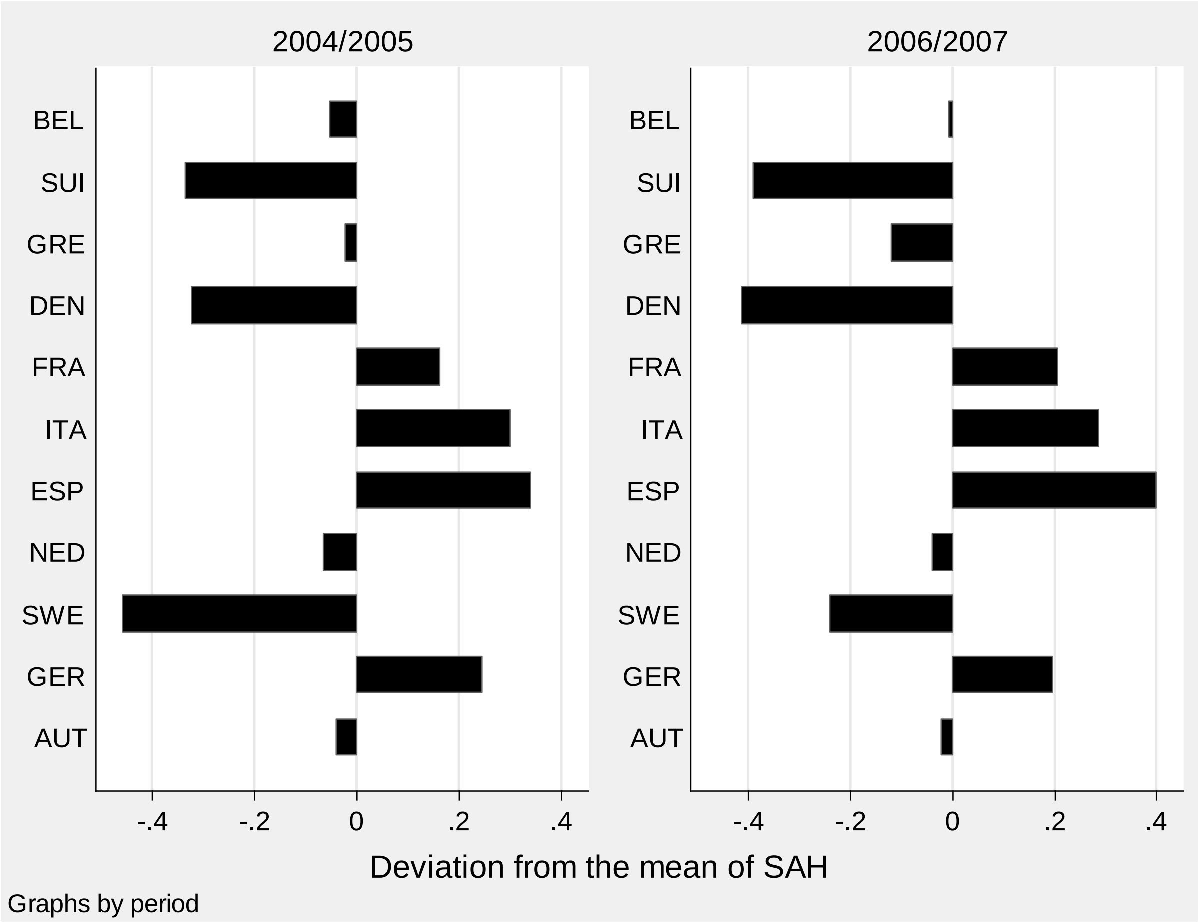
Figure 1: Distribution of self-assessed health by country



Our order condition for the countries is the fraction of individuals with excellent health. Therefore, the healthiest individuals live in Denmark and Sweden. This is in line with the result presented in Jürges (2007). It is obvious that there exists large variation across the countries. While a fraction of 50 % of the Danish population reports very good or better health, the proportion drops under 20 % for Spain. On the contrary, only about 18 % of the Swiss state their health as fair or poor whereas the least healthy population seems to be in Italy and Spain (more than 40 % reporting a health status below good).

In this paper we hypothesize that these reported differences are not only due to differences in true health, but that they are a consequence of variations in the interpretation of the categories. Therefore, in the further empirical analysis, we aim at explaining factors responsible for these differences in the evaluation of self-assessed health across countries. While figure 1 only shows the distribution of self-assessed health categories across European countries, figure 2 represents the deviation from the age-gender standardized mean of SAH.

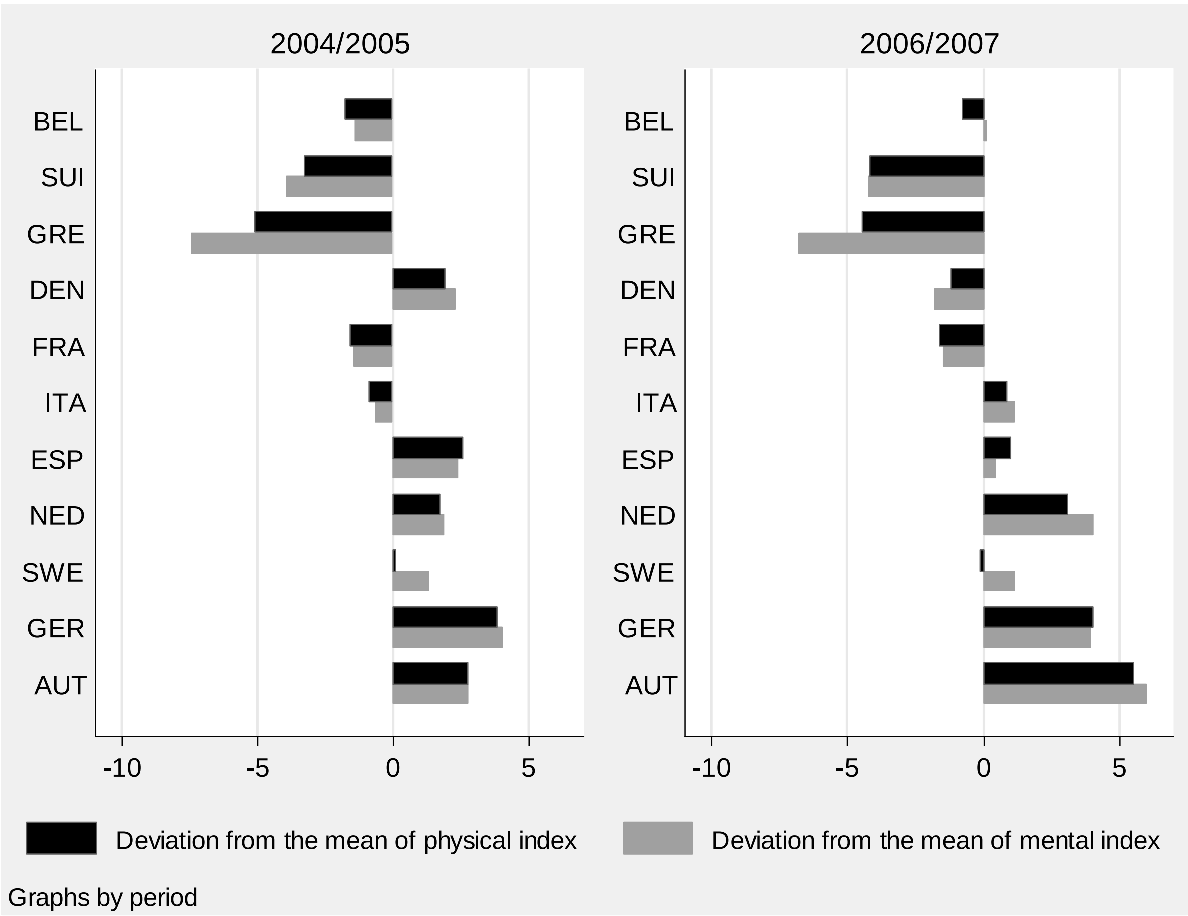
Figure 2: Deviation from the mean of self-assessed health by country



Here, the differences between the countries are distinctly visible. As figure 1 indicates, the countries rating their health lower than the average are France, Germany, Italy and Spain. Interestingly, in the period 2004/2005, Sweden shows the largest negative deviation from the mean. This indicates that Sweden has the healthliest population on average, even healthier than the Danish. The picture changes, however, when the period of 2006/2007 is considered. Here, the magnitude of the deviation of Sweden has come down to a half, a fact not visible through the pooled presentation in figure 1. Between the observation periods, the devations are stable for Belgium, the Netherlands and Austria. Again, this resembles only differences in self-reported health, with no link to the underlying true health.

As discussed before, we construct (quasi-)objective health measures, namely our physical respectively mental condition indices to incorporate a proxy for true health. Hence, in figure 2, the country deviations form the standardized mean of 50 for both indices are presented in Figure 3. Again, a negative variation indicates a better true health status, whereas a positive deviation points to a higher degree of morbidity or inferior mental status. Obviously, there exist large differences compared to the SAH figure. For the period 2004/2005, in Sweden and Denmark, the countries with the best self-assessed health, the picture for the objective health indices is completely different. According to this, we assume that reported health in those countries is overrated compared to the underlying true health. A similar picture results for Austria, while for France and Italy the interpretation is that reported health underrates true health. For the period 2006/2007, the results change slightly. One remarkable alteration is that we observe countries which change from a negative to a positive deviation and vice versa. In detail, both indices deviate in the opposite direction for Denmark and Italy, whereas only the mental respectively the physical index deviates in opposite directions for Belgium and Sweden. Moreover, according to figure 3, true health has dramaticaly declined in Austria and the Netherlands.

Figure 3: Deviation from the mean of mental and physical health index by country



Finally, for most of the countries, we observe a higher variation for the mental condition index. One reason for this may be that for the physical index we use illnesses diagnosed by a physician, whereas for the mental index we use self-reported criteria, which are less strictly defined and as such much more prone to cultural influences. The above findings suggest that the discrepancy between reported and true health has to be analysed in detail. Moreover, the discrepancy can be interpreted as indicator for reporting heterogeneity.

3.4 Estimation approach

When analysing variables with ordered categories such as SAH, ordered response models are often used as estimation procedure. Within this group of models, commonly the distinction between ordered logit or, as in our case ordered probit models is possible, depending on the underlying distribution function. One obstacle to the traditional ordered probit model is the single index or parallel regression assumption (Long, 1997). The coefficient vector is assumed to be the same for all categories of the dependent variable. In detail, this can be interpreted as a shift in the cumulated distribution function through an increase of an independent variable, i.e. the distribution shifts to the right or left, but there is no shift in the slope. By relaxing this assumption and allowing the indices to differ across the outcomes one gets the generalized ordered probit model (Boes, 2007).[9]

In our case, let y be the ordered categorical outcome of SAH, $y \in \{1, 2,\ldots, J\}$. J denotes the number of distinct categories. Underlying the observed variable y is the latent health status of the respondent y^*. While we use panel data, we apply a random effects generalized ordered probit model. For the data at hand, i denotes the cross-sectional unit and t the time dimension:

$$\begin{aligned}
&y_{it}^* = x_{it}'\beta + \varepsilon_{it} \\
&\varepsilon_{it} = u_{it} + \alpha_i \\
&y_{it} = j \Leftrightarrow \tilde{\kappa}_{j-1} + x_{it}'\gamma_{j-1} \le y_{it}^* \le \tilde{\kappa}_j + x_{it}'\gamma_j,\ j = 1,\ldots,5 \\
&E\left[\varepsilon_{it}\right] = 0 \\
&Var\left[\varepsilon_{it}\right] = 1 + \sigma_\alpha^2 \\
&Corr\left[\varepsilon_{it};\varepsilon_{is}\right] = \rho = \frac{\sigma_\alpha^2}{1+\sigma_\alpha^2}
\end{aligned} \tag{1}$$

As it is common for a random-effects specification, the outcome probabilities are conditional on the individual effect α_i. In this context, the βs are the unknown coefficients. While in the

[9] For a general discussion of aspects of heterogeneity in ordered choices and a detailed description of the generalized ordered probit model see Greene and Henscher (2010).

traditional ordered probit model the unknown threshold parameters are constant, the threshold parameters in the generalized model κ_{ij} are individual specific and depend on the covariates:[10]

$$\kappa_{ij} = \tilde{\kappa}_j + x'_{it}\gamma_j , \tag{2}$$

Here, γ_j are the influence parameters of the covariates on the thresholds and $\tilde{\kappa}_j$ represents a constant term. It is important to note that the coefficients of the covariates and the threshold coefficients cannot be identified separately if the same set of variables *x* is used.

$$\begin{aligned} &y_{it} = j \Leftrightarrow \tilde{\kappa}_{j-1} + x'_{it}\gamma_{j-1} \leq y^*_{it} = x'_{it}\beta + \varepsilon_{it} \leq \tilde{\kappa}_j + x'_{it}\gamma_j , \\ &\text{with } j = 1,\ldots,5, \quad t = 1,\ldots,T, \quad i = 1,\ldots,N. \end{aligned} \tag{3}$$

From this, it is clear that $\beta_j = \beta - \gamma_j$.[11] Following Williams (2006), this results in the estimation of *J*-1 binary probit models. First, category 1 versus categories 2,.., *J* is estimated, second, categories 1 and 2 versus 3,.., *J* and so on. For our purpose, this estimation method enables us to control for individual heterogeneity in the β-parameters and hence for heterogeneity across the categories of the dependent variable. Consequently, the advantage of using panel data in combination with a generalization of the ordered probit model is to distinguish between two kinds of heterogeneity. First, unobserved individual heterogeneity is captured by our random effects specification of the ordered probit model. Second, varying cut-points and beta coefficients characterize the observed heterogeneity in the reporting of self-assessed health. As noted before, the β coefficients are individual specific and represent the dependence of the thresholds on the individual response behaviour.

For our analysis, this implies a cut-point shift, when the relative position of these thresholds changes. Instead, if we find a parallel shift in the thresholds, the distribution of SAH shifts completely to the left or the right (index shift). The distinction between both kinds of shifts is of high relevance if the parallel shift cannot be separated from changes in the relative position of the thresholds (Lindeboom and van Doorslaer, 2004). This means that we are not able to separate explanatory variables in those related to reported health and those reflecting true health. To solve this problem, Lindeboom and van Doorslear (2004) suggest distinguishing between socioeconomic explanatory variables and objective health measures. As a consequence, true

---

[10] The order condition in the generalized ordered probit model requires that the predicted probabilities are in the (0; 1) interval.

[11] The standard ordered probit model with the restriction of equal coefficient vector β is nested in the generalized ordered probit model.

health is conditioned by the latter factors and cut-point and index shifts can be identified. In our generalized model, we first test for a cut-point shift related to our mental and physical health index. If the hypothesis of a cut-point shift is rejected, an index shift exists.

The iterative procedure to identify variables that drive the heterogeneity was first proposed by Williams (2006) for cross-section data. In an extension, Pfarr et al. (2010) combine this with the random-effects specification of the generalized ordered probit model by Boes (2007).[12] In detail, we start with a completely unconstrained model (all coefficients varying). For this, several Wald tests are applied on each variable to test for differing coefficients across equations and to constrain the least significant variables. They are further set to have equal effects, and the model is refitted with the identified constraints. This procedure is repeated until only significant variables remain.

4 Results

Table 5 presents the results of the estimation of a generalized ordered probit model for panel data. In the table, we display the results of the four underlying binary models. The first model estimates category 1 (excellent) versus categories 2,.., 5, the second model categories 1 and 2 (excellent and very good) versus 3,.., 5 and so on. The interpretation of a negative coefficient for the model 1-2 versus 3-5 is as follows: the negative value indicates a higher probability to report categories 1 or 2, while a positive coefficient indicates a higher probability of reporting the worse health status.

According to our iterative procedure, which identifies variables violating the parallel lines assumption, we end up with 11 variables to be constrained in the estimation. This means that these variables are assumed to have equal effects across the categories of self-assessed health and hence across the four binary models. In detail, the parallel lines assumption holds for *Gender*, *Marital status*, *Children*, all variables of relative income, *Drinking* and the three variables covering hospital nights. First, females tend to report a worse health status then males. This reflects that health care utilization, health behaviour and the prevalence of various forms of diseases are gender-specific and lead to differences in life-expectancy. Interestingly, as we control for the effects of utilization, behaviour and diseases, we still get the highly significant impact of gender.

[12] The related user-written Stata program regoprob2 is available at the SSC archive.

Table 5: Estimation results of the generalized ordered probit model

SAH	**1 vs. 2-5**		**1-2 vs. 3-5**		**1-3 vs. 4-5**		**1-4 vs. 5**	
	Coeff.	*p* value	Coeff.	*p* value	Coeff.	*p* value	Coeff.	*p* value
Survey Period	0.034	(0.087)	0.068	(0.000)	0.152	(0.000)	0.129	(0.000)
Gender	0.061	(0.000)	0.061	(0.000)	0.061	(0.000)	0.061	(0.000)
Age	0.078	(0.000)	0.082	(0.000)	0.059	(0.000)	0.015	(0.298)
Age^2	-0.046	(0.000)	-0.048	(0.000)	-0.035	(0.000)	-0.006	(0.571)
Marital status	0.051	(0.003)	0.051	(0.003)	0.051	(0.003)	0.051	(0.003)
Foreign	-0.033	(0.652)	0.128	(0.031)	0.162	(0.008)	0.304	(0.000)
Grandchildren	0.026	(0.306)	0.029	(0.160)	0.087	(0.000)	-0.039	(0.248)
Children	-0.015	(0.512)	-0.015	(0.512)	-0.015	(0.512)	-0.015	(0.512)
Very low income	0.095	(0.000)	0.095	(0.000)	0.095	(0.000)	0.095	(0.000)
Low income	0.083	(0.000)	0.083	(0.000)	0.083	(0.000)	0.083	(0.000)
High income	-0.048	(0.082)	-0.048	(0.082)	-0.048	(0.082)	-0.048	(0.082)
Very high income	-0.124	(0.000)	-0.124	(0.000)	-0.124	(0.000)	-0.124	(0.000)
Education	-0.113	(0.000)	-0.146	(0.000)	-0.144	(0.000)	-0.095	(0.000)
Smoking	0.045	(0.047)	0.085	(0.000)	0.079	(0.000)	0.159	(0.000)
Drinking	-0.114	(0.000)	-0.114	(0.000)	-0.114	(0.000)	-0.114	(0.000)
Physical activity	-0.309	(0.000)	-0.356	(0.000)	-0.449	(0.000)	-0.549	(0.000)
Physical health index	0.016	(0.000)	0.023	(0.000)	0.034	(0.000)	0.032	(0.000)
Mental health index	0.033	(0.000)	0.042	(0.000)	0.050	(0.000)	0.052	(0.000)
Doctor visits 1-3	0.367	(0.000)	0.284	(0.000)	0.181	(0.000)	-0.068	(0.265)
Doctor visits 4-11	0.832	(0.000)	0.782	(0.000)	0.722	(0.000)	0.388	(0.000)
Doctor visits >11	1.043	(0.000)	1.107	(0.000)	1.173	(0.000)	0.809	(0.000)
Hospital nights 1-6	0.190	(0.000)	0.190	(0.000)	0.190	(0.000)	0.190	(0.000)
Hospital nights 7-14	0.322	(0.000)	0.322	(0.000)	0.322	(0.000)	0.322	(0.000)
Hospital nights >14	0.576	(0.000)	0.576	(0.000)	0.576	(0.000)	0.576	(0.000)

Table 5 continued

Austria	-0.280	(0.000)	-0.724	(0.000)	-0.559	(0.000)	-0.805	(0.000)
Germany	0.210	(0.000)	-0.130	(0.002)	-0.112	(0.006)	-0.426	(0.000)
Sweden	-0.721	(0.000)	-0.861	(0.000)	-0.541	(0.000)	-0.314	(0.000)
Netherlands	-0.356	(0.000)	-0.327	(0.000)	-0.304	(0.000)	-0.720	(0.000)
Spain	0.353	(0.000)	0.019	(0.682)	0.073	(0.078)	-0.080	(0.155)
Italy	-0.064	(0.231)	-0.057	(0.168)	0.199	(0.000)	-0.031	(0.563)
Denmark	-0.644	(0.000)	-1.019	(0.000)	-0.239	(0.000)	-0.233	(0.000)
Greece	0.359	(0.000)	-0.241	(0.000)	0.237	(0.000)	0.112	(0.069)
Switzerland	-0.227	(0.000)	-0.523	(0.000)	-0.343	(0.000)	-0.311	(0.001)
Belgium	-0.173	(0.000)	-0.414	(0.000)	-0.387	(0.000)	-0.575	(0.000)
_cons	-3.575	(0.000)	-5.402	(0.000)	-7.379	(0.000)	-7.126	(0.000)
ρ	0.413	(0.000)						
N	53931							

p-values in parentheses
Acknowledgement: For those variables highlighted the parallel lines assumption holds.

Regarding the income effects, compared to the reference category (income > 75 % but ≤ 125 % of the country's median equivalent net household income) households with an income lower than 75 % of the median tend to report a poorer health status. For households with a higher income (more than 125 % of median), we find a significantly negative impact. As a negative coefficient directs to report a lower category of SAH, the interpretation is that ceteris paribus high income households tend to report a better health status. Taking the income-health nexus into account, this result is not surprising. A remarkable finding is the effect of drinking behaviour. While our variable encompasses moderate as well as frequent consumption of alcoholic beverages, this may reflect a possible "red wine effect" meaning that moderate (wine) drinking is related to a lower prevalence of suboptimal health (Grønbaek et al., 1999, p. 723). The last group of constrained variables contains hospital nights. Here, we obtain a positive, highly significant gradient with the typical result that more nights in hospital indicate a higher degree of morbidity.

Variables for which the parallel lines assumption is not imposed drive the observed heterogeneity in self-assessed health. The effects of these variables are allowed to vary across the four binary regressions, meaning that the coefficients may differ with respect to magnitude, sign and level of significance. Within the group of socioeconomic variables *Education*, *Smoking* and *Physical activity* show varying influence on the distinct categories of SAH. For the first variable – *Education* – the effect is significantly negative across all equations. The magnitude of the corresponding coefficients differs only slightly. Higher school degrees as well as degrees of vocational training thus lead to a better self-reported health status. The signs of the other two factors – *Smoking* and *Physical activity* – are as expected. We find positive coefficients for (current or past) smokers and negative ones for activities (being engaged in vigorous physical activity like sports or heavy housework at least once a week). The magnitude for both variables increases in absolute terms and is highest for equation 1-4 vs. 5. Hence, poor health is reported more often by smokers, but less often for individuals doing sports or heavy housework. Related to the age structure of the SHARE dataset, the effect of smoking shows the long-lasting impact of adverse health behaviour.

Health care utilization of outpatient care shows large and significant effects. While 1-3 doctor visits in the last 12 months are only significant for the first three equations, more than 4 visits are significant for all regressions. Comparing 1-3 with 4-11 visits, the coefficients of the latter factor are more than twice as high. In addition, the effect is stronger for individuals visiting a doctor more than once a month on average. Using a sample of elderly Europeans, these effects are not surprising and correspond to an increasing morbidity at higher age.[13] Moreover, as the access to medical services is highly regulated for a large share of the population in each country and in addition, regulatory interventions vary across the countries in the sample, doctor visits may represent system effects as well.

Both health condition indices are highly significantly positive over all equations and thus point into the same direction. Physical health encompasses various diagnosed diseases and hence covers aspects of multimorbidity. It is obvious that the coefficients for the mental condition index are always higher than the ones for the physical condition index. We may conclude that individuals suffering from mental disorders report to be more limited with respect to their health than individuals with diagnosed physical diseases. Thus, reporting heterogeneity is driven by physical and mental conditions, but the mental effect dominates. Taking into account the literature on cut-point and index shifts, we are able to interpret the results as follows: Both indices enable us to incorporate proxies of true health and, in consequence, to test for the presence of a cut-point shift or an index shift. As both proxies are varying across the categories, we are able to rule out the possibility of a parallel shift in the thresholds (index shift). Hence, comparing answers on self-assessed health with illness related as well as mental health related questions gives evidence for the hypothesis that heterogeneity is driven by objective health measures.

Based on the SHARE dataset, one aim of this paper is to explore cross-country differences across Europe. Therefore, we include 10 country dummy variables with France as reference category. This enables us to control for cultural characteristics as well as to take peculiarities of the health care systems and the respective health policies into account. As it is clear from the descriptive statistics, we observe large variations between countries and survey periods for SAH. Those countries with the healthiest population (Denmark, Sweden and Switzerland) show a distinct pattern, namely negative and highly significant coefficients for all four equations compared to France. Individuals in those countries are more likely to report a better health status. The influence is highest when deciding between health categories excellent and very good on the one hand, versus good to poor on the other hand. Taking into account Figure 1, this resembles the fact that over 40 % of the people in these countries state to be in the two best health categories. Opposite to these findings, we obtain alternating signs of the coefficients for some countries. Germany for instance shows a positive coefficient for the first equation, whereas the

[13] In our sample, over 23 % of those aged above 65 years have more than 11 visits while this applies to only 12 % for those 65 or younger.

other three have a negative sign. In relation to France, Germany tends to report excellent status less often, while the remaining coefficients show a trend towards reporting the middle category. This comes along with the highest negative impact for the last equation, meaning that Germans state poor health less likely than the French. To conclude, in relation to the reference country, Germans neither report excellent nor poor health status very likely. The findings for Greece are somewhat different, because a positive coefficient for the first equation is followed by a negative for the second, while the last two are positive again. This would imply that Greeks prefer to state very good instead of excellent health, but are less likely to classify themselves into the middle category. We obtain only partly significant results for the two southern countries Italy and Spain compared to the reporting of France. All remaining significant coefficients point to reporting poorer health more often. Not surprisingly, together with Germany, both countries are at the bottom end of the ranking.

To sum up our findings, the observed heterogeneity is typical for when analysing health reporting. We find evidence that health behaviour as well as health care utilization, true health and country characteristics affect reporting behaviour. Hence, the application of models taking observed heterogeneity into account is crucial for the analysis of self-reported health. In addition, unobserved heterogeneity may be present in the data. Using panel specification enables us to test for these unobserved factors. The influence of unobserved heterogeneity is confirmed by the high significance of the correlation of the error terms ρ. Hence, we conclude that among elderly Europeans self-assessed health is influenced by observed as well as unobserved heterogeneity.

5 Summary

Knowledge about the health status of individuals is paramount when health interventions are to be evaluated. Often self-assessed health (SAH) is used as a key measure to this end. However, SAH is prone to inaccuracies due to reporting heterogeneity. These may result in differences of the stated health across population subgroups, even if the underlying true health status is identical. As the elderly typically face the highest level of morbidity and have usually a long history of dealing with their health issues, reporting heterogeneity is a very likely problem in this group. Moreover, it seems of high interest to see how the institutional and cultural setting can influence the divergence of true and self-assessed health. To account for such differences we conduct a comparison across different European countries.

We use the Survey of Health, Ageing and Retirement in Europe (SHARE) for a panel analysis. For eleven countries, we estimate a generalized ordered probit model to identify potential cut point shifts in the health distribution. To account for the true health, in a first stage we estimate indices for mental and physical condition. In a second stage, we include these together with measures for health care utilization, socio-demographic variables and country fixed effects to evaluate their relevance for reporting heterogeneity. While this observed heterogeneity is reflected in the cut-point shifts, we are able to account for unobserved heterogeneity by using a random effects specification.

The results of the generalized ordered probit model indicate that cut-point shifts are present in the reporting of self-assessed health across countries. Significant differences exist. For example, in countries like Germany and Spain, individuals systematically report a lower health status, whereas Dutch respondents show a higher probability to opt for the best category. For both health indices, we find evidence of reporting heterogeneity. This means that a worse objective health status not only leads to a lower perception of own health but also that the impact of the effect varies between the categories of SAH. Moreover, the magnitude of mental health problems exceeds the effect of the pure physical health index. We find further evidence for reporting heterogeneity when looking at aspects like health care utilization and health relevant behaviour.

Simple random effects ordered probit analysis neglects the fact that the classification in the five categories depends on health related variables and on the country of residence. Heterogeneity may be caused by factors like illness experience and demand for health care among elder Europeans. Thus, we observe a gap between true and reported health. Quantifying this gap requires reliable measures of true health. To this end, both aspects, i.e. physical and mental health, have to be taken into account separately. Country effects may reflect aspects like differences in health systems as well as unaccounted cultural variation. For future research, a separation of the health system effect from cultural effects is required. This would allow deriving policy implications focusing on differences in health care systems from an international perspective.

References

Andreyeva, T. 2006. *An International Comparison of Obesity in Older Adults: Effects and Risk Factors.* Santa Monica, CA: RAND Corporation.

Bago d' Uva, T., Van Doorslaer, E., Lindeboom, M., O'Donnell, O. 2008. Does Reporting Heterogeneity Bias the Measurement of Health Disparities? *Health Economics,* **17:** 351-375.

Bago d'Uva, T., Lindeboom, M., O'Donnell, O., Van Doorslaer, E. 2010. Education-related

Inequity in Health Care with Heterogeneous Reporting of Health. Tinbergen Institute Discussion Paper TI2010-122/3, Amsterdam.

Börsch-Supan, A., Jürges, H. 2005. *The Survey of Health, Ageing, and Retirement in Europe - Methodology.* Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).

Boes, S. 2007. *Three Essays on the Econometric Analysis of Discrete Dependent Variables.* University of Zurich: Dissertation.

Bound, J. 1991. Self-Reported versus Objective Measures of Health in Retirement Models. *Journal of Human Resources,* **26:** 106-138.

Butler, J. S., Burkhauser, R. V., Mitchell, J. M., Pincus, T. 1987. Measurement Error in Self-Reported Health Variables. *Review of Economic and Statistics,* **69:** 644-650.

Christelis, D. 2010. Imputations. In Mannheim Research Institute for the Economics of Ageing (Ed.), *Survey of Health, Ageing and Retirement in Europe: Release Guide 2.3.1, Waves 1 & 2* (pp. 27-32). Mannheim: Mannheim Research Institute for the Economics of Ageing.

Coe, N. B., Zamarro, G. 2011. Retirement Effects on Health in Europe. *Journal of Health Economics,* **30:** 77-86.

Disney, R., Emmerson, C., Wakefield, M. 2006. Ill Health and Retirement in Britain: A Panel Data-based Analysis. *Journal of Health Economics,* **25:** 621-649.

Etilé, F., Milcent, C. 2006. Income-Related Reporting Heterogeneity in Self-Assessed Health: Evidence from France. *Health Economics,* **15:** 965-981.

Greene, W. H., Hensher, D. A. 2010. *Modeling Ordered Choices.* Cambridge: Cambridge University Press.

Grønbaek, M., Mortensen, E. L., Mygind, K., Andersen, A. T., Becker, U., Gluud, C., Sørensen, T. I. A. 1999. *Beer, Wine, Spirits and Subjective Health.* .

Hernández-Quevedo, C., Jones, A. M., Rice, N. 2005. Reporting Bias and Heterogeneity in Self-Assessed Health: Evidence from the British Household Panel Survey. HEDG Working Paper 05/04. University of York.

Jürges, H. 2007. True Health vs Response Styles: Exploring Cross-Country Differences in Self-Reported Health. *Health Economics,* **16:** 163-178.

Kerkhofs, M., Lindeboom, M. 1995. Subjective Health Measures and State Dependent Reporting Errors. *Health Economics,* **4:** 221-235.

King, G., Murray, C. J. L., Salomon, J., Tandon, A. 2004. Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research. *American Political Science Review,* **98:** 184-191.

Lindeboom, M., Kerkhofs, M. 2009. Health and Work of the Elderly: Subjective Health Measures, Reporting Errors and the Endogenous Relationship Between Health and Work. *Journal of Applied Econometrics,* **24:** 1024-1046.

Lindeboom, M., van Doorslaer, E. 2004. Cut-Point Shift and Index Shift in Self-Reported Health. *Journal of Health Economics,* **23:** 1083-1099.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks et al: Sage.

Pfarr, C., Schmid, A., Schneider, U. 2010. Estimating Ordered Categorical Variables Using Panel Data: a Generalized Ordered Probit Model with an Autofit Procedure. Discussion Papers in Economics 02/10. Department of Law and Economcis. University of Bayreuth.

Schneider, U., Pfarr, C., Schneider, B. S., Ulrich, V. 2011. I feel good! Gender Differences and Reporting Heterogeneity in Self-assessed Health. *European Journal of Health Economics,* **DOI:** 10.1007/s10198-011-0301-7.

StataCorp. 2009. *Stata: Release 11. Statistical Software - Multiple Imputation.* College Station: StataCorp LP.

Statistisches Bundesamt; Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen; Wissenschaftszentrum Berlin für Sozialforschung 2008. *Datenreport 2008 - ein Sozialbericht für die Bundesrepublik Deutschland.* Berlin: Bundeszentrale für politische Bildung.

van Doorslaer, E., Jones, A. M. 2003. Inequalities in Self-Reported Health: Validation of a New Approach to Measurement. *Journal of Health Economics*, **22:** 61-87.

Williams, R. 2006. Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *The Stata Journal,* **6:** 58-62.

Ziebarth, N. R. 2010. Measurement of Health, the Sensitivity of the Concentration Index, and Reporting Heterogeneity. *Social Science and Medicine,* **71:** 116-124.

Universität Bayreuth
Rechts- und Wirtschaftswissenschaftliche Fakultät
Wirtschaftswissenschaftliche Diskussionspapiere

Zuletzt erschienene Papiere:*

01-11	Pfarr, Christian Schneider, Udo	Choosing between subsidized or unsubsidized private pension schemes: a random parameters bivariate probit analysis
09-10	Kunz, Reinhard Woratschek, Herbert Strelow, Marcel	Future Opportunities and Threats in a Fragmented Sport Media World - Trends and Strategic Implications for Sport Media Enterprises
08-10	Drescher, Christian Herz, Bernhard	Measuring Monetary Conditions in US Asset Markets. A Market Specific Approach
07-10	Woratschek, Herbert Popp, Bastian	Branded Communities as an Alternative Branding Concept to Brand communities: The Case of a German Football Community
06-10	Woratschek, Herbert Horbel, Chris Popp, Bastian	Extending the Service-Profit Chain: The Role of Identification and Co-Creation
05-10	Pfarr, Christian Schneider, Brit S. Schneider, Udo Ulrich, Volker	I feel good! Gender differences and reporting heterogeneity in self-assessed health
04-10	Siebert, Johannes	Aggregate Utility Factor Model: A Concept for Modeling Pair-wise Dependent Attributes in Multiattribute Utility Theory
03-10	Erler, Alexander Drescher, Christian Krizanac, Damir	The Fed's TRAP - A Taylor-type Rule with Asset Prices
02-10	Pfarr, Christian Schmid, Andreas Schneider, Udo	Estimating ordered categorical variables using panel data: a generalized ordered probit model with an autofit procedure
01-10	Herz, Bernhard Wagner, Marco	Multilateralism versus Regionalism!?
05-09	Erler, Alexander Krizanac, Damir	Taylor-Regel und Subprime-Krise – Eine empirische Analyse der US-amerikanischen Geldpolitik
04-09	Woratschek, Herbert Schafmeister, Guido Schymetzki, Florian	International Ranking of Sport Management Journals
03-09	Schneider Udo Zerth, Jürgen	Should I stay or should I go? On the relation between primary and secondary prevention
02-09	Pfarr, Christian Schneider Udo	Angebotsinduzierung und Mitnahmeeffekt im Rahmen der Riester-Rente. Eine empirische Analyse.

* Weitere Diskussionspapiere finden Sie unter
http://www.fiwi.uni-bayreuth.de/de/research/Working_Paper_Series/index.html