

A comparison of different measures of health and their relation to labor force transitions at older ages

Arie Kapteyn¹ and Erik Meijer
University of Southern California

May 5, 2013

Abstract

We discuss three indexes of health that have been proposed in the literature and compare their theoretical and empirical properties, using data from the Survey of Health, Ageing and Retirement in Europe. We then estimate regression models for labor force transitions at older ages and investigate the consequences of using different measures of health. The measures differ both in the underlying statistical model and in the variables included in them. A comparison of distributional properties exhibits marked differences in kurtosis and skewness, while the correlation between the measures is modest. When using the health indexes to explain transitions into retirement, their explanatory power does not appear to vary much. However when explaining transitions into disability the index proposed by Poterba, Venti, and Wise (2011, 2013) provides a better fit than the competing indexes. It appears that this is mainly due to the variables included in the Poterba, Venti, and Wise index, such as health conditions and health care utilization variables, which are either wholly or partly missing from the other indexes.

1. Introduction

Health, be it of an individual or of a population, can be characterized by a large number of indicators. For many purposes it is desirable to summarize the indicators by a limited number of indexes, possibly only one. Several health indexes have been proposed in the literature, varying in statistical methodology and the breadth of variables included in the index. Since health indexes may be constructed for different purposes, there is no need to settle on one preferred index. We are interested in the statistical properties of health indexes and how they are related to observed economic behavior. Health indexes can be classified by a number of different approaches. The simplest approach is to simply ask people to rate their own health on an ordinal scale. A more involved approach relates such self-reports to a number of explanatory variables, such as health conditions or difficulties with activities of daily living and uses a regression approach to determine the weights of each of the explanatory variables in the construction of a health index. A third approach considers health to be a latent construct for which a number of indicators exist. The indicators are then used to somehow estimate the underlying latent variable. In this paper we aim at an empirical comparison of these approaches by taking prominent examples from the literature and comparing them among themselves as well as how well they explain economic behavior.

¹ Corresponding author: Arie Kapteyn, Center for Economic and Social Research, University of Southern California, 12015 Waterfront Dr, Playa Vista, CA 90094, kapteyn@usc.edu. This paper was prepared for the NBER Conference on the Economics of Aging, May 9–11, Carefree, AZ.

We focus on health measures that can be computed for individuals in widely available survey data. The data we use for the empirical part of this paper are from the Survey of Health, Ageing and Retirement in Europe (SHARE), which is modeled after the Health and Retirement Study (HRS) in the U.S. SHARE is a representative sample of individuals 50 and over and their spouses in a number of countries in continental Europe. We use data from waves 1 (2004/2005) and 2 (2006/2007), which include 12 and 14 countries, respectively. We limit ourselves to the 11 countries that are present in both waves. The data contain extensive information about health, as well as many other topics of interest to economists and other social and behavioral scientists. The traditional health measure is self-reported general health (SRH), which has five categories: excellent, very good, good, fair, and poor.² SRH generally correlates strongly with objective measures of health in (rare) instances when both are available for the same individuals. It is a short and easy question and as a result widely available in many data sets. Because of properties and availability, it is generally considered a useful measure for many purposes. However, it is also a crude measure, and perhaps more importantly, it appears to be incomparable across countries without corrections: individuals with similar health as judged by more objective variables give widely different responses on average in different countries (Meijer, Kapteyn, and Andreyeva, 2011; MKA hereafter). Hence, for comparing health across countries, it is not very suitable. Because of this, and the wealth of health data available in SHARE, MKA developed a health index for wave 1 of SHARE that uses much more information, is continuous, has a much higher reliability than SRH, and is comparable across countries. Jürges (2007) at the same time as MKA (but published earlier) also developed a health index for wave 1 of SHARE, with similar goals but a strikingly different methodology. Poterba, Venti, and Wise (2011, 2013) developed a health index for the HRS with yet another methodology, although comparability across countries does not appear to be a core goal of their effort. Our goal in this paper is to study the theoretical and empirical differences between these indexes, so that researchers who want to include a measure of health in their analyses can make a better informed choice as to which index is most appropriate for their analyses, and readers can interpret differences between results from different papers that use different indexes. Section 2 describes the construction of the indexes and highlights differences in methodology and variables used. Section 3 then studies empirical differences, and section 4 studies the differences in explanatory power in a simple model for labor force transitions at older ages. Section 5 concludes.

2. Measuring health: variables and approaches to combining them

This section describes the various health indexes that we aim to compare in the remainder of the paper.

2.1 The methodology used by Meijer-Kapteyn-Andreyeva

MKA estimate models of health that has a MIMIC (Jöreskog and Goldberger, 1975) structure, analogous to the model of Börsch-Supan, McFadden, and Reinhold (1996): they assume that there is a single latent "true" health dimension, that a large number of observed health variables

² This is the "U.S. version", which is also used in the HRS, PSID, and other survey data sets in the U.S. SHARE wave 1 also includes a "European version", with response categories very good, good, fair, bad, very bad.

have this latent health dimension in common, and that true health is in turn related to a set of covariates. This leads to a system of equations

$$y_{ni} = \tau_i + \lambda_i \eta_n + \varepsilon_{ni} \quad (1)$$

$$\eta_n = x_n' \gamma + \zeta_{ni}, \quad (2)$$

where y_{ni} is the i -th observed health measure for individual n , η_n is the true health of the individual, x_n is a vector of covariates, ε_{ni} and ζ_{ni} are error terms, and the other terms are parameters. Thus, (1) is a factor analysis model for the observed health measures and (2) is a regression model for latent health.

This basic model structure is enriched along several lines. First, because most health measures used are binary or ordinal, the linear form (1) is replaced by a binary or ordinal probit equation as appropriate, in which a linear equation like (1) is assumed to hold for an underlying continuous variable, and the relation between this underlying continuous variable and the categorical observed variable is a step function. Second, almost all parameters are allowed to be different across countries, reflecting differences in reporting behavior in the health measures across individuals from different countries (who often are faced with questions phrased in different languages), and institutional differences that might affect the relationship between health and the covariates. The exception is the equation for grip strength, which is an objective measure taken by the interviewer, which MKA assume to be free from differences in reporting behavior. This "anchoring" on grip strength ensures comparability across countries. A further refinement of this is that MKA use a grip strength residual measure instead of observed grip strength (see below) and analyze men and women separately, so that the index is not assumed to be comparable between men and women. Third, missing observations in the health measures are straightforwardly dealt with in this model structure (they are integrated out of the likelihood), so that these do not lead to dropping of observations. There were two (categorical) covariates with significant numbers of missings. For those, missingness was included as an additional category. This is not recommended for the purpose of unbiased estimation of regression models, but is more suitable for prediction purposes, which is what it is used for by MKA.

Some further technical assumptions (such as normal distributions and uncorrelatedness of the error terms) allow estimation by maximum simulated likelihood. With the model estimated, the MKA health index $\hat{\eta}_n$ is the conditional expectation of true health η_n given (all) the observed variables in the model.

2.2 The methodology of Poterba-Venti-Wise

We base the discussion here on the variant of their index discussed in Poterba, Venti, and Wise (2013, PVW hereafter). PVW start with a large set of variables that are assumed to be related to an underlying true health variable. This is similar to MKA, and to stress the similarity, we call the observed variables y_{ni} and true health η_n . PVW run a principal components analysis (PCA) and compute the first principal component $\hat{\eta}_n$, which is their health index. PCA can be defined and interpreted in various ways, but for our current purposes, the easiest interpretation is the

following. Start with (1), and in addition to τ_i and λ_i , treat η_n as an additional parameter to be estimated. Estimate the model by minimizing the sum of squared errors. The resulting estimator of η_n is $\hat{\eta}_n$. No explicit assumptions about distribution or uncorrelatedness of the error terms are made. However, as noted by Wansbeek and Meijer (2000, p. 159), PCA is mathematically equivalent to a restricted form of factor analysis in which the error terms of the different observed variables not only are uncorrelated, but also have the same variance, in other words, equivalent to a form of factor analysis with much stronger assumptions than typically made in applications of factor analysis. On the other hand, as the number of observed variables increases, PCA and factor analysis solutions converge to each other. This finding underlies the estimation of "approximate factor models", which are often used in finance, by PCA. The number of measures used by PVW is large, but much smaller than the number typically used in applications in finance, and several order of magnitude smaller than the number of observations, so there is likely to be some difference between PCA and factor analysis. With the interpretation as a restricted factor analysis model and the additional assumption of normally distributed variables, the principal component index is equal to the conditional expectation of true health given the observed variables used, up to a multiplicative constant. (The conditional expectation is "shrunk" towards zero and thus has lower variance.)

For PCA, binary observed variables can be treated as continuous variables. This is analogous to the treatment of dummy variables as covariates in regression analyses. However, binary variables are obviously not normally distributed, and thus the PCA index is not a conditional expectation, and likely not exactly equal to a scaled version of a conditional expectation that is consistent with the binary nature of the variables.

The description in PVW (esp. their Table 2-1) suggests that they coded all their measures as binary variables, although some of the original variables have more than two categories. Because PVW do not explicitly mention the treatment of missing data, we assume that observations with missing data are dropped. Because they only have one country, whether parameters are identical across countries is not a relevant issue. However, they use multiple waves of the HRS and data for both men and women. They include all of these in a single analysis, because it is found that coefficients are quite similar across waves and between men and women.

2.3 The methodology of Jürges

Jürges (2007) uses a methodology that follows Cutler and Richardson (1997, 1999) and Lindeboom and Van Doorslaer (2004). He also assumes that there is a single latent true health dimension, which we can again call η_n . He assumes that SRH is a categorical reflection of true health, and that a set of other observed variables x_n act as covariates. He estimates an ordinal probit model with SRH as dependent variable and x_n as explanatory variables. At an abstract level, this model structure is equivalent to the MKA model structure, in which the set of dependent variables y_{ni} consists of only one variable, SRH. However, as we will see below, the set of variables that Jürges includes in x_n is very different from the covariates in MKA. The index of Jürges is defined as the predicted linear index from the ordinal probit model, that is, $\hat{\eta}_n = x_n' \hat{\gamma}$, which is the conditional expectation of true health given the covariates (but not

conditional on the dependent variable SRH). Note that SRH is used in the estimation but not in the construction of the index: an individual who reports excellent health has the same value of the health index as another individual with the same values of the covariates but who reports being in poor health. MKA compare their index with an analogous index defined as just the predicted linear index from (2) and find that the linear index has much lower reliability. But in the case of MKA, this amounted to ignoring the information on 25 health measures, whereas in the case of Jürges, this only ignores the information from a single variable, and thus the inefficiency is much smaller.

Like MKA, Jürges allows for cross-country differences in reporting behavior in SRH by allowing the threshold parameters to be country-specific, though not gender-specific. The regression coefficients are assumed equal across countries. For low grip strength, Jürges adds a dummy for missingness to avoid having to drop a sizable fraction of the sample, and to reduce issues of selectivity (missingness of grip strength is strongly related to bad health). For the other covariates, there are no such dummies and we thus assume that records with missing data on these variables are dropped.

2.4 Variables used in constructing the indexes

All three indexes use SRH, albeit in different ways. MKA and Jürges use it as an ordinal dependent variable and estimate a model with an ordinal probit equation. MKA include the information of SRH in the construction of their index, whereas Jürges ignores it in the construction. PVW include a binary indicator for whether SRH is fair or poor in their index. They also include an indicator for whether health was worse in the previous period, which MKA and Jürges do not have. Further, PVW include a dummy for whether health limits work, which MKA and Jürges do not have.

MKA and PVW include mobility, fine motor, and functional limitations (e.g., difficulty walking 100m), which are all binary, except that MKA use climbing stairs as an ordinal variable (0=no difficulty, 1=difficulty climbing several flights but not a single flight, 2=difficulty climbing a single flight). MKA use difficulties with activities of daily living (ADLs, e.g., difficulty dressing) as separate measures, whereas PVW only include a binary indicator for at least one ADL. MKA also include difficulties with instrumental activities of daily living (IADLs, e.g., using a map), which PVW do not use. Jürges does not use any of these difficulties variables.

Both PVW and Jürges include doctor-diagnosed chronic conditions (e.g., ever had diabetes), but the set of conditions is larger for Jürges. PVW also include back pain, which is one of a number of "symptoms" variables. Another set of variables only included by PVW is health care utilization (doctor visit, hospital stay, nursing home stay, home care).

MKA and Jürges include grip strength. Grip strength in middle age has been established as a predictor of health problems and mortality at later ages. However, both MKA and Jürges transform it before usage. MKA argue that grip strength is also a function of "size" (operationalized as height and weight) in a way that is unrelated to health. They correct for this by subtracting a predictive quadratic polynomial in height and weight and use the resulting grip

strength residual as a dependent variable in their analyses. Jürges similarly subtracts age- and sex-specific means and then defines low grip strength as being in the lowest tercile of the distribution of the residual. Jürges also includes an indicator for low walking speed (which individuals who were not subjected to the test assumed to have normal walking speed). MKA do not use walking speed and PVW use neither walking speed nor grip strength. All three indexes use body mass index (BMI), which is derived from self-reported height and weight. PVW use BMI as a continuous variable, which makes it their only variable that is not binary. Based on NIH guidelines for classification, MKA use dummies (as covariates) for being underweight (BMI < 18.5), overweight (25-30), class I obese (30-35), and class II and III obese (> 35). Jürges uses dummies for BMI < 20, 25-30, and 30+.

Finally, MKA include a set of covariates not included by PVW and Jürges: age, whether living with a spouse or partner, household size, education, and household net worth.

In the empirical part of this paper, we compare four indexes based on the ones described here. We have tried to closely approximate the indexes described in the literature, but there are some differences. Thus, although we will refer to the PVW or Jue index, they are not identical to the ones used by PVW or Jürges. However, we believe they are close enough to warrant this notation. The variables we have used in constructing the indexes are listed in Table 1.

We use SRH as a linear index. That is, we treat the category scores 1-5 as being measured on an interval scale. Our MKA index for wave 1 is the same one as in the MKA article. For wave 2, we have used the estimates from the models estimated for wave 1 in MKA and computed the health index for each individual accordingly. For the PVW index, we used the list from Table 2-1 in PVW and tried to mimic these as close as possible in SHARE. However, the wording and response categories occasionally differ from the ones in the HRS, and sometimes variables are not available. Some notable differences are:

- Health problems limit work was not asked in SHARE wave 1. We have replaced it (for both waves) with health problems limiting "activities people usually do".
- Health worse in previous period is not available in wave 1, nor for the refreshment sample in wave 2. We excluded it.
- For "ever experience heart problems", we use the "heart attack" variable in the doctor-diagnosed conditions. However, there is also a "heart problems" symptoms variable, and a high blood cholesterol variable that seems related, neither of which we used in constructing the index.
- Ever experience psychological problems is operationalized as professional (doctor-psychologist) treatment for depression by a family doctor or psychiatrist, which was only asked if the respondent reported having suffered from depression in the last year (new interviews) or since the last interview (re-interviews).
- For ever experience lung disease, we used the "chronic lung disease" variable but did not include the asthma variable.
- For nursing home stay, we only used the question about this in the health care section, not the "household type" variable that indicates living in a nursing home (there were no such cases in wave 1 and very few in wave 2).

Also, because PVW did not address cross-country comparability, we had to make a choice as how to implement it in a cross-country setting. The two ends of the spectrum are completely

country-specific, which by construction would imply that individuals for all countries are equally healthy on average, or joint, which assumes that there are no differences in reporting behavior. We chose the latter, which seems more in line with PVW's treatment of combining all waves and men and women. It would, however, be possible to create an intermediate index, by including some (but not all) interactions between country dummies and the variables that are now used.

The Jürges index was developed for SHARE and thus we should be able to reproduce it quite closely. However, there are still some notable differences in our current implementation (which we will improve in a later version of this paper):

- Jürges uses "heart attack or heart problems", which is a combination of the heart attack doctor-diagnosed condition and the heart problems symptoms variable, whereas we only use the heart attack variable for this.
- Wave 2 includes three additional conditions: other fractures, Alzheimer's or other memory problem, and benign tumor. We have combined these with the "other conditions" variable.
- Instead of the low grip strength dummy based on a residual after correcting for age and sex, we use the MKA grip strength residual (after correcting for height and weight, gender-specific) and include it as a continuous variable.
- We have not included walking speed.
- We used the NIH categorization of BMI, not the one in Jürges (2007).
- We use an ordinal logit instead of an ordinal probit to estimate the model.

Table 1 summarizes the variables we have used in constructing the indexes in this paper, and how we have used them. Observations with missings are generally dropped, with some exceptions, in particular dependent variables in MKA. This difference between MKA and the other indexes, and differences in which variables are used and how many missings they have, lead to differences in the number of respondents for which the indexes are available: for SRH, we have 56,234 observations (across both waves and men and women), for MKA, it is 54,388, for PVW 48,124, and Jue 52,989.

3. Empirical properties of the health measures

The four health measures we consider vary in a number of different respects: they include different variables, are based on different statistical models, and vary in the extent they are assumed or constructed to be internationally comparable. To gain more insight in their comparative properties we present a number of descriptive tables and graphs. All tables and figures will be presented for males and females separately. As noted in the previous section, the number of observations varies by health measure due to the fact that they are based on different sets of underlying variables, which have different patterns of missings. To facilitate comparability we restrict the sample to observations that allow the construction of all four measures.

Table 2 presents sample characteristics of the measures. Since the measures have different scales, the means and standard deviations are not comparable. (Also note that the signs of the indexes are different: MKA is higher for better health, whereas the other indexes are higher for worse

health.) It is of interest however to consider the kurtosis and skewness of the measures. To evaluate the numbers in the table, recall that the kurtosis of a normal distribution is three, while the skewness is zero. PVW and Jue show the highest kurtosis. Kurtosis is particularly high for males. For SRH and MKA, the kurtosis is close to that of a normal distribution. SRH and MKA also show the least skewness; they are both slightly skewed to the right. In contrast, PVW and Jue show considerable skewness to the left.

Figure 1 confirms these observations. To facilitate comparisons, the variables have been standardized so that they are measured on the same scale and better health is always higher. Clearly PVW shows a high peak at about one standard deviation above the mean, while at the same time exhibiting a long left tail. In other words, according to this index many individuals have similar (good) health, but there are also individuals with particularly bad health (those located in the left tail). When using the health measures to explain labor market status or labor market transitions, the PVW may be able to distinguish between various degrees of bad health, while it may be less discriminating in the area of good health. MKA would appear to be better at discriminating between health states across the whole spectrum from very poor to excellent health, but may be less informative when it comes to distinguishing different levels of poor health. We return to this aspect in Section 4.

Although the shape of the densities of the four health measures in Figure 1 reveals considerable differences, it would be conceivable that they still rank the health of individuals in the same way and hence the measures may still exhibit high correlations (although obviously not equal to one). The correlations in Table 3 vary between .54 and .86 (in absolute value). These numbers do not appear to be particularly high. For instance, the correlation between MKA and Jue for males of .63 would imply that in a regression of MKA on Jue only about 40% of the variance would be explained. Thus it appears that the various measures partly measure different things (as was also suggested by the discussion in Section 2), which makes it worthwhile to investigate their relation with respondent characteristics and behavior.

Figure 2 shows the relation between the health measures and age. Once again the measures are standardized so that we can compare their scales. MKA shows the steepest decline with age, while the decline is least for SRH. PVW and Jue appear to be in between these two extremes. It is probably not surprising that SRH shows the smallest gradient with age. In expressing self-reports of health, respondents may very well compare themselves with others of the same age, which would lower the age gradient. In the extreme case that respondents would provide a completely relativistic evaluation (just comparing their age with others in their cohort), the age gradient might be zero. MKA is not only based on ADLs and IADLs, but also on grip strength, which is a physical measure with a pronounced negative age gradient.

A major point of interest is how health status compares across countries. The four measures vary in the extent to which comparability issues are incorporated. SRH and PVW assume complete comparability, in the way we are using them. We could also have constructed the PVW by country, which would have avoided any assumptions about inter-country comparability. For SRH no obvious correction is available in the data, although we could have considered using the vignettes by health domain that have been collected in the SHARE data. This would have been a major task, as it requires not just the modeling of response scale corrections by health domain,

but also a model that relates general health to health by domain. Somewhat similar modeling has been done by Kapteyn et al. (2007) for several domains of work disability, while SHARE vignettes have been used in a number of different domains, but not for overall health; see for instance the special issue of *Social Indicators Research*, edited by Jürges and Van Soest (2012), and Datta Gupta, Kristensen, and Pozzoli (2010). MKA assumes comparability by country of grip strength (or rather a transformation of grip strength that corrects for height and weight), which is used to rescale the within country measures. Jue assumes identical coefficients across countries, but with country-specific cut-off points for the conditional logits that explain SRH by country.

Figure 3 shows standardized means by country and sex. The first observation is that there is considerable correlation between mean female and mean male health across countries for every measure. If a country scores high in female health, it is likely to also score high in male health, and vice versa. The correlations between average male and female health are .95 for SRH, MKA, Jue; and .80 for PVW. Regarding the correspondence between measures, we note that the signs are generally in agreement. If a country is below the mean according to one measure, it usually also is below the mean according to a different measure. There are a few exceptions to this rule. For instance for German males, SRH and PVW suggest that German health is below the mean, whereas MKA pegs it well above it. For both male and female Greeks, MKA suggests a mean health level below the sample mean, while the other measures suggest Greek health to be about average. Table 4 shows the correlations between the country means of the four measures for males and females separately. The correlations of the country means appear of the same order of magnitude as the correlations between the individual measures presented in Table 3.

Our interest is not only in the levels of health according to different measures and how they compare, but also in changes. Table 5 shows the distribution of health changes between 2004 and 2006. As these are changes observed for the same respondents, it is not surprising that mean health changes are negative, since the respondents on average have aged about two years. Once again kurtosis is particularly high for PVW and to a slightly lesser extent for Jue. Skewness appears modest. Figure 4 shows standardized densities of the changes in health measures confirming the fact that the distributions of changes are indeed fairly symmetric. Table 6 shows that all changes are positively correlated, as one would expect. However, correlations are not particularly high.

Figure 5 shows how health changes are related to age. For females, all measures show an accelerating deterioration with age, with the possible exception of SRH. For males that pattern is somewhat less clear. In particular Jue and SHR show little sign of accelerating health deterioration for males. On the other hand, PVW indicates that health deteriorates at an accelerated pace with age.

Figure 6, once again provides a comparison across countries. As one would expect, nearly all measures show a decline, with some minor exceptions. Switzerland appears an outlier with all measures showing improvement in average female health, but not for male health (except according to Jue). Self-reported female health in Greece also shows a rather remarkable improvement. The correlation between average male and female health changes is considerably lower than for levels. The correlation is .17 for SRH, .57 for PVW, and .71 for Jue. Remarkably

the correlation between average female and male health changes is $-.13$ for MKA. Table 7 shows the correlations between the country mean changes of the four measures for males and females separately. The correlations of the country mean changes appear of the same order of magnitude as the correlations between the individual measures presented in Table 6, but they are definitely not the same. For instance, the correlation between average changes in Jue and PVW is $.66$ for males and $.90$ for females. The corresponding correlations for levels in Table 6 were $.43$ and $.41$ respectively. The correlations of average changes in health of men as measured by MKA with the other three measures are lower than for levels. In contrast, for females the correlations for the change measures are higher than for levels.

The descriptive analyses so far have established that the various measures are positively correlated, but at the same time there are substantial differences in distributions, in their relations with age and in how the measures rank countries by health. There is enough scope therefore to investigate how the measures relate to individual characteristics and behavior.

4. Labor force transitions

Within the age group of individuals 50 and older one would expect health to become an increasingly important determinant of transitions into retirement or into some kind of Disability Insurance scheme. To investigate the power of the various health measures in explaining such transitions we estimate linear probability models that explain labor market status in 2006, based on health status in 2004 and changes between 2004 and 2006. One might suspect that health status in 2006 could be endogenous with respect to labor market status in 2006. Empirical evidence, partly based on SHARE, suggests that retirement does not have adverse effects on health. Possibly retirement improves health. See for instance, Coe and Zamarro (2011). To the extent that retirement improves health, we would under-estimate the health change between 2004 and 2006 for those who have retired between waves. This would attenuate the estimated effect of a health change on the probability of a transition from work to retirement. The evidence is somewhat less clear with respect to the effect on DI-receipt, as one might worry about the possibility of self-reports being colored by justification bias (e.g., Bound, 1991). The empirical evidence for the existence of justification bias appears to be limited however, at least in Europe (Kapteyn, Smith, and Van Soest, 2011). In any case, the models to be presented below are not meant to be full-fledged structural models of transitions into retirement or disability. Rather the analyses seek to establish the strength of the links between the various measures and observed transitions.

Table 8 presents transitions in labor market states between 2004 and 2006. Of those working in 2004, 14.4% are retired in 2006. Other less common pathways out of employment are to homemakers, unemployment, and disability. The probability of “unretiring” is very low. Less than 1% of those who were retired in 2004 report to be working in 2006. This contrasts with patterns in the U.S. where the probability of returning to work after retirement appears to be substantially higher (cf. Maestas, 2010). Among the disabled in 2004, 7.6% reports to be working in 2006. These are relatively small numbers though (33 individuals).

To explain transitions we estimate a sequence of models with an increasing number of covariates. The first model only includes 2004 health (“Lagged health”) and the difference in health between 2006 and 2004 (“Health change”) as explanatory variables. The health variables are all standardized to facilitate comparability between the estimated effects by health measure. The second model adds demographics (a third degree polynomial in age); education in three classes; living with a partner) and economic variables (working in the public sector; self-employed; eligible for public pension; eligible for private pension; reached early retirement age; reached normal retirement age; inverse hyperbolic sine of household net worth-ppp adjusted; log household income-ppp adjusted) both for 2004 and 2006. The third model adds country dummies. The models only consider binary choices: we estimate models for retiring vs. working and becoming disabled vs. working. That is, the sample for the first of these only consists of individuals who were working in 2004 and either working or retired in 2006, and the sample of the second only consists of individuals who were working in 2004 and either working or disabled in 2006.

For simplicity we do not present the estimation results for all the added covariates, but concentrate on the coefficients of lagged health and health change. Table 9 presents results for transitions between being employed in 2004 and being retired in 2006 for males. The dependent variable is equal to one if an individual is retired in 2006, so a negative sign implies an increase in the probability of being retired in 2006 if health falls. Both lagged health and health change have a highly significant effect on the probability of retirement. In this model without covariates, R^2 values are low. Adding demographics and economic variables raises the R^2 s substantially and reduces the effect of the health variables. Self-reported health still shows significant effects (and to a lesser extent PVW). Conceivably, the significance of self-reported health is partly the result of justification bias. Adding country dummies further reduces the significance of the health variables, with only lagged self-reported health retaining significance.

Table 10 presents the same sequence of models for females. The conclusions are very similar to those for males. The simple model with just the health variables shows these variables to have highly significant effects. The significance gets reduced when demographics and economic variables are added and goes away once we also add country dummies, once again with the exception of lagged self-reported health.

Table 11 contains estimation results for transitions into disability for males. The R^2 s of the models without covariates are considerably higher than for the model explaining transitions into retirement. The health measures remain significant when we add additional covariates. The size of the regression coefficients of the health variables is hardly affected by the addition of covariates. The R^2 s increase as a result of adding covariates, but the increase is modest, certainly in comparison with the models explaining retirement. All this suggests that for the transition into disability, health is the main determinant with a rather modest role for demographics and socio-economic variables.

Table 12 presents the results for females. Again the health variables are highly significant and hardly affected by the addition of covariates. R^2 s are lower than for males.

Comparing across health measures some tentative conclusions emerge. When it comes to explaining transitions into retirement, there is not much to choose between the various measures. MKA, PVW, and Jue all explain a relatively small percentage of the variance in the transition into retirement. The picture for disability is different. PVW explains a larger percentage of the variance of transitions into disability for both males and females, while the estimated coefficients are larger. It is of interest to speculate why this may be so. As noted in Section 3, PVW has a long left tail and thus appears able to discriminate more finely between health levels at the poor end of the scale.

Another explanation might be that the health care utilization variables in the PVW index add explanatory power. This would be problematic from the standpoint of causal analysis. Presumably, one does not become disabled by going to the doctor, but one goes to the doctor because of a health problem that makes one disabled. Also, a doctor visit is often necessary to become classified as disabled for public disability insurance.

Noting that the doctor-diagnosed chronic health conditions and the health care utilization variables are the main components included in PVW but not in MKA, we investigate whether these are indeed the drivers of the additional explanatory power of the PVW index. We do this by estimating a fourth model, which adds wave 1 health conditions and changes in health conditions between 2004 and 2006 to the MKA-based model 3 (heart attack, high blood pressure, high cholesterol, stroke, diabetes, lung disease, asthma, arthritis, osteoporosis, cancer, ulcer, Parkinson disease, cataracts, hip or femoral fracture, psychological problems, other), and a fifth model, which adds the health care utilization variables in 2004 and changes between 2004 and 2006 (doctor visit, hospital stay, nursing home stay, home care) to model 4. The results for all five models are given in the Appendix, Tables A1 and A2.

Most coefficients of the health conditions in the model for males becoming disabled are small and not significant, with two notable exceptions: stroke (both lagged and change, i.e., both having had one before 2004 and between 2004 and 2006, are significant and large) and psychological problems. These effects remain after inclusion of health care utilization variables. Several of the utilization variables have large and significant coefficients: lagged nursing home stay, lagged home care, and change in home care. For females, the results are slightly different: among the utilization coefficients, only home care and its change are significant at the 5% level, and the effects are much smaller than for males. However, for females, there are more health conditions that are significant, though not all with the expected sign: significant and positive (as expected) coefficients are change in stroke, lagged cancer, and change in cancer. Significant and negative are change in parkinson and lagged hip or femoral fracture. Note, however, that these last two conditions are not included in PVW. The addition of the health conditions increases R^2 values substantially. For males the R^2 increases from .115 to .205, while for females R^2 increases from .053 to .132. Adding health care utilization variables raises the R^2 further, to .281 for males and .145 for females. These values are substantially higher than the R^2 values for the other indexes in Tables 11 and 12.

Summarizing, there is some evidence that the health care utilization components, and to a lesser extent some of the health conditions, give the PVW index some additional explanatory power over the MKA index in explaining transitions into disability.

5. Discussion

As explained in Section 2, the health measures used in this paper are based on very different statistical models and make different assumptions about inter-country comparability. SRH and PVW (in the way we have constructed the index) assume complete comparability; Jue assumes comparability in the construction of the index, but allows for country-specific cut-off points in response scales; MKA makes the least assumptions about comparability, but uses a grip strength related measure to scale the country and sex specific indexes in order to attain comparability. In addition to different assumptions about comparability, the statistical models are different, which for our purposes implies different functional forms.

The most important difference however, probably lies in the choice of variables that are included (or not) in the construction of the indexes, as summarized in Table 1. By design, MKA includes difficulties with mobility, ADLs, and IADLs, plus self-reported health, physical attributes like grip strength and body mass index. In the underlying MIMIC model, “causal” variables include sex, age, living with spouse or partner, household size, education, and net worth. PVW and Jue do not include these socio-economic and demographic variables. On the other hand, Jue includes a long list of health conditions as well as grip strength and body mass index. PVW includes mobility limitations, body mass index, having at least one ADL, a slightly more limited list of health conditions than Jue, and health care utilization variables. PVW is the only index including a pain variable. The importance of including certain variables can be seen by inspecting the model 5 regressions. For instance, in the model using MKA to explain transitions into disability of males, several of the health conditions are highly significant as are nursing home stays and home care. These variables are not included in MKA. Since PVW does include these variables, one would indeed expect PVW to have more explanatory power.

There are several potential exercises that potentially further enlighten these issues. As we have seen, MKA plus health conditions and health utilization gives a better fit than PVW, which in its turn gives a better fit than just MKA. However, the health conditions added in this exercise include some that were not included in PVW. We could additionally look at a regression in which only the ones in PVW are added to the MKA regression. Conversely, we could also run analogous regressions in which we add the additional health conditions and/or variables that are included in MKA but not in PVW to the regressions with the PVW index. Furthermore, MKA estimates a much larger set of parameters because most parameters are allowed to vary across countries, and thus MKA is much more flexible than PVW. We could investigate whether the PVW fit improves further by allowing a similar kind of flexibility in this index through adding interactions between country dummies and the variables currently in the PVW index, or at the extreme, constructing this index separately by country. We leave these exercises for future research.

Do our results imply that one should always include the maximum number of variables in a health index? Not necessarily. Health care utilization rates may vary by country for institutional reasons. So if the goal is to explain differences in health across countries, then one probably does not want to contaminate a health measure with institutional variables. But if one aims at

explaining labor market transitions within a country then including utilization variables may be a good idea.

A different dimension along which the measures differ is their complexity in estimation. Self-reported health is obviously the simplest of the four measures considered, but in a cross-country context its comparability is dubious, as has been argued for instance by Jürges (2007). The next simplest measure is PVW. Principal components analysis is available in many statistical packages and is quite standard. The health index proposed by Jürges is somewhat more complicated, as one has to parameterize the cut-points in the probit analyses. MKA is the most involved, as it attempts to do full justice to the discrete nature of most of the variables and model these as the result of an underlying latent process. Although one might prefer a MIMIC-type approach as used in MKA on theoretical grounds, the empirical results in this paper do not suggest that there is much by way of pay-off in terms of improved explanatory power.

Acknowledgments

We thank Tatiana Andreyeva and Meena Fernandes, who were involved in the data construction for the MKA index that we draw upon.

This paper uses data from SHARE releases 2.0.1 (July 2007) and 2.5.0, as of May 24th 2011. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT- 2001- 00360 in the thematic programme Quality of Life), through the 6th framework programme (projects SHARE-I3, RII-CT- 2006-062193, COMPARE, CIT5-CT-2005- 028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th framework programme (SHARE-PREP, 211909 and SHARE-LEAP, 227822). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064, IAG BSR06-11, R21 AG025169) as well as from various national sources is gratefully acknowledged (see <http://www.share-project.org> for a full list of funding institutions).

References

- Börsch-Supan, A., D.L. McFadden, and S. Reinhold (1996). Living arrangements: health and wealth effects. In D.A. Wise (ed.), *Advances in the Economics of Aging*, University of Chicago Press: Chicago; 193–218.
- Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources* 26, 106–138.
- Coe, N.B. and G. Zamarro (2011). Retirement effects on health in Europe. *Journal of Health Economics* 30, 77–86.
- Cutler, D. and E. Richardson (1997). Measuring the health of the U.S. population. *Brookings Papers on Economic Activity: Microeconomics* 1997, 217–282.

- Cutler, D. and E. Richardson (1999). Your money and your life: The value of health and what affects it. In A. Garber (ed.), *Frontiers in Health Policy Research*, vol. 2, MIT Press: Cambridge, MA; 99–132
- Datta Gupta, N., N. Kristensen, and D. Pozzoli (2010). External validation of the use of vignettes in cross-country health studies. *Economic Modelling* 27(4), 854–856.
- Jöreskog, K.G. and A.S. Goldberger (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* 70, 631–639.
- Jürges, H. and A. van Soest (eds.) (2012). Comparing the well-being of older Europeans. Special Issue of *Social Indicators Research* 105.
- Jürges H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics* 16, 163–178.
- Kapteyn, A., J.P. Smith, and A. van Soest (2007). Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review*, 97(1), 461–473.
- Kapteyn, A., J.P. Smith, and A. van Soest (2011). Work disability, work, and justification bias in Europe and the U.S. In D.A. Wise (ed.), *Explorations in the Economics of Aging*, University of Chicago Press: Chicago; 269–314.
- Lindeboom M. and E. van Doorslaer (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics* 23, 1083–1099.
- Maestas, N. (2010). Expectations and realizations of work after retirement. *Journal of Human Resources* 45, 718–748.
- Meijer, E., A. Kapteyn, and T. Andreyeva (2011). Internationally comparable health indices. *Health Economics* 20, 600–619.
- Poterba, J.M., S.F. Venti, and D.A. Wise (2011). Family status transitions, latent health, and the post-retirement evolution of assets. In D.A. Wise (ed.), *Explorations in the Economics of Aging*, University of Chicago Press: Chicago; 23–69.
- Poterba, J.M., S.F. Venti, and D.A. Wise (2013). Health, education, and the post-retirement evolution of household assets. NBER Working Paper 18695.
- Wansbeek, T. and E. Meijer (2000). *Measurement Error and Latent Variables in Econometrics*. North-Holland: Amsterdam.

Table 1: Variables included in the indexes

	SRH	MKA	PVW	Jue
Self-reported health (US version)	linear	categories	Fair/poor	(a)
Health limits activities (3 categories)			linear	
<i>Mobility, fine motor, and functional limitations</i>				
walking 100 metres		x	x	
sitting two hours		x	x	
getting up from chair		x	x	
climbing stairs (3 categories)		categories	linear	
stooping, kneeling, crouching		x	x	
reaching or extending arms above shoulder		x	x	
pulling or pushing large objects		x	x	
lifting or carrying weights over 5 kilos		x	x	
picking up a small coin from a table		x	x	
<i>Difficulties with activities of daily living (ADL)</i>				
dressing, including shoes and socks		x		
walking across a room		x		
bathing or showering		x		
eating, cutting up food		x		
getting in or out of bed		x		
using the toilet, incl getting up or down		x		
at least one ADL			x	
<i>Difficulties with instrumental activities of daily living (IADL)</i>				
using a map in a strange place		x		
preparing a hot meal		x		
shopping for groceries		x		
telephone calls		x		
taking medications		x		
doing work around the house or garden		x		
managing money		x		

Table 1. Variables included in the indexes (continued)

	SRH	MKA	PVW	Jue
<i>Doctor-diagnosed health conditions</i>				
heart attack			x	x
high blood pressure or hypertension			x	x
high blood cholesterol				x
stroke			x	x
diabetes or high blood sugar			x	x
chronic lung disease			x	x
asthma				x
arthritis			x	x
osteoporosis				x
cancer			x	x
stomach or duodenal ulcer, peptic ulcer				x
parkinson disease				x
cataracts				x
hip fracture or femoral fracture				x
other conditions				x
psychological problems			x	x
<i>Bothered by (symptoms)</i>				
pain in back, knees, hips or other joint			x	
<i>Health care utilization</i>				
doctor visit			x	
hospital stay			x	
nursing home stay			x	
home care			x	
Grip strength		(b)		(b)
body mass index		categories	linear	categories
<i>Demographics and socio-economics</i>				
country		separate		(a)
sex		separate		
age		cubic polynomial		
living with spouse or partner		x		
hhsz		x		
education in 3 classes		categories		
net worth		(c)		

(a) Used in estimating the model but not in constructing the index

(b) Residual after subtracting quadratic polynomial in height and weight and mean of the result (both gender-specific)

(c) Inverse hyperbolic sine of ppp-adjusted net worth

Table 2: Distributional properties of the raw indexes

	SRH	MKA	PVW	Jue
<i>Males</i>				
N	22019	22019	22019	22019
mean	2.92	0.05	-0.37	1.22
sd	1.07	0.49	2.10	1.29
skewness	-0.06	-0.62	1.84	1.17
kurtosis	2.48	3.39	6.61	5.19
<i>Females</i>				
N	25791	25791	25791	25791
mean	3.08	0.00	0.44	1.49
sd	1.06	0.39	2.61	1.39
skewness	-0.17	-0.45	1.24	1.05
kurtosis	2.54	2.94	3.93	4.43
<i>Total</i>				
N	47810	47810	47810	47810
mean	3.01	0.02	0.07	1.36
sd	1.07	0.44	2.42	1.35
skewness	-0.12	-0.53	1.49	1.11
kurtosis	2.50	3.34	4.86	4.74

Table 3: Correlations among the raw indexes

	SRH	MKA	PVW	Jue
<i>Males</i>				
SRH	1.00			
MKA	-0.68	1.00		
PVW	0.64	-0.79	1.00	
Jue	0.54	-0.63	0.69	1.00
<i>Females</i>				
SRH	1.00			
MKA	-0.68	1.00		
PVW	0.66	-0.86	1.00	
Jue	0.56	-0.67	0.73	1.00

Table 4: Correlations among different measures of country averages of health

	SRH	MKA	PVW	Jue
<i>Males</i>				
SRH	1.00			
MKA	0.43	1.00		
PVW	0.86	0.27	1.00	
Jue	0.56	0.62	0.60	1.00
<i>Females</i>				
SRH	1.00			
MKA	0.55	1.00		
PVW	0.92	0.64	1.00	
Jue	0.59	0.85	0.75	1.00

Table 5: Distribution of changes in standardized health indexes

	SRH	MKA	PVW	Jue
<i>Males</i>				
N	6683	6683	6683	6683
mean	-0.17	-0.08	-0.12	-0.06
sd	0.91	0.66	0.78	0.81
skewness	-0.18	-0.26	-0.8	-0.17
kurtosis	3.88	4.83	8.84	7.31
<i>Females</i>				
N	7858	7858	7858	7858
mean	-0.16	-0.07	-0.09	-0.02
sd	0.89	0.63	0.69	0.81
skewness	-0.13	0.1	-0.38	-0.07
kurtosis	3.7	5.04	5.9	4.88
<i>Total</i>				
N	14541	14541	14541	14541
mean	-0.17	-0.08	-0.11	-0.04
sd	0.9	0.65	0.74	0.81
skewness	-0.15	-0.08	-0.62	-0.12
kurtosis	3.79	4.95	7.71	5.99

Table 6: Correlations among the changes in the standardized indexes

	SRH	MKA	PVW	Jue
<i>Males</i>				
SRH	1.00			
MKA	0.54	1.00		
PVW	0.38	0.72	1.00	
Jue	0.22	0.32	0.43	1.00
<i>Females</i>				
SRH	1.00			
MKA	0.46	1.00		
PVW	0.36	0.78	1.00	
Jue	0.20	0.32	0.41	1.00

Table 7: Correlation between different measures of country mean health changes

	SRH	MKA	PVW	Jue
<i>Males</i>				
SRH	1.00			
MKA	0.47	1.00		
PVW	0.35	0.27	1.00	
Jue	-0.06	0.10	0.66	1.00
<i>Females</i>				
SRH	1.00			
MKA	0.36	1.00		
PVW	0.57	0.85	1.00	
Jue	0.55	0.84	0.90	1.00

Table 8: Labor force transitions

Wave 1 labor force status	N	%	Wave 2 labor force status (row %)					
			Retired	Employed	Unemployed	Disabled	Homemaker	Other
Retired	7,294	49.1	91.4	0.9	0.2	1.7	5.0	0.8
Employed or self- employed	4,132	30.0	14.4	76.9	3.0	2.2	2.4	1.2
Unemployed	522	3.7	25.8	21.7	36.9	4.9	7.0	3.7
Permanently sick or disabled	435	3.0	33.2	7.6	1.4	51.1	5.3	1.5
Homemaker	2,114	13.9	21.0	3.0	1.3	3.0	68.3	3.4
Other	38	0.3	34.9	10.7	4.9	17.0	16.3	16.3
Total	14,535	100.0	54.2	25.0	2.6	3.7	13.1	1.5
N			8,113	3,401	355	512	1,963	191

Table 9: Retiring vs. staying employed or self-employed, males

	SRH	MKA	PVW	Jue
<i>Health only</i>				
Health change	-0.0258* (0.0110)	-0.0586*** (0.0148)	-0.0445* (0.0202)	-0.0577*** (0.0151)
Lagged health	-0.0363*** (0.0108)	-0.0940*** (0.0139)	-0.0867*** (0.0199)	-0.0758*** (0.0149)
R^2	0.006	0.028	0.014	0.018
Observations	2050	2050	2050	2050
<i>+demographics, economic</i>				
Health change	-0.0118 (0.00955)	-0.0162 (0.0122)	-0.0246 (0.0175)	-0.0275* (0.0125)
Lagged health	-0.0210* (0.00911)	-0.0221 (0.0122)	-0.0402* (0.0165)	-0.0159 (0.0119)
R^2	0.324	0.324	0.325	0.324
Observations	2019	2019	2019	2019
<i>+country dummies</i>				
Health change	-0.0107 (0.00949)	-0.0156 (0.0123)	-0.0239 (0.0170)	-0.0286* (0.0123)
Lagged health	-0.0120 (0.00937)	-0.0209 (0.0129)	-0.0310 (0.0159)	-0.0138 (0.0117)
R^2	0.339	0.340	0.341	0.341
Observations	2019	2019	2019	2019

Coefficients of demographic and economic variables and country dummies not shown.
Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Retiring vs. staying employed or self-employed, females

	SRH	MKA	PVW	Jue
Health change	-0.0256 [*] (0.0119)	-0.0539 ^{**} (0.0181)	-0.0625 [*] (0.0253)	-0.0599 ^{**} (0.0184)
Lagged health	-0.0414 ^{***} (0.0123)	-0.0724 ^{***} (0.0148)	-0.0720 ^{***} (0.0213)	-0.0761 ^{***} (0.0163)
R^2	0.009	0.018	0.011	0.019
Observations	1714	1714	1714	1714
<i>+demographics, economic</i>				
Health change	-0.0210 [*] (0.0106)	-0.0172 (0.0169)	-0.0377 (0.0228)	-0.0252 (0.0175)
Lagged health	-0.0350 ^{***} (0.0104)	-0.0148 (0.0130)	-0.0405 [*] (0.0184)	-0.0291 (0.0156)
R^2	0.308	0.303	0.305	0.305
Observations	1681	1681	1681	1681
<i>+country dummies</i>				
Health change	-0.0187 (0.0108)	-0.0125 (0.0174)	-0.0362 (0.0227)	-0.0230 (0.0181)
Lagged health	-0.0309 ^{**} (0.0112)	-0.0245 (0.0146)	-0.0351 (0.0187)	-0.0283 (0.0161)
R^2	0.321	0.318	0.319	0.319
Observations	1681	1681	1681	1681

Coefficients of demographic and economic variables and country dummies not shown.
Standard errors in parentheses; ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$

Table 11: Becoming disabled vs. staying employed or self-employed, males

	SRH	MKA	PVW	Jue
<i>Health only</i>				
Health change	-0.0465*** (0.00852)	-0.0644*** (0.0118)	-0.102*** (0.0177)	-0.0564*** (0.0115)
Lagged health	-0.0508*** (0.00824)	-0.0666*** (0.0113)	-0.0879*** (0.0169)	-0.0680*** (0.0130)
R^2	0.068	0.080	0.126	0.072
Observations	1721	1721	1721	1721
<i>+demographics, economic</i>				
Health change	-0.0436*** (0.00821)	-0.0620*** (0.0116)	-0.0979*** (0.0176)	-0.0523*** (0.0112)
Lagged health	-0.0485*** (0.00816)	-0.0659*** (0.0118)	-0.0847*** (0.0171)	-0.0662*** (0.0131)
R^2	0.092	0.100	0.145	0.097
Observations	1704	1704	1704	1704
<i>+country dummies</i>				
Health change	-0.0448*** (0.00825)	-0.0684*** (0.0121)	-0.0990*** (0.0173)	-0.0542*** (0.0112)
Lagged health	-0.0505*** (0.00836)	-0.0715*** (0.0123)	-0.0863*** (0.0169)	-0.0662*** (0.0129)
R^2	0.102	0.115	0.155	0.106
Observations	1704	1704	1704	1704

Coefficients of demographic and economic variables and country dummies not shown.
Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Becoming disabled vs. staying employed or self-employed, females

	SRH	MKA	PVW	Jue
<i>Health only</i>				
Health change	-0.0340*** (0.00693)	-0.0447*** (0.0114)	-0.0788*** (0.0185)	-0.0506*** (0.0119)
Lagged health	-0.0237*** (0.00630)	-0.0312*** (0.00835)	-0.0549*** (0.0148)	-0.0404*** (0.0105)
R^2	0.036	0.031	0.065	0.047
Observations	1485	1485	1485	1485
<i>+demographics, economic</i>				
Health change	-0.0352*** (0.00731)	-0.0477*** (0.0119)	-0.0836*** (0.0192)	-0.0529*** (0.0122)
Lagged health	-0.0248*** (0.00658)	-0.0321*** (0.00891)	-0.0576*** (0.0150)	-0.0409*** (0.0107)
R^2	0.047	0.042	0.077	0.057
Observations	1461	1461	1461	1461
<i>+country dummies</i>				
Health change	-0.0370*** (0.00754)	-0.0515*** (0.0126)	-0.0842*** (0.0192)	-0.0537*** (0.0122)
Lagged health	-0.0282*** (0.00758)	-0.0385*** (0.00979)	-0.0592*** (0.0151)	-0.0428*** (0.0107)
R^2	0.057	0.053	0.085	0.064
Observations	1461	1461	1461	1461

Coefficients of demographic and economic variables and country dummies not shown.
Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

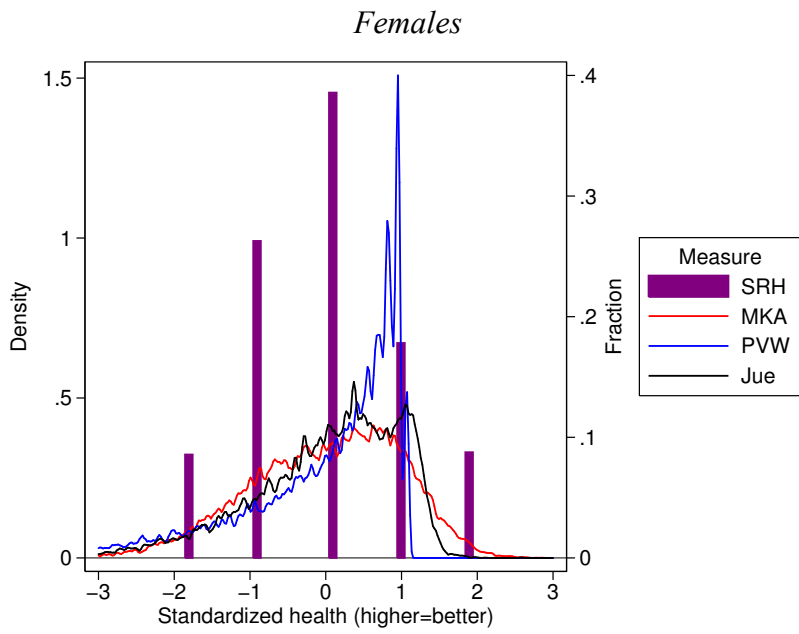
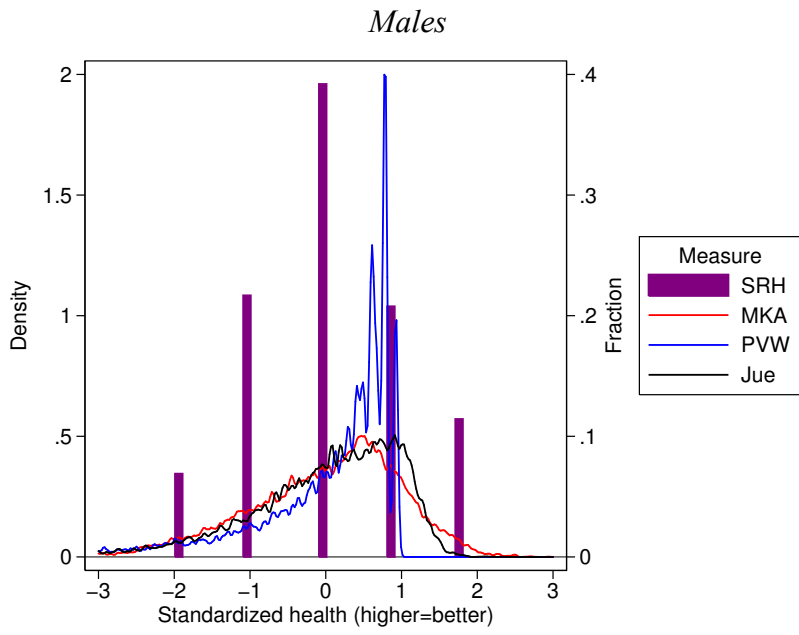


Figure 1. Distributions of the standardized indexes

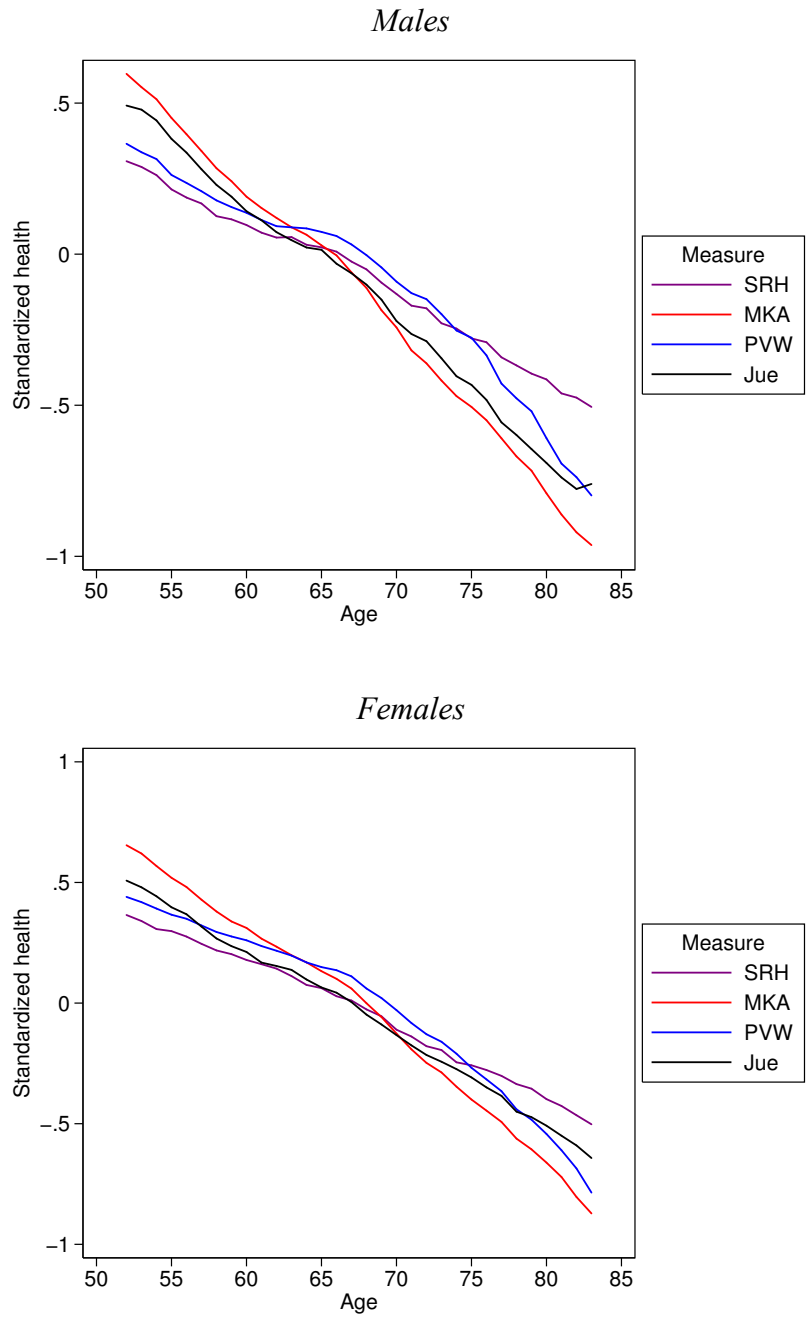


Figure 2. Mean standardized health by age (5-year moving average centered at the given age)

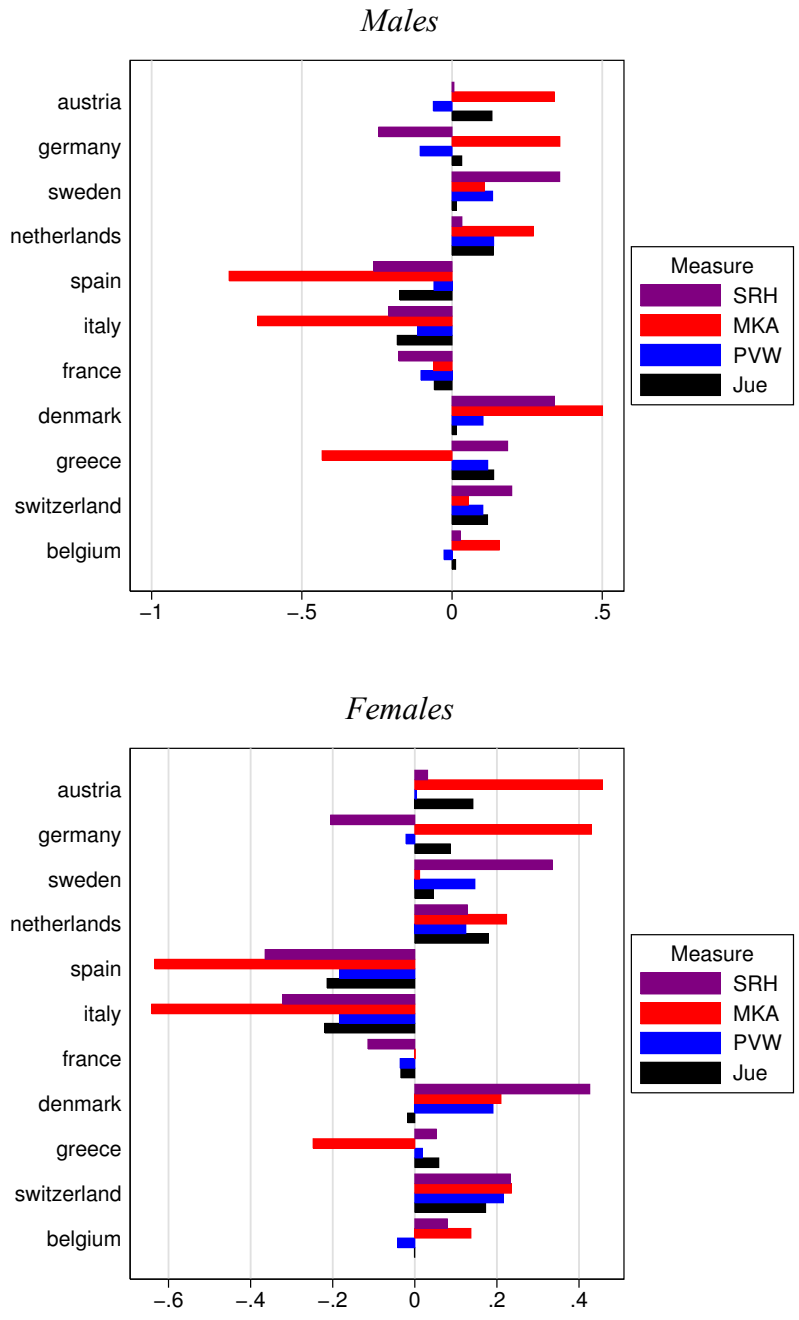


Figure 3. Mean standardized health by country

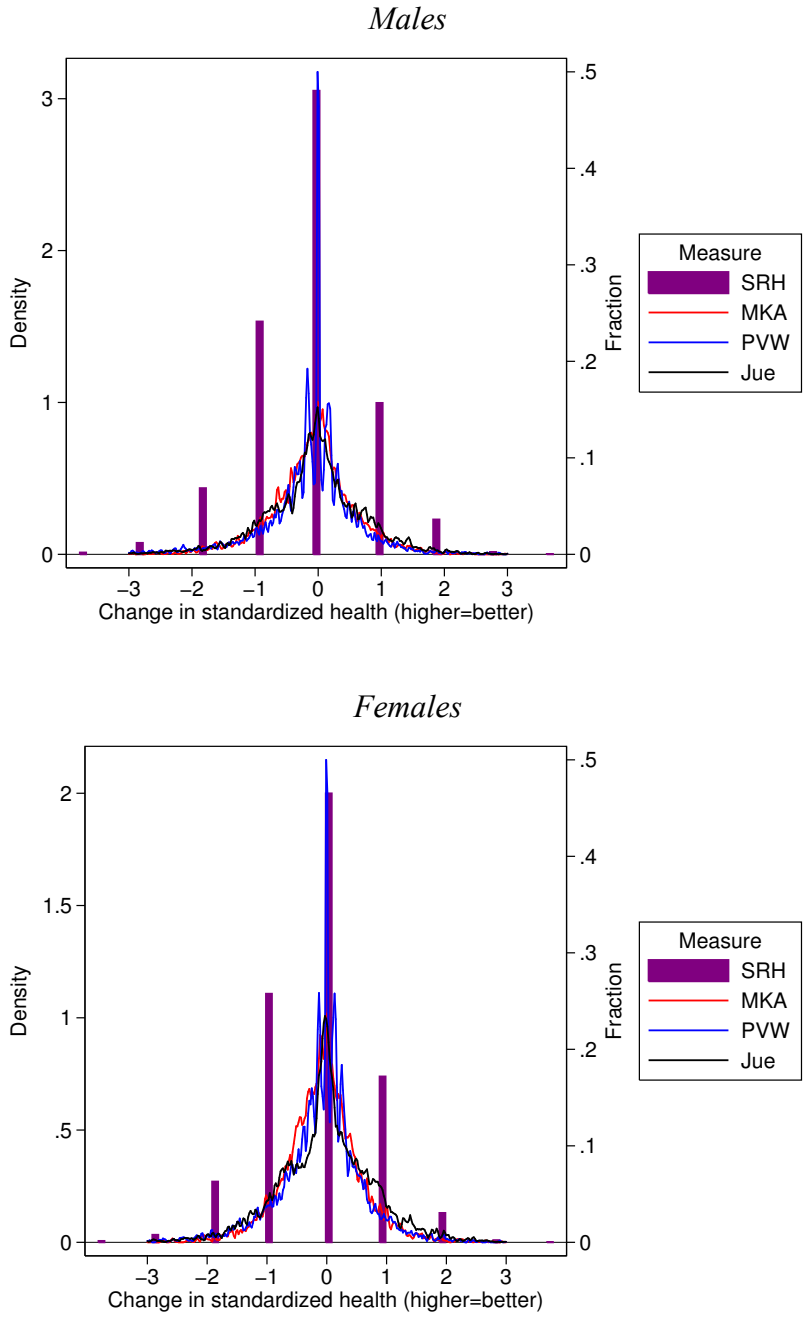


Figure 4. Distributions of the changes in health

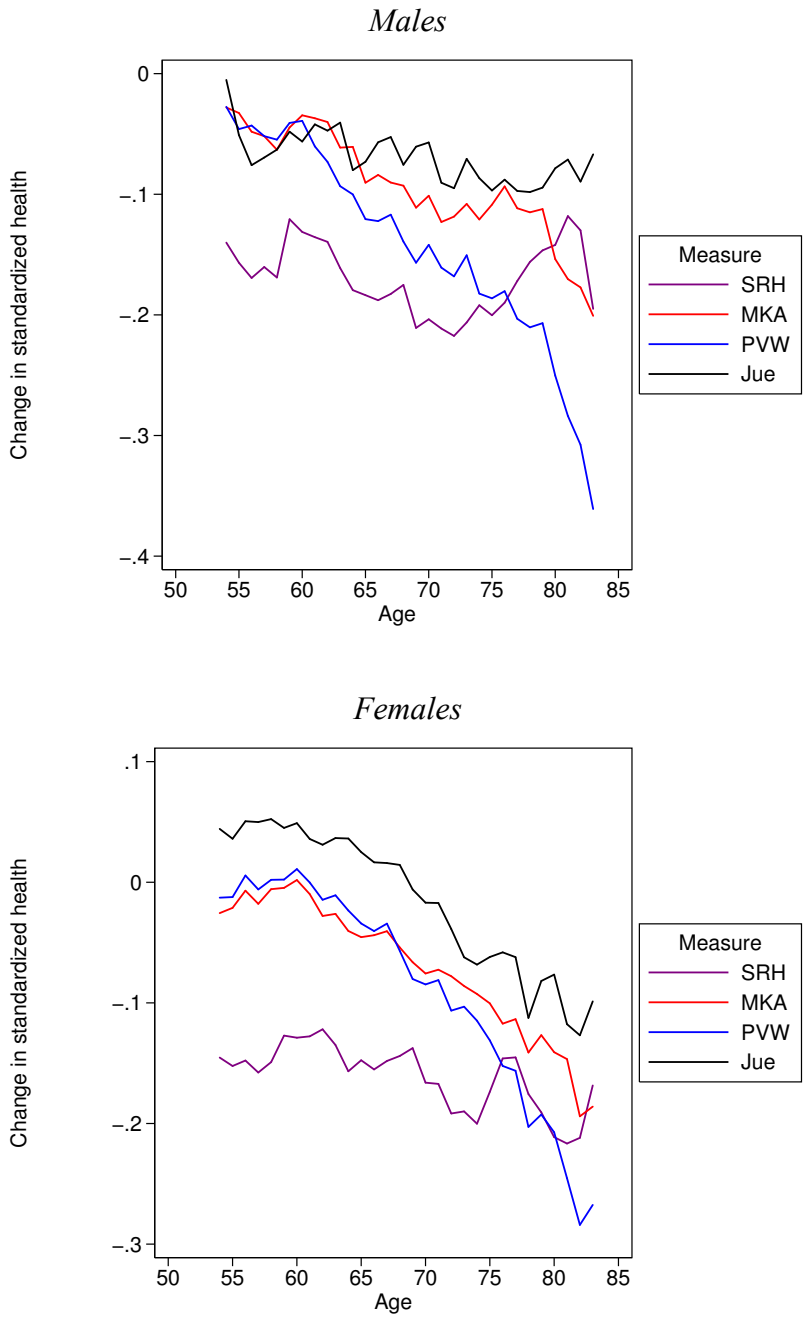


Figure 5. Mean change in standardized health by current age (5-year moving average centered at the given age)

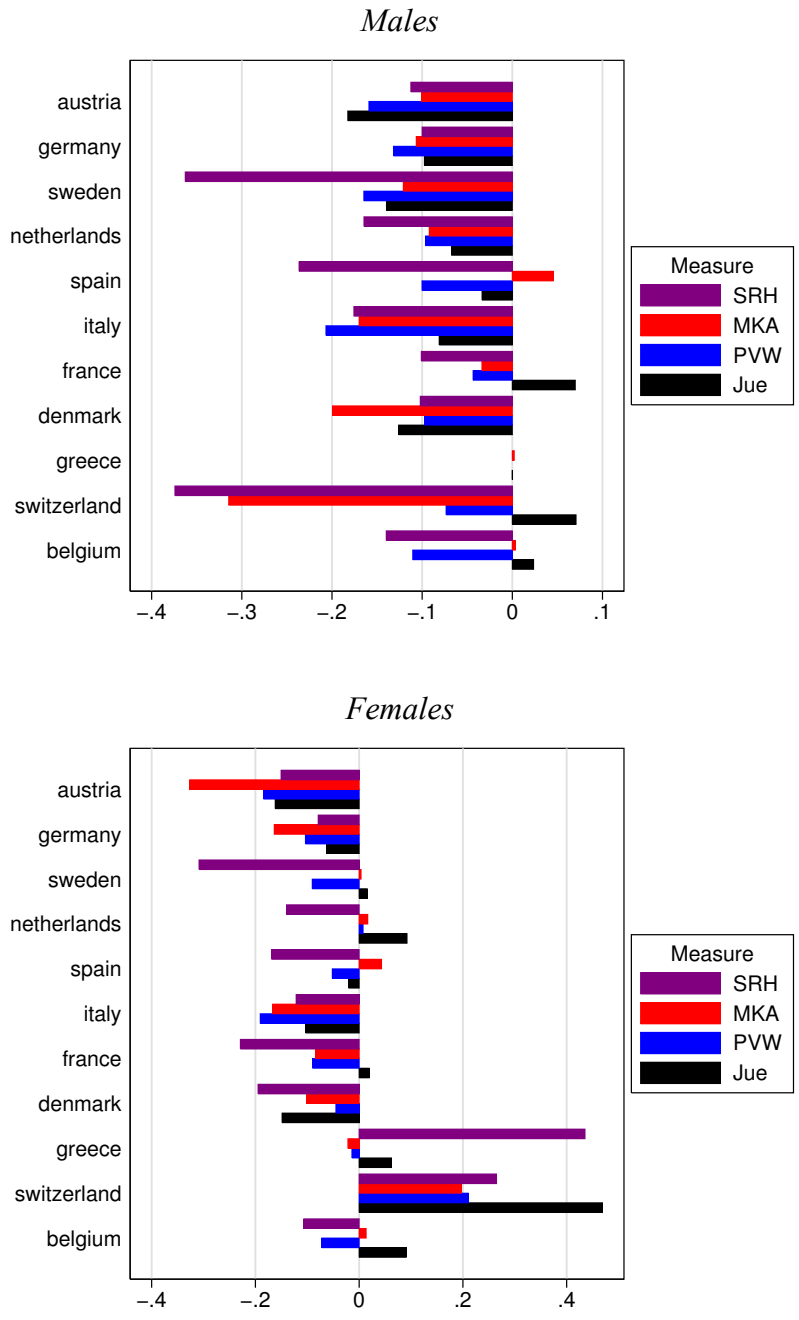


Figure 6. Mean change in standardized health by country

Appendix: The Effects of adding health conditions and health care utilization variables to the MKA index

Table A1: Becoming disabled vs. staying employed or self-employed, MKA index for health, males

	Health only	+demogr, labor, financ	+country dummies	+health conditions	+utilization
Health change	-0.0644*** (0.0118)	-0.0620*** (0.0116)	-0.0684*** (0.0121)	-0.0549*** (0.0110)	-0.0432*** (0.00997)
Lagged health	-0.0666*** (0.0113)	-0.0659*** (0.0118)	-0.0715*** (0.0123)	-0.0496*** (0.0109)	-0.0374*** (0.0100)
w1 heart attack				0.0662* (0.0337)	0.0481 (0.0300)
change heart attack				0.0800 (0.0412)	0.0622 (0.0385)
w1 high blood pressure				-0.0190 (0.0119)	-0.0214 (0.0116)
change high blood pressure				-0.00465 (0.0127)	-0.00837 (0.0130)
w1 high cholesterol				0.0105 (0.0152)	0.0119 (0.0145)
change high cholesterol				0.0126 (0.0125)	0.0129 (0.0122)
w1 stroke				0.322** (0.123)	0.294* (0.120)
change stroke				0.297* (0.140)	0.288* (0.137)
w1 diabetes				0.0121 (0.0303)	0.00873 (0.0298)
change diabetes				0.0129 (0.0274)	0.00969 (0.0231)
w1 chronic lung disease				-0.00403 (0.0426)	0.00560 (0.0393)
change chronic lung disease				0.0125 (0.0521)	0.0255 (0.0510)
w1 asthma				-0.0324 (0.0213)	-0.0334 (0.0191)
change asthma				-0.0706* (0.0345)	-0.0743* (0.0340)
w1 arthritis				-0.00843 (0.0230)	-0.0117 (0.0206)
change arthritis				-0.0174 (0.0204)	-0.0124 (0.0186)

	Health only	+demogr, labor, financ	+country dummies	+health conditions	+utilization
w1 osteoporosis				0.119 (0.103)	0.0944 (0.0873)
change osteoporosis				0.00371 (0.0905)	0.0152 (0.0813)
w1 cancer				0.0800 (0.0484)	0.0547 (0.0496)
change cancer				0.0547 (0.0687)	0.0286 (0.0657)
w1 ulcer				0.0271 (0.0485)	0.0109 (0.0406)
change ulcer				0.0341 (0.0531)	0.0244 (0.0447)
w1 parkinson disease				-0.0313 (0.0284)	-0.0188 (0.0281)
o.change parkinson disease				0 (.)	0 (.)
w1 cataracts				0.0175 (0.0626)	0.0000342 (0.0591)
change cataracts				-0.0585 (0.0682)	-0.0586 (0.0675)
w1 hip or femoral fracture				0.171 (0.169)	0.147 (0.173)
change hip or femoral fracture				0.175 (0.149)	0.139 (0.155)
w1 other conditions				0.0460* (0.0224)	0.0311 (0.0204)
change other conditions				0.0187 (0.0169)	0.00653 (0.0158)
w1 psychological problems				0.154** (0.0582)	0.155** (0.0564)
change psychological problems				0.139** (0.0537)	0.139** (0.0519)
w1 doctor visit					-0.00357 (0.00957)
change doctor visit					-0.00779 (0.00705)
w1 hospital stay					0.0694* (0.0307)
change hospital					0.0325

	Health only	+demogr, labor, financ	+country dummies	+health conditions	+utilization
stay					(0.0234)
w1 nursing home stay					0.887***
o.change nursing home stay					(0.0714) 0
w1 home care					(.) 0.325***
change home care					(0.0895) 0.278***
					(0.0753)
R^2	0.080	0.100	0.115	0.205	0.281
Observations	1721	1704	1704	1704	1704

Coefficients of demographic and economic variables and country dummies not shown.
Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Becoming disabled vs. staying employed or self-employed, MKA index for health, females

	Health only	+demogr, labor, financ	+country dummies	+health conditions	+utilization
Health change	-0.0447*** (0.0114)	-0.0477*** (0.0119)	-0.0515*** (0.0126)	-0.0365** (0.0122)	-0.0353** (0.0120)
Lagged health	-0.0312*** (0.00835)	-0.0321*** (0.00891)	-0.0385*** (0.00979)	-0.0231* (0.0104)	-0.0215* (0.0102)
w1 heart attack				0.0796 (0.0515)	0.0774 (0.0494)
change heart attack				0.0193 (0.0450)	0.0143 (0.0463)
w1 high blood pressure				-0.00952 (0.0107)	-0.00783 (0.0110)
change high blood pressure				-0.00649 (0.0142)	-0.00539 (0.0142)
w1 high cholesterol				-0.00459 (0.0138)	-0.00303 (0.0136)
change high cholesterol				-0.00423 (0.0117)	-0.00385 (0.0116)
w1 stroke				0.229 (0.120)	0.231 (0.121)
change stroke				0.285* (0.135)	0.285* (0.133)
w1 diabetes				0.00902 (0.0204)	0.00599 (0.0216)
change diabetes				0.0677 (0.0572)	0.0636 (0.0573)
w1 chronic lung disease				-0.0384 (0.0287)	-0.0385 (0.0307)
change chronic lung disease				0.0108 (0.0437)	0.00695 (0.0447)
w1 asthma				0.0142 (0.0295)	0.0158 (0.0297)
change asthma				0.0382 (0.0351)	0.0378 (0.0352)
w1 arthritis				0.0172 (0.0171)	0.0202 (0.0172)
change arthritis				0.00994 (0.0157)	0.0116 (0.0156)
w1 osteoporosis				-0.00780 (0.0266)	-0.00257 (0.0264)
change				-0.00920	-0.00503

	Health only	+demogr, labor, financ	+country dummies	+health conditions	+utilization
osteoporosis				(0.0278)	(0.0277)
w1 cancer				0.108*	0.0938*
change cancer				(0.0481)	(0.0471)
w1 ulcer				0.161**	0.145*
change ulcer				(0.0583)	(0.0569)
o.w1 parkinson disease				0.0134	0.0144
change parkinson disease				(0.0481)	(0.0478)
w1 cataracts				0.0371	0.0337
change cataracts				(0.0381)	(0.0370)
w1 hip or femoral fracture				0	0
change hip or femoral fracture				(.)	(.)
w1 other conditions				-0.377**	-0.364**
change other conditions				(0.142)	(0.140)
w1 psychological problems				-0.00566	-0.00939
change psychological problems				(0.0210)	(0.0189)
w1 doctor visit				0.0510	0.0458
change doctor visit				(0.0497)	(0.0442)
w1 hospital stay				-0.0810**	-0.0990**
change hospital stay				(0.0284)	(0.0364)
w1 nursing home				-0.0129	-0.0127
				(0.0280)	(0.0316)
				0.0114	0.0118
				(0.0168)	(0.0163)
				0.0161	0.0168
				(0.0147)	(0.0145)
				0.0403	0.0379
				(0.0213)	(0.0213)
				0.0295	0.0250
				(0.0251)	(0.0252)
					-0.0104
					(0.0117)
					-0.00421
					(0.00771)
					-0.00709
					(0.0271)
					0.00182
					(0.0236)
					-0.122

	Health only	+demogr, labor, financ	+country dummies	+health conditions	+utilization
stay					(0.0794)
change nursing home stay					-0.0424
w1 home care					(0.0505) 0.126*
change home care					(0.0629) 0.121*
Constant	0.0457*** (0.00885)	0.116* (0.0477)	0.143* (0.0586)	0.113 (0.0583)	0.108 (0.0582)
R^2	0.031	0.042	0.053	0.132	0.145
Observations	1485	1461	1461	1461	1461

Coefficients of demographic and economic variables and country dummies not shown.

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$