

**Self-Rated Health Status of the Japanese and Europeans in Later Life: Evidence
from JSTAR and SHARE**

By

Mayu Fujii[†], Takashi Oshio^{††}, and Satoshi Shimizutani^{†††}

October 2012

Abstract

Using panel data from two surveys in Japan and Europe, we examine the comparability of the self-rated health of the middle-aged and elderly across Japan and the European countries and the survey periods. We find that a person's own health is evaluated on different standards (thresholds) across the different countries and survey waves. When evaluated on common thresholds, the Japanese elderly are found to be healthier than their counterparts in the European countries. At the individual level, reporting biases leading to discrepancies between the changes in individuals' SRH and their actual health over the survey waves are associated with age, education, and country of residence.

Key words: self-rated health status; response bias; JSTAR; SHARE.

JEL classification codes: C42, I12.

[†] Research Officer, National Institute of Population and Social Security Research; Address: 6th Floor, Hibiya Kokusai Building, 2-2-3, Uchisaiwai-cho, Chiyoda-ku, Tokyo, 100-0011, JAPAN; Tel: 03-3595-2989 (Ex 5005); Fax: 03-3502-0635; E-mail: fujii-mayu@ipss.go.jp.

^{††} Professor, Institute for Economics Research, Hitotsubashi University; Address: 2-1 Naka Kunitachi-shi, Tokyo, 186-8603, JAPAN; Tel: 042-580-8372; Fax: 042-580-8333; E-mail: oshio@ier.hit-u.ac.jp.

^{†††} Research Fellow, Gender Equality Bureau, Cabinet Office and Consulting Fellow, RIETI; Address: 1-6-1 Nagata-cho Chiyoda-ku, 100-8914; Tel: 03-5253-2111 (Ex 83765); Fax: 03-3581-0408; E-mail: satoshi.shimizutani@cao.go.jp.

1. Introduction

Given the diverse development of health care systems and policies around the world, investigating how population health has varied across different countries over time would help us to better understand factors that affect health and health inequalities. Making such an investigation, however, requires health measures that are comparable across different countries and time points. This study focuses on self-reported health (SRH) and, using two-wave panel data from the Survey on Health, Aging and Retirement in Europe (SHARE) and the Japanese Study on Aging and Retirement (JSTAR) survey, we examine whether SRH can be used to compare the healthiness of the middle-aged and elderly in Japan with that in other countries across different survey waves.

SRH is a subjective health measure collected as an answer to the question “Would you say your health is excellent, very good, good, fair, or poor?”¹ It is one of the most widely used measures of an individual’s general health status in empirical studies (e.g., Sickles and Taubman, 1986; Stern, 1989; Deaton and Paxson, 1998; Smith, 1999; Adams et al., 2003; Contoyannis et al., 2004; Frijters et al., 2005) for several reasons. First, it is a simple single-index measure summarizing a variety of aspects of the general health condition of an individual. Second, information on SRH can be obtained easily and with less cost compared to other health

¹ There are other expressions for self-reported health, such as “self-rated health status,” “self-assessed health status,” “self-perceived health,” and “subjective health status.”

measures such as a biomarker. Furthermore, SRH has been shown to be a good predictor of subsequent mortality (Mossey and Shapiro, 1982; Idler and Benyamini, 1997; Ford et al., 2008) and the use of medical care (van Doorslaer et al., 2000).

While SRH possesses these favorable characteristics, a number of studies have questioned its comparability across different populations. In particular, it has been argued that individuals with different demographic/socioeconomic characteristics may use different standards (thresholds) when classifying their health into the SRH response categories (Sadana et al., 2002; Lindeboom and van Doorslaer, 2004). In other words, reporting biases may exist due to differences in threshold levels used to evaluate one's own health across different populations. For example, individuals living in a certain country may be more likely to avoid extreme answers like "excellent" or "poor" than those in other countries due to cultural differences. It may also be possible that the individuals' tendency to rate themselves as having better (or worse) health has changed over time within a country due to changes in social norms or social conditions.^{2,3}

In an attempt to examine the international comparability of SRH, Jürges (2007) constructs a health index adjusted for differences in threshold levels of SRH across ten European countries

² Even if the threshold levels are the same within and across the populations over time, SRH may be affected by measurement errors. SRH ratings may also depend on the mode of administration (written or verbal), the setting where the ratings were elicited (during a household interview or as part of medical examination), and the sequence of health questions (Crossley and Kennedy, 2002; Zajacova and Dowd, 2011).

³ Besides the country of residence and time points, threshold levels of SRH are shown to depend on sex (Lindeboom and van Doorslaer, 2004; Kapteyn et al., 2007), age (Groot, 2000; Van Doorslaer and Gerdtham, 2003; Lindeboom and van Doorslaer, 2004; Kapteyn et al., 2007), education (Hernandez-Quevedo, et al., 2005; Kapteyn et al., 2007; Pfarr, et al., 2011), and income (Hernandez-Quevedo, et al., 2005). These results, however, are not necessary consistent across different studies.

in the SHARE and explores a discrepancy in the ranking of the countries on population health indicated by the adjusted health index and SRH. The adjusted health index is calculated from the variations in SRH that can be explained by variations in objective health measures.⁴

This study closely follows the empirical method developed by Jürges (2007). However, we extend his work in three ways. First, we perform a comparative study between Japan and ten European countries in the SHARE and examine the health status of the Japanese middle-aged and elderly in an international setting. Second, in contrast to Jürges who concentrates on the first wave of SHARE, we employ the first and second waves of the JSTAR and SHARE surveys in order to examine the comparability of SRH not only across the countries but also across the survey waves. Finally, as an additional analysis, we make an investigation at an individual level on what kinds of demographic /socioeconomic characteristics are associated with reporting biases that lead to discrepancies between changes in actual health and changes in SRH between the survey waves.

Our finding is that a person's own health is evaluated on different thresholds in the different countries and survey waves. When common thresholds are used, the rank of Japan on health becomes higher but the change in health within the country across the survey waves becomes worse. At an individual level, age, education, and the country of residence are associated with

⁴ Other methods of correcting self-reported health are to use vignettes (Kapteyn et al., 2007) or to use the first principal component of many objective health indicators (Poterba et al. 2010; Noguchi, 2011).

discrepancies between the change in SRH and actual health over time.

The remainder of this paper is organized as follows. Section 2 briefly describes the JSTAR and SHARE and variables used in this study. In Section 3, we construct a health index adjusted for differences in threshold levels of SRH due to the country of residence and a survey wave. Section 4 compares the adjusted health index with SRH. In Section 5, we investigate what kinds of demographic /socioeconomic characteristics of individuals are associated with reporting biases that lead to discrepancies between changes in the adjusted health index and changes in SRH between the survey waves. Section 6 concludes the paper.

2. Data description

This study uses two longitudinal data sets on the middle-aged and older people in Japan and Europe, namely, the JSTAR and SHARE survey reports. The JSTAR is the Japanese counterpart of the HRS/ELSA/SHARE, the world-standard longitudinal survey on the middle-aged and the older populations (Ichimura et al., 2009).⁵ The first wave was completed in five municipalities in 2007, and the second wave was conducted in seven municipalities (the five original municipalities plus two new ones) in 2009. The baseline sample consisted of individuals aged 50 to 75, who were randomly chosen within a municipality from its household registration. The

⁵ The Health and Retirement Survey in the US and the English Longitudinal Study of Aging survey in the UK are for the middle-aged and the elderly.

sample size in the first wave was about 4,200 at the baseline, and the response rate was 60% on average. The respondents in the first wave were interviewed again in 2009, and their response rate in the second wave was about 80% (the attrition rate was 20%), with some variations across municipalities.

On the other hand, the SHARE survey completed its first wave in 2004 and the second wave in 2006, and covered ten European countries (Sweden, Denmark, Germany, the Netherlands, France, Switzerland, Austria, Italy, Spain, and Greece). The baseline sample comprised individuals aged 50 and over, and the sample size was about 22,000, with an individual response rate of 48% on average (Börsch-Supan, et al., 2005; Börsch-Supan and Jürges, 2005).

The advantage of using the JSTAR and SHARE surveys for the current study is that both have a wide variety of health measures in common. In particular, both data sets have SRH in the North American version: a five-point scale health measure ranging from “excellent” to “poor”.⁶ The other health measures included in both the JSTAR and SHARE and used in the current study are fifteen different physical conditions self-reported as diagnosed by physicians (heart attack or other heart problems, high blood pressure, high blood cholesterol, stroke or cerebral vascular disease, diabetes, chronic lung disease, asthma, arthritis or rheumatism, osteoporosis, cancer or malignant tumor, stomach, duodenal or peptic ulcer, Parkinson’s disease, cataracts, hip

⁶ In both the JSTAR and the SHARE, there is also the European version of SRH in a five-point scale ranging from “very good” to “very poor”. In this study, we use the North American version. Jürges (2007) used the North American version as well and argued that the empirical results were not affected even when the European version was used.

or femoral fracture, other conditions), whether one was ever treated for depression, whether one's BMI (= weight in kg/squared height in meters) is within the non-normal range, and whether one's grip strength, normalized for height and sex, is in the bottom tertile.⁷ While these other health measures, except for grip strength, are self-reported, they give an objective description of the respondent's health. Hence, we regard them as objective health measures.

We restrict our sample to respondents who were interviewed both in the first and second waves of either the JSTAR or the SHARE surveys and who give information on SRH and the other health measures listed above. We further limit the sample to those who were aged 50–75 in the first wave, so as to make the samples in both data sets comparable; i.e., we exclude from the sample those aged 76 or over in the first wave of SHARE.

Table 1 reports the country-wise proportion of individuals in our survey by sex and age group. The table shows that the sample size varies by country, with Japan having the largest sample size of about 3,000; Greece, Sweden, France, Italy, and the Netherlands having about 1,500; and Spain, Austria, and Denmark having about 1,000 respondents. Switzerland has the smallest sample size of 500. In addition to the sample size, there are some cross-country variations in the age composition. In particular, Japan has the highest percentage of respondents aged over 70 (18.1%) out of eleven countries, while the percentage of those aged 55 or less is

⁷ The non-normal range of BMI is less than 20 (underweight), 25 to 30 (overweight), and more than 30 (obese). We also control for the dummy variable, indicating that the respondent did not complete the grip strength measurement. Besides these measures, Jürges (2007) used a measure of walking speed. However, this measure is not available in JSTAR.

the lowest (17.6%). In contrast to age composition, the differences in sex composition across the countries of the sample are small, although the share of men in Spain and Austria is the smallest, being less than 45%.

Figures 1(a) and 1(b) show the distribution of the SRH across eleven countries in the first and second waves, respectively. The distribution is standardized by age and sex to adjust the cross-country differences in the age-sex composition described in the previous paragraph.⁸

Countries are ordered from the healthiest to the unhealthiest in accordance with the proportion of respondents whose SRH is excellent or very good. While it is generally observed that the respondents from Denmark, Japan, Sweden, and Switzerland are relatively healthy and those in Spain, Italy, Germany, and France are relatively unhealthy, the order of the eleven countries varies across the two waves.

The findings shown in Figure 1 may raise a question on the comparability of SRH across the different countries and time periods as a health measure. As discussed by Jürges (2007), the order of the eleven countries according to SRH is not necessarily consistent with the order per other health measures. For instance, the average life expectancy of persons at birth in Spain,

⁸ The age-sex standardized distribution is the distribution calculated as if every country's sample had the same age-sex composition. For example, the age-sex standardized distribution of the SRH in country A is calculated in the following step. First, we divide the sample into 10 age-sex groups: 5 age groups (age 55 or less, 56 to 60, 61 to 65, 66 to 70, and more than 70) for each sex. Second, we calculate the percentage of individuals reported with poor, fair, good, very good, and excellent health for each of the 10 age-sex groups in country A. Third, for each of the age-sex group in country A, we multiply the calculated percentage of individuals in each of the five SRH categories by the ratio of individuals in that age-sex group in the total sample (i.e., all the eleven countries combined). Finally, we sum these numbers among the whole 10 age-sex groups.

Italy, and France is longer (81, 81.5, and 80.9 years, respectively, in 2007) than that in Denmark (78.4 years in 2007), according to OECD.StatExtracts. Furthermore, the order of the countries in terms of life expectancy of persons at birth does not change between 2007 and 2009. These findings indicate the possibility of reporting biases due to the country of residence and survey waves; that is, individuals in a certain country or wave may have different levels of thresholds in classifying their true health in accordance with SRH response categories.

In what follows, we construct a health index adjusted for differences in threshold levels of SRH by the country of residence and survey waves using nineteen objective health measures available in both the JSTAR and SHARE surveys.

3. Health index adjusted for differences in threshold levels of SRH

To obtain a health index adjusted for potential differences in threshold levels of SRH across countries and survey waves, we use data pooled across the countries and the survey waves of the JSTAR and SHARE and implement the empirical strategy of Jürges (2007). First, we reverse the order of from 1 = excellent to 5 = poor to the order of from 1 = poor to 5 = excellent. Next, we conduct a generalized ordered probit regression of SRH on objective health measures, allowing for the threshold parameters to be dependent on covariates.

The underlying assumption of this method is that there is a continuous latent variable H^*

representing true health, and individuals use certain thresholds c^j ($j = 0, \dots, 5$) to classify their true health into one of the five categories of SRH, represented as $H \in \{1, 2, \dots, 5\}$:

$$H_{it}^* = X_{it}\beta + \varepsilon_{it} \quad (1)$$

$$H_{it} = j \Leftrightarrow c^{j-1} < H_{it}^* \leq c^j \quad (j = 1, 2, \dots, 5) \quad (2)$$

where X_{it} represents a vector of the objective health measures of individual i at survey wave t , ε_{it} is an error term, and $c^0 = -\infty$ and $c^5 = \infty$. In contrast to a simple ordered probit regression, the thresholds are modeled here as a function of the country of residence, the survey wave, and their interactions so that we can address the differences in reporting styles across the countries and survey waves:

$$c_{mt}^j = \gamma_0^j D^m + \gamma_1^j D^t + \gamma_2^j (D^m \times D^t) \quad (3)$$

where c_{mt}^j is the threshold between $H = j$ and $j+1$ of country m at time t , D^m is a set of dummy variables representing the country of residence, D^t is a dummy variable representing the survey wave, and $(D^m \times D^t)$ is a set of their interactions.

Table 2 presents the results of the generalized ordered regressions.⁹ All of the estimated coefficients, expressed in terms of the marginal effect, are statistically significant at the 1% level.

⁹ For a comparison, we also conducted a simple ordered probit regression, a generalized ordered probit regression with only country dummies (i.e., without a wave dummy and interactions between the country dummies and the wave dummy) in the threshold equations, and a generalized ordered probit regression with only country dummies and a wave dummy (i.e., without interactions between the country dummies and the wave dummy) in the threshold equations. However, the values of AIC ($= -2 \times \log \text{likelihood} + 2 \times (\text{number of parameters})$) of the regressions indicate that the generalized ordered probit regression with country dummies, a wave dummy, and their interactions in the threshold equations is a preferred specification.

The absolute value of the estimated coefficient is highest for Parkinson's disease, followed by stroke, heart attack, and chronic lung disease. We calculate the health index as a linear prediction from the generalized ordered probit regression, normalized to 0 for the worst health state and 1 for the best health state across the countries and over the survey periods.

Figures 2(a) and 2(b) show the 25th, 50th, and 75th percentile of the age-sex standardized health index distribution by country in the first and second waves, respectively. The countries are ordered by the median value of the health index. In both the first and the second waves shown in the figures, the median value of the health index is relatively low in Spain, Italy, and Denmark and high in Japan, the Netherlands, and Switzerland. Inequality in health measured from the difference in the 25th and 75th percentile of the health index distribution is largest in Italy, Spain, and the Denmark. Japan is among the countries with the least health inequality.

4. Differences in threshold levels of SRH due to the country of residence and survey waves

Using the health index as calculated in the previous section, we examine differences in threshold levels of SRH due to the country of residence and survey waves. Figures 3(a) and 3(b) present the country- and wave-specific threshold values of the health index by which the individuals in each country and survey wave classify their true health into one of the five SRH categories. The threshold values are calculated in accordance with the distribution of the health

index and SRH. For example, if 20% of all the respondents in a certain country or wave reported that their health was poor or fair, then the threshold value between “fair” and “good” is the value at the 20th percentile of the health index distribution among the respondents in the country or wave.¹⁰ Consistent with the findings of Jürges (2007), in the first wave, the threshold value between fair and good was lowest in Sweden and highest in Germany. In other words, the respondents in Sweden were most likely to report good or better health, while those in Germany had the opposite characteristic. Furthermore, the threshold values varied across not only countries but also waves. For example, in Japan, the threshold value between fair and good was 0.817 in the first wave, but it dropped to 0.785 in the second wave. As a result, in the second wave, Japan became the country whose respondents were the second most likely to report good or better health.

With the health index at hand, we compute SRH adjusted for the differences in threshold levels due to country of residence and survey waves. More specifically, to adjust for the differences in SRH due to differences in threshold values, we set the threshold values between the SRH categories common across the eleven countries and over the two survey waves and use these common threshold values to classify each respondent into the five SRH health categories based on the values of the person’s health index. The common threshold value between certain

¹⁰ There are some countries where the threshold value of the health index from very good to excellent is one. This means that the health index of all of the respondents whose SRH is “excellent” is equal to one in these countries.

adjacent health categories (e.g., between fair and good) is set as the average of all the country- and wave-specific SRH threshold values between the adjacent health categories presented in Figures 3(a) and 3(b).

Figures 4(a) and 4(b) show the distribution of adjusted SRH in the first and second waves, respectively. In contrast to what we observed in Figures 1(a) and 1(b), Denmark and Sweden became moderately unhealthy countries. The respondents in these countries appeared to be healthier in SRH in spite of their moderately low health index [Figures 2(a) and 2(b)], because the threshold value of the health index by which they classified themselves as in good or better health was also low [Figures 3(a) and 3(b)]. Even after adjusting SRH for differences in threshold levels of SRH, the respondents in Japan were among the healthiest in both the first and second waves.

Figures 5(a) and 5(b) show whether the respondents in each country overrated or underrated their health in the first and second waves, respectively. The horizontal axis represents the percentage of respondents who reported their health as very good or excellent from a country- and wave-specific threshold, and the vertical axis the proportion of those who would have reported their health as very good or excellent from the common threshold. The respondents in countries located above the 45-degree line underrated their health, while those in the countries below the 45-degree line overrated their health. Many countries were located above the

45-degree line, indicating that respondents tended to underrate their health. However, countries such as Denmark and Sweden were located below the 45-degree line in both the first and second waves. Interestingly, the respondents in Japan changed their reporting behavior between the first and second waves: they underrated their health in the first wave, while overrated their health in the second wave. Japan was located very close to the 45-degree line in both waves, however, indicating that the respondents in Japan overrated or underrated their health much less than those in other countries do.

In Figures 6(a) and 6(b), we highlight the discrepancies between the changes in SRH and the SRH adjusted for differences in threshold levels of SRH over the period of the two survey waves. The figures show that, in many countries, while the percentage of respondents who reported that their health was very good or excellent from a country- and wave-specific threshold decreased from the first wave to the second, the proportion of those who would have reported that their health was very good or excellent from the common threshold slightly increased from the first wave to the second. Furthermore, the change in health between the two waves was less significant when the adjusted SRH was used. Therefore, we can argue that using the unadjusted SRH to compare health of the middle-aged and elderly over time leads to different conclusions than those obtained using the adjusted SRH.

5. Discrepancies between changes in unadjusted and adjusted SRH over two waves

Taking the advantage of panel data, we further explore the discrepancies between the changes in the unadjusted SRH and the SRH adjusted for differences in threshold levels of SRH over the period of the two survey waves. At an individual level, the discrepancies occurred if, between the two survey waves, (1) the unadjusted SRH was upgraded while the adjusted SRH remained unchanged or deteriorated, (2) the unadjusted SRH remained unchanged while the adjusted SRH deteriorated, (3) the unadjusted SRH was downgraded while the adjusted SRH remained unchanged or improved, and (4) the unadjusted SRH remained unchanged while the adjusted SRH improved. We call cases (1) and (2) an upward-biased response to a change in one's own health, and cases (3) and (4) a downward-biased response. On average, for the eleven countries, about 26.01% of the respondents made an upward-biased response to a change in their health between the two survey waves, while slightly more than 31.29% respondents made a downward-biased response.

To investigate the determinants of biased responses to a change in health between the two waves, we conduct a multi-nominal probit regression. Here, the dependent variable is a categorical variable taking a value of 1 if the respondent made an upward-biased response, 2 if the respondent made a downward-biased response, and 3 otherwise. The independent variables include sex, age dummies, education dummies, and country dummies. To take into account the

possibility that an individual's experience of a significant life event between the two waves causes biased responses to a change in health, we also use as independent variables a dummy variable indicating that the respondent retired between the first and second waves, and another indicating that the respondent was separated with his/her spouse between the first and second waves.

Table 3 shows the results of the multi-nominal probit regression. The base outcome is "otherwise," namely, the case when neither the upward nor the downward response is made. The results in Column 1 indicate that individuals aged 55 to 60 and the Japanese and Italians were also more likely to make an upward-biased response to changes in their health, while individuals with higher education and the Swedish, Spanish, French, and Greek were less likely to do so. We also observe that individuals who were separated with their spouse between the two waves were more likely to make an upward-biased response, although significant only at the 10% level.

The results in Column 2 indicate that individuals with higher education and the Swedish, Dutch, Spanish, and French were more likely to make an upward-biased response, while those aged 55 to 60 were less likely to do so. Hence, there is evidence of biased responses to a change in health between the two waves due to individuals' age, education, and country of residence.

We further examine the determinants of contradictory responses to a change in health

between the two waves; that is, the cases where an individual upgraded (downgraded) his/her SRH in response to deteriorated (improved) health conditions. More specifically, we implement another multi-nominal probit regression, where the dependent variable is a categorical variable taking a value of 1 if the respondent's SRH was upgraded when his/her adjusted SRH deteriorated between the two waves (sharing 4.3%), 2 if the respondent's SRH was downgraded when his/her adjusted SRH improved between the waves (sharing 5.43%), and 3 otherwise. The independent variables for this regression are the same as used in Table 3. The regression results are presented in Table 4. We observe that the coefficients that are significant in Table 3 turn less significant and/or lower in size (except for the significance level of the coefficient on the dummy for separated), indicating that contradictory responses are not closely associated with individual attributes.

6. Conclusion

This study investigated whether we could compare SRH across the different countries and time periods as a measure of general health status of the elderly. Using panel data from the JSTAR and the SHARE surveys and following the analytic strategy of Jürges (2007), we calculated the SRH adjusted for differences in threshold levels of SRH due to country of residence and survey waves, and explored the difference in an individual's health status indicated by SRH and

adjusted SRH.

Our findings suggest that SRH of the elderly may not be comparable across different countries and survey waves. Indeed, in a given survey wave, the respondents in Sweden and Denmark tended to overrate their health, while those in the other countries were more likely to underrate their health. Moreover, the degree of overrating/underrating of the respondents within a country varied over the survey waves: the respondents in Japan even changed their tendency from underrating to overrating during the two waves. Thus, differences in the adjusted SRH across the countries and time periods did not coincide with differences in the unadjusted SRH. From the adjusted SRH, the Japanese elderly can be considered healthier than the elderly in most of the SHARE countries.

We also found that, at an individual level, reporting biases that led to discrepancies between the changes in SRH and the adjusted SRH over the survey periods were associated with an individual's age, education, and country of residence. Hence, taking SRH at face value and comparing it across the different individuals without taking into account their characteristics can be misleading. For example, it is possible that individuals who tend to make an upward-biased response in reporting their health may adapt easily to deteriorated health rather than lowering their self-assessment of health (Oswald and Powdthavee, 2008).

While the results seem suggestive, there are some limitations to this study. First, we make a

strong assumption that only two factors, the country of residence and survey waves, affect the thresholds of SRH. However, as suggested by previous studies, the thresholds could also depend on other demographic and socioeconomic factors such as sex, age, and education. Further questions such as why individuals in a certain country tend to overrate or underrate their health, or why individuals with certain characteristics (including the country of residence) tend to make biased responses to a change in their health over time, remain to be studied. Further research is required to address these limitations.

Acknowledgements

The inspiration for this paper came from the International Social Security (ISS) project, Phase VII, organized by the National Bureau of Economic Research (NBER). We are very grateful to the participants of the conference in Munich and Rome in 2012 for their very useful comments.

The views expressed in this paper are those of the authors, and not of any institution we belong to.

References

- Adams, P., M. D. Hurd, D. McFadden, A. Merrill, and T. Ribeiro (2003) "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status", *Journal of Econometrics*, Vol. 112, No. 1, pp. 3-56.
- Börsch-Supan, A., A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist, and G. Weber eds. (2005). *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*, Mannheim: MEA.
- Börsch-Supan, A. and H. Jürges eds. (2005) *The Survey of Health, Ageing and Retirement in Europe: Methodology*, Mannheim: MEA.
- Contoyannis, P., A. M. Jones, and N. Rice (2004) "The Dynamics of Health in the British Household Panel Survey", *Journal of Applied Econometrics*, Vol. 19, No. 4, pp. 473-503.
- Crossley, T. F. and S. Kennedy (2002) "The Reliability of Self-assessed Health Status", *Journal of Health Economics*, Vol. 21, No. 4, pp. 643-658.
- Deaton, A. S. and C. H. Paxson (1998) "Ageing and Inequality in Income and Health", *American Economic Review, Papers and Proceedings*, Vol. 88, No. 2, pp. 248-253.
- Ford, J., M. Spallek, and A. Dobson (2008) "Self-rated Health and a Healthy Lifestyle Are the Most Important Predictors of Survival in Elderly Women", *Age and Ageing*, Vol. 37, No. 2, pp. 194-200.

- Frijters, P., J. P. Haisken-DeNew, and M. A. Shields (2005) "The Causal Effect of Income on Health: Evidence from German Reunification", *Journal of Health Economics*, Vol. 24, No. 5, pp. 997-1017.
- Groot, W. (2000) "Adaption and Scale of Reference Bias in Self-assessments of Quality of Life", *Journal of Health Economics*, Vol. 19, No. 3, pp. 403-420.
- Hernandez-Quevedo, C., A. M. Jones, and N. Rice (2005) "Reporting Bias in Self-assessed Health: Evidence from the British Household Panel Survey", HEDG Working Papers, No. 05/04.
- Idler, E. and Y. Benyamini (1997) "Self-rated Health and Mortality: A Review of Twenty-seven Community Studies", *Journal of Health and Social Behavior*, Vol. 38, No. 1, pp. 21-37.
- Ichimura, H., H. Hashimoto, and S. Shimizutani (2009) "JSTAR First Results: 2009 Report", RIETI Paper Series, No. 09-E-047.
- Jürges, H. (2007) "True Health vs Response Styles: Exploring Cross-Country Differences in Self-Reported Health", *Health Economics*, Vol. 16, No. 2, pp. 163-178.
- Kapteyn, A., J. P. Smith, and A. Van Soest, (2007) "Vignettes and Self-reports of Work Disability in the United States and the Netherlands", *American Economic Review*, Vol. 97, No. 1, 461-473.
- Lindeboom, M. and E. van Doorslaer (2004) "Cut-point Shift and Index Shift in Self-reported

Health”, *Journal of Health Economics*, Vol. 23, No. 6, pp. 1083-1099.

Mossey, J. and E. Shapiro (1982) “Self-rated Health: A Predictor of Mortality among the Elderly”, *American Journal of Public Health*, Vol. 72, No. 8, pp. 800-808.

Noguchi, H. (2011) “Shakaiteki Keizaiteki Youin to Kenkou to no Ingasei ni taisuru Shokosatsu - `Shakaihoshoujittaichousa’ oyobi `Kokuminseikatsukisochoosa’ wo Mochiita Jitushou Bunseki- (Considerations on a Causal Relationship between Socioeconomic factors and Health: Empirical Analysis Using the `National Survey on Social Security and People’s Life’ and the `Comprehensive Survey of Living Conditions of the People on Health and Welfare’),” *Kikan Shakaihoshou Kenkyu*, Vol. 46, No. 4, pp. 382-402.

Oswald, A. J. and N. Powdthavee (2008) “Does Happiness Adapt?: A Longitudinal Study of Disability with Implications for Economists and Judge”, *Journal of Public Economics*, Vol. 92, No. 5/6, pp. 1061-1077.

Pfarr, C., A. Schmid, and U. Shneider (2011) “Reporting Heterogeneity in Self-assessed Health among Elderly Europeans: The Impact of Mental and Physical Health Status”, MPRA Paper, No. 29900.

Poterba, James, Steven Venti and David Wise (2010). “Family Status Transitions, Latent Health and the Post-Retirement Evolution of Assets,” NBER Working Paper 15789.

Sadana, R, C. D. Mathers, A. D. Lopez, C. J. L. Murray, and K. M. Iburg (2002) “Comparative Analyses of More Than 50 Household Surveys on Health Status”, in Murray C. J. L., J. A. Solomon, C. D. Mathers, and A. D. Lopez eds., Summary Measures of Population Health. Concepts, Ethics, Measurements and Applications. Geneva: World Health Organization, pp. 369-386.

Sickles, R.C. and P. Taubman (1986) “An Analysis of the Health and Retirement Status of the Elderly”, *Econometrica*, Vol. 54, No. 6, pp. 1339-1356.

Smith, J. P. (1999) “Health Bodies and Thick Wallets: The Dual Relation between Health and Economic Status”, *Journal of Economic Perspectives*, Vol. 13, No. 2, pp. 145-166.

Stern, S. (1989) “Measuring the Effect of Disability on Labor Force Participation”, *Journal of Human Resources*, Vol. 24, No. 3, pp. 361-395.

van Doorslaer, E., A. Wagstaff, H. van der Burg, T. Christiansen, D. De Graeve, U.G. Gerdtham, M. Gerfin, J. Geurts, L. Gross, U. Hakkinen, and J. John (2000) “Equity in the Delivery of Health Care in Europe and the U.S.”, *Journal of Health Economics*, Vol. 19, No. 5, pp. 553-584.

Van Doorslaer, E. and U. G. Gerdtham (2003) “Does Inequality in Self-assessed Health Predict Inequality in Survival by Income? Evidence from Swedish Data”, *Social Science and Medicine*, Vol. 57, No. 9, pp. 1621-1629.

Zajacova, A. and J. Dowd (2011) "Reliability of Self-rated Health in US Adults", *American Journal of Epidemiology*, Vol. 174, No. 8, pp. 977-983.

Table 1: Sample size by country, sex, and age group

	Total	Sex		Age group				
		Male	Female	-55	55-60	60-65	65-70	70-
JP	2833	51.4	48.6	17.6	22.5	20.4	21.4	18.1
AT	1014	43.9	56.1	21.4	21.0	27.0	17.1	13.5
DE	1349	48.0	52.0	25.9	19.3	24.5	20.3	9.9
SE	1636	47.0	53.0	22.7	26.9	21.7	16.7	12.0
NL	1501	46.0	54.0	27.8	25.3	20.0	14.5	12.4
ES	1037	44.5	55.5	25.2	18.5	18.4	21.3	16.6
IT	1504	43.9	56.1	20.8	23.1	24.3	17.8	14.0
FR	1524	45.3	54.7	29.0	22.3	18.7	16.3	13.6
DK	966	48.7	51.3	31.1	23.3	18.4	15.0	12.2
GR	1782	47.3	52.7	30.0	21.2	16.4	18.4	14.0
CH	553	48.6	51.4	26.6	23.0	21.2	15.2	14.1
Total	15699	47.1	52.9	24.5	22.5	20.8	18.1	14.0

Notes: JP = Japan, AT = Austria, DE = Germany, SE = Sweden, NL = Netherlands, ES = Spain, IT = Italy, FR = France, DK = Denmark, GR = Greece, CH = Switzerland

Table 2: Generalized ordered probit regression of self-rated health with country-wave-specific thresholds (Pooled across eleven countries and two survey waves)

	Coefficient estimates
Heart attack or other heart problems	-0.677** (0.022)
High blood pressure	-0.289** (0.013)
High blood cholesterol	-0.125** (0.016)
Stroke or cerebral vascular disease	-0.824** (0.043)
Diabetes	-0.525** (0.021)
Chronic lung disease	-0.666** (0.037)
Asthma	-0.321** (0.033)
Arthritis or rheumatism	-0.565** (0.018)
Osteoporosis	-0.424** (0.026)
Cancer or malignant tumour	-0.631** (0.039)
Stomach, duodenal or peptic ulcer	-0.351** (0.032)
Parkinson's disease	-1.152** (0.124)
Cataracts	-0.211** (0.028)
Hip or femoral fracture	-0.445** (0.068)
Other condition	-0.578** (0.017)
Ever treated for depression	-0.325** (0.024)
Low grip strength	-0.189** (0.014)
Grip strength test not completed	-0.466** (0.028)
BMI not normal (not within 20 to 25)	-0.154** (0.013)
Observations	31398
Log likelihood	-38819.791
AIC	77967.58
Note: ** $p < 0.01$.	

Table 3: Multi-nominal probit regression of upward/downward-biased responses to a change in health between two waves

	Upward-biased response	Downward-biased response
Female	0.008 (0.007)	0.003 (0.008)
Age (Baseline: 55 or younger)		
55-59	0.022* (0.011)	-0.018+ (0.011)
60-64	0.006 (0.011)	0.00001 (0.012)
65-69	0.005 (0.012)	0.013 (0.012)
70-	0.015 (0.010)	-0.009 (0.011)
Final educational achievement (Baseline: Elementary school)		
Lower secondary school	-0.011 (0.010)	0.011 (0.011)
Upper secondary school	-0.021* (0.010)	0.025* (0.011)
College or above	-0.034** (0.011)	0.029* (0.013)
Country (Baseline: Austria)		
Japan	0.078** (0.018)	-0.017 (0.018)
Germany	-0.001 (0.019)	0.004 (0.020)
Sweden	-0.055** (0.016)	0.076** (0.020)
Netherlands	-0.003 (0.018)	0.049* (0.020)
Spain	-0.044* (0.020)	0.072** (0.024)
Italy	0.033+ (0.019)	-0.001 (0.020)
France	-0.040* (0.018)	0.072** (0.021)
Denmark	0.01 (0.020)	0.025 (0.022)
Greece	-0.043* (0.017)	-0.004 (0.019)
Switzerland	-0.019 (0.023)	0.017 (0.026)
Event between 2 waves :		
Retired	0.014 (0.014)	-0.012 (0.015)
Separated	0.058+ (0.031)	-0.032 (0.030)
N	14718	14718
Log L	-15737.105	-15737.105

Note. The results presented are marginal effects. Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

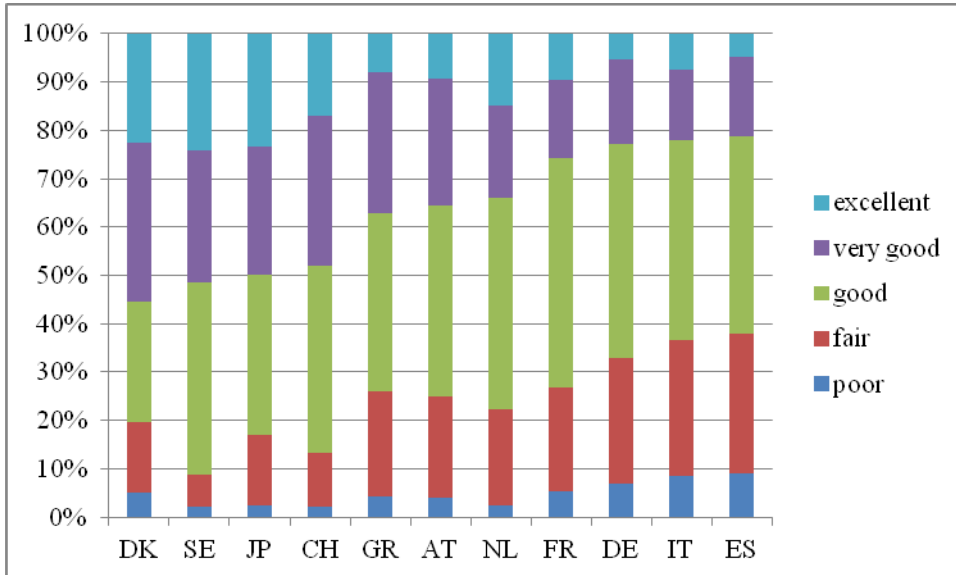
Table 4: Multi-nominal probit regression of reversed responses to a change in health between two waves

	Upward-reversed response	Downward-reversed response
Female	0.002 (0.003)	0.004 (0.004)
Age (Baseline: 55 or younger)		
55-59	0.008 (0.005)	-0.005 (0.005)
60-64	-0.003 (0.005)	-0.009 ⁺ (0.005)
65-69	0.006 (0.006)	-0.001 (0.006)
70-	0.012 ⁺ (0.005)	-0.004 (0.006)
Final educational achievement (Baseline: Elementary school)		
Lower secondary school	0.001 (0.005)	0.003 (0.006)
Upper secondary school	-0.001 (0.005)	0.006 (0.006)
College or above	0.002 (0.005)	0.006 (0.006)
Country (Baseline: Austria)		
Japan	0.037** (0.010)	-0.025** (0.006)
Germany	0.011 (0.010)	-0.015* (0.007)
Sweden	0.009 (0.007)	0.013 (0.009)
Netherlands	0.002 (0.009)	-0.007 (0.008)
Spain	-0.007 (0.009)	-0.003 (0.010)
Italy	0.016 (0.010)	-0.006 (0.008)
France	-0.007 (0.008)	0.002 (0.009)
Denmark	0.01 (0.010)	-0.016* (0.008)
Greece	-0.023** (0.006)	-0.033** (0.006)
Switzerland	-0.002 (0.011)	-0.016 ⁺ (0.009)
Event between 2 waves :		
Retired	-0.009 (0.006)	0.004 (0.007)
Separated	0.028 ⁺ (0.017)	-0.029** (0.010)
N	14718	14718
Log L	-5603.8301	-5603.8301

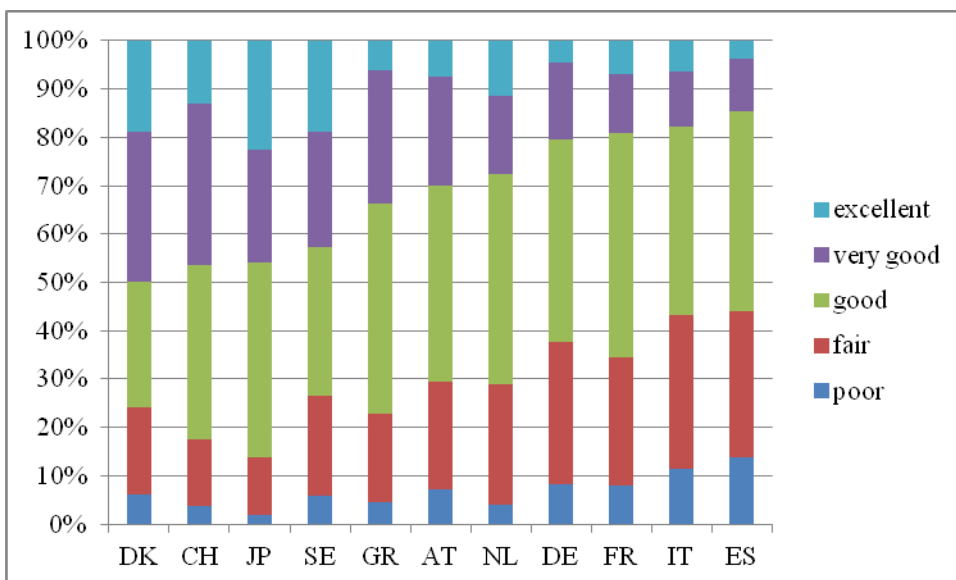
Note. The results presented are marginal effects. Standard errors in parentheses. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Figure 1: Self-rated health by country and wave

a) First wave



b) Second wave

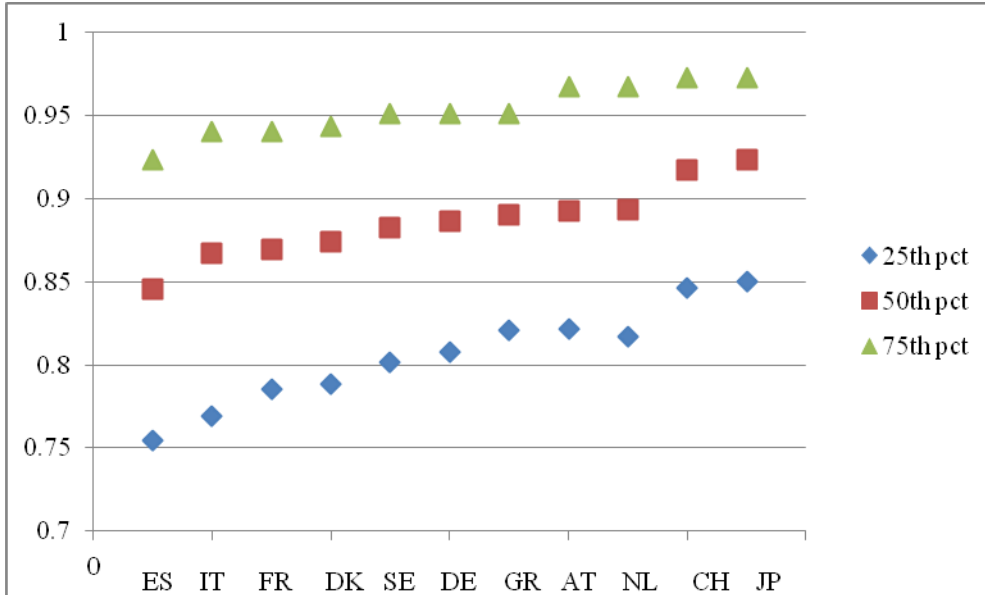


Note. Age-sex standardized. Countries ordered by the level of the fair to good threshold level.

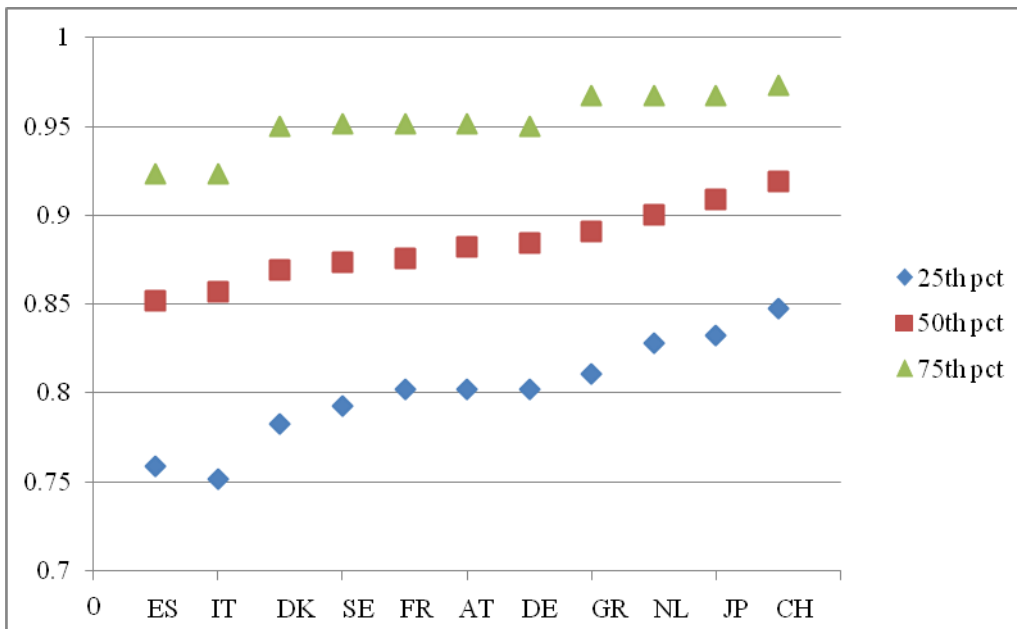
See Note on Table 1 for country codes.

Figure 2: Distribution of the health index by country and wave

a) First wave



b) Second wave

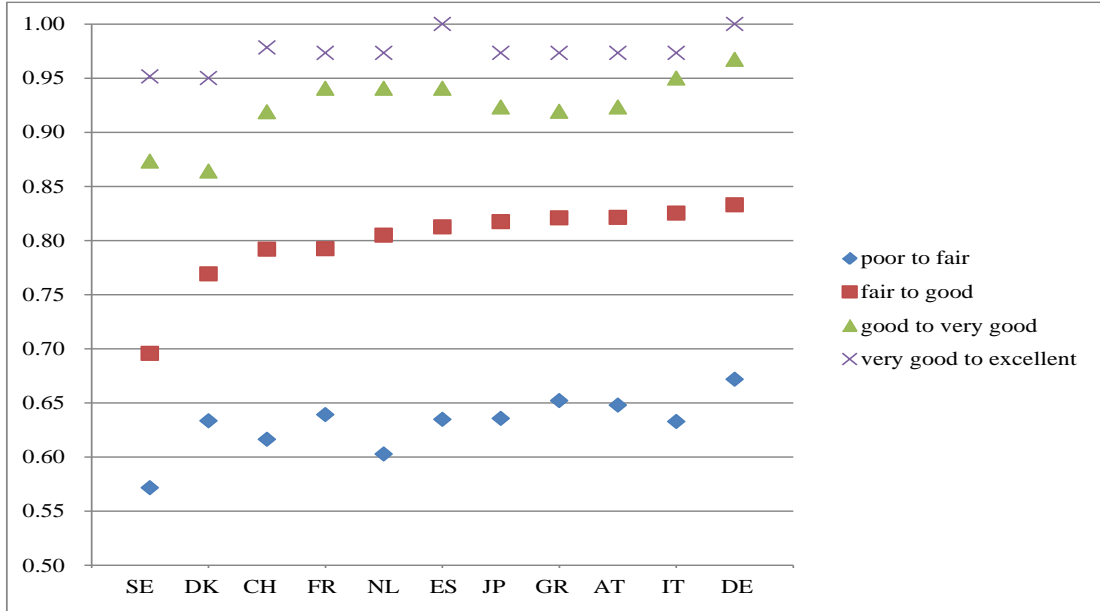


Note. Countries ordered by the value of the health index at 50th percentile.

See Note on Table 1 for country codes.

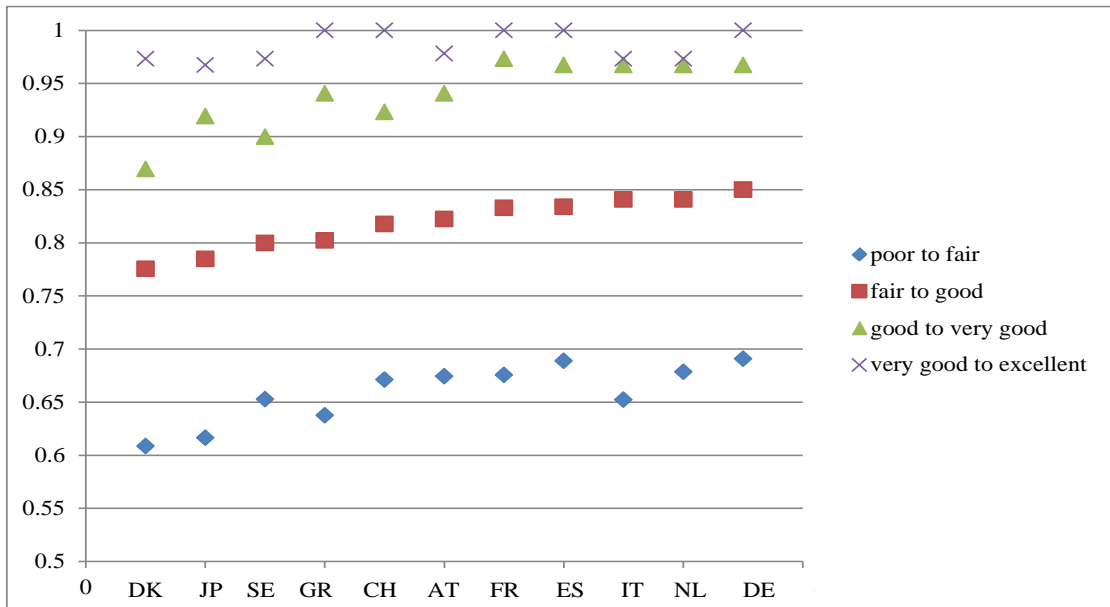
Figure 3: Health index threshold values by country and wave

a) First wave



b) Second

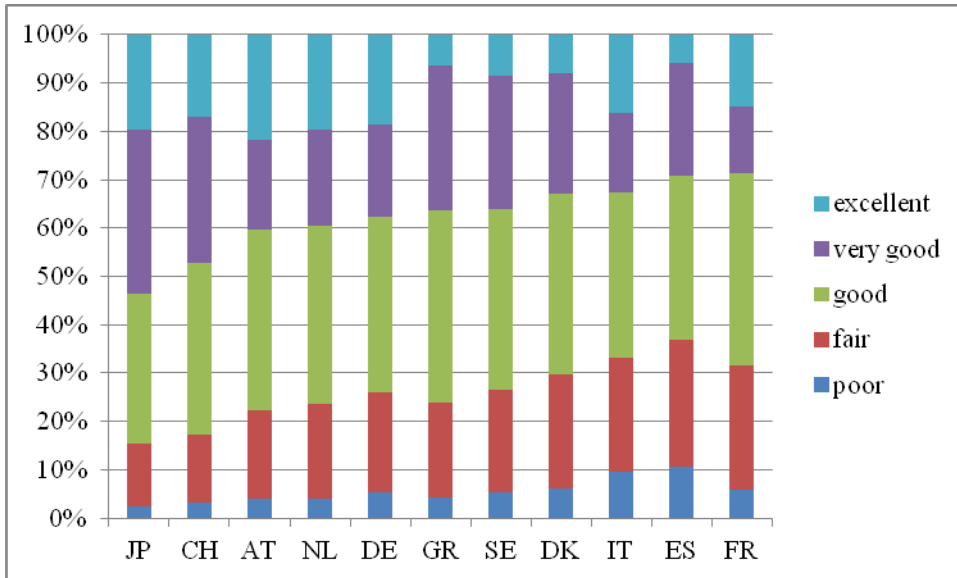
wave



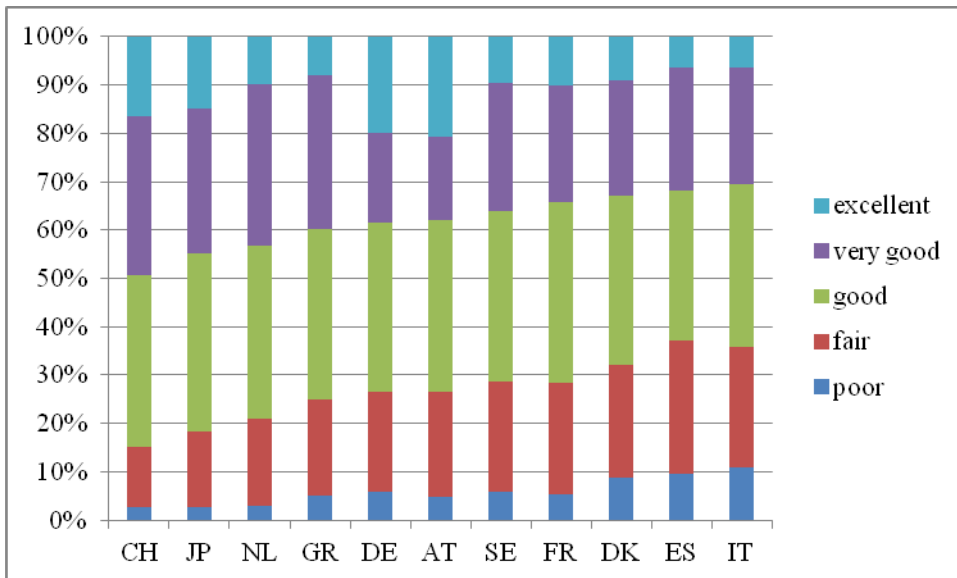
Note. Countries ordered by the level of the fair to good thresholds. See Note on Table 1 for country codes.

Figure 4: Self-rated health adjusted for differences in threshold levels of SRH by country and wave

a) First wave



b) Second wave

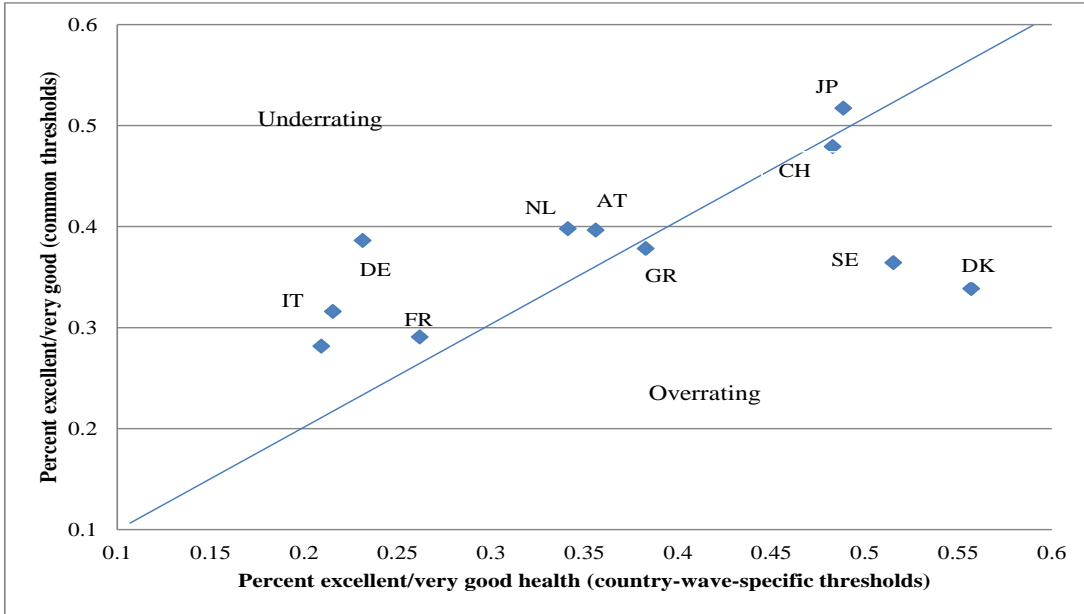


Note. Age-sex standardized. Countries ordered by the proportions of individuals in excellent/very good health.

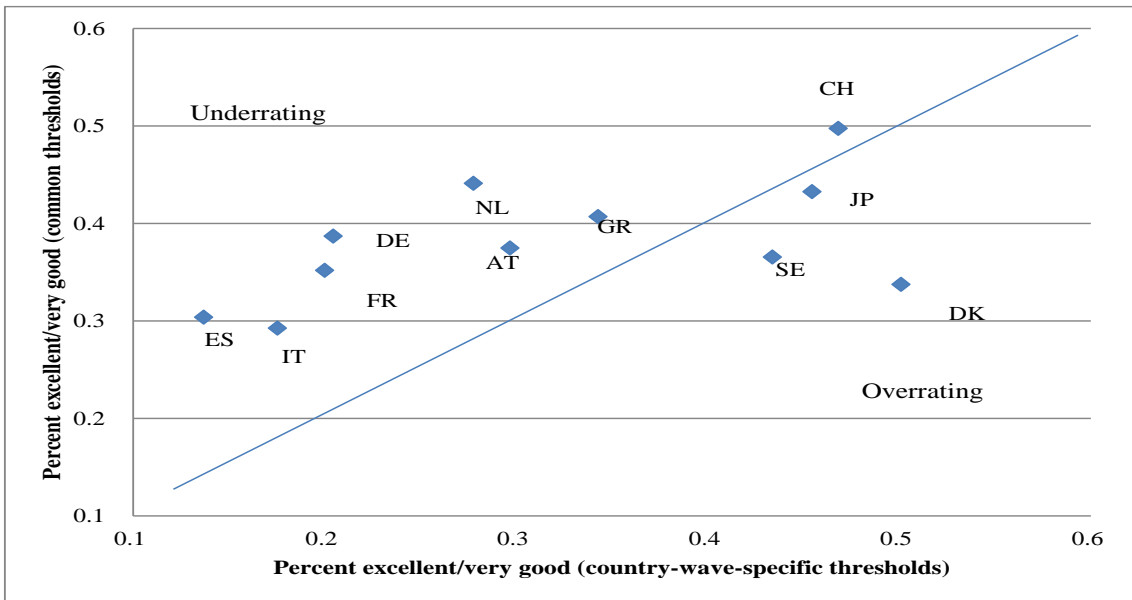
See Note on Table 1 for country codes.

Figure 5: Overrating and underrating health by country and wave

a) First wave



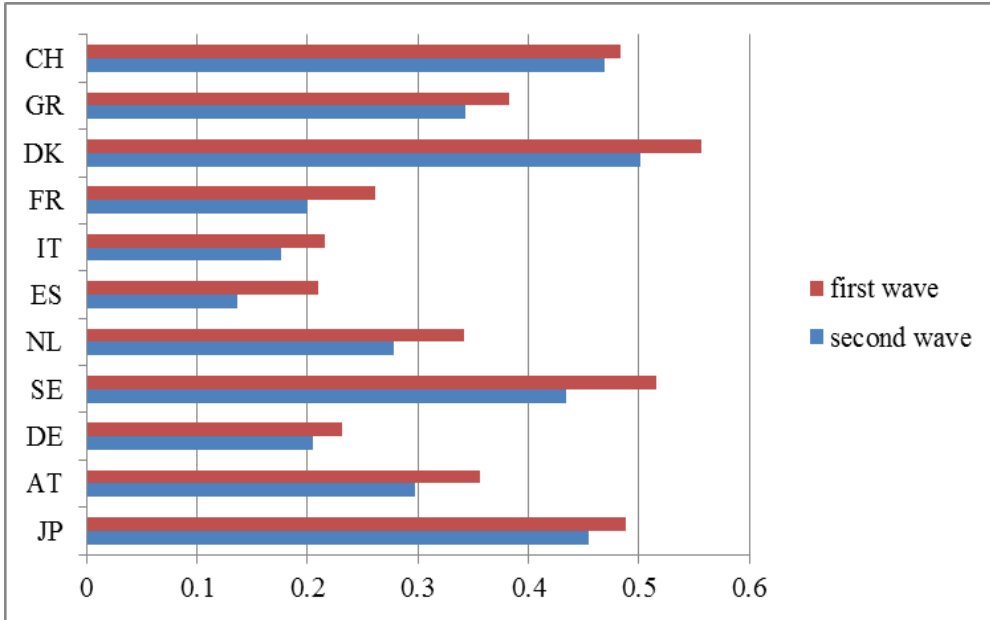
b) Second wave



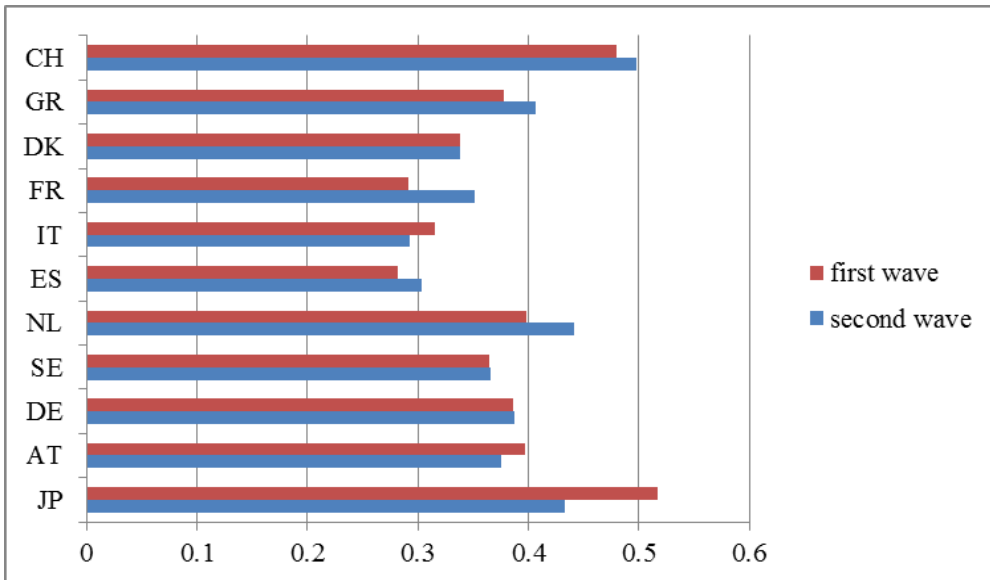
Note. See Note on Table 1 for country codes.

Figure 6: Percentage of excellent/very good health in self-rated health by country and wave: unadjusted and adjusted for differences in threshold levels of SRH

a) Self-rated health: unadjusted



b) Self-rated health: adjusted for differences in threshold levels of SRH



Note. Countries ordered by the level of the fair to good threshold. See Note on Table 1 for country codes.