



# HHS Public Access

Author manuscript

*J Health Soc Behav.* Author manuscript; available in PMC 2015 December 03.

Published in final edited form as:

*J Health Soc Behav.* 2014 September ; 55(3): 320–341. doi:10.1177/0022146514541446.

## How Health Conditions Translate into Self-Ratings: A comparative study of older adults across Europe

Melissa Hardy, Francesco Acciai, and Adriana Reyes

Department of Sociology & Population Research Institute The Pennsylvania State University

### Abstract

Using data from the Survey of Health, Ageing and Retirement in Europe, we examine how respondents translate morbidity and disability into self-rated health (SRH), how national populations differ in SRH, and how normative and person-specific reporting styles shape SRH. We construct proxy variables that allow us to specify cultural differences in reporting styles and individual differences in relative rating behavior. Using generalized logistic regression, we find that both of these dimensions of subjectivity are related to SRH; however, their inclusion does not significantly alter the connection between SRH and the set of disease and disability indicators. Further, country differences in SRH persist after controlling for all these factors. Our findings suggest that observed country differences in SRH reflect compositional differences, cultural differences in reporting styles, and perceptions of how health restricts typical activities. SRH also seems to capture underlying but unmeasured health differences across populations.

---

Self-rated health (hereafter SRH) provides an overall assessment of a multidimensional construct by combining the physical, mental, and social aspects of health in a single ordinal variable (Idler et al 1999). SRH, an item frequently included in large national surveys, has been discussed in conceptual and empirical reviews and compared to other health indicators (Kramers 2003; Wilson and Cleary 1995; Jylhä 2009). SRH, which has demonstrated stability, consistency, and good test-retest reliability, is strongly related to a wide set of health outcomes, including general morbidity (Bayliss et al. 2012; Benyamini et al. 2000), reported symptoms (Idler and Kasl 1995; Verbrugge and Jette 1994), health care utilization (Miilunpalo et al 1997), and mortality (DeSalvo et al. 2005; Idler and Benyamini 1997).

There is general agreement that the main determinant of SRH is physical health (Manderbacka, Lundberg, and Martikainen 1999) and that this connection holds in countries with both homogeneous and ethnically diverse populations (Idler and Benyamini 1997). Further, the view that self-rated health is a relatively stable but unobserved characteristic is implicit in the ordinal models used in much of the quantitative research, as is the assumption that people map this underlying construct to an ordinal scale of adjectives in a consistent way across the scale. However, once we begin to compare across countries, the understanding of cross-national differences depends on how one parses country differences

---

<sup>7</sup>Because coefficients are exponentiated, they must be multiplied to obtain the country-specific effects. For example, each additional chronic condition increases the odds of some restrictions by 34.9% for Danes; among Germans, each additional condition increases the odds by 63.8% (1.349\*1.215); the effect of an additional chronic condition on the odds of 'severe restrictions' does not differ for Danes and Germans (the nonsignificant coefficient .900 on the bottom right).

in health status versus country norms in how underlying health conditions may be translated into SRH (Jylhä et al. 1998).

When people respond to questions about SRH, they are making subjective evaluations by deciding where to place themselves in a set of predefined health categories. If we can assume that people have equivalent health information, that they weigh this information in the same way, and that their translations of this information onto a 5-point scale are consistent across the response set, then estimates of group differences from ordinal models can be taken largely at face value. However, we know that people with the same reported conditions, symptoms, and limitations rate their health differently, a divergence which suggests unobserved heterogeneity in health information, variation in the evaluative frameworks, or individual bias (e.g., pessimism or optimism) in choice of adjective (Jylhä 2009).

The cognitive process that produces these ratings relies on what people know about their own health and how people think about what health means. Health information can reflect contact with the health care system and the level of health literacy. What ‘health’ means, however, is clearly subjective. Further, the subjective nature of these deliberations--how people weigh the information they have and how they understand their own circumstances--can have both cultural and personal components. The cultural component can incorporate the social and physical environment people negotiate on a daily basis, including the shared construction of what ‘good’ health means (Knäuper and Turner 2003; Jylhä 2009). Such understandings provide the content of different health ratings, which inform the respondent’s selection of an adjective.

In this paper, we use data on eleven European countries from the Survey of Health, Ageing and Retirement in Europe (SHARE) to examine whether and how cross-national differences in SRH are influenced by health information, functional limitations, health-related restrictions in typical activities, and two dimensions of subjective rating behavior. Rather than bifurcate the scale of SRH, we use the full 5-category range, which allows us to identify nuances in relationships that may be missed when variation in SRH is collapsed. We use generalized logit models to accommodate the ordinality of SRH while relaxing the proportionality assumption, which allows us to see how relationships may be depend on where on the scale they are evaluated (Williams 2006). Finally, we assess how SRH reflects judgments that include both a social/cultural and an individual component. To do so, we create two proxy variables. The first proxy incorporates information on country-specific response styles and allows us to assess how country differences in rating behaviors contribute to observed country differences in SRH. The second proxy indicates individual adherence to these response styles (relative to others in similar circumstances) and allows us to assess whether person-based differences in rating behaviors influence SRH once we control for a series of widely used health indicators of disease, symptoms, functional limitations, disability, depression, and cognition.

## Background

The populations of European countries are among the oldest in the world. By 1975, close to 15 percent of the populations in Sweden, Austria, and Germany were aged 65 and older, and by 2000, eight of the ten 'oldest' countries in the world were European (UN 2007). Nevertheless, comparative studies of the relative health of older European populations have been slow to emerge. To date, survey-based comparative health research has focused primarily on diseases, disabilities, and mortality (Bambra 2011). But different measures produce different health rankings across countries. For example, life expectancy at age 60 is longest in France, and shortest in Denmark (WHO 2013). Using the physical component summary measure of health, Italy ranks at the top, the Netherlands at the bottom (Ware et al. 1998). If we focus on surviving from age 15 to 60, Switzerland has the best odds for men (France the worst), while Italy has the best odds for women (Greece the worst) (WHO 2013). Certainly all these measures reference important health outcomes; all fit well within a medical model of health; and all invite an increasingly micro-level focus on biology, genetics, and kinetics. But if we are concerned with the general health of populations, where we define 'health' as something more than the absence of disease or functional impairment, then studies of more subjective measures of health also provide important information (Idler, Hudson and Leventhal, 1999; Ferraro, Farmer and Wybraniec, 1997).

A major difficulty in making cross-country comparisons using SRH lies in the ambiguity of interpreting country differences. For example, where access to health care is unequal and quality of health care inconsistent, health information (such as disease diagnoses or risk factors) also will be unevenly distributed and unevenly understood across regional populations. Further, if people translate this health information into SRH using different evaluative frameworks, and if these frameworks have a cultural component, then these inconsistencies in both information and reporting styles will appear as country differences in self-rated health.

While considerable inequality in health care access and literacy severely complicate studies of SRH that include the developing world, among higher income countries, these differences are greatly reduced. Among the European countries in this study, all have national health care systems, which do not eliminate all problems of access or understanding, but do represent a national commitment to population health. Remaining inequities in the distribution or comprehension of health information is likely to overlap with differences in education and financial resources, which is one important reason to include socioeconomic characteristics in models of SRH (Manderbacka et al. 1999; Mansyur et al. 2008). Even so, some populations are more diverse than others, and these cultural differences can produce within-country variation in SRH (Viruell-Fuentes et al. 2011). To address the cultural component of rating behavior, researchers have concentrated on trying to adjust for group differences in reporting styles in comparative research (e.g., Jürges 2007), differential item functioning among subgroups (Grol-Prokopczyk, Freese, and Hauser 2011; King et al 2004), and differences in 'positional objectivity' when comparing regions with very unequal access to health care (Sen 2002).

## Subjective Dimensions of Self-Rated Health

Research into how people construct their health ratings has relied largely on in-depth interviews (Kaplan and Baron-Epel 2003). These smaller studies support the view that people formulate their health ratings using biomedical (disease) information as well as functional assessments and that their judgments combine both individual and social elements (Nettleton 1995). However, the relative importance of disease or impairment factors appears to differ across the range of responses. For example, for those in worse health, diagnosed diseases figured prominently in their ratings, while those in better health weighed difficulty with routine activities more heavily (Kaplan and Baron-Epel 2003).

Adjusting for differences in how underlying health is mapped to a 5-category scale raises two questions. The first—“how sick must you be to rate your health as ‘fair’ rather than ‘good’”—is an issue of thresholds. People may differ in where thresholds are placed when choosing an adjective in the ordinal scale, and the placement of these thresholds may be, in part, governed by country differences in how people talk about their health. If left unaddressed, country differences in response styles would be confounded with country differences in underlying population health (Jürges 2007).

A second question involves how specific conditions are weighed in constructing a composite rating of overall health. Health economists have favored using standardized health indices to account for differences in how specific health conditions are translated into summary assessments (Cutler and Richardson 1997; Jürges 2007). Each disease or performance measure can then be weighted relative to its importance in shifting health appraisals from better to worse. For example, a diagnosis of Parkinson’s disease may have a relatively large ‘disability weight,’ since people with Parkinson’s disease consistently rate their health lower than those without the disease other things equal (Jürges 2007). However, the impact of any disease also may vary by country. For example, a common diagnosis may be weighed less heavily than a rare one, as a common one may have better developed strategies of disease management and social and physical environments that are more accommodating. Further, individuals are likely to deviate from the normative rating and weigh certain symptoms or diagnoses more heavily if they interfere more with their quality of life.

The influence of any unobserved variation in underlying health (some dimension of severity, for example) or in the nature of translation (people’s tendency to be more or less ‘optimistic’ in their health evaluations) therefore remain. An emerging strategy for dealing with the latter involves the use of vignettes (King et al. 2004; Grol-Prokopczyk et al. 2011; Salomon et al. 2004), which suggests that people anchor their ratings in different ways. These differences can be measured by asking them to rate hypothetical health profiles, although the use of vignettes requires sets of additional questions and assumptions. Even so, using various research approaches to improve our understanding of how different subjective dimensions shape self-ratings can enrich our conceptual and comparative frameworks.

## Data and Methods

We use the Survey of Health, Ageing and Retirement in Europe (SHARE), which includes data for eleven European countries: Austria, Belgium, Denmark, France, Germany, Greece,

Italy, Netherlands, Spain, Sweden, and Switzerland. Based on national probability samples of a total of 25,736 respondents, these 2004 data were collected to be representative of the target populations in each country—those born in 1954 or earlier not living in institutions or abroad who speak the official language(s) of the country. Because countries employ different sampling strategies, we use weights provided by SHARE to account for variations in survey design and ensure that our estimates are representative (Börsch and Jürges 2005).

Our dependent variable is the European version of self-rated health (WHO 1996; Murray et al. 2002). Respondents were asked, “Would you say your health is: very good (5); good (4); fair (3); bad (2); very bad (1)?” Unlike the U.S. version, which allows ratings of poor, fair, good, very good, and excellent, the European version elaborates the lower end of the distribution, while collapsing the higher end. Even so, the two scales have been shown to produce concordant responses (Jürges, Avendano, and Mackenbach 2008). Our independent variables include standard demographic characteristics, measures of socio-economic status, cognitive functioning, self-reported morbidity factors, dummy variables for countries (using Denmark as the reference country), proxies for our two dimensions of subjectivity, and controls for various survey design features, such as the placement of the SRH question in the survey and the presence of a second person during the interview. We also specify an error structure that adjusts for observation clustering and violations of model assumptions, such as homoscedasticity.

Age is calculated by differencing the month and year of birth from the month and year of interview. The ages of our respondents range from 50 to 104 years old, with a mean of 65 and only minor differences in means across countries; we rescale age from the minimum value of 50 and use a dummy variable for gender, coded 1 for female respondents. We include 3 categories of marital status: married or in a partnership; widowed; divorced, separated, or never been married. Never marrieds were added to this last category after statistical tests indicated their similarity with the other statuses in this group once age was controlled.

Measures of economic status and education have been harmonized across countries. Education is based on the 7-point ISCED (International Standard Classification of Education) scale and ranges from 0 to 6, with 0 indicating no formal education and 6 the highest level (post-tertiary) of education. SHARE constructed a measure of gross household income (with imputed values for missing cases) using a detailed inventory of income sources and amounts (Börsch-Supan and Jürges 2005). To make monetary values comparable, both income and net worth are adjusted for purchasing power parity, which accounts for the different currencies and price structures across European countries. All monetary values are expressed in 2005 Euros and standardized to German prices. Since income is collected at the household level, we adjusted for household composition and then log transformed.<sup>1</sup> Household net worth sums material and financial assets minus debt. We

---

<sup>1</sup>Cross-country comparisons require constructing comparable monetary measures. Because of differences in currencies and price levels across countries, monetary values are adjusted by combining purchasing power parities with market exchange rates in each country using 2005 Germany as the price standard and the Euro as the currency. We also divide by the square root of the number of persons in the household to reflect economies of scale in consumption for multiple person households (Jürges 2007; Vignoli and De Santis 2009).

use the same set of adjustments for net worth as for income. Our final SES indicator is a binary variable coded '1' if the respondent is employed, which provides a necessary control for income.

Rating one's own health is also a semantic exercise; therefore, verbal skills are particularly important to this process. Language skills such as reading and writing are correlated with education and can be indicative of differences that have persisted across adulthood. Cognitive performance is also related to language and analytic skill, but may be influenced by underlying health conditions, severity of symptoms, or medications. We include five measures of cognitive function. *Verbal Fluency* tallies the number of different animals the respondent can name in one minute and ranges from 0 to 90<sup>2</sup> (Ardila, Ostrosky-solis, and Bernal 2006). *Memory* is based on the number of words the respondent can recall from a list of ten words read by the interviewer (Fournet et al. 2012) and ranges from 0 to 10. *Numeracy* (Chapman and Liu 2009) measures mathematical performance and indicates the number of correct responses to a set of four questions testing the ability to calculate and problem solve. We specify three categories of performance: none (no correct answer), low (only 1 correct answer), and medium-high (2 or more correct answers). In addition, respondents are asked to rate their reading and writing skills needed for their daily lives as excellent, very good, good, fair, or poor.

Six self-reported variables capture different dimensions of illness and disability. In each case, respondents review lists of conditions and report which items apply to them. The count (0 to 12) of *Symptoms* during the past six months addresses how health problems are experienced by the respondent (e.g., pain in joints; chest pain; difficulty breathing). *Chronic conditions* is the number of doctor-diagnosed conditions (from a list of 12) reported. *Depression* indicates whether the respondent reports ever experiencing depressive symptoms (again listed) for a period longer than two weeks.

Indicators of disability address different levels of severity. *ADL limitations* refer to the need for assistance with activities of daily living, and *IADL limitations* indicate assistance is needed with instrumental activities of daily living. Both are included as binary variables.<sup>3</sup> *Functional limitations* are the reported number (a maximum of 10) of mobility limitations, such as walking 100 meters, climbing several flights of stairs without resting, or reaching or extending your arms above shoulder level.

Finally, we generate two proxy variables as indicators of the subjective filters respondents may apply when they rate their health. These two variables represent a decomposition of the variation in *Self-rated activity restrictions*, or the extent to which respondents say they have been 'limited because of a health problem in activities people usually do' over the past six months. Response options are: not limited; limited, but not severely; and severely limited.

<sup>2</sup>The range of reported values for verbal fluency is wider than one would expect; however, the median value is 18 and only 1 percent of respondents have a score of 39 or higher, with .1 percent of respondents reporting 56 or more, and .05 percent naming 60 or more. We used the log transform of this variable both because of the skewed distribution and because we believe relative verbal fluency captures the relationship better than word count. Results are not affected by these relatively large values.

<sup>3</sup>The distributions of both ADL and IADL are highly skewed, with about 10 percent or less of the total sample reporting any limitations. In the series of models we tested, the distinction between those with and without limitations was the significant one; adding information on how many limitations did not improve model fit.



This question differs from the other questions about limitations as it does not specify the possible types of restrictions. Instead, respondents must determine whether and rate how their health restricts them in their normative activities, thereby adding an important experiential dimension to their functional reports. Our first country-specific proxy variable (CSP) incorporates country-specific rating behaviors by using the expected level of health-related restrictions in typical activities for people with the same set of observed characteristics living in the same country. These expected values capture country-specific representations of how health problems interfere with how people live their lives. The second proxy variable, the person-specific deviation (PSD) from this expected value, is an indicator of personal response style, or whether the respondent is more or less positive about their circumstances relative to what would be expected.

## Approach

We use a three-stage procedure to analyze how self-rated health is structured across our eleven European countries. In the first stage, we use generalized logistic regression (Williams 2006) with self-rated health (SRH) as the dependent variable. Generalized logit analysis allows an alternative ordinal modeling approach when some of the independent variables do not meet the proportionality assumption. Ordered logit provides one set of coefficients under the assumption that the effect of an independent variable is the same as one shifts comparison pivot points up or down the scale. Generalized logit, on the other hand, allows the coefficient of the independent variable to change across the response range, such as relaxing the assumption that the gender difference in the odds of reporting very good versus worse categories (good, fair, bad, and very bad) is the same as the gender difference in the odds of reporting very bad versus better categories (bad, fair, good, and very good), for example. It is more efficient than multinomial logit because multiple coefficients are not estimated when the proportionality assumption is met. It is more flexible than ordered logit because estimates are not averaged across the response range when the proportionality assumption is violated.

The first model estimates country-specific thresholds controlling for age. The second adds other demographic characteristics, SES, and cognitive measures. The third model adds self-reported indicators of illness and disability. We also tested the hypothesis that health indicators had country-specific associations with SRH, but the model without interactions was preferred. In this stage, we are interested in: (1) whether and how country differences in SRH change once these compositional differences are controlled; and (2) how the various health indicators are associated with SRH across the range of responses.

In the second stage, we decompose *Self-rated activity restrictions* into two proxy variables — country specific ratings (CSP) and person specific deviations from these ratings (PSD). CSP is the predicted value of *Self-rated activity restrictions* based on a generalized logit model that includes: demographic, SES, and morbidity measures; country; and interactions between country and symptoms, country and chronic conditions, and country and functional limitations. It captures how people in a given country with shared characteristics typically evaluate the connection between their health and activity restrictions.

PSD, which is based on the error term, is a measure of relative perceived severity, or how respondents rate their levels of restriction relative to CSP, or the country-specific norm for a synthetic reference group of people with the same demographic, SES, disease, and disability characteristics. In this stage, we are interested primarily in capturing how respondents from different countries may assign different weights to various health indicators in rating their restrictions. We then use these estimates and the observed individual characteristics to create CSP and PSD.

In stage three we estimate two additional generalized logit models for SRH in which we sequentially add these two proxy variables, thereby specifying two sorts of subjective filters. CSP captures self-rated activity restrictions while incorporating country-differences in rating behavior. PSD provides information on person-specific rating behavior that is more or less negative than the (country-specific) norm. In this final stage, we focus on whether CSP is related to SRH controlling for compositional differences in demographic, SES, and health indicators; whether the associations between SRH and health indicators are thereby reduced; and whether including CSP changes country differences in SRH. Lastly, PSD's association with SRH (controlling for all other factors) and the final structure of country coefficients will indicate whether SRH captures national differences in overall health beyond what differences in disease, disability, and evaluative frameworks generate.

## Results

Descriptive statistics for the variables used in this analysis are reported in Table 1. We include country specific statistics as well as those for the pooled sample. For all countries, 'good' was the modal category for the dependent variable, but percentages ranged from more than 50 percent in the Netherlands to about 40 percent in Spain, Italy, and Greece, to 36 percent in Sweden. Spain and Italy had relatively low proportions reporting 'very good' health and relatively high proportions reporting 'fair' and 'bad' health.

Average education was higher in Denmark and Germany and lower in Spain and Italy. Most respondents were currently married. Average income was highest in Switzerland, twice as high as incomes in Greece. Net worth was also highest in Switzerland and lowest in Greece, but here the ratio was closer to 3:1. The highest rates of employment occurred in Sweden and Switzerland (about 40%), and the lowest rates were in Austria and Italy (about 20%). Spain and Italy were lowest on verbal fluency, memory, and numeracy. The number of symptoms and chronic conditions averaged between 1 and 2, lowest in Switzerland; between 7 and 12 percent reported at least one ADL, and 4 to 9-plus percent reported at least one IADL. Across all countries, about one-in-four respondents reported ever experiencing depressive symptoms, with the lowest proportion in Greece and the highest proportion in France. Whereas functional limitations were higher in Spain and Italy, self-rated activity restrictions were most severe in the Netherlands; more than 70 percent of Greeks claimed no self-rated activity restrictions.

### Stage 1: Generalized Logit Results for Three Models

We turn now to the results of the generalized logit models. Table 2 includes parameter estimates for our first 3 models. Exponentiated coefficients are organized within five panels.



The top panel reports coefficients for the ‘very bad’ versus ‘better’ health, where ‘better’ includes the four more favorable response categories. For variables that meet the proportionality assumption (e.g., logged income and logged net worth), these coefficients apply across the full set of ordered comparisons; therefore, we report just one coefficient per variable in the top panel. For variables that violate the proportionality assumption, we report the full series of four coefficients as the comparison shifts from ‘very bad’ versus ‘better’ to ‘worse’ versus ‘very good’; in the top panel, these coefficients are shaded in gray. Some independent variables may satisfy the assumption in some models, but not in others. We use bold-italics type to indicate whether a specific coefficient is significantly different from that reported in the top panel, since non-proportionality does not mean that coefficients for every comparison are distinct. We use asterisks to denote whether the estimate is significantly different from zero. Goodness-of-fit statistics are reported at the bottom.

The coefficients for country controlling for age (model 1) capture country differences in the odds of ‘better’ versus ‘worse’ health. Values greater than 1 indicate better health than the reference country, Denmark; values less than 1 indicate worse health than Denmark. Only Sweden violated the proportionality assumption, which means that the differences between the remaining countries and Denmark were consistent across the range of responses. In all countries except Switzerland, reported health status was worse than in Denmark. These differences were largest between Italy and Denmark, but not significant between Belgium and Denmark. Swedes more often reported worse health than Danes at the lower end of the scale, but at the upper end of the scale (comparing worse to ‘very good’ health), Swedes reported significantly better health than Danes.<sup>4</sup>

For models 2 and 3, we highlight patterns of relationships. When we added demographic and SES variables, both Italy and Sweden displayed non-proportional effects. Women and widows/widowers reported worse SRH. Higher income and higher net worth was associated with better SRH. All these effects were proportional, and all were consistent with previous studies. Age, education, and employment status had non-proportional effects, but were also as expected, with people of older ages reporting worse SRH in all but the first comparison (very bad to better). Employment status made the most difference at the lower end of the scale, with the employed much less likely to rate their health as very bad; differences in health ratings associated with employment status narrowed as we move up the scale. More educated people were more likely to rate their health as very good suggesting that education mattered more at the upper end of the scale, distinguishing those with ‘best’ SRH. The cognitive variables all showed positive relationships with SRH.

Self-reported indicators of disease and disability included in Model 3 markedly improved model fit, and each indicator exhibited a unique relationship with SRH. Those reporting

---

<sup>4</sup>To illustrate interpretation of proportional and nonproportional effects, consider the coefficients for ‘Italy’ in models 1 and 2. In model 1, the odds of rating health as better than ‘very bad’ are about 63% lower for Italians than for Danes; because the effects is proportional, the odds of rating health better than ‘bad’ are also 63% lower for Italians than Danes, as are the odds of rating health better than ‘fair,’ and so forth. In model 2, however, the Italy-Denmark contrast is nonproportional; therefore, the odds of rating health better than ‘very bad’ are not significantly different for Italians than Danes nor are the odds of rating health better than ‘bad;’ they are 17% lower for Italians for rating health better than ‘fair,’ and 44.2% lower for rating health better than ‘good.’ In model 1, estimates indicate consistently lower odds of reporting ‘better’ versus ‘worse’ health for Italians versus Danes; in model 2, however, the odds for Italians and Danes do not differ at the bottom of the scale, but as we move up the scale, the gap between Italians and Danes grows larger.

more symptoms, more chronic conditions, depression, ADLs, IADLs, and functional limitations rated their health as worse. The effect of reporting chronic conditions grew stronger as the comparison shifted toward the positive end of the scale, suggesting that having one or more chronic conditions may have been relatively common among those with ‘very bad,’ ‘bad,’ or ‘fair’ SRH, while the absence of chronic conditions was better at discriminating those reporting ‘very good’ health. Remaining disease and disability indicators satisfied the proportionality assumption, suggesting that they influenced reports in a similar way across the range of responses.

Now that the variability in self-rated health linked to disease and disability was directly specified, some of the relationships described earlier were altered. While the effect of income appeared to have strengthened somewhat, net worth was no longer significant. The effect of education was now proportional (perhaps because of its overlap with chronic conditions), with the relationships between SRH and cognitive indicators largely unchanged. The coefficient for employment status was halved, indicating that employment was concentrated among those with fewer diseases and disabilities.

Two of the more intriguing differences occurred for age and female. Comparing those with similar levels of morbidity, older people were less likely to rate their health as ‘very bad,’ ‘bad,’ or ‘fair,’ but also less likely to rate their health as ‘very good.’ The reversal in the direction of this net relationship was consistent with the earlier research on the influence of reference group on self-rated health. Once we compared younger to older respondents who shared disability or disease diagnoses, we saw that younger respondents viewed their health more negatively. To the extent that older respondents referenced their age peers in making their ratings, certain health problems may have been more common; therefore, having these problems may not have seemed inconsistent with rating ones health as ‘good’ overall (Jylhä 2009). In contrast, for younger respondents, age contemporaries were less likely to share their health conditions, which may have led them to a more negative rating. Similarly, once we compared men and women with similar levels of disease and disability, women rated their health better than men except at the high end of the distribution, where they were equally likely as men to rate their health as ‘very good.’

After controlling for compositional differences in demographic, socioeconomic, cognitive, disease, and disability characteristics, countries can be sorted into four major groups, with Switzerland, Belgium, and Greece reporting the ‘best’ self-rated health.<sup>5</sup> Denmark, Spain, the Netherlands, and France were in the second tier, with Denmark at the top and France at the bottom. This tier was made of two intersecting clusters; in fact, although Denmark and France were significantly different from each other, they were both not significantly different from Spain and the Netherlands. Austrians rated their health as worse than the French, putting them in the third tier, but better than the Germans, who were at the bottom of the scale. Sweden and Italy, for which the proportionality assumption was not met, were more difficult to assign to only one tier, as their health ratings compared to Denmark’s – and

---

<sup>5</sup>To establish the grouping of countries for the different models, we performed a sequential series of Wald tests to determine whether country pairs were significantly different in SRH once compositional differences in populations had been controlled. On occasion, specific countries may straddle rungs in this hierarchy, an outcome which we report for the reader to consider.

indirectly to those in other countries – differently across the response set. Specifically, Swedes were like Germans or Austrians in that they were more likely to use the low end of the scale and rate their health more negatively than Danes; however, this did not hold at the upper end of the scale. In the last comparison (worse vs. very good), Swedes had higher odds than the Danes of rating their health as ‘very good.’ For Italians, the comparative shift was in the other direction: they were among the least likely to say ‘very bad,’ more similar to the middle tiers in claiming ‘bad’ or ‘fair,’ and more like the Austrians in avoiding the rating of ‘very good.’ In other words, Italians seemed to gravitate toward the intermediate categories.

## Stage 2: Decomposing Self-Rated Activity Restrictions into two Proxy Variables

Our goal in this part of the analysis is to capture as best as we can the country differences in self-rated activity restrictions as well as the country differences in the way respondents weigh the multiple dimensions of disease and disability in their ratings. Because the question refers to the level of difficulty people experience in doing what ‘people usually do,’ respondents should have in mind some normative notion of typical activities, something that also may differ across cultures; therefore, we include interaction terms to test for these differences. We tested various extension of this model, including memory, BMI, rural/urban, and additional interactions before settling on this model.<sup>6</sup>

Results are reported in Table 3, which is organized to help the reader interpret the coefficients. In the top panel above the dashed line we report coefficients that are averaged across all respondents and that refer to the comparison “not limited” versus “moderately or severely limited.” Below the dashed line, on the left, we show the effect of symptoms, chronic conditions, and functional limitation for Denmark, our reference country. The underlined coefficients represent the country-versus-Denmark differences for respondents with no symptom, chronic condition, or functional limitation. On the right, we report the interactions between country and symptoms, conditions, and functional limitations, which indicate differences in how these health problems are weighted by respondents in each country, again with Denmark as a reference. Finally, below the solid line we report the coefficients of the variables that do not respect the proportionality assumption. These coefficients refer to the comparison “not or moderately limited” versus “severely limited”. In the top panel we again highlight the non-proportional coefficients.

With the exception of ‘Female,’ the set of demographic, SES, depression, and disability indicators all met the proportionality assumption. The severity of self-rated activity restrictions did not increase with age, but women were less likely than men to report activity restrictions, other things equal; the gender gap increased for the second contrast, with men even more likely to rate their restrictions as severe. All SES measures were significant, with the more educated and those with higher income and net worth more likely to report no or

---

<sup>6</sup>We decided not to include the cognitive measures in this stage for several reasons. Only two of the five variables—verbal fluency and self-reported reading skills—were significant; these two language measures were mediating the effect of education; including them improved the fit of the model only on the margin (by .0026 in pseudo- $R^2$ ); and the conclusions (without the nuances of mediating effects) did not change. We therefore chose the more parsimonious model.

only moderate restrictions. In contrast, respondents who reported ever being depressed, ADLs, or IADLs rated their restrictions as more severe.

The coefficients for symptoms, chronic conditions, and functional limitations were associated with more severe self-rated activity restrictions among Danes. While the effects of chronic conditions and functional limitations were proportional, the influence of symptoms was weaker when predicting severe self-rated activity restrictions. Because we included three sets of interaction terms that involve country, the underlined coefficients on the left for the ten included countries indicate how the odds of reporting more serious self-rated activity restrictions differ between a given country and Denmark among people with no symptoms, chronic conditions, or functional limitations. Respondents in Spain, Italy, France, Greece, and Belgium were much less likely than Danes to report moderate or severe self-rated activity restrictions. Ratings of respondents from the remaining countries did not differ from those of Danes. Respondents in the Netherlands were the most likely and the Spanish were the least likely to report severe restrictions.

Coefficients for the three sets of interaction terms are reported in the right-hand columns and indicate how the relationship between symptoms (or chronic conditions or functional limitations) and self-rated activity restrictions differed by country, with the relationship among Danes being the point of comparison.<sup>11</sup> Symptoms appeared to be more important in Sweden, less important in the southern European countries. In contrast, chronic conditions were less important among the Swedes, more important in Austria and Germany. Functional limitations were most important for Austrians.

In sum, self-rated activity restrictions tended to be rated more severely as health conditions mounted. Once health information was included, men reported more severe self-rated activity restrictions than women. Country differences were evident in both the distribution of self-rated activity restrictions and in the way various factors figured into those ratings. Residents in southern Europe were more likely to report no or less severe self-rated activity restrictions, while in the Netherlands self-rated activity restrictions were more likely to be seen as severe.

Using the results from this model, we generated predicted values of self-rated activity restrictions to create CSP, which incorporates cultural inter-subjectivity in evaluating the severity of self-rated activity restrictions caused by health. The second proxy variable, PSD, is the difference between the observed and predicted values of *self-rated activity restrictions*, or the person-specific deviation. Summary statistics for these two variables are reported at the bottom of Table 1. CSP can range from 0 to 2, with higher values indicating more severe restrictions. Mean values for PSD within country were very close to zero, as expected. For PSD, values greater than zero indicate that the respondents assessed their restrictions as more severe than typical for respondents from the same country with the same demographic, socioeconomic, and morbidity characteristics.

### Stage 3: Incorporating Two Aspects of Subjectivity

In table 4 we repeat the results for model 3 (with the full set of observed independent variables) and add results for models 4 and 5 (which stage in the proxy variables); we

organized these results as we did in Table 2. The fit statistics for model 4 versus model 3 indicate CSP performed a mediating role, since very little new variance was explained. Worse SRH was associated with a higher expected severity of activity restrictions. Also, once CSP was controlled, the association between employment and SRH was somewhat weaker, and the association between SRH and symptoms, ADLs, and IADLs was somewhat stronger. With CSP in the model, coefficients for countries registered the most notable changes, suggesting that country-specific response styles were indeed reflected in the country differences in SRH reported in Model 3.

Model 5 added PSD and allowed us to test whether relative differences in personal rating behaviors—how respondents rate their restrictions relative to the country’s normative response for those with like characteristics—were related to SRH once we controlled for our set of covariates. The significant improvement in model fit indicated that these relative ratings of activity restrictions also were reflected in SRH. People who rated their activity restrictions as less severe than their counterparts also provided a more positive rating of SRH. This attempt to specify (between-group) culturally shared rating behaviors and the relative (to the ‘synthetic reference group’) severity of personal rating behavior also produced a somewhat amended ranking of countries on SRH.

### The changing pattern of country differences in SRH

Figure 1 summarizes country rankings on SRH as we moved through the various stages of our analysis. The three columns refer to country rankings based on models 1, 3, and 5 respectively. Average ratings decline (SRH is worse) as we move from top to bottom. The star-line delimiters are the boundaries on the different tiers of countries: in general, those within a tier were not statistically different from each other; however, across tiers we did find significant differences. We indicate the occasional fuzziness in defining the separate tiers with shadowed text. Both Sweden and Italy moved across tiers (although in opposite directions) as the comparison shifted from ‘very bad’ versus better health to lower ratings of health versus ‘very good.’ We used the weighted probabilities of the ratings to approximate their location had their coefficients been the same across all comparisons. We remind the reader of their atypical status by using italics and parentheses.

Beginning in the first column, countries were distributed across six tiers, with the Swiss rating their health the best, and Germans and Spaniards followed by Italians rating their health least favorably. The boundary between the second and third tiers was fuzzier, with Denmark, Sweden, and Belgium not statistically different from each other; Belgium not statistically different from the Netherlands, but different from Greece; and Denmark and Sweden different from the Netherlands. The Austrians and French, who offer equivalent ratings, were in the middle of the distribution. This ranking was adjusted for differences in the age composition of the population, but in many ways repeated the descriptive information we saw in Table 1. This first column answers the question: How would countries compare with regard to SRH if their age distributions did not differ?

Column 2 depicts the rankings based on model 3, in which we controlled for demographic, SES, cognitive, disease, and disability indicators, all of which could explain why some respondents rate their health as better or worse. We again found six tiers, but Switzerland no

longer stood alone at the top; it was joined by Greece and Belgium, suggesting that one reason the Swiss rated their health relatively favorably was because as a population, they had higher SES and lower levels of health problems and disabilities. But now that we are comparing respondents from these different countries with similar SES, disease, and disability profiles, respondents from Switzerland, Greece, and Belgium rate their overall health as better than respondents from other countries; and respondents from Germany rate their overall health as worse than their counterparts in other countries, with Austrians above them, and Italians above the Austrians, but below the French.

In column three we look at the country rankings that correspond to model 5, in which we added controls for expected levels of activity restrictions, country-specific reporting styles, and person-specific deviations. By doing so, we assessed the extent to which the country rankings we observed in model 3 reflected country differences in reporting styles. Three changes in rankings are worth noting. First, our six tiers have been reduced to four, suggesting that some of the country differences we identified in earlier models were due to country-specific response styles. Second, Greece and Belgium were once again distinct from Switzerland, suggesting that response styles in the former two countries were relatively favorable; once we took that more positive framework into account, these two countries shifted to the second tier. Third, Spain and the Netherlands switched tiers: Spain shifted to a worse relative rating and the Netherlands moved firmly into the second tier. This switch indicated that while Spaniards might be more positive when providing an overall rating, the response style in the Netherlands was more negative relative to their counterparts in other countries. What remain as country differences in SRH may reflect country differences in a variety of factors, including aspects of disease severity or progression that provide unique information relative to the indicators we specified, subjective aspects of health that may be difficult to capture in surveys but nevertheless inform ratings, or features of health infrastructure that shape disease management as well as artifacts of language (Viruell-Fuentes 2011) or other sources of measurement error.

## Discussion

Scholars differ in whether they view the subjective dimension of SRH as an asset or a nuisance (Jylhä et al. 1998). Variation in SRH that is unexplained by ‘health facts’ may be regarded as experiential differences in health, equally valid but somewhat ambiguous in meaning. In fact, people’s ratings might reflect health awareness not captured by the specific health indicators, but nevertheless relevant to a rating of overall health. In general, we collect more information on the presence of diagnosed conditions or limitations than on the severity of discomfort. Whereas the former have an observational component, the latter relies on judgment, since people experience illness in different ways. Side effects of medications, different sorts of complications of multi-morbidity, and different pain thresholds are just a few of the factors involved in overall health that are difficult to fully capture without incorporating some aspect of individual subjectivity.

In this paper, we exploited the distinction between self-reported (from a list of conditions) and self-rated (an overall evaluation) measures of health to gain leverage in studying subjectivity in ratings behavior. In decomposing the variance in self-rated activity



restrictions, we estimated two dimensions of subjectivity addressed in the theoretical, but generally not in the empirical literature on SRH—the inter-subjectivity of social groups (in this case country) and the relative severity of respondents' own subjective ratings, which we measured as their deviation from these norms. We found population-based evidence consistent with what has been reported in a small number of studies based on in-depth interviews (Kaplan and Baron-Epel 2003; Nettleton, 1995). Both dimensions of subjectivity appear to be operating; however, they do little to blunt the direct effects of the disease and disability indicators (Manderbacka et al. 1999). Instead, they suggest that in making their ratings people consider how specific health conditions affect their routine activities, and they apply both a cultural (country-specific) and a personal lens in doing so. We also gained insight into how people translate their circumstances into these ratings, thereby learning more about country differences in how health is experienced.

Once compositional and subjective factors were controlled, SRH across countries was reshuffled, in part because respondents from southern Europe rated the activity limitations due to their health conditions more moderately than did respondents from the rest of Europe, especially central European countries. Other studies have noted that although both men and women living in Mediterranean countries generally report worse health than those who live in Continental or Scandinavian countries, they are not more likely to be hospitalized, and they have higher survival rates and longer life expectancy (Knoops et al., 2004). We leave as speculative at this point whether these differences reflect physical and social accommodation, diets and health behaviors (Trichopoulos and Lagiou 2004), political regimes (Huijts, Perkins, and Subramanian 2010), or some other set of factors, although these are issues we continue to investigate.

We recognize that our study is limited by a number of factors. Although the overall number of cases we analyze is quite large, these cases are unevenly distributed across the eleven countries, and somewhat different strategies were used to generate the samples. Surveys were conducted in languages that differ across countries and in some cases within countries. We used the full scale of SRH rather than a binary indicator, but even five categories mask considerable within category heterogeneity, with most respondents placing themselves in the middle three categories. Also, CSP and PSD do not relate unambiguously to response styles, or dispositions, or positive outlooks. Instead, they consist of random error, unmeasured health indicators, and the subjective components of interest to us. One can imagine other indicators or different specifications that might be relevant. For example, we considered the possibility that the number of ADLs as well as the presence of any ADL might be an important distinction in health ratings. In models that included fewer health indicators than our current models, results supported this view; however, once we shifted to the full specification in model 3, the uniquely important component about ADLs for SRH was whether one or more were reported. Finally, cross-national research allows us to compare across national contexts, which includes language as a central feature. These comparisons necessarily rely on survey translations that aim at concept equivalence, but may not always achieve it (Viruell-Fuentes et al. 2011).

Extensions of this research can address these general questions within a longitudinal framework, using a larger number of countries, or by defining smaller geographic (and

potentially more homogeneous) units, since country and culture are not the same, and by investigating improved measures of subjective dimensions of health that might allow better tests and further refinement of these ideas. What these findings do suggest is that one reason SRH differs is that people invoke different approaches to this semantic exercise (Nettleton 1995). If we acknowledge that people's ratings reflect their overall judgment, then identifying the factors that shape rating behavior can provide us with a better understanding of interpersonal and group differences in SRH. That health conditions and disability indicators are important components of SRH supports the argument that SRH is a valid health measure, although differences in positional objectivity can affect both health and health information (Sen 2002).

Any evaluation may reflect a rating process that is both interpersonally shared and person-specific. These subjective components need not undermine the validity of SRH, since people occupying the same position may well evaluate conditions, symptoms, or limitations in different ways; or they may use different translational rules in choosing the appropriate adjective. In this study, we attempted to specify two subjective dimensions of the rating process—shared rating norms and deviations from those norms that are person-specific. We attributed some part of country differences in SRH to each of these dimensions, with country rankings becoming more compact when we did so. Still, differences remain.

SRH may reflect additional information or some aspect of unmeasured robustness or fragility in health that is perceived by the respondent but not reflected in the set of health indicators used in the models. Attempting to parse these schemes introduces considerable complexity, but may allow us to distinguish differences in what people are rating as well as how they make rating decisions. In doing so, we can productively explore ways to better specify the subjective elements of health appraisals as we expand our measures of the clinical elements of health status. In this way, we can increase our confidence in what self-rated health reports are telling us. This paper takes a small step in that direction.

## REFERENCES

- Ardila A, Ostrosky-solis F, Bernal B. Cognitive testing toward the future: The example of semantic verbal fluency (Animals). *International Journal of Psychology*. 2006; 41(5):324–332.
- Bambra, Clare. Health inequalities and welfare state regimes: theoretical insights on a public health 'puzzle'. *Journal of Epidemiological Community Health*. 2011; 65:740–45.
- Bayliss, Elizabeth A.; Eliis, Jennifer L.; Ann Shoup, Jo; Zeng, Chan; McQuillan, Deanna B.; Steiner, John F. Association of patient-centered outcomes with patient-reported and ICD-9-based morbidity measures. *Annals of Family Medicine*. 2012; 10:126–133. [PubMed: 22412004]
- Benyamini Y, Idler EL, Levental H, Levental EA. Positive affect and function as influences on self-assessments of health: expanding our view beyond illness and disability. *Journal of Gerontology Psychological Sciences*. 2000; 55B:107–116.
- Borawski EA, Kinney JM, Kahana E. The meaning of older adults' health appraisals: Congruence with health status and determinant of mortality. *Journal of Gerontology: Social Sciences*. 1996; 51B:S157–S170.
- Börsch, Supan A.; Jürges, H., editors. *Health, Ageing and Retirement in Europe Methodology*. Mannheim Research Institute for Economics of Ageing; Mannheim: 2005.
- Chapman GB, Liu J. Numeracy, frequency, and Bayesian reasoning. *Judgment and Decision Making*. 2009; 4:34–40.

- Cutler DM, Richardson E. Measuring the health of the U.S. population. *Brookings Papers on Economics Activity, Microeconomics*. 1997;217–71.
- DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. *Journal of Internal Medicine*. 2006; 12(3):267–275.
- Ferraro KF, Farmer MM, Wybraniec JA. Health trajectories: Long-term dynamics among black and white adults. *Journal of Health and Social Behavior*. 1997; 38:38–54. [PubMed: 9097507]
- Fournet N, et al. Evaluating short-term and working memory in older adults: french normative data. *Aging & Mental Health*. 2012; 16(7):922–930. [PubMed: 22533476]
- Grol-Prokopczyk, Hanna; Freese, Jeremy; Hauser, Robert M. Using Anchoring Vignettes to Assess Group Differences in General Self-Rated Health. *Journal of Health and Social Behavior*. 2011; 52:246–61. [PubMed: 21673148]
- Hennessey CH, Moriarty DG, Zack MM, et al. Measuring health-related quality of life for public health surveillance. *Public Health Rep*. 1994; 109(5):665–672. [PubMed: 7938388]
- Huijts T, Perkins JM, Subramanian SV. Political regimes, political ideology, and self-rated health in Europe: a multilevel analysis. *PLoS ONE*. 2010; 5(7):e11711. [PubMed: 20661433]
- Idler E, Hudson SV, Leventhal H. The meanings of self-ratings of health: A qualitative and quantitative approach. *Research on Aging*. 1999; 21(3):458–476.
- Idler, Ellen L.; Kasl, SL. Self-ratings of health: Do they also predict change in functional ability. *Journal of Gerontology*. 1995; 50B(6):S344–S353.
- Idler, Ellen L.; Benyamini, Yael. Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*. 1997; 38:21–37. [PubMed: 9097506]
- Jürges, Hendrik. True Health vs Response Styles: Exploring Cross-Country Differences in Self-Reported Health. *Health Economics*. 2007; 16:163–78. [PubMed: 16941555]
- Jürges, Hendrik; Avendano, Mauricio; Mackenback, Johan P. Are different measures of self-rated health comparable? An assessment in five European countries. *European Journal of Epidemiology*. 2008; 23(12):773–781. [PubMed: 18814040]
- Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified model. *Social Science & Medicine*. 1999
- Jylhä M, Guralnik J, Ferucci Luigi, Jokela Jukka, Heikkinen Eino. Is Self-Rated Health Comparable Across Cultures and Genders. *Journal of Gerontology: Social Sciences*. 1998; 53B:S144–152.
- Kaplan G, Baron-Epel O. What lies behind the subjective evaluation of health status? *Social Science and Medicine*. 2003; 56:1669–1676. [PubMed: 12639584]
- King, Gary; Murray, Christopher J. L.; Salomon, Joshua A.; Tandon, Ajay. Enhancing the Validity and Cross-Cultural Comparability of Survey Research. *American Political Science Review*. 2004; 98:191–207.
- Knäuper B, Turner PA. Measuring health: improving the validity of health assessments. *Quality of Life Research*. 2003; 12:342–352.
- Knoops, Kim T.B.; de Groot, Lisette C.P.G.M.; Kromhout, Daan; Moreiras-Varela, Olga; van Staveren, Wija A. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly european men and women: The HALE project. *JAMA*. 2004:1433–1439. [PubMed: 15383513]
- Kramers PG. The ECHI project: health indicators for the European Community. *European Journal of Public Health*. 2003; 13(3):101–106. [PubMed: 14533758]
- Mackenbach JP, Van Den Bos J, Joung IM, Van de Mheen H, Stronks K. The determinants of excellent health: Different from the determinants of ill-health? *International Journal of Epidemiology*. 1994; 23:1273–1281. [PubMed: 7721531]
- Manderbacka K, Lundberg O, Martikainen P. Do risk factors and health behaviours contribute to self-rating of health? *Social Science & Medicine*. 1999; 48:1713–20. [PubMed: 10405010]
- Mansur C, Amick BC, Harrist RB, Franzini L. Social capital, income inequality, and self-rated health in 45 countries. *Social Science & Medicine*. 2008; 66(1):43–56. [PubMed: 17913321]
- Miilunpalo S, Vuori Oja, Pasanen M, Urponen H. Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology*. 1997; 25(2):517–528. [PubMed: 9180644]

- Nettleton, S. The sociology of health and illness. Polity Press; UK: 1995.
- Salomon JA, Tandon A, Murray CJ. Comparability of self rated health: cross sectional multi- country survey using anchoring vignettes. *BMJ*. 2004;328–358.
- Sen, Amartya. Health: perception versus observation. *British Medical Journal*. 2002; 324:860–61. [PubMed: 11950717]
- Trichopoulos D, Lagiou P. Mediterranean diet and overall mortality differences in the European Union. *Public Health Nutrition*. 2004; 7(7):949–951. [PubMed: 15482623]
- Verbrugge LM, Jette AM. The disablement process. *Social Science and Medicien*. 1994; 38:1–14.
- Vignoli, Daniele; De Santis, Gustavo. Individual and contextual correlates of economic difficulties in old age in Europe. *Population Research and Policy Review*. 2010; 29:481–501.
- Viruell-Fuentes, Edna A.; Morenoff, Jeffrey D.; Williams, David R.; House, James S. Language of Interview, Self-Rated Health, and the Other Latino Health Puzzle. *Am J Public Health*. 2011; 101(7):1306–1313. [PubMed: 21164101]
- Ware JE Jr, Gandek B. Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. *Journal of Clinical Epidemiology*. 1998; 51(11):903–912. [PubMed: 9817107]
- WHO.. Health interview surveys: towards international harmonization of methods and instruments. Vol. 58. WHO Regional Publications for Europe; Copenhagen: 1996.
- Williams, Richard. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*. 2006; 6:58–82.
- Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995; 273(1):59–65. [PubMed: 7996652]

Country Rankings on Predicted SRH from Model 1	Country Rankings on Predicted SRH from Model 3	Country Rankings Based on Predicted SRH from Model 5
Switzerland *****	Switzerland	Switzerland *****
Denmark (Sweden)	Greece Belgium *****	Belgium Greece
Belgium *****	Denmark Spain (Sweden)	Denmark Netherlands (Sweden)
Netherlands Greece *****	Netherlands *****	*****
Austria France *****	France ***** (Italy)	France Spain Austria (Italy)
Germany Spain *****	***** Austria *****	*****
(Italy)	Germany	Germany

**Figure 1.**  
Comparison of Country Rankings Based on SRH.

*Note:* Ranking based on the estimated coefficients reported in Tables 2 and 4 as well as Wald difference tests. Star-line delimiters define the tiers of countries with similar average ratings of SRH. Sweden and Italy have nonproportional effects, so their placement is an approximation based on the relevant coefficients.

**Table 1**

	Austria (N=1,772)	Germany (N=2,810)	Sweden (N=2,850)	Netherlands (N=2,655)	Spain (N=2,086)	Italy (N=2,407)	France (N=2,733)	Greece (N=2,519)	Switzerland (N=879)	Belgium (N=3,492)
	1.9	2.3	1.8	0.7	2.7	2.0	1.6	1.1	0.5	1.5
	7.6	11.3	7.4	5.1	11.8	11.6	7.2	5.6	2.9	6.0
	30.3	32.3	26.5	25.1	32.3	38.9	28.9	29.6	16.1	23.7
	43.0	43.0	36.1	50.6	42.2	39.8	48.8	40.6	48.2	49.3
	17.2	11.1	28.2	18.5	11.0	7.7	13.5	23.1	32.3	19.5
	65.0 (10.2)	64.9 (10.1)	65.0 (10.8)	63.7 (10.1)	65.3 (10.5)	65.6 (10.2)	65.1 (10.7)	64.2 (9.9)	64.9 (10.8)	65.3 (10.2)
	2.9 (1.3)	3.3 (1.1)	2.6 (1.6)	2.6 (1.3)	1.6 (1.4)	1.7 (1.2)	2.2 (1.8)	2.0 (1.5)	2.6 (1.2)	2.7 (1.5)
	54.8	55.3	53.9	54.3	57.6	57.0	55.3	53.5	53.9	54.4
	58.1	61.8	64.8	70.0	65.4	64.1	65.1	68.7	66.8	70.0
	19.3	18.1	20.2	14.0	13.7	13.0	16.8	10.3	17.2	13.3
	22.6	20.1	15.0	16.0	20.9	22.9	18.1	21.0	16.0	16.7
	31,317 (29,183)	34,627 (37,553)	31,099 (24,017)	34,396 (30,726)	17,640 (23,292)	18,761 (20,801)	30,326 (33,692)	16,973 (17,226)	39,986 (35,920)	26,838 (33,879)
	148,529 (269,391)	163,127 (375,579)	138,927 (286,661)	216,282 (596,914)	240,385 (766,347)	180,678 (608,235)	269,590 (629,579)	147,302 (256,538)	415,935 (925,468)	254,149 (625,482)
	21.0	28.6	40.9	30.3	23.7	18.8	27.7	26.5	39.0	22.0
	21.6 (9.7)	19.7 (7.1)	22.8 (7.4)	19.4 (6.1)	14.7 (5.8)	13.5 (6.0)	19.4 (7.8)	14.3 (4.7)	19.9 (5.9)	19.3 (6.3)
	3.5 (2.1)	3.6 (1.9)	3.9 (2.0)	3.7 (2.0)	2.4 (1.8)	2.5 (1.9)	3.0 (1.9)	3.2 (1.8)	3.9 (2.0)	3.2 (2.0)
	4.5	4.4	1.8	3.7	20.0	13.1	10.2	4.6	2.6	5.2
	7.8	13.8	11.6	10.4	29.8	23.3	18.5	17.6	8.4	14.7
	87.7	81.8	86.6	85.9	50.2	63.6	71.3	77.8	89.0	80.1
	3.9 (1.0)	3.6 (1.0)	4.3 (0.9)	3.5 (1.0)	2.9 (1.2)	3.1 (1.2)	3.8 (1.2)	3.3 (1.2)	3.9 (0.9)	3.9 (1.0)
	3.8 (1.0)	3.4 (1.0)	4.2 (1.0)	3.4 (1.1)	2.7 (1.2)	2.9 (1.2)	3.5 (1.3)	3.1 (1.2)	3.7 (1.0)	3.6 (1.2)
	1.3 (1.4)	1.5 (1.6)	1.6 (1.7)	1.2 (1.5)	1.8 (1.9)	1.7 (1.7)	1.6 (1.7)	1.3 (1.5)	1.0 (1.2)	1.6 (1.6)

*J Health Soc Behav*. Author manuscript; available in PMC on 2015 December 03.



	Overall (N=25,736 <sup>2</sup> )	Denmark (N=1,533)	Austria (N=1,772)	Germany (N=2,810)	Sweden (N=2,850)	Netherlands (N=2,655)	Spain (N=2,086)	Italy (N=2,407)
CHRONIC EVER DEPRESSED (%)	1.6 (1.5)	1.6 (1.5)	1.3 (1.3)	1.5 (1.4)	1.5 (1.4)	1.3 (1.3)	1.7 (1.5)	1.8 (1.6)
Yes	27.4	25.8	17.8	23.2	30.1	30.8	31.8	27.1
ADL LIMITATIONS (%)	10.1	9.4	9.1	9.5	8.1	6.9	9.8	11.7
One or more	8.3	8.7	8.6	8.0	7.3	5.9	8.7	9.3
IADL LIMITATIONS (%)	1.5 (2.2)	1.2 (1.9)	1.6 (2.1)	1.5 (2.1)	1.2 (1.9)	1.2 (2.0)	1.9 (2.4)	1.8 (2.3)
One or more	56.4	54.4	51.9	48.5	55.0	55.4	58.8	58.6
FUNCTIONAL LIMITATIONS	29.9	32.5	34.0	34.9	30.4	25.0	36.7	28.3
SELF-RATED ACTIVITY RESTRICTIONS (%)	13.7	13.1	14.1	16.6	14.6	19.6	4.5	13.1
None	0.57 (0.48)	0.59 (0.48)	0.62 (0.48)	0.69 (0.50)	0.59 (0.47)	0.64 (0.50)	0.45 (0.36)	0.55 (0.48)
Moderate	-0.002 (0.53)	-0.001 (0.53)	0.002 (0.53)	-0.006 (0.54)	0.006 (0.57)	0.001 (0.63)	0.004 (0.46)	-0.002 (0.53)
Severe								
CSP								
PSD								

<sup>1</sup> Means and standard deviations (in parenthesis) for continuous variables, and percentages for categorical variables.

<sup>2</sup> This is the number of observations used for all subsequent estimations.

**Table 2**Exponentiated Coefficient from Generalized Logit Models for SRHealth<sup>1,2,3,4</sup> – Stage 1 results

	Model 1	Model 2	Model 3
<b>VERY BAD</b>			
Country (Denmark)			
Austria	0.668 ***	0.751 ***	0.590 ***
Germany	0.455 ***	0.567 ***	0.437 ***
Sweden	0.553 **	0.416 ***	0.378 ***
Netherlands	0.850 *	1.207 **	0.877
Spain	0.443 ***	1.141	0.974
Italy	0.372 ***	1.453	1.603 *
France	0.665 ***	0.881	0.829 **
Greece	0.849 *	1.634 ***	1.379 ***
Switzerland	1.900 ***	2.080 ***	1.437 ***
Belgium	0.941	1.194 *	1.222 **
Age	0.943 ***	0.999	1.031 ***
Education		0.938	1.045 **
Female		0.908 **	2.123 ***
Marital Status (Married/Partnership)			
Divorced/Separated/Never married		0.896	0.983
Widowed		0.889 *	1.079
ln(Income)		1.034 *	1.072 ***
ln(Worth)		1.347 ***	1.142
Employed		6.642 ***	3.226 ***
ln(Verbal fluency)		1.480 ***	1.712 ***
Memory		1.059 ***	1.026 *
Numeracy (Null)			
Low		1.280 **	1.273 **
Medium-High		1.549 ***	1.353 ***
Self-reported reading skills		1.240 ***	1.220 ***
Self-reported writing skills		1.285 ***	1.108 **
ln(Symptoms)			0.453 ***
Chronic conditions			0.884 **
Ever depressed (yes/no)			0.809 ***

	Model 1	Model 2	Model 3
ADL limitations (yes/no)			0.697 ***
IADL limitations (yes/no)			0.599 ***
Functional limitations			0.734 ***
Constant	264.154 ***	9.490 ***	46.190 ***
<b>BAD</b>			
Sweden	0.663 ***	0.487 ***	0.458 ***
Italy		1.128	1.061
Age		0.990 *	1.023 ***
Education		0.973	
Employed		<b>3.241</b> ***	2.165 ***
Female			1.815 ***
Chronic conditions			<b>0.804</b> ***
Constant	38.713 ***	1.667 ***	6.855 ***
<b>FAIR</b>			
Sweden	0.756 ***	0.574 ***	<b>0.565</b> ***
Italy		<b>0.830</b> *	<b>0.625</b> ***
Age		<b>0.983</b> ***	<b>1.012</b> ***
Education		1.028	
Employed		<b>2.204</b> ***	1.808 ***
Female			<b>1.357</b> ***
Chronic conditions			<b>0.625</b> ***
Constant	5.656 ***	0.258 ***	1.042
<b>GOOD</b>			
Sweden	<b>1.603</b> ***	<b>1.412</b> ***	<b>1.627</b> ***
Italy		<b>0.697</b> ***	<b>0.558</b> ***
Age		<b>0.967</b> ***	<b>0.991</b> *
Education		<b>1.081</b> ***	
Employed		<b>1.377</b> ***	<b>1.164</b> *
Female			<b>1.058</b> ***
Chronic conditions			<b>0.526</b> ***
Constant	0.512 ***	0.026 ***	0.087 ***
R <sup>2</sup>	0.0542	0.1142	0.2550
Log pseudolikelihood	-31488	-29490	-24803

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>
BIC	63169	59437	50156
N	25736	25736	25736

\*  
p<0.05

\*\*  
p<0.01

\*\*\*  
p<0.001

<sup>1</sup> All models include control variables for survey design effects and calibrated individual weights.

<sup>2</sup> We report robust standard errors that correct for both heteroskedasticity and clustering (Froot, 1989; Rogers, 1993).

<sup>3</sup> Bold-italics type indicates whether a specific coefficient is significantly different from the comparable shaded coefficient reported in the top panel.

<sup>4</sup> The coefficients of variables that do not meet the proportionality assumption are shaded gray.

**Table 3**

Exponentiated Coefficients for Generalized Logit Model for Self-Assessed Activity Limitations<sup>1,2,3,4</sup> – Stage 2 results

<b>NOT LIMITED</b>		<b>Coefficients</b>		
Age		1.003		
Education		0.933	***	
Female		0.831	***	
Marital Status (Married/Partnership)				
Divorced/Separated/Never married		1.114		
Widowed		0.914		
ln(Income)		0.950	**	
ln(Worth)		0.810	*	
Employed		0.610	***	
Ever depressed (yes/no)		1.232	***	
ADL limitations (yes/no)		1.876	***	
IADL limitations (yes/no)		2.311	***	
Constant		0.438	***	
<b>Interactions between Country and . . .</b>				
ln(Symptoms) (Denmark)		3.085	***	
Chronic conditions (Denmark)		1.349	***	
Functional Mobility Limitations (Denmark)		1.385	***	
Country (Denmark)		Country (Denmark)		
Austria		0.791		
Germany		1.066		
Sweden		1.154		
Netherlands		1.229		
Spain		0.499	***	
Italy		0.443	***	
France		0.565	***	
Greece		0.323	***	
Switzerland		0.919		
Belgium		0.506	***	
		Austria	0.862	1.203 *
		Germany	1.031	1.215 **
		Sweden	1.360 *	0.857 **
		Netherlands	1.235	0.985
		Spain	0.706 *	0.955
		Italy	0.564 ***	1.013
		France	0.836	0.903
		Greece	0.588 **	1.016
		Switzerland	0.774	0.865
		Belgium	1.045	0.983
				1.124 *
				0.955
				0.997
				1.001
				0.939
				1.037
				1.052
				0.982
				1.116
				1.001
<b>MODERATELY LIMITED (for nonproportional effects)</b>			Chronic conditions <sup>5</sup>	Functional Limitations <sup>7</sup>
Netherlands		2.204	***	

NOT LIMITED	Coefficients		
Spain	<b><i>0.134</i></b> ***		
France	<b><i>0.944</i></b>		
Belgium	<b><i>0.718</i></b> *		
Female	<b><i>0.576</i></b> ***		
ln(Symptoms) (Denmark)	<b><i>1.992</i></b> ***	Austria	<b><i>0.916</i></b>
		Germany	<b><i>0.900</i></b>
			<b><i>1.129</i></b> *
Constant	0.049***		
R <sup>2</sup>	0.2806		
Log pseudolikelihood	-17695		
N	25736		

\* p<0.05

\*\* p<0.01

\*\*\* p<0.001

<sup>1</sup> All models include control variables for survey design effects and calibrated individual weights.

<sup>2</sup> We report robust standard errors that correct for both heteroskedasticity and clustering (Froot, 1989; Rogers, 1993).

<sup>3</sup> Bold-italics type indicates whether a specific coefficient is significantly different from the comparable shaded coefficient reported in the top panel.

<sup>4</sup> The coefficients of variables that do not meet the proportionality assumption are shaded gray.

<sup>5</sup> Coefficients in this column estimate differences between the given country and Denmark in the effect of ln(symptoms) on the odds of reporting some (moderate or severe) limitations versus none.

<sup>6</sup> Coefficients in this column estimate differences between the given country and Denmark in the effect of Chronic conditions on the odds of reporting some (moderate or severe) limitations versus none.

<sup>7</sup> Coefficients in this column estimate differences between the given country and Denmark in the effect of Functional Mobility Limitations on the odds of reporting some (moderate or severe) limitations versus none.



**Table 4**Exponentiated Coefficient from Generalized Logit Models for SRHealth<sup>1,2,3,4</sup> – Stage 3 results

	Model 3	Model 4	Model 5
<b>VERY BAD</b>			
Country (Denmark)			
Austria	0.590 ***	0.591 ***	0.583 ***
Germany	0.437 ***	0.458 ***	0.440 ***
Sweden	0.378 ***	0.379 ***	0.373 ***
Netherlands	0.877	0.943	0.969
Spain	0.974	0.762 **	0.710 ***
Italy	1.603 *	1.367	1.411
France	0.829 **	0.749 ***	0.735 ***
Greece	1.379 ***	1.099	1.059
Switzerland	1.437 ***	1.391 ***	1.423 ***
Belgium	1.222 **	1.090	1.103
Age	1.031 ***	1.031 ***	1.031 ***
Education	1.045 **	1.032 *	1.043 **
Female	2.123 ***	1.933 ***	1.870 ***
Marital Status (Married/Partnership)			
Divorced/Separated/Never married	0.983	1.003	1.022
Widowed	1.079	1.061	1.057
ln(Income)	1.072 ***	1.059 ***	1.063 ***
ln(Worth)	1.142	1.094	1.093
Employed	3.226 ***	2.901 **	2.761 **
ln(Verbal fluency)	1.712 ***	1.698 ***	1.639 ***
Memory	1.026 *	1.026 *	1.027 *
Numeracy (Null)			
Low	1.273 **	1.284 **	1.308 **
Medium-High	1.353 ***	1.370 ***	1.397 ***
Self-reported reading skills	1.220 ***	1.223 ***	1.207 ***
Self-reported writing skills	1.108 **	1.108 **	1.108 **
ln(Symptoms)	0.453 ***	0.536 ***	0.529 ***
Chronic conditions	0.884 **	0.914 *	0.874 **
Ever depressed (yes/no)	0.809 ***	0.841 ***	0.845 ***
ADL limitations (yes/no)	0.697 ***	0.818 *	0.801 *

	Model 3	Model 4	Model 5
IADL limitations (yes/no)	0.599 ***	0.735 **	0.714 **
Functional limitations	0.734 ***	0.797 ***	0.790 ***
CSP		0.432 ***	0.360 ***
PSD			0.248 ***
Constant	46.190 ***	70.039 ***	143.808 ***
<b>BAD</b>			
Sweden	0.458 ***	0.455 ***	0.459 ***
Italy	1.061	<b>0.883</b>	<b>0.850</b>
Age	1.023 ***	1.024 ***	1.023 ***
Employed	2.165 ***	1.933 ***	1.847 ***
Female	1.815 ***	1.672 ***	1.642 ***
Chronic conditions	<b>0.804</b> ***	0.853 ***	0.815 ***
PSD			0.243 ***
Constant	6.855 ***	9.745 ***	17.655 ***
<b>FAIR</b>			
Sweden	<b>0.565</b> ***	<b>0.558</b> ***	<b>0.575</b> ***
Italy	<b>0.625</b> ***	<b>0.529</b> ***	<b>0.477</b> ***
Age	<b>1.012</b> ***	<b>1.013</b> ***	<b>1.013</b> ***
Employed	1.808 ***	1.643 ***	1.688 ***
Female	<b>1.357</b> ***	<b>1.280</b> ***	<b>1.304</b> ***
Chronic conditions	<b>0.625</b> ***	<b>0.668</b> ***	<b>0.666</b> ***
PSD			0.316 ***
Constant	1.042	1.405 *	1.650 **
<b>GOOD</b>			
Sweden	<b>1.627</b> ***	<b>1.592</b> ***	<b>1.642</b> ***
Italy	<b>0.558</b> ***	<b>0.493</b> ***	<b>0.475</b> ***
Age	<b>0.991</b> *	<b>0.991</b>	<b>0.991</b>
Employed	<b>1.164</b> *	<b>1.082</b>	<b>1.105</b>
Female	<b>1.058</b>	<b>1.011</b>	<b>1.028</b>
Chronic conditions	<b>0.526</b> ***	<b>0.555</b> ***	<b>0.565</b> ***
PSD			<b>0.418</b> ***
Constant	0.087 ***	0.117 ***	0.115 ***

	<b>Model 3</b>	<b>Model 4</b>	<b>Model 5</b>
R <sup>2</sup>	0.2550	0.2556	0.2928
Log pseudolikelihood	-24803	-24783	-23544
BIC	50156	50124	47688
N	25736	25736	25736

\*  
p<0.05

\*\*  
p<0.01

\*\*\*  
p<0.001

<sup>1</sup> All models include control variables for survey design effects and calibrated individual weights.

<sup>2</sup> We report robust standard errors that correct for both heteroskedasticity and clustering (Froot, 1989; Rogers, 1993).

<sup>3</sup> Bold-italics type indicates whether a specific coefficient is significantly different from the comparable shaded coefficient reported in the top panel.

<sup>4</sup> The coefficients of variables that do not meet the proportionality assumption are shaded gray