How Health Conditions Translate into Self Assessments:

A Comparative Study of Older Adults across Europe

Melissa Hardy

Francesco Aciai

Adriana Reyes

Department of Sociology
&
Population Research Institute
The Pennsylvania State University

Prepared for the 2013 Annual Meeting of the Population Association of America

April 11-13, 2013, New Orleans, Louisiana.

How Health Conditions Translate into Self Assessments:

A Comparative Study of Older Adults across Europe

Introduction

Self assessed health measures are frequently available in large national surveys and provide one method of comparing health status across time and place. These measures have been found to be strongly related to reported symptoms, diagnosed conditions, prospective mortality, and other indicators of health. At the same time, because respondents are making a subjective evaluation of their health by deciding where to place themselves in a set of predefined categories, group comparisons of self-reported health confound differences in the underlying health status with differences in how people scale their health relative to the response categories. A personal health assessment translates information known by the respondent into an ordered sequence of adjectives. That translation may have a cultural component, as well, which would be reflected in residual group differences. In this paper, we use the Survey of Health, Aging, and Retirement in Europe (SHARE) to examine country differences in self-reported health and the extent to which they can be explained by SES and demographic characteristics in conjunction with three categories of health and cognition measures: objective measures, which capture performance; self-reported measures, which require the respondent to report details of specific diagnoses, symptoms, or experiences; and self-assessed measures, which require the respondent to indicate relative severity. We also take advantage of an experimental design feature that randomly assigned respondents to groups who rate their health before versus after they answered a battery of questions about chronic conditions, symptoms, and activity limitations. This design feature allows us to investigate whether country differences in self reported health are affected when respondents are primed by their own accounts of the specifics of their health conditions.

Health, Health Reports, and Subjectivity

A major dilemma in studying health disparities across populations lies in the nature of health measurement. Population estimates are most accurate when they are based on large national samples, but

cost effective measures of health in these large samples generally must rely on information reported by the respondent or, in some cases, administrative data supplied by a health care provider. In either case, the health information is filtered through mediating mechanisms. Information provided by respondents includes symptomatology, activities, and diagnosed conditions. Reports of symptoms require respondents to describe what they are feeling by linking these feelings to words that capture both the nature of that experience and the degree of severity. Reports of activity limitations ask respondents to rate their ability to do certain things. Reports of diagnosed diseases tap more authoritative assessments by clinicians, although these reports also involve 'naming' various sets of symptoms according to diagnostic criteria. One difference in this latter category is the presumption that these diagnoses are based on both biomarker tests and reported symptoms, although the former might be performed in response to the latter. Because diagnosing diseases requires physician visits and laboratory tests, these reports are confounded with access to and utilization of health care. In countries with national health care, access may be less problematic, although the quality of the health care and the extent to which preventive health care is habituated may continue to differ even among these countries. The alternatives to these self reports and self assessments are clinical assessments, screenings, and biodata, all of which are far too expensive to perform on large samples. Studies that have compared these various types of data conclude that self-rated health is a significant predictor of health outcomes, such as mortality and morbidity (Benyamini, Idler, Levental, & Levental, 2000), chronic conditions, functional limitations, and symptoms (Idler & Kasl, Self-ratings of health: Do they also predict change in functional ability, 1995) (Verbrugge & Jette, 1994); that this connection holds in countries with both homogeneous and ethnically heterogeneous populations (Idler & Benyamini, Self-rated health and mortality: A review of twenty-seven community studies, 1997); and that it corresponds with epidemiological indicators such as physician visits (Miilunpalo, Vuori, Pasanen, & Urponen, 1997).

As an overall assessment of a multidimensional construct, self-rated health provides a relatively inclusive measure of general health, in that it covers physical, mental, and social aspects of health (Idler, Hudson, & Leventhal, 1999). However, as a summary measure, differences in self-rated health may

reflect differences in respondents' sensitivity to changes that occur or problems that arise along any of these dimensions. For example, those who rate their health at the low end of the scale seem to weight their current disease status more than those who report more positive health, and those who report better health considered risk factors and positive indicators as important. Functional status and vitality are important for everyone (Benyamini, Leventhal, & Leventhal, 2003). Also, the pattern of interaction between respondent and interviewer appear to correlate with self-reports of health. In particular, reports of 'excellent' health have been linked to extended interactions (e.g., exchanges prompted by inconsistencies in other health reports) and may suggest bias toward the upper end of the scale (Gabarski, Schaeffer, & Dykema, 2011).

Because age affects both health and how health may be evaluated (Groot, 2000), age may be reflected in ratings inconsistently across the response set; for example, older people may have wider boundaries for mid-range ratings of health and more restricted boundaries for extreme values of health. This question of differential boundaries may also apply across countries as a reflection of cultural factors or differences in comparative frameworks. A number of studies finding country differences in how self-rated health is reported have addressed this question with a small number of countries (Appels, Bosma, Grabauskas, Gostautas, & Sturmans, 1996) or with combinations of national and regional surveys (Bardage, et al., 2005), or included all ages (Olsen & Dahl, 2007), or with surveys that do not include additional detailed information on health. More recent studies have approached this issue through attempts at standardization. For example, one study attempts to match self-assessments of health to an index of disability constructed from more detailed health information (Jurges, 2007) for respondents across a number of European countries. Using prevalence of chronic conditions and physical health measures reported in the survey, Jurges constructs disability weights based on an ordered regression of these predictors on self-rated health. Comparing the distribution of self-rated health across countries after adjusting these ratings to a common translation of disability to health categories (as if all respondents assigned the same adjective for the same 'disability' value), Jurges concludes that Danes and Swedes tend to overrate their health, while Germans underrate, and Austrians and Greeks are consistent

with these 'typical' ratings. What is assumed in this approach is that those who share a given disability value also share the same 'true' underlying health status. However, standardizations like this one conclude that 'true' health has no subjective component; that similar conditions may be experienced in different ways, with different consequences of day-to-day living or that conditions and individual discomfort thresholds combine to create symptoms—these become components of measurement error. However, people may be aware of systemic changes that defy simply description

In this paper, we take a different approach. One reason respondents may offer different adjectives to describe similar health conditions may involve the frames they use for their responses. These framing differences can occur along a number of dimensions. First, some respondents may choose an adjective based on a general sense of how they 'feel,' while others may reflect on the various dimensions of their health and provide some kind of weighted summary. In the former case, reports may less reliably capture differences in their underlying health conditions. In the latter case, respondents are performing their own scaling exercise, matching their health specifics to the scale. If respondents are asked to perform this task in a consistent way and if this framing effect exists, then we should see a difference in health assessments between the two groups.

A second reason respondents may differ in their reports may be due to their understanding of the semantic differentials inherent in the scale. The reliability of the scale depends on a shared understanding of the meanings, and the meaningful differences, across categories. Those with better language skills may be better able to parse these distinctions. Similarly, respondents with better numeracy skills may be more sensitive to the implied underlying metric of these categories and think more carefully about this numeric frame of reference. Although these skills are likely to be correlated with education, assessments of respondents' facility with words and numbers provide additional dimensionality to these constructs.

Third, respondents from different demographic groups may have different reference groups they invoke to decide on the appropriate rating. Studies have shown that the scaling of health tends to be adaptive, with younger respondents assigning more negative ratings than older respondents in similar circumstances. But since we are looking only at those aged 50 and older, will age operate in the same

way. In other words, once we compare younger to older respondents with similar responses to the detailed health questions, will older respondents view their health more favorably? Studies have also reported more favorable health assessments from men and from the married (cite), which have been attributed to the higher levels of chronic illnesses among women and greater financial and social resources among the married.

Fourth, presumably respondents have additional information about severity or discomfort they experience, and these differences may be reflected in their health ratings as well. This subjective aspect of health ratings often creates unease, since we have few opportunities to validate these sensations and differentiate their severity. Pain scales are a good example of this conundrum. Patients are routinely asked to equate their level of pain either with a number (from none at all to unbearable) or to faces that express increasing levels of distress. People experience pain differently, and assumptions of standardized levels of pain that reflect group averages will be poor reflections of those on the fringes of the distributions. The better approach is to have an idea of how a person scales pain, to use vignettes for example; however, even this approach does not eliminate the problem.

Finally, people from different countries may have different reference groups they invoke to decide on the appropriate rating. Whether that manifests in country differences at the outset or whether these country differences are restructured as we parse the population will provide some insight into this phenomenon. Do country differences reflect compositional differences in respondents' socioeconomic status, age, cognitive performance, self-reported and self assessed measures of conditions, symptoms, and activity limitations, as well as verbal and numeric skill? Does this rank ordering reflect other accepted country differences in health, such as life expectancy? And if people in different countries are inclined to view their health in more or less positive terms even when these additional sources of health variation are considered, why might that be the case?

Data and Variables

We use data from the Survey of Health, Ageing and Retirement in Europe (SHARE) that includes survey results for eleven European countries in wave 1: Austria, Germany, Sweden, Netherlands, Spain, Italy,

France, Denmark, and Greece. Based on national probability samples, these data were collected in 2004 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 of the country. The data used in the analysis derive from wave 1 (2004) and includes 25,736 people. The overall response rate in wave 1 is 85.3%, with a range going from 73.7% in Spain, to 93.3% in France.

The initial survey includes two versions of the scale respondents may use to describe their health. The first version elaborates the lower end of the scale, allowing respondents to report 'very bad,' 'bad,' 'fair,' 'good,' and 'very good.' The second scale condenses lower categories and elaborates higher categories by allowing respondents to describe their health as 'poor,' 'fair,' 'good,' 'very good,' and 'excellent.' To avoid asking the two questions back-to-back, one scale is provided prior to the set of questions that ask respondents to self-report and self-assess other aspects of their health; the second scale is provided subsequent to this battery of particular health questions. Respondents are randomly assigned to two groups, which are distinguished by which version of the scale comes first.[1] Studies have found these two measures to be comparable (Jurges et al. 2007); however, others have argued that the European version is preferable in a cross-European context (WHO 1996; Murray et al. 2002), and thus has been commonly used (Verropoulou, 2009).

We include standard demographic variables such as age, gender, and marital status. Age is calculated from month and year of birth, and month and year of interview. The age range in our sample runs from 50 to 104 years old. The mean age is 65, with little variation between countries. In the analysis we subtract 50 to the original value of age and it around the grand mean. We use a dummy coded 1 for female respondents, and 0 for males. Women represent about 55% of the sample, with no significant differences across countries. We recoded the original 7 categories of marital status into 3 categories: married or in a partnership; divorced, separated, or never been married[2]; or widowed. Most people (about 65%) are married or in a partnership; almost 20% are widowed; while only about 15% are either divorced,

[1] We are attempting to unravel the scale/semantic differences in another paper.

[2] Never marrieds were added to this category after statistical tests indicated their similarity with the other statuses in this group.

separated, or single. Denmark, Sweden, and Switzerland have a slightly different distribution, with divorced, separated, or single more numerous than widowed.

In addition to demographic variables, we include measures of economic status and education that have been harmonized across countries. For the level of education, responses are arrayed on the 7-point ISCED (International Standard Classification of Education) scale, which is maintained by the United Nations Educational, Scientific and Cultural Organization[3]. The range is from 0 to 6, with 0 indicating no formal education, and 6 being the highest level of education (post-tertiary education). Throughout our analysis, education is also mean-centered. SHARE has generated a measure of household gross 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, we also use a PPP[4] index, which accounts for different currencies and costs of living across European countries. All monetary values are expressed in 2005 Germany Euros. Since income is collected at the household level, we also adjust for household composition using the square root to reflect the economies of scale in consumption (OECD 2006; OECD 2008; Jürges 2007; Vignoli and De Santis 2009) and then take the natural logarithm. Household net worth is generated as material and financial assets minus debt. We use the same adjustment here as we did for income, i.e., imputation, PPP values, the square root equivalence scale, and natural log. The final indicator in this set is employment status, included as a binary variable.

Performance-based Assessments. We have three cognitive measures that were scored on the basis of embedded tests. The *Verbal Fluency* score tallies the number of different animals the respondent can name in one minute. Ranging from 0 to 90, we use the natural log transform of this measure. *Memory* is based on the number of words the respondent can recall from a list read by the interviewer. Values range from 0 to 10. *Numeracy* measures mathematical performance and is created from 4 different questions testing respondents ability to calculate correctly. We defined three categories of

[3] For more information: http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

[4] Purchasing Power Parity

performance: none (no correct answer), low (only 1 correct answer), and medium-high (2 or more correct answers).

Self-reported Health Indicators. We have six variables in this set. *Symptoms* count the number of conditions (e.g., pain in your joints; chest pain; difficulty breathing; dizziness) experienced during the past six months. The range is from 0-12. *Chronic disease* is the number of doctor-diagnosed chronic diseases (from a card listing 12 possible diseases) reported by each person. *Depression* indicates whether the respondent has experienced depressive symptoms (read by the interviewer from a list of recognized symptoms) for a period longer than two weeks. Only respondents reporting no symptoms are coded '0'. *ADL limitations* describes the number of limitations with activities of daily living (ADL), and *IADL limitations* are the number of limitations with instrumental activities of daily living (IADL). We use binary version of both IADLs and ADLs. *Mobility limitations* indicate the number of limitations with mobility, arm function & fine motor function reported by each individual. Ten options are presented to the respondent, e.g., walking 100 meters; climbing several flights of stairs without resting; reaching or extending your arms above shoulder level), and the range is from 0 to 10.

Self-assessed health indicators. Our final three measures ask respondents to report their level of skill or the amount of difficulty they experience. *Self-assessed limitations* refer to respondents' reports of the extent they have been '*limited because of a health problem in activities people usually do'* over the past six months. Their response options are: severely limited; limited, but not severely; not limited. This question differs from the other questions about limitations as it does not specify what the possible limitations are; instead, respondents' own judgment is tapped. In every country but Germany, more than half of the respondents consider themselves not limited; about 30% belong to the middle category, with some variation between countries, and about 15% report severe limitations. Noteworthy, however, is that both Italy and Spain have a more positive distribution than the overall sample for this self-assessed measure of limitations, while they are rate worse on as far as the three more objective measures of limitations (ADL, IADL, and mobility) are concerned. *Self-assessed reading (writing) skills* ask respondents to rate their skill relative to those '*needed in their daily lives*' on a scale with five options:

Excellent; Very good; Good; Fair; Poor. In our analysis we use these variables as if they were continuous. Table 1 provides details on question wording and response categories for the full set of variables.

[Table 1 about here]

Approach

We use the self reported health status, European version, as our dependent variable. Because of its ordinal nature, we use ordered logistic regression to estimate our models. This approach estimates multiple thresholds and a single set of coefficients that assess the influence of the covariate on the odds of responses that traverse sequential thresholds. An underlying assumption of this model is that of proportionality. In other words, the effect of the covariate must be the same across all sequential comparisons in the response set. In tests of that assumption, we discovered that not all the independent variables met this assumption. Our final estimates are therefore generalized logistic regression coefficients using partial proportionality. The table of regression coefficients for out models therefore includes the proportional estimates (for those predictors meeting the assumption) following by scale-specific regression coefficients (for the remaining three comparisons) for variables which violate the assumption. In this way, generalized logit with partial proportionality is a hybrid approach, blending the efficiency of ordered logit with the flexibility of multinomial logit.

Results

Descriptive statistics for the variables used in this analysis are reported in Table 2. We include country specific statistics as well as those for the pooled sample. Our dependent variable—self-reported health---allows five categories ranging from 'very bad' to 'very good.' For all countries, 'good' is the model category, but the overall distributions are somewhat different. For example, the proportion reporting 'good' health ranges from slightly over half in the Netherlands to closer to one-third in Sweden, with about 40 percent in Spain, Italy, and Greece. The lower proportion in Sweden is complemented by a relatively large percentage of 'very good,' also the case in Greece, whereas Spain and Italy have relatively low proportions reporting 'very good' health and relatively high proportions reporting 'fair' and 'bad'

health. The 'flag' variable illustrates the equal split between those who report their health status before versus after the battery of more detailed questions through random assignment.

[Table 2 about here]

Demographic and socioeconomic status indicators also show some cross-country variation. On average, respondents are age 65, with somewhat more women than men in the samples. Average education (measured on the 7-point scale) is higher in Denmark and Germany and lower in Spain and Italy. Most respondents are currently married. Average income (adjusted per person) is highest in Switzerland, with average income in Greece less than half as high as among the Swiss. Similarly, accumulated wealth is also highest in Switzerland and lowest in Greece, but here the ratio is closer to 3:1. The highest rates of employment occur in Sweden and Switzerland (about 40%), and the lowest rates are in Austria and Italy (about 20%).

Assessments of cognitive health. These measures—Verbal fluency, memory, and numeracy—are scaled relative to the number of correct answers provided by the respondent (see Table 1 for additional details). One might expect that countries would be arrayed on these measures in much the same way as they are arrayed on education or income, and for some countries that is the case. For example, Spain and Italy are consistently at the lowest end of the distribution in both SES characteristics and on verbal fluency, memory, and numeracy. However, Germany, which ranks first in education falls mid-range on verbal fluency, memory, and numeracy, and Sweden ranks high on verbal fluency and memory, but mid-range on education and income.

Self-reported measures of health. We include symptoms, number of chronic illnesses, whether the respondent has experienced depressive symptoms, and the numbers of ADLs , IADLs, and mobility limitations (e.g., lifting or carrying something heavier than 10 pounds; climbing several flights of stairs) as self-reported measures, since respondents' reports were in reference to specific lists of conditions and activities. The average number of symptoms and the average number of chronic diseases reported was between 1 and 2; between 7 and 12 percent reported at least on ADL, and 4 to more than 9 percent

reported at least one IADL. Across all countries, about one-in-four respondents reported experiencing depressive symptoms, with the lowest proportion in Greece and the highest proportion in France.

Self-assessed measures of health. In this final set of variables, respondents are asked to report their assessments of their ability to do certain things. The self-assessed limitations ask them to rate the amount of difficulty they have doing everyday tasks. The self-assessed reading and writing asks them to rate their skills in these two areas. Reports of mobility restrictions seem most severe in the Netherlands and less severe in Greece, where more than 70 percent report no difficulties in performing day-to-day tasks; Spain has the smallest proportion of people who report severe limitations. Self-assessed reading and writing skills are similar, but not identical to earlier distributions on cognitive related tasks. Respondents from Sweden rate their skills highest, while respondents from Spain and Italy rate their skills lowest.

[Table 3 about here]

We turn now to the results of the logit models with partial proportionality.[5] Table 3 includes parameter estimates for 6 models beginning with the simplest model specifying only country differences and the second model adding the dummy variable for the design feature. As we move from left to right, the models include additional covariates, beginning with demographic and SES measures, then adding assessments of cognitive health, then self-reported health, then additional self-assessments of mobility limitations and reading/writing skills. Our interest is fourfold. First, to what extent does health status reflect socioeconomic differences, and do we continue to observe these stratification effects even when we have controlled for the full range of health specifics. Second, how is health status linked to cognitive performance as measured through performance and through self assessments, and are the self assessments redundant with the performance measures? Third do the three categories of health measures and the items within each set capture unique dimensions of self-assessed health and where we have different levels of measures for one concept (e.g., ability to function), are there sizeable overlaps or does each measure capture something distinct? Fourth, to what extent are the initially observed country differences in health

[5] Given the possibility that respondents are 'nested' within countries, we considered the possibility that the error structure should allow for within-country correlation. We tested this hypothesis and found that the results were consistent across the two approaches.

status preserved, reshaped, or eliminated as we progressively specify key dimensions of population heterogeneity? In this analysis, we used dummy variables for each country as proxies for country differences and used Denmark as the excluded category. Our starting point is the distribution included in Table 2. We list the independent variables in the left-most column. The first set of estimates describes how the odds of reporting 'very bad' to 'better' health differ as the value of the independent variable changes 1 unit. For variables that respect the proportionality assumption, this coefficient describes the net effect of the independent variable across the full range of comparisons: 'very bad' and 'bad' to 'fair,' 'good,' or 'very good;' 'very bad,' 'bad,' or 'fair,' to 'good' or 'very good;' or 'very bad' through 'good' relative to 'very good.' Given their proportional effects, we do not repeat these coefficients. However, for variables that do not meet the assumption for any of the models, we report the coefficients that are specific to each contrast. In some cases, the difference in these coefficients will be small, but in other cases, changes in the coefficient will provide us with additional information about the relationship. [6] Finally, these models assume that self-assessed health is a continuous latent variable and that respondents invoke thresholds to synchronize their perceived health with the response set. How these thresholds differ across groups and across countries is our interest.

When we look at just country differences in Model 1, we reproduce the within-country distributions in a different metric: here we compare the log-odds of worse or better health relative to Denmark's response distribution. Only Sweden and Greece violate the proportionality assumption, which means that the differences between the remaining countries and Denmark are consistent across the range of responses. In all countries except Switzerland, reported health status is worse than in Denmark. These differences are largest between Italy and Denmark, smallest between the Netherlands and Denmark. When we look more closely at Sweden and Greece, we see that at the lower end of the scale (comparing 'very bad' to better health) Germans are more likely to say 'very bad' whereas Swedes are not

[6] Unlike regular regression models, in which the error term acts as a residual term, with variance equal to unexplained variance in the dependent variable, logit models assign a value to the error variance; consequently, comparing coefficients across specifications is less straightforward. These adjustments will be included in the next draft; however, our results to date indicate that the comparisons we report here essentially unchanged.

significantly different from Danes. This pattern holds as we begin to shift the comparison farther up the scale, and when we compare 'fair' or worse health to 'good' or better health, Greeks and Swedes look about the same, both more likely to report worse than better health. When we look at the top of the scale, contrasting those reporting 'very good' health to lower ratings, Swedes are more likely than Danes to report 'very good' health, and Greeks do not differ from Danes.

When we add the dummy variable for the pre- post- design feature (Model 2), only Sweden violates the proportionality assumption, but the pattern of coefficients remains the same as in Model 1. Looking at the country coefficients for all but Greece and Sweden, we see they are slightly lower than in Model 1. The coefficient for Greece now indicates a consistent tendency toward more negative health reports, suggesting that Greeks were more sensitive to the placement of the question.

In model 3 when we add the demographic and SES variables, both Italy and Sweden display non-proportional effects here and in subsequent models. As we expect, older respondents report worse health; being married, employed, and more educated as well as having higher income and wealth are associated with reports of better health; and we see no gender differences. Remaining country differences are somewhat smaller, but the pattern has also changed. Greeks are now more likely to report better health than Danes. The pattern for the comparison of Swedes and Danes is the same, although the tendency toward reporting more negative health at the lower end of the distribution is smaller and the tendency to report 'very good' health is somewhat larger among the Swedes. Differences between Italians and Danes occur in the mid- and higher range of responses, with no difference in reports of 'very bad' health. The coefficient for the design variable is also a bit larger, with the odds of reporting better health 23 percent higher among those who reported their health status after the battery of specific questions.

Model 4 adds the assessments of cognitive health, all of which indicate that those in better cognitive health with higher verbal and numeracy skills report better health on average. Once we control for these cognitive dimensions, the effects of gender and age become non-proportional. Older ages are associated with lower likelihoods of reporting 'very good' health and higher likelihood of reporting 'bad' to 'good' health. Women are both less likely to report 'very bad' or 'bad' health, but also less likely to

report 'very good' health. As for country differences, we see that the contrast between Danes and Greeks is larger, with the odds of reporting better versus worse health about 50 percent higher for Greeks. Respondents from Belgium are also tilted toward better health than the Danes, but no difference between Danes and respondents from Spain, France, or the Netherlands.

When self-reported health indicators are included in Model 5, we again see an expected pattern, with those reporting more symptoms, more chronic illnesses, depression, ADLS, IADLS, and limitations in physical functioning reporting worse health, with the effect of reporting chronic diseases getting stronger as we move toward the positive end of the scale. Comparisons of country coefficients show a mixture of higher, lower, and similar coefficients to the previous model, with the design effect continuing to increase as we control for additional reported health conditions.

Finally, in model 6, we have additional self-assessed health items that rate reading and writing skills as well as difficulty in doing the sorts of things that 'people usually do.' Over the sequence of models, we see that some of the demographic and SES variables no longer register unique effects. For example, it appears that the advantages of marital status and wealth are mediated through the self-reported measures. What is more interesting is how the pattern of country differences changed as we controlled for respondent heterogeneity. When we looked at these differences net of SES, many of these differences were reduced, but by the time we controlled for the various dimensions of cognitive, physical, and emotional health, country effects were generally as large as were reported in the first model, the design effect was stronger, but there was no evidence of an interaction between the design effect and country.[7]

[Figure 1 about here]

In Figure 1, we illustrate these country differences in predicted health distributions using the results from the final model. Although we do not include a comparison of the pre- post- design effects, the odds of reporting better health is 36 percent higher for those responding after they were asked the detailed health battery compared to those who provided their assessments prior to this module.

[7] We tested for differential effects by country and found no evidence that was the case.

Discussion

We began our investigation hypothesizing that self assessed health may differ across respondents and among countries for a variety of reasons. The literature suggests that SRH corresponds to various health outcomes, thereby promoting the validity of the measure for capturing differences in underlying health. We add to this literature in three ways. First, we expand the range of health indictors to include self-reported, self-assessed, and a performance-based measure of memory and demonstrate that all of these various measure account for some unique variability in SRH. Although some of the effects of socioeconomic status are mediated through these health differences, we continue to see the stratifying effects of income and education on SRH. Second, we hypothesized that respondents may invoke different frames for reference when they rate their health. We suggested three bases for these different frames. First, we argued that people may differ in the amount of detail they regard as salient to an overall assessment, reporting how they generally 'feel' rather than trying to scale on the basis of conditions, symptoms, limitations, or illness. The design feature of SHARE in which half of respondents assess their health before they answer this battery of detailed health questions, with the remainder answer afterward, allows us to test whether the rehearsal of health details may lead to different outcomes. Indeed, it does. We observed a difference from the simplest models forward, and the impact of that rehearsal only strengthened as we controlled for additional covariates. In addition, this design effect operated consistently across the range of responses, and reminding respondents what was (and was not) a health problem for them produced improved health ratings. In contrast to what was previously reported, these older respondents do not appear initially biased toward more positive ratings. Instead, it appears that having been asking a variety of questions, many accompanied by lists of specific health problems they could report, leads respondents to rate their health better than those who responded without this immediate context.

A second factor we hypothesized as influential to their assessments was their facility with language. Higher levels of verbal fluency predict better health ratings, as does the ability to read and write. It appears that at least part of the initial effect of education is mediated through these skills, but in

the final model, all these skill factors—verbal fluency, numeracy, reading skill, and writing skill—make unique contributions to explaining how SRH differs across respondents.

Third, we suggested that countries may provide difference frames of reference that respondents invoke when they assess their health. The differences we initially observe remain, in somewhat reordered fashion, after we have controlled for demographic, SES, and health specifics. That self-assessment may be tapping another dimension of underlying health status is suggested by the performance of the set of independent variables that reflect mobility, functionality, and performance of day-to-day activities. We include four self-reported measures that capture aspects of this general domain: symptoms, ADLs, IADLs, and mobility. All are associated with lower health ratings. Then we add a fifth measure—a self-assessment of 'difficulty' in doing daily tasks. Distinguishing among those who report no versus moderate versus severe limitations adds to the predictive power of the model; coefficients for the self-reported measures are somewhat diminished, so there is some overlap in their connection to SRH. But each on also has a unique effect, suggesting that SRH may a composite of many different health dimensions.

//////////

References

Angel, R., & Cleary, P. D. (1984). The effects of social structure and culture on reported health. *Social Science Quarterly , 53* (3), 464-473.

Angel, r., & Gronfein, W. (1988). The use of subjective information in statistical models. *American Sociological Review , 28* (12), 1229-1238.

Angel, R., & Thoits, P. (1987). the impact of culture on the cognitive structure of illness. *Culture, Medicine and Psychiatry , 11*, 465-494.

Appels, A., Bosma, H., Grabauskas, V., Gostautas, A., & Sturmans, F. (1996). Self-rated health and mortality in a Lithuanian and a Dutch population. *Social Science and Medicine , 42*, 681-689.

Bardage, C., Pluijm, S., Pedersen, N., Deeg, D., Jylha, M., Noale, M., et al. (2005). Self-rated health among older adults: a cross-national comparison. *European Journal of Ageing , 2*, 149-158.

Benyamini, Y., Idler, E. L., Levental, H., & Levental, E. A. (2000). Positive afect and function as influences on self-assessments of health: expanding our view beyond illness and disability. *Journal of Gerontology Psychological Sciences , 55B,* 107-116.

borawski, E. A., Kinney, j. M., & Kahana, E. (1996). The meaning of older adults' health appraisals: congruence with health status and determinants of mortality. *Journal of Gerontology , 51B* (3), S157-S170.

Chipperfield, J. G. (1993). Incongruence between health perceptions and health problems. *Journal of aging and health , 5* (4), 475-496.

Groot, W. (2000). Adaption and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics , 19*, 403-420.

House, j. S., Lepkwski, J. M., Kinney, A. M., Mero, R. P., Kessler, R. C., & Herzog, R. A. (1994). The scial stratification of aging and health. *Journal of Health and Social Behaviors , 35*, 213-234.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior , 38*, 21-37.

Idler, E. L., & Kasl, S. (1991). Health perceptions and survival: Do global evaluations of health status really predict mortality? *Journal of Gerontology , 46* (2), 855-865.

Idler, E. L., & Kasl, S. L. (1995). Self-ratings of health: Do they also predict change in functional ability. *Journal of Geronotology , 50B* (6), S344-S353.

Idler, E., Hudson, S. V., & Leventhal, H. (1999). The meanings of self-ratings of health: A qualitative and quantitative approach. *Research on Aging , 21* (3), 458-476.

Johnson, R. J., & Wolinsky, F. D. (1993). The structure of health status among older adults: Disease, disability, functional limitations, and perceived health. *Journal of Health and Social Behavior , 34*, 105-21.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causese of disease. *Journal of Health and Social Behavior , Extra Issue*, 80-94.

Miilunpalo, S., Vuori, O., Pasanen, M., & Urponen, H. (1997). 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 , 25* (2), 517-528.

Olsen, K. M., & Dahl, S.-A. (2007). Health differences between European countries. *Social Science and Medicine , 64*, 1665-1678.

Verbrugge, L. M., & Jette, A. M. (1994). The disablement process. *Social Science and Medicine , 38*, 1-14.

Wolinsky, F. D., & Johnson, R. J. (1992). Perceived health status and mortality among older men and women. *Journal of Gerontology , 47* (6), S304-S312.

Zimmer, Z., Natividad, J., Lin, H.-S., & Chayovan, N. (2000). A cross-national esamination of the determinants of self-assessed health. *Journal of Health and Social Behavior , 41* (December), 465-481.

Table 1: Definitions of Independent Variables Included in the Study

	Question	Original Range	Type
Verbal fluency	Now I would like you to name as many different animals as you can think of. You have one minute to do this. Ready, go.	0-90	Performance based
Memory	TEN WORDS LIST LEARNING DELAYED RECALL. A little while ago, I read you a list of words and you repeated the ones you could remember. Please tell me any of the words that you can remember now?	0-10	Performance based
Numeracy	Mathematical performance - the higher the better (generated from cf012 - cf015)	1-5	Performance
Depression ever	Has there been a time or times in your life when you suffered from symptoms of depression which lasted at least two weeks? 1. Depression 2. Pessimism 3. Suicidal 4. Guilt 5. Sleep 6. Interest 7. Irritability 8. Appetite 9. Fatigue 10. Concentration 11. Enjoyment 12. Tearfulness	Yes/No	Self-reported

Symptoms	BOTHERED BY SYMPTOMS For the past six months at least, have you been bothered by any of the health conditions on this card? 1. Pain in your back, knees, hips or any other joint 2. Heart trouble or angina, chest pain during exercise 3. Breathlessness, difficulty breathing 4. Persistent cough 5. Swollen legs 6. Sleeping problems 7. Falling down 8. Fear of falling down 9. Dizziness, faints or blackouts 10. Stomach or intestine problems, including constipation, air, diarrhoea 11. Incontinence or involuntary loss of urine 12. Other symptoms, not yet mentioned	0-11	Self-reported

Chronic conditions	Number of chronic diseases reported by each individual Has a doctor ever told you that you had any of the conditions on this card? 1. A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure 2. High blood pressure or hypertension 3. High blood cholesterol 4. A stroke or cerebral vascular disease 5. Diabetes or high blood sugar 6. Chronic lung disease such as chronic bronchitis or emphysema 7. Asthma 8. Arthritis, including osteoarthritis, or rheumatism 9. Osteoporosis 10. Cancer or malignant tumor, including leukemia or lymphoma, but excluding minor skin cancers 11. Stomach or duodenal ulcer, peptic ulcer 12. Parkinson disease 13. Cataracts 14. Hip fracture or femoral fracture 97. Other conditions, not yet mentioned	0-12	Self-reported
ADLNO	It describes the number of limitations with activities of daily living (ADL). Six activities are included: Dressing, including putting on shoes and socks Walking across a room Bathing or showering Eating, such as cutting up your food Getting in and out of bed Using the toilet, including getting up or down	0-5	Self-reported

GALI limitations	For the past six months at least, to what extent have you been limited because of a health problem in activities people usually do? 1. Severely limited 2. Limited, but not severely 3. Not limited	3 cat	Self-rated
IADLNO	It describes the number of limitations with instrumental activities of daily living (IADL). Six activities are included: Using a map to figure out how to get around in a strange place Preparing a hot meal Shopping for groceries Making telephone calls Taking medications Doing work around the house or garden Managing money, such as paying bills and keeping track of expenses	0-6	Self-reported
Mobility	It corresponds to the number of limitations with mobility, arm function & fine motor function reported by each individual. Please tell me whether you have any difficulty doing each of the everyday activities on card 9. 1. Walking 100 metres 2. Sitting for about two hours 3. Getting up from a chair after sitting for long periods 4. Climbing several flights of stairs without resting 5. Climbing one flight of stairs without resting 6. Stooping, kneeling, or crouching 7. Reaching or extending your arms above shoulder level 8. Pulling or pushing large objects like a living room chair 9. Lifting or carrying weights over 10 pounds/5 kilos, like a heavy bag of groceries 10. Picking up a small coin from a table	0-10	Self-reported

SR reading skills	How would you rate your reading skills needed in your daily life? Would you say they are.... 1. Excellent 2. Very good 3. Good 4. Fair 5. Poor	1-5	Self-rated
SR writing skills	How would you rate your writing skills needed in your daily life? Would you say they are.... 1. Excellent 2. Very good 3. Good 4. Fair 5. Poor	1-5	Self-rated

Table 2: Descriptive Statistics for Dependent and Independent Variables by Country

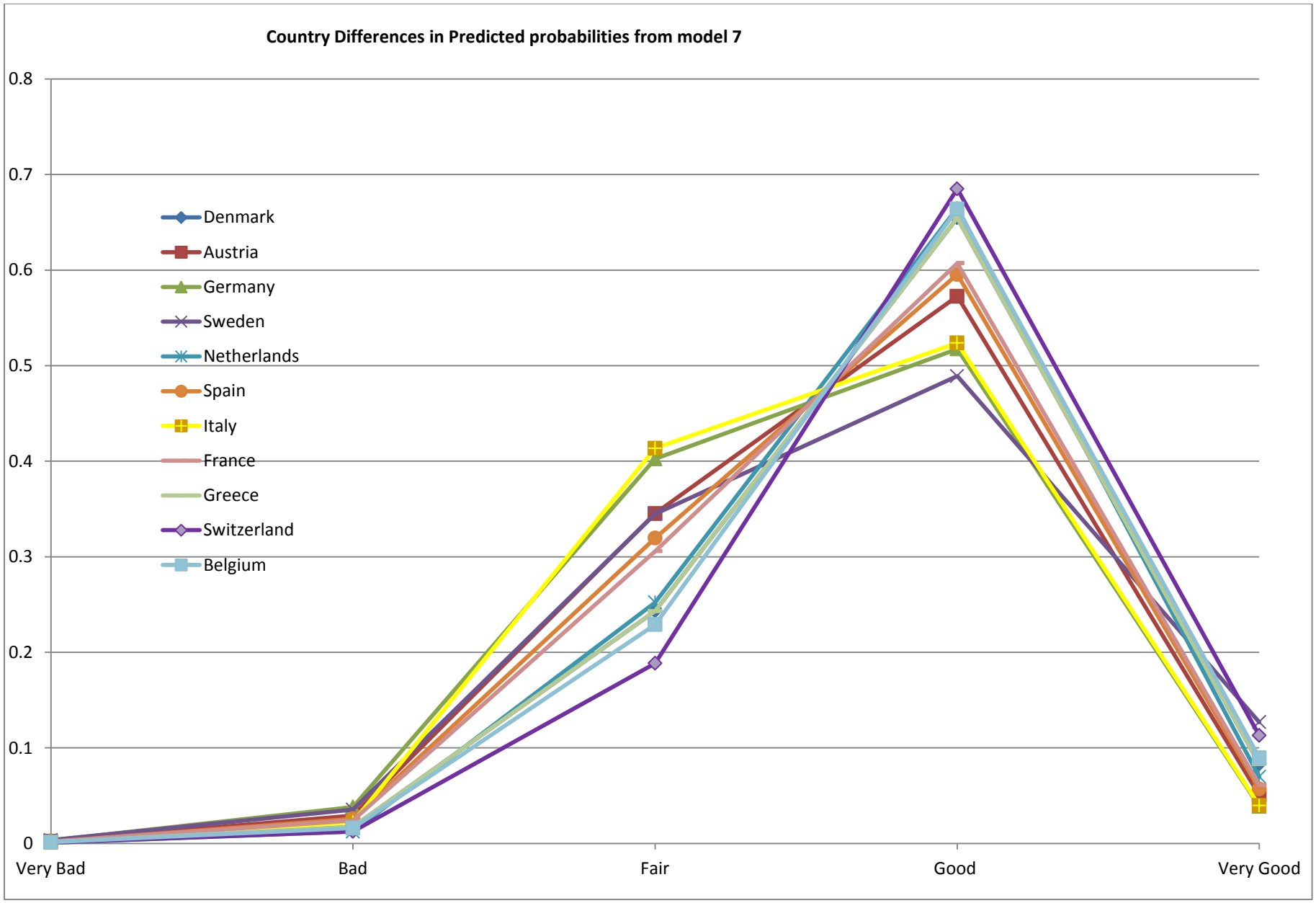
	All	Den	Aus	Ger	Swe	Neth	Spain	Italy	Fra	Gre	Swit	Belg
SRH_EU												
Very Bad	2.0	2.9	1.9	2.3	1.8	0.7	2.7	2.0	1.6	1.1	0.5	1.5
Bad	9.4	5.6	7.6	11.3	7.4	5.1	11.8	11.6	7.2	5.6	2.9	6.0
Fair	31.4	22.4	30.3	32.3	26.5	25.1	32.3	38.9	28.9	29.6	16.1	23.7
Good	43.8	44.0	43.0	43.0	36.1	50.6	42.2	39.8	48.8	40.6	48.2	49.3
Very Good	13.4	25.1	17.2	11.1	28.2	18.5	11.0	7.7	13.5	23.1	32.3	19.5
FLAG												
1 = After	49.6	50.6	49.2	49.9	49.7	50.5	49.7	48.4	49.7	50.4	49.9	51.9
AGE	65.0 (10.3)	63.9 (10.4)	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)
EDUCATION	2.4 (1.5)	3.2 (1.4)	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)
FEMALE	55.6	54.0	54.8	55.3	53.9	54.3	57.6	57.0	55.3	53.5	53.9	54.4
MARITAL STATUS												
Married/Partp	64.4	62.3	58.1	61.8	64.8	70.0	65.4	64.1	65.1	68.7	66.8	70.0
D/S/NM	15.8	20.7	19.3	18.1	20.2	14.0	13.7	13.0	16.8	10.3	17.2	13.3
Widowed	19.8	17.0	22.6	20.1	15.0	16.0	20.9	22.9	18.1	21.0	16.0	16.7
INCOME	27,770	31,927	31,317	34,627	31,099	34,396	17,640	18,761	30,326	16,973	39,986	26,838
WORTH	205,604	172,178	148,529	163,127	138,927	216,282	240,385	180,678	269,590	147,302	415,935	254,149
EMPLOYED												
Yes	26.3	38.2	21.0	28.6	40.9	30.3	23.7	18.8	27.7	26.5	39.0	22.0
VERBAL FLUENCY	17.8 (7.4)	21.4 (6.9)	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)
MEMORY	3.1 (2.0)	4.1 (1.9)	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)
NUMERACY												
Null	8.9	4.6	4.5	4.4	1.8	3.7	20.0	13.1	10.2	4.6	2.6	5.2
Low	18.0	13.7	7.8	13.8	11.6	10.4	29.8	23.3	18.5	17.6	8.4	14.7
Medium-High	73.1	81.7	87.7	81.8	86.6	85.9	50.2	63.6	71.3	77.8	89.0	80.1
SYMPTOMS	1.6 (1.6)	1.5 (1.7)	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)
CHRONIC #	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)	1.6 (1.4)	1.4 (1.4)	1.1 (1.2)	1.7 (1.5)

EVER DEPRESSED												
Yes	27.42	25.82	17.75	23.19	30.14	30.83	31.76	27.09	33.79	14.9	21.45	32.9
ADL LIMITATIONS												
Yes	10.1	9.4	9.1	9.5	8.1	6.9	9.8	11.7	11.6	6.8	6.7	11.6
IADL LIMITATIONS												
Yes	8.3	8.69	8.61	7.95	7.27	5.9	8.74	9.32	9.2	5.62	4.17	8.42
MOB 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)	1.5 (2.1)	1.5 (2.0)	0.9 (1.6)	1.4 (2.1)
SR LIMITATIONS												
None	56.4	54.4	51.9	48.5	55.0	55.4	58.8	58.6	61.2	70.6	65.7	61.1
Moderate	29.9	32.5	34.0	34.9	30.4	25.0	36.7	28.3	23.6	23.6	25.0	24.4
Severe	13.7	13.1	14.1	16.6	14.6	19.6	4.5	13.1	15.3	5.8	9.3	14.5
SR READING SKILLS	3.6 (1.1)	4.0 (1.0)	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)
SR WRITING SKILLS	3.5 (1.2)	3.8 (1.2)	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)

Table 3: Coefficients for Logit Models with Partially Proportionality

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
VERY BAD						
Country (Denmark)						
Austria	-0.446***	-0.441***	-0.246**	-0.269***	-0.515***	-0.533***
Germany	-0.792***	-0.790***	-0.779***	-0.720***	-0.915***	-0.809***
Sweden	-0.596***	-0.594***	-0.425*	-0.634***	-0.830***	-0.942***
Netherlands	-0.131	-0.131	0.012	0.036	-0.217**	0.686*
Spain	-0.832***	-0.831***	-0.280***	-0.006	-0.153	-0.405***
Italy	-1.016***	-1.011***	-0.026	0.297	0.239	0.333
France	-0.455***	-0.453***	-0.139*	-0.065	-0.165*	-0.336***
Greece	-0.172*	-0.169*	0.263***	0.401***	0.232**	0.001
Switzerland	0.546***	0.549***	0.734***	0.706***	0.354***	0.338***
Belgium	-0.136*	-0.137*	0.121	0.180**	0.225**	0.083
Flag		0.160***	0.205***	0.201***	0.265***	0.307***
Age			-0.036***	-0.008	0.028***	0.034***
Education			0.238***	0.140***	-0.026	0.043**
Female			-0.038	0.349*	0.758***	0.570***
Marital Status (Married/Partnership)						
Divorced/Separated/Never married			-0.202***	-0.134*	-0.028	0.015
Widowed			-0.198***	-0.149**	0.074	0.05
ln(Income)			0.065***	0.048**	0.079***	0.059***
ln(Worth)			0.494***	0.404***	0.162	0.07
Employed			1.722***	1.930***	1.231***	0.939**
ln(Verbal fluency)				0.549***	0.602***	0.473***
Memory				0.081***	0.038***	0.027*
Numeracy (Null)						
Low				0.377***	0.315***	0.248**
Medium-High				0.688***	0.455***	0.312***
ln(Symptoms)					-0.810***	-0.609***
Chronic conditions					-0.131**	-0.107*
Ever depressed (yes/no)					-0.216***	-0.160***
ADL limitations (yes/no)					-0.350***	-0.162*
IADL limitations (yes/no)					-0.580***	-0.234*
Mobility limitations					-0.323***	-0.214***
Self-reported limitations (None)						
Moderately limited						-1.284***
Severely limited						-2.880***
Self-reported reading skills						0.187***
Self-reported writing skills						0.102**
Constant	4.604***	4.525***	4.802***	3.467***	4.809***	5.257***
BAD (only nonprop Bs)						
Sweden	-0.440***	-0.439***	-0.278**	-0.464***	-0.640***	-0.750***
Italy			-0.270**	0.052	-0.124	-0.211

Employed			1.086***	1.189***	0.820***	0.571***
Age				-0.015***	0.020***	0.025***
Female				0.137*	0.621***	0.474***
Education					0.011	
Chronic conditions					-0.219***	-0.170***
Netherlands						0.363**
Severely limited						-2.616***
Constant	2.728***	2.649***	2.950***	1.832***	2.894***	2.961***
FAIR (only nonprop Bs)						
Sweden	-0.348***	-0.347***	-0.183*	-0.324***	-0.428***	-0.562***
Italy			-0.514***	-0.251**	-0.532***	-0.780***
Employed			0.752***	0.805***	0.607***	0.508***
Age				-0.020***	0.012***	0.013***
Female				-0.020	0.363***	0.255***
Education					0.110***	
Chronic conditions					-0.468***	-0.396***
Netherlands						-0.017
Severely limited						-1.996***
Constant	0.934***	0.853***	1.060***	0.076	0.989***	0.606***
GOOD (only nonprop Bs)						
Sweden	0.377***	0.379***	0.581***	0.514***	0.603***	0.472***
Italy			-0.662***	-0.446***	-0.601***	-0.785***
Employed			0.486***	0.352***	0.140*	0.113
Age				-0.036***	-0.009	-0.008
Female				-0.149**	0.104	0.024
Education					0.151***	
Chronic conditions					-0.647***	-0.578***
Netherlands						-0.192*
Severely limited						-1.258***
Constant	-1.313***	-1.397***	-1.362***	-2.050***	-1.464***	-2.098***


Country Differences in Predicted probabilities from model 7
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Very Bad
Bad
Fair
Good
Very Good
Denmark
Austria
Germany
Sweden
Netherlands
Spain
Italy
France
Greece
Switzerland
Belgium