3.3 Cross-Country Differences in General Health
Hendrik Jüges

Introduction

This contribution looks at variations in self-assessed general health between SHARE countries. Self-reports of health have proved to be useful indicators of an individual's health, for example as predictors of mortality (Idler and Benyamini 1997). However, the comparability of self-reported measures across groups of individuals has been questioned in a number of studies (e.g. Groot 2000; Sen 2002; Lindeboom and Van Doorslaer 2003). One major concern with self-assessed health is that respondents do not perceive the health self-assessment scale given to them as absolute. Individuals with the same true health status may have different reference levels against which they judge their health. For instance, respondents may be likely to report "very poor" health only if they feel they are much less healthy than others of the same sex, age, education, or income. A common finding is that older respondents tend to have a "milder" view of their health, i.e. they tend to rate their health as better than otherwise comparable younger respondents (Groot 2000; Van Doorslaer and Gerdtham 2003). Thus self-reported health of young and old respondents may not be directly comparable, and the observed decline in self-reported health by age may underestimate the decline in true health. In fact, the effect of changing reference levels seems to be so strong that it is taken into account in some formulations of the self-assessed general health question (used e.g. in the BHPS): respondents are explicitly asked to self-report their health relative to other people of their own age.

In cross-cultural studies like SHARE, there are additional concerns. Respondents from different countries and cultures may not only have different reference levels of health, but response categories may also have different connotations. Self-reported health categories are verbal representations of different health states, which may not mean the same thing to all respondents. For instance, "excellent" is a term that is used in everyday parlance in the Anglo-Saxon world, but Germans would often consider "ausgezeichnet" as an ironic exaggeration, in particular if used in the context of health. A comparison of self-reported general health across countries has to take such differences in habitual language use into account.

Figure 1 shows the proportion of SHARE respondents who report to be in very good or excellent health by country. According to their subjective assessment, the healthiest elderly live in Denmark, Sweden, and Switzerland. Nearly 50% of the Danish respondents and more than 40% of the Swedish and Swiss respondents report to be in very good or excellent health. The least healthy are French, German, Italian, and Spanish elderly. The proportion of respondents in very good or excellent health is at around 20% and thus only about half as large as in the healthy countries.

In the light of these large cross-country differences, it is natural to ask if they can be taken at face value. If we find that Danes are much more likely to report excellent health than Germans, does that mean that Danes are really that much more healthy than Germans? Or are they simply more likely to report excellent health, even if they have about the same true level of health? Because much of the added value of SHARE comes from multi-disciplinary, cross-country comparisons, many of our future analyses depend on the existence of a good comparable summary measure of the respondents' overall health. SHARE has been especially designed to produce such a measure.
The purpose of this contribution is to show how SHARE can be used to compute a health measure that is adjusted for possible cross-cultural bias of the kind described above. We also demonstrate the usefulness of this adjustment in a simple policy example where we study the cross-national relationship between health care expenditures and self-reported health. The example shows clearly that self-reports taken at face value can produce spurious results.

In addition what is presented in this contribution, SHARE offers a second, complementary way to purge our data from cross-country reporting bias. We have collected information on anchoring vignettes from a subset of respondents. Anchoring vignettes are short descriptions of people in different states of health which respondents are asked to rate on the same scale as they are asked to rate their own health (see King et al. 2004). Comparison of vignette ratings and own ratings will allow to correct each respondents' self-assessment for possible effects of response styles.

To ensure comparability with a large number of other surveys, SHARE contains two different versions of the self-reported health question. Both are 5-point scales. One ranges from „excellent“ to „poor“ (used e.g. in the U.S. Health and Retirement Survey), the other ranges from „very good“ to „very poor“ (used e.g. by WHO in numerous studies). To ease the exposition, we concentrate on the former version. Moreover, for the text and Figures, we treat the scale as dichotomous (very good or better versus good or worse health). However, the analysis described below was made using the original 5-point scale. Detailed results are shown in Tables 3A.9-3A.11 in the Appendix to this chapter.

Combining SHARE Health Data in a Single Index

The basic assumption underlying our analysis is that there is such thing as a „true“ and comparable health status. This implies that one must be willing not to accept the respondent's own judgements as absolute (Sen 2002). Conceptually, we consider true health as a continuous, latent (i.e., unobservable) variable. When respondents answer survey questions about their health, they assess their true health (possibly with measurement error; see Crossley and Kennedy 2002) and project this value onto the scale provided. Equivalent econometric formulations are the ordered logit or probit models. Differences in language use that affect the relationship between true health and self-assessed health
can be interpreted as differences in the so-called thresholds or cutpoints between adjacent health categories.

As described in detail in other contributions of this volume, SHARE contains a wide array of information on health problems: self-reported diagnosed chronic conditions, mental problems, physical symptoms (especially pain), or functional limitations. We also perform measurements and tests like grip strength, gait speed, and various cognitive tests. We use all available information in SHARE to compute a continuous health index for each individual. The idea of this index is to combine in a single number not only the prevalence of a large variety of conditions and limitations but also the effect of these conditions and limitations on the respondents’ health. The health index is scaled such that it has a value of 0 for the respondent with the worst observed health and a value of 1 for respondents without any conditions, symptoms, or limitations ("perfect health"). The presence of a condition or limitation reduces the value of the index by a specific amount. This amount (called disability weight) differs between conditions and symptoms and reflects their effect on health. For instance, Parkinson’s disease has a larger weight than diabetes. The weights are assumed to be the same for each respondent (and hence the same across countries). We use disability weights that are specific to the SHARE population.

Figure 2 shows the distribution of the health index by country. The countries are sorted by the median value, shown as a circle, with the most healthy country (Switzerland) on the left and the least healthy country (Spain) on the right of the graph. The upper and lower bars indicate the 90th and the 10th percentiles of the health distribution, respectively. Health inequality (measured by the ratio of the 90th to 10th percentile) is largest in Spain (1.77) and smallest in Switzerland (1.38). It is interesting to compare Figures 1 and 2. First, there are some changes in the countries’ ranks. For instance, Sweden drops from 2nd to the 7th, while the Netherlands rises from 6th to 3rd rank. Both countries are now in a larger group with very similar median health.

![Figure 2: Distribution of standardised health index, by SHARE country](image-url)
Cross-Country Differences in Response Styles

The next step of the analysis is to relate the health index values to the respondents' self-reported health levels. The idea is that each individual reports very good or excellent health only if his or her health index value passes a specific threshold value. In other words, these thresholds indicate how healthy respondents must be in order to state that they are, say in very good rather than in good health. As mentioned in the introduction, it is possible that this threshold varies systematically with the respondents' characteristics, for example age. Here, we are specifically interested in cross-country variations in thresholds. We compute country-specific reporting thresholds as the exact quantiles of the country-specific health index distribution that correspond to the proportion of respondents that report up to a specific health level. For example, 48.5% of all Danish respondents reported to be in very good or excellent health. The Danish reporting threshold between "good" and "very good" is thus computed as the 48.5th percentile of the Danish health index distribution, which is .77. Consider Germans as another example. Only 21.1 percent of them reported to be in very good or excellent health. The 21.1th percentile of the German health index distribution is .84. Germans need to be much healthier than Danes to claim that they are in very good health. In terms of disability weights, the difference is about one half heart attack.

The results for all countries are shown in Figure 3, ranked according to their computed good-to-very-good threshold. Figure 3 reflects differences in reporting styles across SHARE countries and can be used to predict the self-reported health level of a respondent of a specific health index in each SHARE country. For example, someone with a health index value of .79 would be predicted to report very good or better health in Denmark or Sweden but good or worse health in all other SHARE countries.

![Figure 3](image)

*Figure 3* Health index cutpoints between good and very good health, by SHARE country

Self-Reported and Adjusted Health Levels

Given the health index and the reporting thresholds, it is straightforward to compute adjusted distributions of self-reported health. We simply need to use the same thresholds for each respondent. This could be some arbitrary value (such as .79, as we just used to explain in Figure 3), some specific country's value or the (unweighted) average across all
countries. Here, we use the SHARE average, that is we compute which health level a respondent would report, given his or her health index, if he or she behaved like the average SHARE respondent. Specifically, each respondent is assigned to very good or excellent health if his or her health index is above .805, and to good or worse health if below .805.

Figure 4 compares self-reported health levels with adjusted health levels. The x-axis shows the proportion of respondents in very good or better health given their country-specific reporting style. The y-axis shows the proportion of respondents in very good or better health if everyone showed the same reporting behaviour. Respondents in countries to the left of the 45°-line systematically undervalue their health compared to the SHARE average, respondents in countries to the right systematically overvalue their health. Considering what we have already seen above in Figure 3, the results are not surprising. Scandinavians have a more positive attitude towards their health. Germans, Dutch, and the Swiss are less positive. In the remaining countries (Mediterranean and Austria), differences between reported and adjusted health levels are unsystematic.

To illustrate what our adjustment of self-reported values achieves, consider again Denmark and Germany. Although there are huge differences in the distributions of self-reported health between Danes and Germans (nearly 27.4 percentage points), the difference in adjusted health levels are negligible (1.2 percentage points), and probably much more realistic. However, accounting for different response styles does not equalise all health differences. For instance, Spain remains at the bottom of the health distribution. Cross-country differences in self-assessed health thus partly reflect variations in reporting thresholds, but the data do also suggest some real between-country differences in physical health.

**Demonstrating the Value of SHARE: A Simple Policy Example**

Let us finally show the value of our adjustment for different reporting styles in a simple policy example. One of the major strengths of SHARE is the cross-country dimension of the data, which allows to exploit international differences in institutions for policy analyses. Let us assume that we are interested in the relationship between health care expenditures
and some simple but comprehensive measure of health. We start the analyses by looking at the relationship between health care expenditures in 2003 (as percentage of GDP – the data are taken from the OECD) and the proportion of elderly who are in very good or excellent self-assessed health (see left Panel in Figure 5). It appears as if there is no clear relationship between health expenditures and health outcomes. It might be positive, but very weakly. The picture changes if we consider our corrected self-reported health measure (see right Panel in Figure 5). The relationship between expenditures and health becomes positive. Linear regression analysis suggests that a one percentage point increase in health care expenditures is associated with a (statistically significant) 4.2 percentage point increase in the proportion of very healthy respondents. This result is robust in the sense that dropping any single (supposedly influential) country from the analysis does not change our finding that health care expenditures are more positively related to a health measure that is adjusted for differences in reporting styles. Of course, this simple example cannot replace a full-blown policy analysis, and it clearly cannot tell us whether 1% of GDP are well spent when it increases the proportion of very healthy elderly by 4.2 percentage points. However, it shows that a correction for cross-national differences in reporting styles does affect results significantly. One of SHARE's main assets is that it allows such corrections and prevents spurious policy conclusions.

**Summary**

This contribution looks at differences in self-reported health across SHARE countries and corrects these differences for differences in reporting styles. The main results are:

- Self-reported general health shows large cross-country variations. According to their self-reports, the healthiest respondents live in the Scandinavian countries and the least healthy live in Southern Europe.

- These differences are only partly reflected by differences in true health (measured by the prevalence of chronic conditions, by functional limitations, and objective health measures such as grip strength, and walking speed).
- Another part of the cross-country variation in self-reported health must be attributed to differences in reporting styles. The SHARE data allow to compute health measures that take differences in reporting styles into account. Such comparable measures are a necessity in cross-national, multidisciplinary analyses.

- If differences in reporting styles are taken into account, cross-country variations in general health are reduced but not eliminated.

It should be noted that longitudinal data will greatly benefit the kind of analysis presented in this contribution. First, we will be able to study changes in self-reported health at the onset of chronic diseases cross-nationally. Second, since we will be able to study the relationship of self-rated health and mortality cross-nationally. This will both significantly improve our understanding of the determinants of self-assessed health in different countries.

References