



Full length article

Adjusting for heterogeneous response thresholds in cross-country comparisons of self-reported health



Teresa Molina

University of Hawaii at Manoa, Department of Economics, 2424 Maile Way Saunders Hall, Honolulu 96822, HI, United States

ARTICLE INFO

Article history:

Available online 1 September 2017

JEL Classification Codes:

J14
I10
I15

Keywords:

Self-reported health
Anchoring vignettes
Aging surveys

ABSTRACT

Comparisons of subjective scale measures across countries can be distorted by the use of different response thresholds in different countries. Anchoring vignettes are a survey tool designed to address this problem, and are becoming widely used to adjust comparisons of subjective measures across groups, primarily within countries or across relatively similar high-income countries. This paper expands the existing literature by comparing six domains of self-reported health across the United States, England, and two countries with very different income levels, cultures, and geographic locations – Indonesia and China. In the raw data, respondents from the U.S. and England appear to be in worse health than their Indonesian and Chinese counterparts, but across the majority of health domains, this relationship completely reverses once I account for threshold differences.

© 2017 Elsevier B.V. All rights reserved.

Introduction

As populations around the world age, it is becoming increasingly important to learn about the quality of life of their growing elderly populations. Recent interest in aging has led to the proliferation of large-scale surveys of the elderly across the globe. These surveys collect a wealth of health measures, including chronic conditions, self-reported health status, and even biomarkers (BMI, lung capacity, blood pressure, etc.). With the vast amount of information available, it is not immediately clear exactly which variables can help make for more meaningful cross-country comparisons of elderly health status. For instance, though prevalence rates of chronic conditions may seem like appealing measures of health, measurement error in the form of under-diagnosis may correlate with income levels across countries. Similarly, although life expectancy is often used as an all-encompassing objective measure of health, it can often be more of a reflection of infant and child mortality rather than of the health of the elderly (Cutler et al., 2006): populations with higher life expectancy do not necessarily have healthier elderly populations. Finally, biomarkers may be too specific to capture the multidimensional nature of health and are often infeasible to measure for large populations.

Self-reported health offers an alternative which could complement the measures discussed above. One common self-reported

measure is an answer to the following question: *In general, how healthy do you feel?* – to which respondents can choose *excellent*, *very good*, *good*, *fair*, or *poor*.¹ General self-reported health measures like these may be better suited to encompass the multidimensional nature of health, and moreover, have been found to be significantly related to objective measures like mortality, even after controlling for demographic and socioeconomic covariates (Idler and Benyamini, 1997; DeSalvo et al., 2006). A more specific variant of this measure is based on questions about health in particular domains: mobility, pain, cognition, affect, breathing, and sleeping. For example, the mobility domain question asks, “Overall in the last 30 days, how much difficulty did you have moving around?” – to which respondents can respond *none*, *mild*, *moderate*, *severe*, or *extreme*. These questions are being asked in longitudinal aging surveys around the globe, including the United States, England, Indonesia, and China, the four countries I analyze and compare in this paper.

There are many reasons we may be interested in cross-country comparisons of health, including their potential to contribute to policy evaluations of different institutions across the globe. Unfortunately, despite the many advantages discussed above, comparisons of raw self-reported health measures can be problematic. First of all, questions and responses may

¹ Different surveys use different scales. For example, *very good*, *good*, *fair*, *bad*, *very bad* (in certain waves of ELISA) or *very healthy*, *somewhat healthy*, *somewhat unhealthy*, and *unhealthy* (in IFLS).

E-mail address: tmolina@hawaii.edu

translate differently in different languages. Moreover, even if all translations were identical, it is still possible that people from different countries interpret response choices and questions in different ways. Whether individuals perceive themselves as having “no difficulty” or “mild difficulty” with a particular health domain could rely heavily on the health-related expectations and reference groups that shape their interpretations of these terms. Within countries, sub-groups that have lower expectations for health (like the elderly) or are exposed to reference groups with lower average health levels (those of lower socioeconomic status) have been shown to use different response thresholds in their evaluations of health (Salomon et al., 2004; Dowd and Todd, 2011; Bago d’Uva et al., 2008a,b; Bago d’Uva et al., 2011; Molina, 2016). Populations of different countries do not only differ in terms of their demographic composition; they also come from different cultural and historical backgrounds and face different economic and geographic conditions, which likely affect their expectations about health and the reference groups they use when evaluating their own health. If people across countries use different response scales, it is impossible to tell, simply by comparing raw self-reports, whether observed differences reflect true underlying health disparities or simply differential reporting behavior. As a result, naive comparisons of self-reported health can lead to misleading policy conclusions, for instance regarding the relationship between national health expenditures and health levels (Sadana et al., 2002; Jürges, 2007).

In this paper, I use anchoring vignettes to deal with the use of different response thresholds, often referred to in this literature as differential item functioning (DIF).² The same set of 18 vignettes are included in the Indonesian Family Life Survey (IFLS), the Health and Retirement Study (HRS) from the United States, the English Longitudinal Study of Aging (ELSA), and the China Health and Retirement Longitudinal Study (CHARLS).³ These vignettes describe the health conditions of hypothetical individuals and ask survey respondents to rate the health of the individuals described. Used with various statistical techniques, these anchoring vignettes allow researchers to test for reporting heterogeneity across groups and adjust for this heterogeneity in comparisons of health levels across populations.

First developed by King et al. (2004), anchoring vignettes have been applied to cross-country comparisons in a wide range of topics, from political efficacy (King et al., 2004) to work disability (Kapteyn et al., 2007; Angelini et al., 2012). Several studies that share this paper’s focus on the well-being of the elderly have found strong evidence for cross-country reporting heterogeneity in the evaluation of satisfaction with social contacts (Bonsang and van Soest, 2012b), income and job satisfaction (Bonsang and van Soest, 2012a), and satisfaction with healthcare effectiveness (Sirven et al., 2012).⁴ All of these studies find that the ranking of countries by naive reported satisfaction levels differs substantially from the ranking of countries by satisfaction levels adjusted for reporting heterogeneity.

This paper focuses on a slightly different feature of elderly well-being: self-reported health in six domains (mobility, pain,

cognition, affect, sleeping, and breathing).⁵ Early studies of cross-country reporting heterogeneity in either general or domain-specific health have investigated heterogeneous reporting behavior without the use of anchoring vignettes (Jürges, 2007) or without a formal econometric model (Salomon et al., 2004). Although there are now many studies that use anchoring vignettes in the framework of a hierarchical ordered probit (HOPIT) model to study *within-country* reporting heterogeneity in domain-specific health (Bago d’Uva et al., 2008a,b; Dowd and Todd, 2011; Mu, 2014; Molina, 2016), evidence on *cross-country* reporting heterogeneity is still quite limited. By drawing on four separate but harmonized datasets, I am able to make comparisons across a culturally, geographically, and socioeconomically diverse group of countries.

This paper contributes to the existing literature by studying reporting heterogeneity in domain-specific health across countries that vary drastically in terms of geographic location, cultural and historical background, and levels of economic development. This diversity is particularly important in light of the existing cross-country literature’s almost exclusive reliance on the Survey of Health, Aging, and Retirement in Europe (SHARE). Although countries within Europe are certainly heterogeneous across many dimensions, their geographic proximity and related historical backgrounds limit the extent of the heterogeneity that can be studied using the SHARE.

I utilize vignettes from Indonesia and China, two countries far less well-explored in the vignettes literature, and adjust for heterogeneous response thresholds to make valid comparisons with the United States HRS and England’s ELSA, two aging surveys upon which the Indonesian and Chinese vignette modules were based. Consistent with the related literature, I find strong evidence for response heterogeneity across these four countries. Although respondents in the United States and England appear to be less healthy than their Indonesian and Chinese counterparts in the raw data, I find that accounting for the use of different thresholds completely reverses this result: adjusted self-reported health for the Americans and English is significantly better than that of the Indonesian and Chinese. Thresholds may vary across these countries for a number of reasons, including differences in average health levels, language, culture, or socioeconomic status. Understanding the individual contributions of each of these mechanisms, though outside the scope of this paper, promises to be an interesting next step for future research.

Vignette methodology

In this section, I outline how anchoring vignettes are used to correct for the use of different response thresholds in comparisons of subjective scale measures.

Empirical model

Like King et al. (2004) and many others that have followed, I use a hierarchical ordered probit (HOPIT) model. Consider a subjective response to the question, “Overall in the last 30 days, how much of a problem did you have with bodily aches or pains?” The categorical response to this question, which I will denote Y_i , can take a value from $j = 1$ to 5, where $j = 1$ represents “no difficulty,” $j = 2$ “mild difficulty,” $j = 3$ “moderate difficulty,” $j = 4$ “severe difficulty,” and $j = 5$ “extreme difficulty.” Y_i is determined by a latent variable Y_i^* , which is a function of individual respondent characteristics and an error term:

⁵ It should be noted that in this paper I deal with domain-specific self-reported health, not the general self-reported health measure more commonly used in existing literature. Anchoring vignettes for general health status are not available in these datasets.

² Originally a term used in the education testing literature, King et al. (2004) use this term to refer to heterogeneous response thresholds across individuals.

³ All four of these surveys either focus on the elderly or have large sample sizes of older individuals. As a result, my sample sizes are much larger than those in the widely used Survey of Health, Aging, and Retirement in Europe (SHARE), which also includes vignettes. As I describe on the next page, these four countries also represent much greater geographic, cultural, and socioeconomic diversity than the countries in the SHARE.

⁴ Though not the focus of their paper, van Soest and Vonkova (2014) also offer evidence for reporting heterogeneity in reported concentration levels of the elderly across countries.

Download English Version:

<https://daneshyari.com/en/article/5100157>

Download Persian Version:

<https://daneshyari.com/article/5100157>

[Daneshyari.com](https://daneshyari.com)