ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Banks, James; Smith, James P.

Working Paper

International comparisons in health economics: Evidence from aging studies

IZA Discussion Papers, No. 6297

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Banks, James; Smith, James P. (2012) : International comparisons in health economics: Evidence from aging studies, IZA Discussion Papers, No. 6297, Institute for the Study of Labor (IZA), Bonn,
https://nbn-resolving.de/urn:nbn:de:101:1-201203145710

This Version is available at:
https://hdl.handle.net/10419/58782

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 6297

International Comparisons in Health Economics: Evidence from Aging Studies

James Banks
James P. Smith

January 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

International Comparisons in Health Economics: Evidence from Aging Studies

James Banks
Institute for Fiscal Studies
and University of Manchester

James P. Smith
RAND Corporation
and IZA

Discussion Paper No. 6297
January 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 6297
January 2012

ABSTRACT

International Comparisons in Health Economics: Evidence from Aging Studies*

We provide an overview of the growing literature that uses micro-level data from multiple countries to investigate health outcomes, and their link to socioeconomic factors, at older ages. Since the data are at a comparatively young stage, much of the analysis is at an early stage and limited to a handful of countries, with analysis for the US and England being the most common. What is immediately apparent as we get better measures is that health differences between countries amongst those at older ages are real and large. Countries are ranked differently according to whether one considers life-expectancy, prevalence or incidence of one condition or another. And the magnitude of international disparities may vary according to whether measures utilize doctor diagnosed conditions or biomarker-based indicators of disease and poor health. But one key finding emerges – the US ranks poorly on all indicators with the exception of self-reported subjective health status.

JEL Classification: I0, H0

Keywords: international health, labor markets

Corresponding author:

James P. Smith
RAND
1776 Main Street
P.O. Box 2138
Santa Monica, CA 90407
USA
E-mail: smith@rand.org

* The research was funded by a grant from the National Institute on Aging (R37-AG025529). James Banks is grateful to the Economic and Social Research Council for co-funding through the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS.

Introduction

International comparisons of health outcomes are growing in importance as a method of gaining insight into the determinants and consequences of health status. In part this is because some institutions - such as the organization of health insurance, the provision of health care, welfare states or work place arrangements – vary more systematically and perhaps exogenously across countries than within countries. But also this is because of the emergence of a set of high quality and comparable international health and economic panel data in which economists have played a central role in the design process. These panel surveys allow econometric modeling with an international comparative lens that was previously not possible.

These international surveys and the subsequent economic research that has emerged from them have predominantly dealt with health and economic transitions around, and after, the retirement years. As well as being the years of life during which the exit from the labor force eventually takes place, these are also the years where the majority of transitions from good health to poor health, and then subsequently to death, take place.

Quite naturally then, these economic and health levels and transitions at older ages and the nature of the dual interactions between them are the main focus of our essay. Western countries, even those at relatively similar levels of development and national income, have a quite diverse set of policies in place to deal with the retirement and health behaviors of their individual populations and have experienced a quite diverse set of economic and health outcomes at the population level. This juxtaposition of a diverse set of policies alongside an equally diverse set of health, economic, and

labor force outcomes has been a signal to economists that research centering on the inter-relationships between them ought to be fruitful if the not inconsiderable methodological and data issues surrounding such international comparisons can be overcome.

The scope of our essay is limited in at least two important ways. First, we focus our attention primarily on international comparisons within industrialized Western countries. There is an excellent set of surveys by economists on developing countries where other issues, such as malnutrition for example, play a critical role (Strauss and Thomas, 1998 and 2008). Second, as mentioned above, our focus also rests primarily on health in mid and later life and not in the context of early life development, except of course when these early life health conditions play an important contributory role to later life health.

This essay is divided into four sections. The next section briefly describes the set of new aging surveys that are facilitating comparative international studies by economists of health related issues at older ages. Section 2 presents data and compares patterns of mortality and prevalence-based measures of health outcomes for a set of European countries and the United States based on that data. Section 3 moves the discussion on to discuss the onset of disease and incident mortality, where analyses and modelling are much more amenable to interpretations aimed at detecting possible causal effects. This section also discusses alternative analytical methods that have been used to explain the health differences that emerge. Section 4 highlights our main conclusions.

1. New data on aging populations

Until relatively recently, the major limitation to conducting comparable international research on health using micro-data has been that there was almost no such data, certainly if one adds the proviso of comparably defined data. The Panel Study of Income Dynamics (PSID) had some excellent European counterparts with the British Household Panel Study (BHPS) and the German Social-Economic Panel (GESOP) being the most prominent examples but until recently these all-age panel studies had little information on health. And many of the other European panels had sustainability issues due to lack of funding and high attrition. For example, the European Community Household Panel (ECHP) was a panel survey with economic and health information that lasted only from 1994-2001 in part due to the cumulative impact on non-retention in the study.[1]

Of course, many Western countries have been collecting population health surveys for some time as part of their yearly health monitoring systems. Good examples would include NHANES and NHIS in the United States and the Health Surveys for England (HSEs) in England, but they exist in other countries as well, albeit usually for a shorter period of time. Yet several factors have limited the value of such data to economists interested in health comparisons. Firstly, whilst the health information can often be state of the art with biological samples and anthropometric measures alongside detailed self-reported data on the main indicators of health status, health insurance, and health care utilization, the economic information in these surveys is typically quite limited. Second, these surveys are typically cross sectional rather than longitudinal. Third, the more detailed measures collected are rarely comparable, either across

[1] See http://circa.europa.eu/irc/dsis/echpanel

countries or even within countries over time since health monitoring agencies have typically placed less emphasis on consistent time series of data than economic agencies did. Finally, academic epidemiological surveys are often not national in scope and are typically not even publicly available given the very slowly changing traditions of that discipline regarding data sharing.

Thus, there was somewhat of a data vacuum for those interested in modeling the relationship between economic and health outcomes around the world. In the most notable example, perhaps, rapid rates of population aging around the world in particular led to a demand from policymakers and academics alike for analysis that would help understand the linked economic and health trajectories as individuals age on a comparable basis around the world. The data constraints on such analysis were clear (see National Academy of Science (2001) for example) and, as a result, a set of high quality harmonized data sets has now evolved in order to detail the changing health, economic status, work, and family relations of immediately pre- and post retirement populations. Importantly, economists played a prominent but not exclusive leadership role in designing and implementing these studies. As a result, these aging surveys have been designed specifically to enable researchers to monitor and model impacts of new health and retirement programs and policies that may affect the incentives of the older population when it come to how long they continue to work, how they obtain and finance their health care and whether it is effective, and whether they are able to achieve adequate incomes during their retirement years.

This set of surveys started with the Health and Retirement Study (HRS) in the United States. HRS was originally a panel study of those 51-61 years old in 1991 using a

two-year periodicity to monitor economic and health transitions (particularly retirement and the onset of poor health conditions) in the subsequent years and the manner in which these economic and health domains mutually influence each other at older ages.[2] The scope of the study has expanded significantly in the subsequent 20 years, in part by adding older and younger birth cohorts so that HRS now attempts to be continuously population representative of Americans who are at least 50 years old.

HRS has spawned international surveys in 24 other countries that share a common scientific and policy mission with a mutual desire to harmonize some of their main survey content. Internationally comprable surveys currently include the English Longitudinal Survey of Ageing (ELSA) in England, TILDA in Ireland, 15 countries in the SHARE continental European network,MHAS in Mexico, and six surveys in Asia—IFLS in Indonesia, KLoSA in South Korea, CHARLS in China, LASI in India, HART in Thailand, and JSTAR in Japan. Plans are also under way for comparable studies in Brazil and Argentina so that a South American counterpart is on the horizon. The HRS set of international surveys now cover countries representing more than half of the world's population. All surveys have adopted a two year periodicity and cover populations with either a lower age limit of 50 in the Western countries or age 45 in the Asian countries as work and health transitions often begin at an earlier age there.

The explicit aim of these studies is to have significant comparable content so that cross-national analytical studies can be conducted while allowing scientific innovation at the country level. But the content also has to reflect the reality and policies of each country. Finally, all participating countries in this aging network have committed and

[2] The HRS has been primarily funded by the Behavioral and Social Science branch (BSR) at the NIA with significant supplemental funding from the Social Security Administration.

adhered to widespread and quick release of data into the public domain, both within their own country and to the international scientific community. This was a major departure from the norm in some countries and may well be the most important legacy of this network.

2. Health Outcomes in International Comparative Research:

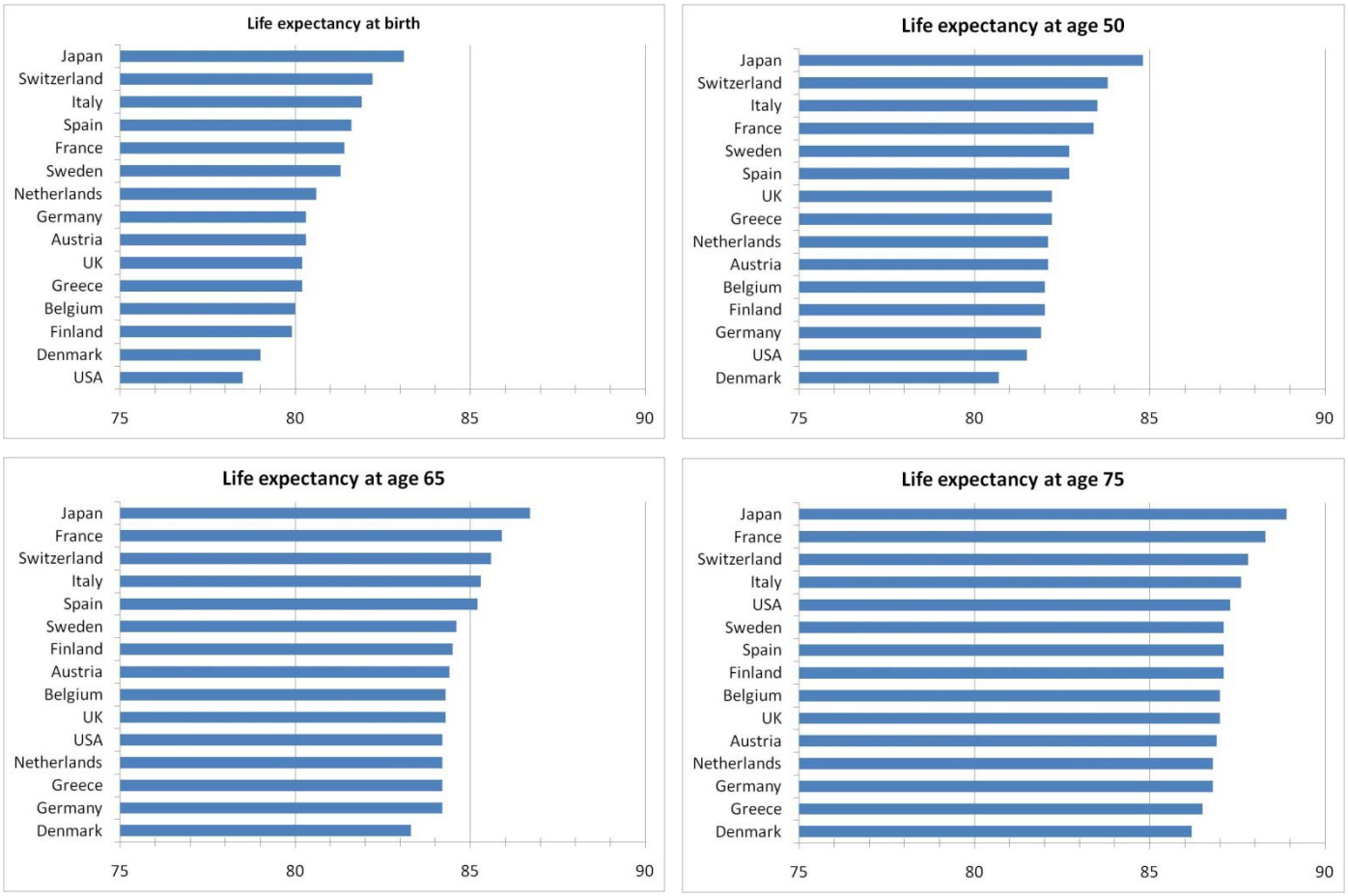
2.a. Mortality

By far the health outcomes that have received the most attention in international comparative research by economists and others are various measures of mortality. Indeed, mortality has several important virtues as a health outcome especially in an international context. Since it is one of the primary indicators of national well-being, most countries have devoted considerable effort over decades through their collection of vital statistics to accuracy of its measurement. As a result, mortality measurement has improved steadily over time. Mortality is the most objective of health measures and has the considerable if obvious virtue of meaning precisely the same thing in all countries of the world. Since reliable life tables exist for most Western Industrialized countries, this health outcome can be defined within all segments of the life cycle and indeed for single year of age if necessary. Since causes of death vary considerably over life from infancy to old age, separating out these unique stages of life is critical. Through the efforts of demographers (e.g. the Berkeley Human Mortality Database[3]) and international health organizations (such as WHO[4]), reasonably long time series of mortality and life-expectancy data are now readily available especially for the Western Industrialized

[3]http://www.demog.berkeley.edu/~bmd/overview.html
[4]http://www.who.int/topics/life_expectancy/en/

countries. Finally for many countries in these data basis, cause of death is also available, which considerably extends the range of hypotheses that can be tested.

Figure 1: International differences in life-expectancy by age



Source: World Health Organization, 2009. Note: Raw data for these figures are presented in Table A1

To illustrate the patterns, Figure 1 charts levels of life expectancies for fifteen Western Industrialized countries evaluated at four specific ages- birth, 50, 65, and 75. Even within these high income western countries, there is considerable variation amongst them in the remaining length of life. What stands out most starkly, however, is that despite being the highest income country on the list, the United States is at the

bottom of the group at birth and only one step above the bottom at age 50. However, the relative position of the US steadily improves thereafter until by age 75 America ranks in the top tier of this group of countries.

A recent report of the National Academy of Science in the United States, a panel on which we were both members, conducted a comprehensive study of the reasons for international differences in life-expectancy in high income Western countries as well as the reasons for observed differences in the trends over time (Crimmins, Preston, and Cohen, 2011). A particular focus of that report concerned what might explain the low expectancy of three poorly performing countries - the United States, Denmark, and the Netherlands. According to that study, the strongest evidence pointed to smoking as a key culprit, accounting for 78% of the gap in female (and 41% for men) life expectancy between the US and other high income countries in 2003. Of all the factors, smoking was most consistent with international differences in levels and timing of mortality trends.

Another factor thought to play a significant if secondary role was obesity and perhaps variation in exercise although comparable international measurement of exercise is difficult so it is hard to establish a firm position on the role of exercise. Factors that were not found to be all that important were differences in social ties and integration and inequality in spite of the prominent role that these two factors have played in the epidemiological literature. These types of findings that challenge the set of existing priors of epidemiological research (for example a large negative role for income inequality- see Wilkinson, 1996) are a good example of the tensions and value created by bringing economists into these areas of traditional epidemiological research (Deaton

and Paxson (2004) and Banks, Berkman and Smith, 2011). Finally, the much maligned health care system in the United States in light of its high cost relative to other western countries was actually thought to have improved mortality outcomes in America relative to these other countries and to have played a significant role in the steady rise in the ranking of American life expectancy with age that was revealed in Figure 1.

Despite its strengths, life expectancy is not the same as current mortality and has its limitations as a health outcome – life expectancy is related to the past deaths of cohorts at older ages and does not necessarily reflect the current or future mortality prospects of today's cohorts at older ages. This could be especially important when evaluating the role of risk factors to future mortality and also when comparing across countries when the past histories of cohorts may differ. In addition, most of the international comparative research with some aspect of mortality as the health outcome has relied on time series aggregate data for countries so has been largely macro in nature with the strengths and limitations of that style of research. The list of explanatory variables that can be considered is typically rather small and heterogeneous effects among important sub-groups in the population difficult to isolate. Individual level mortality outcomes, as opposed to aggregate life expectancy, can be one of the outcomes analysed with the new set of aging surveys, but that requires relatively long panels that are not scarred by excessive rates of attrition.

2.b. Subjective Health Status

Despite its importance as a central measure of health in international comparative research, the almost complete reliance on mortality as the health index has

been a problem as well. People may die at very old ages, but that may be little consolation if the quality of their lives were marred by long episodes of multiple illness and disease and difficulty in functioning in the everyday activities of life. After mortality, perhaps the most widely used health index in international comparative research has been some form of subjective health status- individuals' self-evaluation of their overall health. Many indexes have been used to compare the health of nations but a very common method is based on subjective scales of general health status (GHS). The appeal of this type of index is partly based on an attempt to summarize a very multidimensional concept such a health in an intuitively appealing and simple manner. There is also a considerable body of evidence that, at least within a country, these GHS scales may even out perform more clinically based measures in predicting future health outcomes (Ware et al, 1978 Erdogan-Ciftci et al, 2010, Kippersluis et al, 2010).

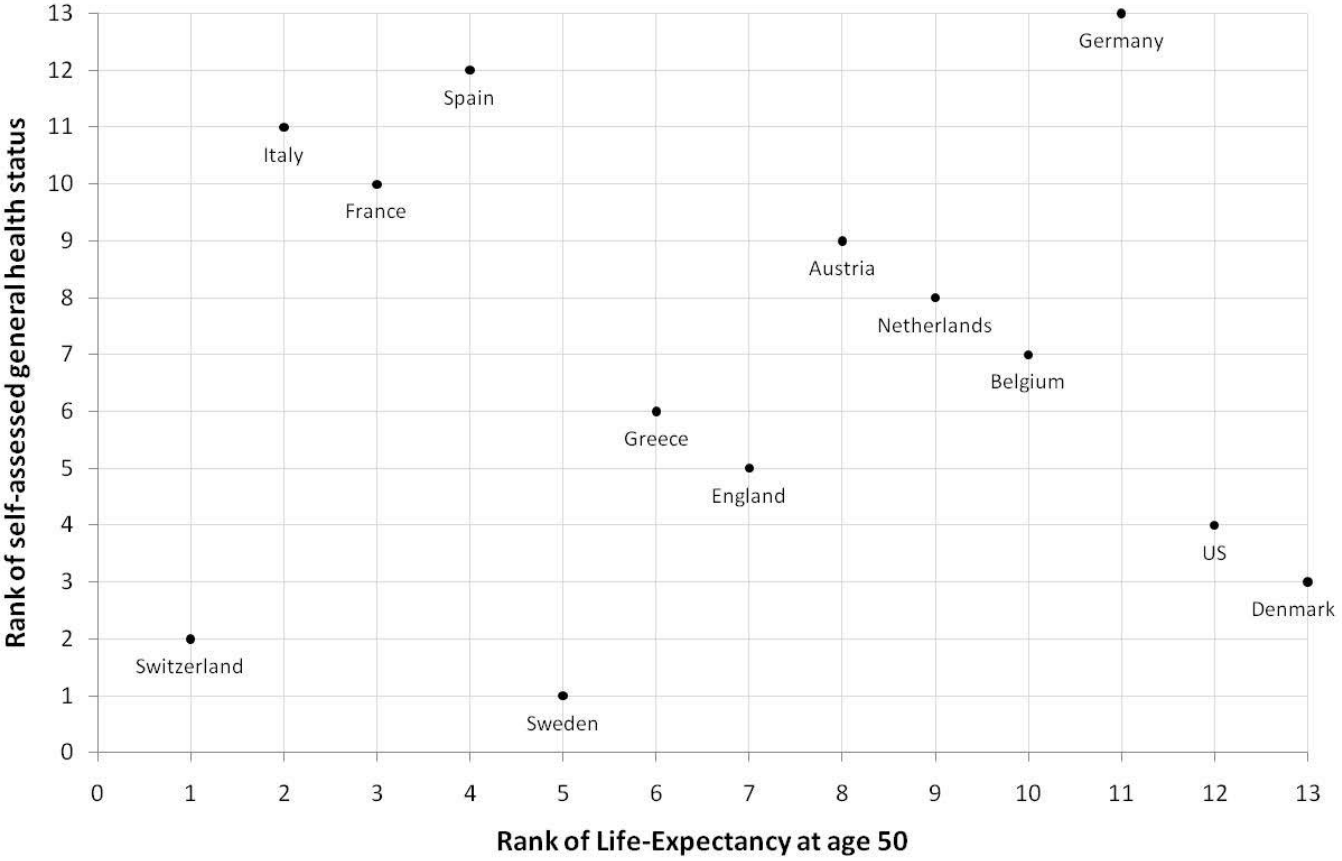
While there are several variants, a standard metric attempts to capture that most global aspect of general health status by asking people to evaluate their health on a five-point scale—excellent, very good, good, fair and poor, a scale that was included in HRS, ELSA, and SHARE. But even for industrial Western European countries at roughly similar level of development, data based on subjective scales about the general state of one's health apparently produce unusually large variation in health as shown in Table A2 in the Appendix which shows the distribution of answers to the Self-Assessed General Health Question amongst those aged 55-64 years in the US and twelve European countries. In fact there is even a serious question of whether or not such subjective health scales even rank countries correctly based on their health. For

example, the data on general health status in Table A2 stand in sharp contrast to WHO data on life expectancy presented in Figure 1 above.

To illustrate, Figure 2 shows the correlation in country rankings of life-expectancy at age 50 and a simple cardinal index of self-assessed health constructed from the data presented in Table A2. Some countries lie close to the diagonal of this figure, i.e. they are ranked similarly in both self-assessed health and life-expectancy. But the correlation is far from perfect. One cluster of countries – Italy, France and Spain – rank very low in the subjective health distribution but high in the life-expectancy distribution. And a number of other countries – the US and Denmark in particular, but also Belgium and Sweden – have a much more positive subjective assessment of their health than would be indicated by their life-expectancy. Overall, the correlation between the two widely used and basic measures of overall health appears to be non-existent (-0.08).

To give another example, using self-reported health scales American men rate themselves as healthier than their English counterparts(Banks, Marmot, Oldfield, and Smith. (2009))Nor are the differences between the two countries trivial—the proportion of English men reporting bad health is eight percentage points higher than it is in the US. Controlling for education or income does not eliminate the contradiction—in every education-income, a higher fraction of American men report good health than do their English counterparts.

Figure 2: Self-Assessed Health and Life-Expectancy



Notes: Life-expectancy data from World Health Organization 2009; Self-Assessed General Health Status data from ELSA, HRS and SHARE, 2004, individuals aged 55-64 (see notes to Table A2)

This apparent ranking of age adjusted American health as being superior to that of the English stands in sharp contrast to evidence that we will show in the next sections based on other more specific measures of health such as the prevalence of specific diseases. The apparent contradiction between these two standard measures of health status—self reported disease prevalence and self-reported health status— in particular raises questions for international comparisons of which provides the more reliable index and why the contradiction exists in the first place. The same type of problem occurs in many other applications within and outside of health - rating of functional difficulties or

work disability, overall subjective wellbeing or life satisfaction, or assessments of poverty and needs are a few examples (Kapteyn et al, 2007).

One method that economists have used with increasing frequency for international comparisons of subjective health is the use of anchoring vignettes (King et al, 2004). These are motivated by the premise that answers to any type of subjective questions by themselves cannot separate two distinct phenomena, namely true 'objective' differences between groups and differences created by the use of different response scales by each group. While the possibility of response scale differences amongst people certainly applies to within country analysis as well, it is viewed as a more critical issue for cross-country research where systematic cultural differences loom much larger.

It is the separation of objective differences from response scale differences that is the primary goal of the vignettes methodology. Vignettes are simple short stories of three to five sentences in length that describe a hypothetical person say in their health and respondents in surveys are asked to evaluate the health of the hypothetical person in the same manner and using the same scale as when they are asked to evaluate their own health. Variation in responses to vignettes can then be used to identify differences in response scales across people in different countries. With response scale differences identified, we can then go back to individual responses about their own health and adjust them to put them onto a common response scale. The remaining variation in health across countries after applying a common response scale would then constitute 'objective' variation in health across countries.

The two uses of the word ‘same’ in the previous paragraph highlights the two assumptions necessary to use vignettes for identification of response scale differences (Van Soest et al, 2011a, 2011.c). These two critical assumptions are referred to as Response Consistency and Vignette Equivalence. The assumption of Response Consistency states that individuals use the same set of subjective thresholds when they describe the vignette persons as they used on themselves while Vignette Equivalence refers to the assumption that all respondents interpret the vignettes as meaning the same thing. Van Soest et al (2011c) attempt to test response consistency by giving respondents (without their explicit knowledge) a subset of vignettes that directly mirror their own situation. These ‘replica’ vignettes were placed within a much larger set of other vignettes. The elements of self-description in the vignette come from a series of questions in a previous module relating to the respondent’s health. If the replica vignettes did in fact replicate the person well enough, then respondents should answer questions about themselves and the replica vignette persons the same. Van Soest et al (2011.c) performed a series of tests and found the vignettes worked well in terms of response consistency assumption in some domains (sleep, mobility, and affect) but not others (concentration and memory). Similarly, Bago D’Uva et al (2009) show that cognitive function and mobility vignettes don’t perform well when validated against objective measures of those dimensions.

The main distinguishing characteristic appears to rest on whether the problems in a particular domain can be captured in relatively few straight forward ways in a vignette (getting to sleep, duration of sleep, feeling tired when awakening) or is much more complex that that with too many aspects of a problem left unsaid (cognitive ability).

Vignettes are potentially a very useful tool but should not be viewed as settled science. A good deal more experimentation and testing are needed to have vignettes that capture the essence of the issues in some domains. The challenge for applying vignette-based methods to general health status is that it is a summary of many multidimensional domains of health simultaneously. At this stage of the science and development of vignettes, subjective measures of general health status as the primary health outcome in international comparative research should probably be avoided. It is not known at this point whether different reporting thresholds can explain the contradiction between better self-reported health in the US compared to England in spite of much higher prevalence and incidence of disease in the US.

In addition, the exercise of aiming to use vignettes to 'correct' subjective differences in order to conform with objective differences in measured health across countries should not be taken to mean that these subjective differences do not matter for behavior, nor that subjective assessments of health are not a legitimate outcome for analysis. Consider two people with precisely the same measured health or health behavior yet where one puts himself at a different point on a subjective scale than another – two individuals with high blood pressure, for example, but where one describes his health as poor and the other as very good. Some analytical consideration of why such a situation should occur would seem to be warranted and in this sense, individuals' responses to the vignette questions become objects of interest in their own right. To what extent do they capture differences in group-specific norms or even some measure of how an individual's 'expectation' of their health, and if they do to what extent do these norms and expectations drive health subsequent behaviors?

Taking the example of reporting differences across income or education groups within a country, it seems entirely plausible that individuals coming from a high income or education group, whose parents, family members and friends have always been fit and in good health, will assess a given set of poor-health conditions as 'worse' than a respondent of lower socioeconomic position for whom such poor health may be more of an everyday encounter. But in the international context, where issues such as cultural and linguistic differences in subjective scales cannot be easily swept under the carpet, variation in vignette responses across countries could be due to either of these sets of issues, or more likely some combination of the two. As such, there is clearly room for more work to be done on the topic of vignettes for subjective health comparisons.

2.c. Morbidity: Self-reported Prevalence of Disease

A fundamental measure of health status relates to the presence of disease. For prior lifetime prevalence, the new aging surveys all collect data on individual self-reports of specific diseases of the general form 'Did a doctor *ever* tell you that you had …'.The specific diseases that are included vary somewhat amongst the surveys but normally include diabetes, hypertension, heart disease, heart attack, stroke, chronic lung diseases, and cancer. To illustrate, Table 1 lists disease prevalence as obtained from the HRS, ELSA, and SHARE surveys for four of these diseases- diabetes, stroke, cancer, and lung disease among those ages 55-64. Once again, while we see considerable variation across this set of Western Countries in disease prevalence in middle age, the eye-catching statistic is that America ranks at the bottom in all four types of disease prevalence. These results confirm the findings of Avendano et al (2009) who used the

same data but a slightly different age group and set of countries (with the European countries pooled into one region) to look at differences in health and in the extent of the health-wealth gradient between each of the three regions.

Table 1
Unadjusted prevalence of disease (%) – Ages 55-64

	Diabetes	Stroke	Cancer	Lung Disease
Austria	7.59	2.65	3.74	2.20
Denmark	6.88	3.44	5.53	6.41
France	9.56	2.21	4.00	4.09
Germany	9.51	3.25	6.33	4.09
Greece	7.75	1.43	1.31	2.70
Italy	8.76	2.11	4.28	4.73
Netherlands	6.94	3.16	5.19	6.02
Spain	12.63	1.61	3.64	4.56
Sweden	7.37	2.95	4.36	2.68
Switzerland	5.24	0.96	5.85	3.39
Belgium	7.63	2.68	6.55	4.80
England	6.23	2.15	6.26	5.19
USA	14.93	4.43	8.50	7.21

Source: HRS, ELSA, and SHARE. 2004 data; All respondents aged 55-64

In a recent paper (Banks, Muriel and Smith, 2010), we looked more deeply into these large disparities in health focusing on the United States and England, looking at more extensive data on disease prevalence for those 55-64 years old and those ages 70-80. A number of points are worth mentioning here. Firstly, minority populations (African-Americans and Hispanics in the US and the non-white population in England) were excluded from the analysis and the results were unaffected – the differing size and health of minority populations is not the reason for the much higher rates of disease in America. Second, given the sharp age gradient in disease, prevalence of all diseases is of course much higher among those in their seventies. However, the same country ranking prevails – for cancer, diabetes, and heart problems, American excess disease is equally

large (in absolute value) in the older of the two age groups considered. Before discussing attempts at explaining such country variation in disease prevalence we briefly pause once again to assess how much of these differences are real.

2.d. Morbidity: Biomarker based evidence

The conventional approach to obtaining information on presence of disease by simply asking respondents whether or not they have been diagnosed with a set of important chronic diseases can present problems for an international comparative analysis. Undiagnosed disease varies considerably even within western industrialised countries (Gakidou et al, 2011) and the problem is much worse if one also considers developing countries. In addition, thresholds for disease diagnosis may vary across countries and over time since disease thresholds are being periodically revaluated usually with a lowering of disease thresholds (Smith, 2007.b).

Collection of biomarkers data – the broad name used to encompass data taken from biological samples (typically blood, saliva, urine or hair) along with anthropometric measurements and performance tests – has been proposed in order to address this issue. Indeed, the need for biomarkers in international health surveys is now well understood (National Academy of Science, 2007). In response to this need, many of the new international health/retirement surveys (HRS, ELSA, TILDA, IFLS, CHARLS, LASI) have included or will include biomarkers in the set of information to be collected. Biomarkers common to most or all surveys are measurements of blood pressure, height, weight and waist circumference, measures of glycosylated hemoglobin (HbA1c) to indicate Diabetes, lipids (Total and HDL cholesterol) and inflammatory markers related

to the risk of cardiovascular disease (C-reactive Protein and Fibrinogen). In developing countries where low iron deficiency is a major health risk, tests for hemoglobin are also conducted and other surveys have also collected further biomarkers in areas of particular research interest to the survey teams. Finally, in a very limited set of these surveys (HRS, ELSA, TILDA), genetic material is now being collected as well but the scientific return from this investment is not yet known. These surveys vary and continue to debate whether these markers are best obtained through dry blood spots (less expensive, easier to obtain, but with an unstable set of current providers) or venous blood (McDade, 2011).[5]

In addition to these measurements and biological samples, physical performance and cognition tests are usually administered in these surveys with the same overall goal of obtaining objective measures of these dimensions of health. Physical performance tests typically include grip strength, walking speed, balance, lung function test and timed chair rises, and cognitive function tests include short form and adaptive tests for aspects such as memory, numerical ability, verbal fluency, and fluid intelligence (Banks et al, 2011).

Of all the countries discussed above, at the present time the biomarker data are most complete and comparable for the ELSA and HRS samples. In Banks et al (2006) we used this to show conclusively that the health disparities between the two countries

[5]There is still much that has to be done in the adoption of biomarkers for international comparisons. Continental Europe, as represented by the SHARE study, has lagged behind the other international surveys in the development of biomarker data.And it is also known that differences in labs could affect evaluation of the assays,which is much more likely to be a problem for international comparisons where protocols and labs differ across surveys. But plans are underway to address both these issues as the field moves forward. Biomarkers are scheduled for introduction into the data collection for some subset of the SHARE countries and plans are being made for the same blood or saliva samples to be sent to the multiple labs used in different countries that are employed in the major international surveys in order to test lab effects.

followed the self-reported doctor diagnosed diseases as opposed to the self-reported general health status. In the population aged 40 to 70, the fraction with HbA1c levels of greater than 6.5% which is the level indicating diabetes was 6.4% of the sample in the US compared with 3.8% in England. And looking at C-reactive Protein, the fraction with levels of CRP that are classified as indicating a high risk of subsequent cardiovascular disease was 40.1% and 30.4% in the US and England respectively. In these cases, and also in the case of the other biomarkers considered (hypertension, cholesterol and fibrinogen) the biomarker indicators pointed to the same conclusion as the prevalence data in Table 1 - a significant excess disease among the Americans compared to the English.

As far as an economic analysis is concerned, the greatest value of biomarkers may not be simply to obtain a more accurate assessment of the presence of a disease given some threshold (and by doing so obtaining important knowledge about undiagnosed disease). Instead it may prove to be their value as sub-clinical indicators of poor health, i.e. their role as predictors of future onsets of disease prior to a clinical diagnosis, that holds an exciting potential for future analysis. One of the distinguishing characteristics of a microeconometric as opposed to an epidemiological modelling of the dynamics of health outcomes at older ages is an emphasis on an integrated understanding of risk factors, individual information and expectations, and individual behavioural choices in the health and economic domains. Whilst biomarkers may not represent risk factors in a strictly exogenous sense, they may nevertheless provide useful information on health risks, particularly when combined with respondents' subjective assessments of their health and subjective assessments of the probabilities of future

health, mortality and disability. As a simple example, consider two individuals with a similar ‘biomarker’ risk of future heart disease but with differing subjective assessments of the chances that their health will limit their future ability to work. Following these two individuals in longitudinal data and looking for differing reactions to subsequent health events would be instructive with regard to distinguishing the effects of health changes from health ‘shocks’. More generally, building objective biomarker-based health data into dynamic models may be a way of summarizing health histories parsimoniously for the purposes of international comparative analysis. And putting together expectations and risks into a dynamic structural model of health and economic behaviour, whilst a challenging agenda methodologically, would seem to offer interesting potential for a more causal analysis of international differences in many economic decisions such as the timing of retirement, intervivos transfers or bequests, and asset accumulation and decumulation in anticipation of potential health and long-term care costs.

2.e. Morbidity: Accounting for prevalence differences

Even using a comparative international approach, prevalence data does not naturally lend itself to identifying possible causal factors that may account for differences in illness. A disease prevalence outcome is inherently a cumulative measure of a past history, the precise timing of which may not be known. Analysis that aims to identify causal factors is much better targeted at incidence analysis either for disease or mortality, which we will discuss below. Nevertheless, multivariate analysis of international differences in prevalence of disease can still play an important role in

ruling out potential ‘explanations’ of country differences, particularly those where there are differences in the level and distribution of known risk factors for general health or specific diseases.

In addition to demographic controls such as age, gender, and ethnicity, the minimum set of risk factors in a disease prevalence analysis usually include some measure of SES-education and/or income- and behavioral risk factors such as smoking, obesity, exercise, and drinking. Two standard risk factors for disease prevalence relate to measures of smoking behavior and excess weight, such as being overweight or obese. Their frequent use is partly a result of reasonably comparable international measurement as a set of simple BMI thresholds have been almost universally adopted in Western countries. A different and somewhat lower set of thresholds are increasingly being adopted in Asian countries (Lee and Smith, 2011).

Table 3 lists levels of obesity, and current smoking in our set of Western Industrialized countries. In the right hand panel of the table, we rank the countries according to these two risk factors, according to the prevalence of two of the more major diseases (using the data in Table 1 above) and according to Life Expectancy at age 50. This table illustrates well the complications introduced to the picture when disease prevalences are added to general health outcomes. Perhaps not surprisingly the country rankings of the two diseases are quite different, both to each other and to the ranking according to life expectancy or, for that matter, self-reported general health (presented earlier). For example, while Spain ranks second worst in diabetes it ranks highest in cancer with their mortality rank at age fifty more closely resembling their cancer ranking.

TABLE 2: International Differences in obesity and smoking rates and country rankings in disease and Life-Expectancy at age 50

	Fraction Obese	Currently Smoking	Country rank on:				
			Obesity	Smoking	Diabetes	Cancer	LE 50
Austria	23.7	24.2	11	8	5	3	8
Denmark	13.6	35.9	1	12	3	8	13
France	16.6	19.0	5	12	11	4	11
Germany	16.9	24.9	6	9	10	11	3
Greece	21.7	31.9	9	2	8	1	6
Italy	17.7	22.8	7	7	9	5	2
Netherlands	16.0	27.4	4	11	4	7	9
Spain	23.3	19.7	10	3	12	2	4
Sweden	15.5	19.7	3	3	6	6	5
Switzerland	15.1	27.1	2	10	1	9	1
Belgium	19.9	20.6	6	5	7	12	10
England	29.0	19.0	12	1	2	10	7
USA	30.9	20.6	13	5	13	13	12

Source: ELSA, HRS, SHARE, 2004 Data; Individuals aged 55-64.
Note: Fraction Obese (BMI>30) is based on Body Mass Index (BMI) computed from self-reported height and weight in all countries except for England where BMI is based on measured height and weight.

Once again, the most comprehensive analysis of the role of potential risk factors in international differences in disease prevalence has also involved the England-US comparison. In our initial paper on this issue (Banks et al, 2006), for each disease controls were put in place for education, income, smoking, alcohol drinking, obesity and being overweight all measured contemporaneously, as well as a summary indication of smoking history. While these risk factors all have statistically significant associations in the expected direction with the disease outcomes collectively they can explain very little of the English-American differences in disease prevalence. In part, as Table 2 illustrates, this is because the distribution of conventional known risk factors is not all that different between these two countries. This point of very common conventional risk factors is central in guiding the search toward less conventional explanations and behaviors.

In subsequent work, the search for explanations went to other less conventional risk factors. Banks, Berkman, and Smith (2010) assessed whether psychosocial aspects of social relationships and social participation accounted for US- English differences in morbidity and life expectancy. Using HRS and ELSA measures of social support and social networksthey found from cross-sectional analyses and 3- to 5-year follow-ups in mortality that current differences in social networks and social integration between the US and England did not explain differences in either current morbidity or subsequent mortality. Once again, this is at least in part because observed differences in social networks and integration between these two countries are small.

Banks et al, (2010b) were able to show that higher rates of diabetes among Americans could be largely accounted for by including measures of waist circumference as well as BMI differences, particularly for women – controlling for waist circumference explained about three-quarters of the country differences for women and a little less than half among men. But there is a real sense in which this is not an explanation since contemporaneously raised waist circumference for a given weight could be as much an indication of the presence of diabetes or related conditions as it is a risk factor or cause.

A final related, but distinct issue, which may generate some of the observed international differences in prevalence concerns differential screening for disease, with cancer perhaps the best example. Screening is reported to be more aggressive for breast and prostate cancer in the US than in the UK and Western Europe. Howard et al. (2009) document that two year mammography rates among women 50-64 years old were 77.7% (using HRS) compared to 46.2% in Western European SHARE countries with considerable variation amongst the SHARE countries. Similarly in the same age groups

PSA tests in the prior year for men were 27.1% in Europe compared to 42.2% in the US. Preston and Ho (2010) document that these screening differences between America and Europe are long-standing.

Screening may produce higher prevalence and incidence not only through earlier diagnosis but also through over-diagnosis. Due to earlier detection of not so dangerous cancers, this could plausibly contribute to lower incident mortality (i.e. a lower probability of death within a particular time-period following the diagnosis of a new condition) in the US compared to Western Europe which would also raise US prevalence. Treatment may be more effective in reducing mortality if present at earlier stages of disease and five year survival rates appear to be higher in US than in Europe. Although controversial, Preston and Ho (2010) report significant declines in prostate cancer mortality in America compared to Europe after PSA approval by the FDA.

3. Comparing incident disease onsets and subsequent mortality

3.1 Measuring Diseases Incidence in Aging Panels

Since prevalence questions are asked in each wave of these new longitudinal aging surveys, incidence as well as prevalence of disease can be calculated among panel participants as new onsets of disease are revealed each time a successive wave of the panel is collected.[6] The issue of whether these are reasonably accurate measures of incidence depend in part on rates of retention in the panels since attrition may be related to disease onset. Analysis of the impact of non-retention has been recently conducted for the HRS and ELSA surveys (Michaud et al, 2011 and Banks et al, 2011.a). These studies reported that as long as temporary attritors (those who leave the survey for a wave or

[6]While we use the term incidence and onset, a more precise term would be new detection.

two) are brought back into the survey that the evidence for non-random retention is weak. In addition, estimates of disease incidence in these surveys reasonably closely match those obtained from first differencing age-adjusted prevalence rates in the main population based health surveys in both countries.

Table 3, taken from Banks, Muriel and Smith (2010) shows that, as with prevalence, four year incidence of disease is also generally higher in America than in England amongst those in the 55-64 year old age group. Americans in this age group not only suffer from higher past cumulative disease risk as indicated by their higher disease prevalence; they also experience higher rates of new new disease onset compared to the English. Estimates of incidence are much closer in the 70-80 year age group, but still with an edge toward higher incidence among Americans, particularly for the cardiovascular conditions.

Table 3
Estimates of Disease Onset between 2002 and 2006 —England and US

Ages 55-64	Stroke	Lung	Cancer	HBP	Diabetes	Heart	Heart Attack
England	1.70	2.00	2.99	10.17	3.33	2.61	1.85
US	2.07	3.08	4.26	10.03	6.00	6.25	3.31
Ages 70-80	**Stroke**	**Lung**	**Cancer**	**HBP**	**Diabetes**	**Heart**	**Heart Attack**
England	4.68	2.78	4.80	9.83	4.44	4.80	3.38
US	5.51	3.89	5.88	8.31	4.66	9.28	5.42

Sources: Banks, Muriel, and Smith 2010. England—2002 English Longitudinal Survey of Ageing (ELSA) United States—2002 Health and Retirement Survey (HRS)

3.2 Measuring Mortality and Mortality Incidence in Aging Panels

A parallel set of issues arises in measuring mortality and incident mortality in these new aging surveys – do they provide reasonably accurate measures of mortality at the individual level? As we discussed above, such an outcome is crucial for international comparisons since it allows researchers to focus on a health measure that is unambiguously comparable across countries without being tied to using crude aggregate indicators such as life-expectancy.

There are essentially two ways by which mortality status can be identified in these aging panels. The first is when, as a consequence of attempts to interview respondents for the next survey wave, the survey organization finds that the respondent is deceased. The second is by matching survey participants to a country's National Death Index, which typically includes information about date and cause of death of all respondents regardless of their participation in subsequent waves of the survey, thus removing worries about the sensitivity of analytical findings (for certain empirical specifications at least) to issues relating to the level or nature of attrition in the survey.

Once again, the two most advanced of these aging surveys in terms of mortality linkage are HRS and ELSA. Age specific mortality data from each survey, collected through linkage of the sample information to the national death registers, has been compared to the full set of age-specific death rates from each countries life tables (for more detail see Banks, Muriel, and Smith, 2010). HRS data on respondent mortality are remarkably close to those obtained from the American life table. At all ages 50 and above, the two mortality curves closely overlap with the only difference being the larger random component in the survey data especially at older ages when numbers of living

respondents in the HRS sample become relatively thin. Over the whole of the 50+ age range, there does not appear to be any systematic difference between the national death registry and HRS based estimates of the four year probability of survival by age.

The situation in England is somewhat different, with the correspondence being less close. After age 65, mortality among ELSA respondents is somewhat lower than mortality in the English life tables issued by the Office for National Statistics. The most likely explanation for this discrepancy is that compared to HRS, ELSA is an immature survey in the sense that it has not yet reached population representative steady state. ELSA's baseline sample in 2002 was drawn from the non-institutionalized population thereby leaving out those residing in nursing homes whose mortality prospects especially at older ages are higher than average. A similar bias existed in the original HRS sample of older respondents (i.e. the AHEAD sample of those ages 70-80 who were initially recruited in 1993). Since respondents are subsequently followed into nursing homes in both the HRS and ELSA samples we speculate that this bias no longer exists in the older HRS sample but is only just beginning to diminish in ELSA.

A related issue concerns the usefulness of these new aging surveys to conduct analysis of subsequent mortality by disease prevalence and ultimately by disease incidence. To illustrate this potential, Table 4 displays six year mortality, or more specifically survival rates, organized by disease prevalence in 1998 in HRS and ELSA for two age groups- ages 55-64 and ages 70-80.[7]

[7]Although the first ELSA baseline took place in 2002, ELSA respondents were recruited from prior waves of the Health Survey for England at which time health conditions were recorded and permissions were collected for the linkage to death registers. Thus when simple prevalences are the only baseline variable required, earlier data on respondents, in this case from the 1998 HSE, can be used as a starting point for analysis.

For the younger of the two age groups, survival rates for those without disease are rather similar in the two countries and the picture is rather mixed for those with disease depending on the disease in question. Mortality rates are higher in England for those with lung disease or cancer, but lower for those with hypertension or stroke. Note that, particularly for this younger old-age group, the issue discussed earlier relating to cancer screening, whereby more aggressive screening leads to the discovery of more minor cancers which may be less likely to be fatal over the following six years, could be a partial explanation of the US-England difference in this domain and in lung disease as well.

Table 4
Fraction surviving to 2004, by Disease Prevalence in 1998

		Ages 55-64			Ages 70-80		
1998 Prevalence		HRS	ELSA	Diff	HRS	ELSA	Diff
Hypertension	Yes	91.06	92.83	-1.77	77.02	74.10	2.92
	No	95.17	94.05	1.12	78.65	74.82	3.83
Stroke	Yes	75.95	81.05	-5.10	60.90	49.17	11.73
	No	94.33	94.07	0.26	79.22	76.45	2.77
Diabetes	Yes	85.54	85.12	0.42	69.05	59.84	9.21
	No	94.60	94.05	0.55	79.55	75.67	3.88
Lung Disease	Yes	85.46	81.08	4.38	62.50	45.83	16.67
	No	94.22	93.90	0.32	79.67	75.33	4.34
Cancer	Yes	83.73	71.74	11.99	69.58	62.79	6.79
	No	94.22	94.16	0.06	79.24	74.78	4.46

Note: HRS and ELSA microdata. All individuals participating in 1998 survey regardless of subsequent participation in future waves

For those in their seventies more systematic differences emerge. Survival rates over this six year period for those without each of these specific diseases are around

three to five percentage points higher in the US compared to England compared to the US, with mortality therefore being three to five percentage points lower in the US. And the differences in survival rates for those with the prevalent conditions at baseline are considerably greater – survival rates in the US are between seven and seventeen percentage points for those with the more severe conditions at baseline, most likely reflecting the more aggressive (and considerably more expensive) treatment of illness in the US. The near universality of Medicare in these ages in the US and attributes of treatment in Medicare compared to other state older age health insurance programs in European countries may provide an explanation. This should be viewed now as only an hypothesis since there is little direct evidence on this at this point. Thus it is only the different prevalence of conditions at baseline which is leading to the comparable aggregate mortality rates across the two countries for this age group.

4. Methods and Explanations for International Health Differences.

In many respects, these new international aging surveys will be ideal resources with which to test key economic and policy relationships in the determinants of health outcomes as well as the impact of health shocks on economic resources. Policies vary considerably across Europe especially compared to the United States. For example, compared to the US, many European countries set up policies that aim at lessening the impact of negative health shocks on income, the result of which should be smaller negative income reductions but larger labor market reductions due to a new health shock.

Following the treatment in Smith (1999), Adams et al (2003) and in particular Smith, (2007.a), current realizations of both economic status and health can be thought to reflect a dynamic history in which both health (H_t) and SES (Y_t) are mutually affected by each other as well as by other relevant forces. Most of the relevant ideas can be summarized by the following equation:

$$(1)\ H_t = \alpha_0 + \alpha_1 H_{t-1} + \alpha_2 Y_{t-1} + \alpha_3 \Delta \hat{Y}_t + \alpha_4 X_{t-1} + u_{1t}$$

where X_{t-1} represents a vector of other possibly non-overlapping time and non-time varying factors influencing health and SES and u_{1t} are stochastic shocks to health.These models typically include in X_{t-1} a vector of baseline health conditions of the respondent—self-reported general health status, the presence of a chronic condition at baseline, and the extent of functional limitations scale and a standard set of behavioral risk factors (currently a smoker, number of cigarettes smoked), whether one engaged in vigorous exercise, and BMI and a standard set of demographic controls—birth cohort, race, ethnicity, sex, region of residence, and the like.

Much interest however lies in the SES measures that include education, household income and wealth, baseline levels of and innovations in household wealth and in income. In this framework, we can estimate whether past values of SES predict health ($\alpha_2 \neq 0$).[8]In the tradition of Granger causality, one view is that this provides a test of no direct causal path (conditional on H_{t-1}) between Y_{t-1} and H_t. More precisely, this reduces to a joint test that there is not a direct causal link between Y_{t-1} and H_t and no unobserved common factors that could create an ecological correlation betweenY_{t-1} and H_t. Since it is in general impossible to rule out the possibility of all such unobserved

[8]Foran insightful debateabouttheconditionsunderwhether coefficients are zeroor stationary also reveals something aboutcausality,seethepaperbyAdamsetal.(2003)andthecommentson that paperin thesame volume.

factors, caution is warranted in going beyond the language of prediction with these parameters on lagged values. $\alpha_2$may be better viewed simply as the ability of past values of SES to predict future health onsets. The reason is thattheremaywellbeunobserved factorscorrelatedbothwithpastSESandhealth,evenafter we have conditioned on the measured components.

A key parameter $\alpha_3$measures the effect ofnewinnovationsofSES ($\Delta\hat{Y}_t$) on health. The term $\Delta\hat{Y}_t$representsthat partofthe between period changein SES (ΔY_t) thatisan innovation. To estimate α3,werequireexogenousvariationinSES($\Delta\hat{Y}_t$)notinducedby health.Inparticular,thisimpliesthatitis notappropriatetousethefullbetween-periodchanges SES (ΔY_t) to estimate these effects since such variationhopelesslyconfoundsfeedbackeffects. For example, a new health problem could reduce labor supply and hence income so that regressing the change in health on the change in income would have it all backward.

These arguments apply equally well if our heath outcome is a new onset of disease or new incident mortality. But it is important to note that evidence regarding one health outcome does not imply that we should expect this result to transfer to a different health outcome, especially in light of our earlier discussion about the differences in these health outcomes across countries Even if economic resources have a weak relation to health onsets at older ages, this is not inconsistent with financial resources mattering a lot for an outcome such as mortality. Access to financial resources may be essential in dealing with consequences of health problems after they occur. Using a common analytical strategy of conditioning on past health and SES histories it would be preferable to separately examine both all cause mortality and mortality by individual

cause. Some causes of mortality are more likely to be amenable to individual behaviors as well as to economic resources.

At the same time, analysis of incidence of disease or incident mortality poses some serious practical challenges to these new emerging international panel aging surveys. Predominantly it requires panels to be sufficiently long both in terms of allowing enough realizations of health outcomes to occur as well as providing a time-series window of sufficient length that innovations in economic shocks or changes in policy can be witnessed. Additionally there needs to be only moderate levels of sample attrition since innovations in health and SES can only be observed for continuing sample members, unlike register based information such as deaths or cancers which can be observed on an ongoing basis for all original baseline sample members. At this point only HRS with ten waves (covering twenty years) in its original cohort fully meets those criteria. As ELSA moves into and beyond its fifth wave (covering ten years), it is beginning to become long enough to meet these criteria and the first international comparative analysis will soon be possible.

To illustrate the potential of this approach, Banks, Muriel, and Smith (2010) conducted a wealth- mortality analysis using the original HRS birth cohort, i.e. those 51-61 years old in 1992. As discussed above, to isolate causal effects of wealth on mortality, it is necessary to isolate innovations in household wealth that were not caused by health. However, those original HRS respondents had by then been interviewed for eight rounds over a fourteen year time period during a time span when large and heterogeneous wealth changes occurred, mostly induced by asset price changes in housing and stock markets through the 1990s. Banks, Muriel, and Smith (2010) report

that using this long panel of respondents in the United States that mortality between 2002 and 2006 was not related to changes in wealth that they experienced during the prior ten year period 1992-2002. This supported the view that the primarily pathway driving the correlation between health and wealth at older ages is that poor health leads to a depletion of household wealth. The potential value of carrying out such an analysis in a different policy environment where there are different arrangements for financing of health care costs and the loss of income associated with health related retirements is immediately apparent.

4.2 Estimating policy effects with international comparative data

Arguably the most important aim of applied economic and social research is to investigate the causal impacts of policy institutions, defined in the broadest sense, on economic, social and health outcomes. After all, while other factors such as individual characteristics or heterogeneity may explain more of the variance in outcomes they are typically outside the domains of control of policymakers.

Applied micro-economists interested in causal policy analysis typically follow one of two possible approaches, sometimes referred to as ex-ante and ex-post policy evaluation. Firstly, they could choose to develop and estimate a structural model of individual, household or firm behavior and use the fundamental behavioral parameters to investigate, ex-ante, how behavior would change in counterfactual policy environments. Alternatively, they could exploit some kind of policy experiment or quasi-experiment where outcomes are compared between control and treatment groups following some policy change. In the 'gold-standard' policy studies, policy interventions are designed

specifically with evaluation in mind and data are collected on treatment and control groups in the manner of a randomized control trial. Alternatively, historical policy changes or natural events that only affected certain groups can be used to define a quasi-experimental policy evaluation methodology. There are simply more opportunities to conduct such analysis in an international comparative framework.

Whilst it is not our purpose here to discuss the strengths and weaknesses of these approaches, it is certainly the case that their strengths and weaknesses depend on the particular policy application under consideration. For policy questions such as those related to low-frequency life-cycle processes such as health, ageing and pensions the weaknesses in each approach are often noticeable. In one case, intervention studies such as policy experiments would often need to wait decades until the appropriate outcome variables were observed, not a useful timescale for applied policy analysis. In addition, the use of policy changes that happened long enough ago for long-run outcomes to be subsequently observed is typically hampered by a lack of micro data collected around the time of the policy change.

In the other, the current state of dynamic structural models of life-cycle behavior in the field of health, ageing and pensions is such that simplifications in either the behavioral assumptions, the stochastic environment facing individuals, or the characterization of key policy variables need to be made in order for the model solution and estimation to be feasible. Such simplifications, whilst not removing the ability of the model to generate important insights, limit the extent to which such models are used by applied policy analysts. For these reasons there has therefore been an increasing interest

in using international comparative research as a way of understanding the effect of policy on aging and health outcomes.

The concept of using geographical variation to understand policy effects is not confined to international comparisons and as such is not at all new. It has, after all, become relatively common to use regional or state variation to identify effects of interest when looking at labor market or health outcomes – for some of many examples on health or health-related behaviors see Adda and Cornaglia (2006, 2010) or Skinner et al (2006). Graduate students are now routinely trained to carry out and interpret such studies and the conditions under which such state-level comparisons can provide credible evidence of policy effects are well known. Consequently, one of the best ways to begin think about the burgeoning international comparisons literature is simply as an extension of this method. For a good methodological overview of the key issues set out in the context of international differences see Kapteyn (2010).

The analogy of countries as states highlights three important areas where operational and practical considerations arise when carrying out internationally comparative research. First, there is the problem of collecting or bringing together comparable measurements for the regions under study, which we have discussed extensively above. This is a key issue which has limited the development of international comparative work on health in the past but is now arguable on the wane. It is also an issue that is certainly more acute in an international study as opposed to a state-level comparison since many datasets collected at the national level will automatically have common protocols applied to all geographic areas.

Second, there is the issue of the connectedness or otherwise of markets and policy environments across the regions being studied. In this case, international comparisons have an advantage over state-level studies since there are arguably fewer spillover effects across countries than across states, and the endogenous allocation of individuals into specific policy environments is presumably less of an issue when the costs of moving (broadly defined) are greater.

Finally there are fewer, if any, common components of policy across countries. In contrast, individuals being compared across states or regions within a country will be facing many common policy variables (such as federal taxes, welfare programs or nationally provided public goods). Whilst this means that fully characterizing the policy environment can be more of a challenge to researchers, if done successfully it can mean that the extent of variation in an international comparative study is much greater than in a state-level comparison.

What are the possibilities? Hands down the most influential use of international comparisons in economics as a research strategy in recent decades is the Gruber-Wise analysis of social security programs around the world on age of retirement. (Gruber-Wise, 1999, 2004, 2007, 2010). Why was the Gruber-Wise project so influential on subsequent research and why has there not been a comparable international project in health economics to date? One reason is that the crucial policy parameter that was the focus of that work - the 'tax on work' was relatively easy to conceptualize and implement empirical. The 'tax' on work, which depends a lot on national policy, varies more across countries than it does within a country. Even more importantly, it also

varies more over time within other European countries than it did in the United States where changes in the main federal social security program have been almost impossible.

The ability to 'explain' major differences in the retirement ages across countries movedthe debate away from country specific explanations which tended to paralyze the debate , e.g. Europeans and Americans are different, or the box of labor where the young and old are competing for the same number of jobs. Of course, even in this retirement research complications arose in the follow-up work where it was increasingly recognized that that there were multiple programs at play- not just social security. Country differences in the generosity of disability programs which tended to be another route to pre-retirement exit from the labor force would be an excellent example.

The extension of this international work into health economics has not really yet happened to the same extent. Garber and Skinner (2008), in their study of the efficiency of US health care, document extensive international variation in health-care spending and utilization which suggests there are possibilities for a fruitful comparative analysis based on micro-data. One reason that such an analysis has not happened is that, as we have been discussing throughout our paper, sufficient data on outcomes and covariates are only just becoming available on a comparable basis. In addition, however, it is far more difficult to characterize the central policy parameters to the same extent. For example, with publicly provided health insurance, it is not simply the deductable and the co-insurance rate that matters. In many countries with national health insurance, many aspects of non-price rationing come into play- delays and queues are the norm. As we mentioned above, other factors such as screening rates for cancer vary significantly across countries can be explored in more depth. Variation in the quality of practice

likely varies much more across countries so that it is another important area for future research. Perhaps, the lowest hanging fruit concerns the importance of having any insurance which is universal in most European countries, but not in the US.

Conclusions

In this paper we have attempted to provide an overview of the burgeoning literature that uses micro-level data from multiple countries to investigate health outcomes, and their link to socioeconomic factors, at older ages. Since the data on which much of this literature is at a comparatively young stage, and the demands placed on that data are somewhat great (namely, a long time-series of longitudinal data on detailed variables in a number of health and economic domains), much of the analysis is at a very early stage and limited to only a handful of countries, with analysis for the US and England being by far the most common.

One thing that is immediately apparent as we get better measures is that health differences between countries amongst those at older ages are real and large. Countries are ranked differently according to whether one considers life-expectancy, prevalence or incidence of one condition or another. And the magnitude of international disparities may vary according to whether measures utilize doctor diagnosed conditions or biomarker-based indicators of disease and poor health. But one key finding emerges – until very old age the US ranks poorly on all health indicators with the exception of self-reported subjective health status.

The as yet unmet challenge of this research agenda is to come up with 'causal' explanations of these international health differences, whether these relate to

international differences in current or past behaviours or, more constructively, current or past policy and institutional variation. Up to now, what analysis there is has been better at ruling out potential explanations than ruling them in, a fact which we attribute to the still relatively short time-series dimension of the data that are available and the still rather small set of countries for which that data exist. Both these factors are changing rapidly and the opportunities for future research are expanding accordingly. There seems no doubt that studies in health economics basing their analytical methods on international health comparisons will provide important research findings in the future.

References

Adams, P., M. Hurd, D.McFadden, A.Merrill, and T.Ribeiro, "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status," *Journal of Econometrics*, 2003, 112, pp. 3-56.

Adda, J. and F. Cornaglia, 2006, "Taxes, Cigarette Consumption, and Smoking Intensity," *American Economic Review*, vol. 96(4), pages 1013-1028, September.

Adda, J. and F. Cornaglia, 2010, "The Effect of Bans and Taxes on Passive Smoking," *American Economic Journal: Applied Economics*, vol. 2(1), pages 1-32, January.

Avendano, M., M. Glymour, J. Banks, and J.Mackenbach. (2009). Health disadvantage in US adults Aged 50-74: Are poor Europeans healthier than Americans? *American Journal of Public Health* 99:540-48.

Banks J., M. Marmot, Z. Oldfield, and J.P. Smith. (2006). Disease and disadvantage in the United States and in England. *JAMA* 295(17):2037-45.

Banks, J., M. Marmot, Z. Oldfield, and J.P. Smith. (2009). SES and health on both sides of the Atlantic. Pp. 359-406 in Developments in the Economics of Aging, edited by D. Wise. Chicago, IL: University of Chicago.

Banks, J., A. Muriel, and J.P. Smith. (2010). Disease prevalence, disease incidence, and mortality in the United States and in England. *Demography* 47 (Supplement):S211-31.

Banks, J., A. Muriel, and J.P. Smith. (2011a). "Attrition and ageing surveys: Evidence from ELSA and HRS." *Longitudinal and Life Course Studies.*Volume 2. No. 2 pp. 101-126.

Banks, J., M. Kurmari, J.P. Smith, and P. Zaninotto. (2011b). What explains the American disadvantage in health compared to the English? The case of Diabetes. Forthcoming in *Journal of Epidemiology and Community Health.*

Banks, J., L. Berkman, and J.P. Smith (2011c). Do cross country variations in social integration and social interactions explain differences in life expectancy in industrialized countries? In E. Crimmins, S. Preston, and B. Cohen (eds.), International Differences in Mortality in Older Ages: Dimensions and Sources. Washington DC: National Academy Press, pp. 210-267.

Bago d'Uva, T., E. D. Van Doorslaer, M. Lindeboom, O. O'Donnell, and S. Chatterji (2008). Does Reporting Heterogeneity Bias the Measurement of Health Disparities? *Health Economics*, 17(3), 351{375.

Bago d'Uva T, M Lindeboom, O O'Donnell, E van Doorslaer, (2009), Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity. HEDG Working Paper 09/30, forthcoming in *Journal of Human Resources*.

Case, A., D. Lubotsky, and C. Paxson. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review* 92(5):1308-34.

Crimmins, E, S. Preston, and B. Cohen (eds.). 2011. Explaining Divergent Levels of Longevity in High Income Countries. Washington DC: National Academy Press,

Deaton, A. and Christina Paxson. 2004. Mortality, Income, and Income Inequality over Time in Britain and the United States. In Perspectives on the Economics of Aging, pp, 247-286. David Wise editor. University of Chicago Press.

Erdogan-Ciftci, E, Doorslaer, E.K.A. van, Bago d'Uva, T.M. & Lenthe, F.J. van (2010). Do self-perceived health changes predict longevity? *Social Science & Medicine*, 71(11), 1981-1988.

Gakidou, E. Mallinger, L, Abbott-Klafter, J. et al. Management of Disabetes and associated cardiovasuclular risk factors in seven countries: a comparison of data from national health examination surveys, Bull World Health Organ 2011; 89; 172-183

Garber, Alan M., and Jonathan Skinner. 2008. "Is American Health Care Uniquely Inefficient?" *Journal of Economic Perspectives*, 22(4): 27–50. doi**:**10.1257/jep.22.4.27

Gruber, J.and Wise, D. Social Security Programs and Retirement Around the World: Introduction and Summary. in Jonathan Gruber and David A. Wise, eds., Social Security programs and retirement around the world. Chicago: University of Chicago Press, 1999.

Gruber, J.and Wise, D. Social Security Programs and Retirement Around the World: Mirco-Estimation. Chicago: University of Chicago Press, 2004

Gruber, J.and Wise, D. "Social Security Programs and Retirement Around the World: Fiscal Implications of Reform. Chicago: University of Chicago Press, 2007

Gruber, J.and Wise, D. "Social Security Programs and Retirement Around the World: The Relationship to Youth Employment. Chicago: University of Chicago Press, 2010.

Howard, D.H. L. Richardson, and K. Thorpe. "Cancer Screening and Age in The United States And Europe."*Health Affairs*November December 2009. Vol. 28 No 6. pp 1838-1847.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org

Kapteyn, A., van Soest, A. and Smith, JP. "Vignettes and Self-Reported Work Disability in the US and the Netherlands," *American Economic Review* 97(1):461-473, March 2007.

Kapteyn, A 2010. What Can We Learn From and About Global Aging. *Demography* 47 (Supplement):S191-210

King, G., C. Murray, J. Salomon, and A. Tandon. 2004. Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research, *American Political Science Review*98(1), 567-583.

Kippersluis, J.L.W. van, O'Donnell, O.A., Doorslaer, E.K.A. van & Ourti, T.G.M. Van (2010). Socioeconomic Differences in Health over the Lifecycle in an Egalitarian country. *Social Science & Medicine*, 70(3), 428-438.

Lee, J, and Smith JP "The Effect of Health Promotion on the Diagnosis and Management of Diabetes," 2011, to be published in *Journal of Epidemiology and Community Health*,

McDade, T. The State and Future of Blood-Based Biomarkers in the Health and Retirement Study.2011.*Forum for Health Economics and Policy*. Berkeley Economic Press Vol. 14: Issue. 3

Michaud, P.C. Kapteyn, A.. Smith, JP and van Soest, A.2011"Effects of Attrition and Non-Response in the Health and Retirement Study," *Longitudinal and Life Course Studies.*Volume 2. No. 2.145-169.

National Academy of Sciences (2007), *Biosocial Surveys*, Washington: The National Academies Press

Preston, S and Ho, J. (2010). Low Life Expectancy in the United States: Is the Health Care System at Fault? In E.M. Crimmins, S.H. Preston, and B. Cohen (Eds.) International Differences in Mortality at Older Ages: Dimensions and Sources. Washington, DC: The National Academies Press.

Skinner, J.,D. Staiger, E. Fisher, 2006, "Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction" *Health Affairs*, web exclusive (February 2006).

Smith, JP "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status," *Journal of Economic Perspectives* 13(2):145-167, Spring 1999

Smith, JP. "The Impact of Socioeconomic Status on Health over the Life-Course," *Journal of Human Resources* 42(4):739-764, Fall 2007a.

Smith, JP "Nature and Causes of Male Diabetes Trends, Undiagnosed Diabetes, and the SES Health Gradient," PNAS; Proce*edings of National Academy of Sciences* 104(33):13,225-13,231,August14, 2007b.

Strauss, J. and D. Thomas. 1998. "Health, nutrition and economic development", *Journal of Economic Literature*, 36(3):766-817.

Strauss, J. and D. Thomas. 2008. "Health over the life course", in T.P. Schultz and J. Strauss (eds.), Handbook of Development Economics, Volume 4, Amsterdam: North Holland Press.

Van Soest, A. Delaney, L., Harmon, C., Kapteyn, A, and Smith, JP. "Validating the Use of Vignettes for Subjective Threshold Scales," July 2011.a*Journal of the Royal Statistical Society- Series A.* Volume 174- Issue 3.

Van Soest, A. Kapteyn, A, Andreyeva, T. and Smith, JP. "Self Reported Disability and Reference Groups," NBER working paper # 17153 June 2011b, Forthcoming. Investigations in the Economics of Aging. David Wise editor- University of Chicago Press.

Van Soest, A. Kapteyn, A, Smith, JP, and Vonkova, H. June 2011c. Anchoring Vignettes and Response Consistency. Rand Working Paper.

Ware, Davies-Avery, A and Cathy Donald, C., "General Health Perceptions," R-1987/5 HEW, RAND. Santa Monica California 1978.

Wilkenson, Richard G., Unhealthy Societies: The Afflictions of Inequality. London: Routledge, 1996.

Appendix

Table A1: Life Expectancy by Country and Age

Country	Age 0	Age 50	Age 65	Age 75
Austria	80.3	32.1	19.4	11.9
Belgium	80.0	32.0	19.3	12.0
Denmark	79.0	30.7	18.3	11.2
Finland	79.9	32.0	19.5	12.1
France	81.4	33.4	24.9	13.3
Germany	80.3	31.9	19.2	11.8
Greece	80.2	32.2	19.2	11.5
Italy	81.9	33.5	20.3	12.6
Japan	83.1	34.8	21.7	13.9
Netherlands	80.6	32.1	19.2	11.8
Spain	81.6	32.7	20.2	12.1
Sweden	81.3	32.7	19.6	12.1
Switzerland	82.2	33.8	20.6	12.8
UK	80.2	32.2	19.3	12.0
USA	78.5	31.5	19.2	12.3

Source: WHO 2009 Life Tables

Table A2
Self-Assessed Health by Country—Ages 55-64

	Germany	Spain	Greece	Italy	Netherlands	Sweden	Austria
Excellent	4.3	6.1	7.1	8.2	15.5	24.3	9.5
Very good	17.7	17.4	34.9	15.6	17.8	25.8	28.2
Good	44.4	43.3	37.3	43.8	43.4	37.5	38.4
Fair	26.1	25.2	18.1	25.8	20.0	9.1	18.7
Poor	7.6	8.0	2.6	6.6	3.4	3.3	5.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SAH score	46.3	47.1	56.5	48.3	55.6	64.7	54.5
Rank of e50	11	4	6	2	9	5	8
Rank of SAH	13	12	6	11	8	1	9

	France	Denmark	Switzerland	Belgium	United States	England
Excellent	10.9	23.2	15.7	12.2	16.1	14.5
Very good	19.9	29.6	31.6	25.0	32.0	30.3
Good	46.2	24.0	40.1	41.2	27.6	30.5
Fair	17.8	16.3	8.8	16.7	17.2	17.7
Poor	5.2	6.8	3.8	5.0	7.1	6.9
Total	100.0	100.0	100.0	100.0	100.0	100.0
SAH score	54.3	61.5	61.7	55.7	58.2	56.9
Rank of e50	3	13	1	10	12	7
Rank of SAH	10	3	2	7	4	5

Source: HRS, ELSA, and SHARE; 2004.

Notes: Rank of e50 taken from data in Table M; SAH score calculated by allocating 1 for each percent of the population reporting Excellent health, 0.75 for each percent reporting Very Good, 0.5 for good, 0.25 for fair and 0 for Poor. So an index of 100 would indicate the whole population reporting excellent health, and an index of 0 would indicate everyone reporting poor health.