Volume Title: Discoveries in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-14609-X (cloth); 978-0-226-14609-6 (cloth); 978-0-226-14612-6 (EISBN)

Volume URL: http://www.nber.org/books/wise13-1

Conference Date: May 9-11, 2013

Publication Date: June 2014

Chapter Title: Comment on "A Comparison of Different Measures of Health and their Relation to Labor Force Transitions at Older Ages"

Chapter Author(s): Steven F. Venti

Chapter URL: http://www.nber.org/chapters/c12973

Chapter pages in book: (p. 151 - 156)

Comment

Steven F. Venti

The recent emergence of cross-national surveys based on a common research design has stimulated research on a number of important topics. Users of these data can take advantage of cross-national variation in public pension, health care, and disability programs to answer questions that were previously difficult to address using data from a single country. In many areas of comparative international research, health plays a key role. Because health is multidimensional it is often useful for researchers to summarize health in a single index that can be used to "explain" differences in nonhealth outcomes. A single index summarizing health is also valuable in its own right as a measure of what is happening to the health of the population. Kapteyn and Meijer (KM hereafter) have taken on the challenge of how to best summarize health in a single index using cross-national data. Their starting point is to compare the properties and performance of three indexes of health that have been proposed in the literature. Although they stop short of endorsing a particular index, their analysis provides a great deal of information about how to construct an improved index.

There are three issues that complicate the construction of a health index that can be used in cross-national comparisons. The first is the choice of a statistical model to summarize information on a variety of health measures (survey responses to questions about health difficulties and conditions) into a single index. The three statistical methodologies considered by the authors appear, at least superficially, to be quite different: a MIMIC model proposed by Meijer, Kapteyn, and Andreyeva (2011; MKA hereafter), an index based on self-reported health status (SRHS) adjusted for country effects proposed by Jürges (2007), and a principal components approach used by Poterba, Venti, and Wise (2013; PVW hereafter).

The second issue is the choice of which of many available health measures to include in the construction of an index. The original MKA and Jürges models use different subsets of the health measures available in SHARE for eleven European countries. The PVW model uses still another different subset of health measures available in the HRS for the United States. In both surveys there are a wide variety of health-related measures to choose from. Perhaps the most frequently used measure is self-reported health status (SRHS), a simple five-point scale describing the range of health from poor to excellent. As has been widely noted, country differences in SRHS reflect not just unobserved true health, but also systematic country reporting effects unrelated to true health. This shortcoming limits the usefulness of SRHS

Steven F. Venti is the DeWalt H. Ankeny '21 and Marie Ankeny Professor of Economic Policy and professor of economics at Dartmouth College and a research associate of the National Bureau of Economic Research.

For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see http://www.nber.org/chapters/c12973.ack.

Table 3C.1 Number of health measures used to construct each index

Health measure	MKA	PVW	Jürges
SRHS	1	2	no
ADL/IADL	**15**	1	no
Functional limitations	9	**9**	no
Health conditions	no	8	**16**
Utilization	no	4	no
Back pain	no	1	no
Grip strength	yes	no	yes
BMI	yes	yes	yes

in comparative analyses. As an alternative, a number of recent studies have developed indexes based on other, presumably more objective, health measures that are less susceptible to reporting bias. These include self-reported ADLs (e.g., difficulty bathing) and IADLs (e.g., difficulty shopping), self-reported functional limitations (e.g., difficulty climbing stairs), the prevalence of health conditions (e.g., has a doctor ever told you that you have diabetes?), or indicators of health utilization (e.g., number of doctor visits). These measures, though arguably more objective than SRHS, are still not immune to country-specific reporting bias. There are still other health measures such as mortality or grip strength for which systematic reporting errors are even less likely to be an issue. Table 3C.1 shows the health measures used to construct each of the three indexes KM analyze. There is some overlap, but the core subset of health measures used in each study is quite different.

The third issue that complicates the construction of a health index is how to account for country-specific reporting bias. As noted, cross-national variation in respondent-reported health measures may arise from two sources: differences in genuine health and systematic differences in the way residents of each country answer questions. Respondents asked to judge whether a task is "difficult" may apply different thresholds in different countries. Moreover, "difficult" may have different connotations when translated into different languages. The key is to distinguish genuine health effects from reporting bias. The most widely used way to do this is to assume that some health measures are "objective" and thus not affected by systematic reporting differences among countries. Given this assumption, one can interpret cross-national variation not explained by objective measures as reporting bias. The analysis by Jürges recognizes that SRHS is subjective, but implicitly assumes that the set of health measures used to predict SRHS are objective. Chief among these predictors are respondent reports of health conditions based on questions such as: "Has a doctor ever told you that you have high blood pressure?" If countries differ in access to health care or have different diagnostic thresholds for high blood pressure, then responses may systematically differ across countries or, to put it another way, persons in different

countries with the same true blood pressure will provide different answers to the question. The MKA analysis assumes most health measures, other than grip strength, may not be objective. They use grip strength (adjusted for height and weight) as a benchmark to assess reporting bias in other health measures. The problem with this approach is that grip strength reflects only one dimension of genuine health. If mortality, walking speed, biomarkers, or some other (assumed) objective measure is used to identify reporting bias, then the results might be quite different. The third methodology (PVW) simply does not account for reporting bias.

I have no way of judging if one set of assumptions about the objectivity of particular measures of health is superior to another. However, one way to gauge the extent of the general problem is to use all three methodologies—each with a different adjustment for reporting bias—to produce health indexes and see if there are differences in these indexes. Similarity of the three indexes would suggest that reporting bias is not a major problem. KM reestimate the three indexes using the SHARE data, but preserving differences in the choice of variables made in the original studies. Summary measures indicate that the health indexes based on the three approaches are broadly similar. All display the expected downward sloping age profile and the distributions have similar shapes. However, the correlations between the indexes are between 0.54 and 0.86 and the figures show some variation in standardized health by country. That all three methods of accounting for reporting bias do not produce the same level of health in each country may indicate that reporting bias is an issue. However, the country differences may also be the consequence of using very different sets of health measures in the three models as well as differences in genuine health.

The KM implementation of the PVW model combines all countries in a single principal components model. This implementation assumes cross-country comparability. The assumption can be relaxed by estimating separate principal component models for each of the eleven SHARE countries to determine if the factor loadings are similar across countries. I obtain these estimates using a list of twenty-five health measures that is close, but not an exact match, to the list used by KM in their pooled version of the PVW model. If health measures have large reporting bias components, I would expect the factor loadings to vary among countries. In table3C.2 I show the correlations between the factor loadings on each health measure for each pair of countries. All entries in the table exceed 0.9 and most exceed 0.95 indicating that the relationship between each health measure and the overall index (the first principal component) is similar across countries. This suggests that country-specific response bias is probably not a major concern, at least for this subset of health measures.

KM evaluate the performance of the three models by including the lagged index and the change in the index as explanatory variables in simple linear models of the transition into retirement or disability. They find that none

Table 3C.2 **Cross-country correlations for loadings on twenty-five health measures**

	Austria	Germany	Sweden	Netherlands	Spain	Italy	France	Denmark	Greece	Switzerland	Belgium
Austria	1	0.984	0.967	0.945	0.956	0.961	0.963	0.942	0.938	0.949	0.948
Germany		1	0.973	0.945	0.969	0.975	0.964	0.948	0.966	0.975	0.956
Sweden			1	0.973	0.965	0.953	0.967	0.976	0.938	0.947	0.961
Netherlands				1	0.945	0.912	0.963	0.976	0.905	0.939	0.963
Spain					1	0.974	0.963	0.959	0.972	0.965	0.959
Italy						1	0.965	0.939	0.974	0.944	0.957
France							1	0.962	0.940	0.941	0.985
Denmark								1	0.923	0.940	0.968
Greece									1	0.960	0.939
Switzerland										1	0.945
Belgium											1

of the indexes have a statistically significant effect on retirement if demographic, economic, and country effects are included in the model. Although these estimates provide no information about the relative power of the three indexes, I did find it surprising that health is unrelated to retirement. I estimated a similar regression (not reported) employing the PVW index, using similar (but not exact) controls for the United States using HRS data for 2004 and 2006. The resulting estimates were statistically significant and larger in magnitude than those reported for European countries in tables 3.9 and 3.10 of KM. The source of the apparent difference between the European and United States' results is worthy of further analysis.

The authors next evaluate the performance of the three health indexes in a model of the transition to disability. All three indexes are statistically significant in all versions of the model (with and without covariates) and the R^2s are considerably higher than in the retirement models. Interestingly, the PVW index has greater explanatory power than the other indexes. Is this outcome the result of the statistical model or the choice of health measures used to construct the index? Noting that the PVW model includes measures of health conditions and health utilization, but the MKA model does not, the authors reestimate the disability equation supplementing the MKA index with health conditions and utilization variables. They find both sets of variables provide considerable explanatory power, although many of the individual estimates of health conditions are not statistically significant, especially for males.

These results suggest that disparities in performance between the three health indexes have more to do with the list of health measures used to construct the index rather than on the statistical model used. The utilization of health services and, to a lesser extent, the prevalence of health conditions help to explain the transition to disability. Presumably these measures contain information about genuine health that is not contained in other measures such as ADLs, IADLs, or functional limitations. So why not use them to construct an index regardless of which statistical model is used? The authors suggest that using these variables would be problematic for causal analysis. They argue: "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." Their concern is justified, but going to the doctor may still be a useful, though imperfect, indicator of underlying health, particularly when used in a statistical model designed to extract the common factor (genuine health) from a large number of noisy health measures. A related concern raised by the authors is that utilization variables such as doctor visits or hospital stays confound institutional differences—in health care systems, in diagnostic standards, and so forth—with differences in genuine health. But genuine health and all of its other proxies, including ADLs, IADLs, diagnoses, grip strength, and even mortality, are to one degree or another, also "contaminated" by institutions. I thus find it difficult to draw

the line between those health measures that are too closely related to institutions and those that are not. Their line of demarcation is that it may be problematic to use an index based in part on utilization variables to explain differences in health across countries, but using the index to explain retirement transitions is probably acceptable. I would argue for wider use of utilization variables because differences in institutions tell us something about genuine health differences.

In sum, my reading of the results is that country-specific response bias—known to be a serious problem if SRHS is used to measure health status—is probably less an issue for health measures such as ADLs, IADLs, and health conditions. Second, the choice of statistical model is relatively unimportant. Although there is more work to be done in this area, I suspect that the three models will perform similarly if they are based on the same set of health measures. The choice of a statistical model will then depend on other criteria. The PVW and Jürges models have the advantage of being easy to compute. However the MKA model explicitly yields estimates of reporting bias for each of the health measures (given the assumption that grip strength is reported without error). A third finding is that the estimated health index is sensitive to the choice of health measures used to construct the index. As far as what variables should be included, it is clear that there is no "one size fits all" solution. Despite some legitimate concerns raised by the authors, my reading of their results is that measures of health utilization and the prevalence of health conditions provide valuable information about genuine health and, for most purposes, should be included in the health index.

The recent availability of cross-national surveys such as the HRS, ELSA, and SHARE have allowed analysts to exploit new sources of variation in trying to address key issues in the areas of retirement, public pensions, disability, and well-being (among others). But the benefits of cross-national data are not without some difficult challenges, one of which is how to process the health content in these surveys. This chapter makes an important contribution to our understanding of how to construct a health index that concisely summarizes the data and is comparable across countries.

References

Jürges H. 2007. "True Health vs Response Styles: Exploring Cross-Country Differences in Self-Reported Health." *Health Economics* 16:163–78.

Meijer, E., A. Kapteyn, and T. Andreyeva. 2011. "Internationally Comparable Health Indices." *Health Economics* 20:600–19.

Poterba, J. M., S. F. Venti, and D. A. Wise. 2013. "Health, Education, and the Post-Retirement Evolution of Household Assets." NBER Working Paper no. 18695, Cambridge, MA.