Reporting heterogeneity across the Atlantic

A vignette study to identify differences in health between the US and England, based on self-reported health domains.

Thomas Slager (311829)

Supervisor: Dr. Teresa Bago d'Uva

Co-readers: Prof. Dr. Eddy Van Doorslaer & Dr. Pilar García-Gómez

October 2012

Preface

Before you lies my master thesis, written during the year 2012, as my final piece of work for my masters degree in Health Economics at the Erasmus University in Rotterdam. It has been an interesting and long school year, in which I have been introduced to a lot of new interesting perspectives on health care. The combination of comparing different countries on demographics and economic variables, making comparisons between countries and using quantitative methods in stata interested me a lot. In the meantime, I also managed to get a job at the Statistics Netherlands (CBS) from September 2012, of which I am very happy and proud. I can safely say I will be busy with quantitative methods for quite some time then!

I have to say the journey of completing this thesis has enriched me with a lot of knowledge in health economics. The journey, however, was not without challenges, as I initially had to finish before September and the course load during the year was pretty heavy at some times. It was therefore impossible for me to follow a normal thesis trajectory. However, thanks to my supervisor, it has become possible to finish without too big of a delay. This brings me to the next point. I have a great deal of thanks to give to my supervisor, Teresa Bago d'Uva, whose constructive criticism, knowledge and of course guidance proved to be essential in finishing this thesis, which I eventually finished two months later than planned due to the start of my working career.
Finally, I hope this thesis can be of use within the world of health economics and vignettes.

Thomas Slager, October 2012.

Summary

There exists a lot of difference in the way in which different populations value their own, self-assessed health. These differences in reporting (reporting heterogeneity) are most notable between countries and have different reasons.

By using surveys including anchoring vignettes, comparability between health measures between countries and populations is made possible. Within these surveys, respondents not only report their own health, but also report health of a hypothetical person with a certain health state. As the vignettes are identical for all respondents, individual variation in the rating of health must be due to reporting heterogeneity.

Differences in health have been indicated for six general health domains (breathing, sleeping, pain, memory/cognition, depression and mobility) between England and the US, using data from the English Longitudinal Survey of Ageing (for England) and the Health retirement Study (for the US). The data includes information on respondents' self-assessed health, together with vignette ratings, rated by those individuals. The data is used to estimate an ordered probit model and a HOPIT model for all health domains. By doing this, an examination is possible between the two models, and the impact of reporting heterogeneity is measured.

It is clear that from the results that reporting heterogeneity is existent between the two countries when correcting for a selection of socio-demographic indicators. From a population point of view, English residents are generally better off than American individuals in the sense that they have a lower probability of reporting more health problems across all health domains except sleep, where no difference is found. Also, the conclusion can be drawn that true health differences between the countries are larger when considering heterogeneity in reporting and that certain health effects of socio-demographics could be underestimated or not found at all when not taking reporting heterogeneity into account. Further, certain effects of some socio-demographic factors are present for America but not for England, and vice versa.

Concluding, it is necessary to correct for reporting heterogeneity when making comparisons in health between the US and England. When not correcting for this bias, true health differences are not generated.

Summary Dutch

Populaties beoordelen hun eigen gezondheid op verschillende manieren. Door het gebruik van zogenaamde vignetten, wordt de vergelijkbaarheid van gezondheid tussen verschillende populaties en landen mogelijk gemaakt. Door het gebruik van vignetten in enquêtes, rapporteren respondenten hierdoor niet slechts hun eigen gezondheid, maar rapporteren zij ook de gezondheid van een hypothetisch persoon met een bepaalde gezondheidstoestand. Omdat de vignetten voor iedereen gelijk zijn kan de individuele variatie in de waardering van eigen gezondheid worden gemeten.

Verschillen in gezondheid zijn bepaald voor zes algemene gezondheidsdomeinen (ademen, slapen, pijn, geheugen/cognitie, depressie en mobiliteit) tussen Engeland en de VS, met behulp van gegevens uit twee nationale enquêtes; de English Longitudinal Survey of Ageing (voor Engeland) en de Health retirement Study (voor de VS). Deze enquêtes bevatten informatie over de eigen gezondheid van respondenten en de vignet ratings van die personen. De gegevens worden gebruikt om parameters voor een geordend regressie model en een HOPIT regressie model voor alle gezondheids-domeinen te schatten. Hierdoor is een vergelijking mogelijk tussen de resultaten van de twee modellen en kan de impact van het verschil in rapporteren tussen de landen worden bepaald.

Uit de resultaten komt naar voren dat er verschillen bestaan in de rapportage van eigen gezondheid tussen de twee landen bij het corrigeren voor een selectie sociaal-demografische indicatoren. Uit populatie-oogpunt zijn Engelsen over het algemeen beter af dan Amerikanen; zij hebben een lagere kans op rapportage van meer gezondheidsproblemen in alle domeinen, op slapen na, waar geen verschil wordt gevonden. Ook kan de conclusie worden getrokken dat de ware gezondheidsverschillen tussen de landen groter zijn bij het in acht nemen van heterogeniteit en dat bepaalde effecten van de sociaal-demografische factoren kunnen worden onderschat of niet worden gevonden wanneer niet wordt gecorrigeerd. Verder bestaan bepaalde effecten voor Engeland wel maar niet voor Amerika en vice versa. Er kan geconcludeerd worden dat het noodzakelijk is om te corrigeren voor heterogeniteit bij het vergelijkingen van gezondheid tussen Amerika en England. Wanneer hier niet voor wordt gecorrigeerd, worden daadwerkelijke gezondheidsverschillen niet gegenereerd.

Contents

1. Introduction

Health surveys, used to measure health on national level, depend heavily on self-assessed health measures. However, there exists evidence that different populations and sub-groups of people report their own health differently (Salomon et al. 2004, Etilé & Milcent 2006, Bago d'Uva et al. 2008a & 2008b). These differences are most apparent between countries and within poor countries, as income and health variety is greater there (Lynch et al. 2004). Health disparities are due to many factors, like a difference in ability to estimate own health, adaptation to disease or illness, different expectations of health and other biases.

Because of this 'reporting heterogeneity', self-reported health measures between populations are not fully comparable. The following passage in an article by Murray et al. (2001) embraces the subject effectively: *"The challenge is to ascertain how much of the difference is determined by real differences in health and how much is due to differences in the way individuals report on their health relative to different norms and expectations"* (Murray et al. 2001:2).

Because of the presence of reporting heterogeneity, a certain health state based on a categorical scale, can in this way for example be seen as "extreme" for a 50-year-old, while it is considered only "moderate" for a 80-year-old (Salomon et al. 2004). Therefore, it is important for health researchers to correct for these differences in self-reported health and to make health more comparable. This is a crucial element to compare self-reported health measures between populations and population sub-groups.

A way to correct for reporting heterogeneity is by including "anchoring vignettes" in surveys. Respondents not only report their own health, but also report health of hypothetical persons with certain health states (the vignettes), which represent fixed levels of health. An example of a vignette in the health domain mobility is: "*Tom has a lot of swelling in his legs due to his health condition. He has to make an effort to walk around his home as his legs feel heavy"* (Disability Vignette Survey 2007). These vignettes are thus used alongside ratings of own health on a comparable, categorical scale, so that variation in rating can be related to differences in reporting behaviour and heterogeneity can be identified (King et al. 2004). Vignettes are being used in a growing number of national household surveys, like the English Longitudinal Study of Ageing (ELSA), the Health and Retirement Study (HRS) and the Survey of Health, Ageing and Retirement in Europe (SHARE) (Bago d'Uva et al. 2011).

A frequent comparison in terms of health is made between the US and England, as these countries have a strong link in socio-economic position and health. Furthermore, both countries are Western, the data for these countries is best comparable and health expenditures and health systems vary tremendously. Cross-country comparison should therefore measure variation, along comparable socio-demographic indicators, effectively. Possibly as a consequence of differences in health systems, accessibility and spending, there exist large health disparities between these two countries in terms of objective health measures like mortality, morbidity and disease prevalence. The US population is unhealthier than Western European countries (including England) regarding objective health measures, with the exception of self-reported subjective health status. In other words: Americans do not rate their own health as worse, despite evidence from objective health measures proving the contrary (Banks et al. 2006, Banks & Smith 2012). This poses an interesting discrepancy; especially given the US is the highest income country.

This thesis embraces the subject of comparing the elder populations within England and the United States on the basis of their self-assessed health. Different countries, especially in Europe, have been compared to each other in terms of self-assessed health, but a good comparison between England and the US has yet to be made using vignettes and general health domains. Thus, the purpose of this thesis is to check if the inconsistency in health measures can be (partly) explained by reporting heterogeneity between the US and England, using anchoring vignettes. Also, the effects of certain demographics on both countries are compared.

The main research question in this thesis is: " *To what extent are the differences in self-assessed health between England and the US caused by reporting heterogeneity, and in what amount does this difference account for real health disparities between the two countries?*

To compare the two populations, the Health and Retirement Study (HRS) for the US and the English Longitudinal Survey of Ageing (ELSA) for England are used.

This thesis is divided into five chapters. Chapter two will discuss the theory and knowledge known to date surrounding the subject. Chapter three will discuss the data and methods used, whilst chapter four will cover the results. Finally, in chapter five, the discussion and conclusion is presented, along with theoretical implications and a number of strengths and limitations.

2. Theoretical framework

2.1 Socio-economic status and self-assessed health

Population health should be a major concern on every country's agenda; studies have identified big health inequalities between and within countries (Van Doorslaer & Koolman 2004, Van Ourti 2003) and suggested a relationship between socio-economic status (SES) and self-reported health (Kunst et al. 1995, Adams et al. 2004, Frijters et al. 2005). The latter can be seen as an important issue, as people with a high SES live longer, live in better health and suffer less from disability and illness. People with a low SES in turn die younger and suffer more from illness and disease (Demakakos et al. 2008:331, Marmot 2005). It is therefore important to know population (sub-group) health numbers to be able to distinguish groups and make them more comparable with regards to self-assessed health. To generate these population (sub-group) numbers, respondent surveys are used wherein respondents value their own health, often along with other socio-demographic characteristics.

Self-reported health measures are useful and informative instruments, which can be obtained easily and at low cost by large-scale national or household surveys.
It is self explanatory that the data, which is collected, has to be comparable. When incomparable health measures are used, analysis results can be biased. A striking example of such a bias is the fact that the Aboriginals report better self-assessed health than the Australian population, despite living in poorer conditions and suffering from a higher mortality rate (Mathers & Douglas 1998). We would therefore opt to invest more in the Australian population, rather than the Aboriginals, however morbidity and mortality is higher amongst the latter group.
There exists evidence that different socio-economic groups of people tend to estimate their own (self-assessed) subjective health differently (Salomon et al. 2004, Etilé & Milcent 2006, Bago d'Uva et al. 2008a & 2008b, Crimmins et al. 2010). This difference in reporting is dependent of conceptions of health in general, and varies because of different expectations of health, risk aversion, adaptation to disease or/and illness, financial incentives to report ill health and complexity of the survey itself (Bago d'Uva et al. 2008a). An example of different

conceptions of health could occur if higher educated people are better informed about a certain health condition or treatment, so they are less tolerant of a given health state. In terms of expectations of health, a respondents' use of health categories like "some difficulties" and "great difficulties" depends on their norms and expectations for health on a particular health domain (Murray et al. 2001). For example, a health state such as "you have some difficulties with getting out of bed" within a mobility health domain can in this way be seen as severe for a highly educated 50-year-old, while it is considered only a minor problem for a low educated 80-year-old.

Adaptation refers to the amount in which an individual can adapt to a certain disease or illness. This way, ill patients tend to report much higher quality of life measures than would be expected (Groot, 2000). This adaptation can be seen as a cause for differences in scale reference and can influence conclusions in decision-making surrounding cost-effectiveness and respondents' reporting of self-assessed health (Groot, 2000).

The question therefore is raised if subjective health estimates are measured without a certain bias. It has been shown that differences for instance exist in self-reporting between different socio-economic groups (Etilé & Milcent 2006). Further evidence exists that different older populations report their health differently (Melzer et al. 2004). Higher educated older Europeans for example tend to value a certain heath state more negatively for the same state (Bago d'Uva et al. 2008b). This bias is illustrated in a hypothetical example in figure 1 on the next page, where two individuals, L and H, with low (L) and high (H) education rate their own health as the same based on a categorical scale (both moderate health), while in truth, it is different. This is due to differences in response scales; take for example the health categories on lines L and H. L and H possess different references of certain health levels. Take, for example, moderate health: L's reference of moderate health is at a lower true health than H's reference of moderate health. For individuals L and H to report the same health category, like moderate health, true health levels (see the true health line on the right) H^*_H for H and H^*_L for L should be possessed by the individuals (see figure 1). There however exists a big difference in true health between H and L, indicated by the difference between H^*_H and H^*_L. The difference in response scales between L and H is the reason why different health statuses are perceived and why true health can differ when

different individuals report the same health statuses. Of course, this is a hypothetical example and different situations are possible.

Concluding, it is safe to say it is necessary to correct for reporting heterogeneity between populations, in order to observe and compare true 'objective' health differences.

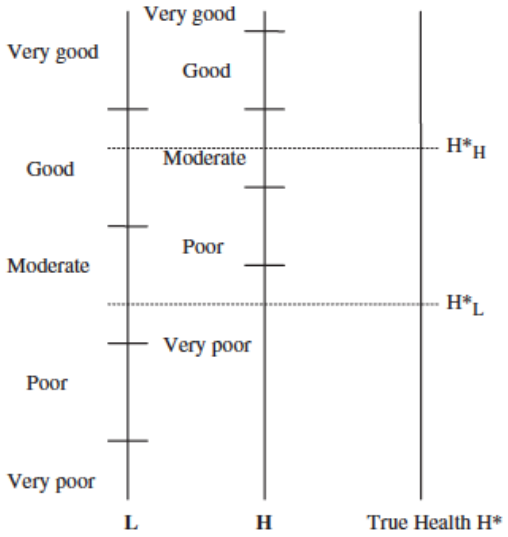


Figure 1: Self-reported health for high and low educated individuals (Source: Bago d'Uva et al. 2008b)

2.2 Vignettes

Self-reported health measures are frequently used in health economics, as they are powerful predictors for a wide range of outcomes, like mortality and medical care use (Bago d'Uva et al. 2008b). However, because of reporting heterogeneity, disparities exist between groups and self-reported health measures between populations are not fully comparable. The following passage in an article by Murray et al. (2001) embraces the subject effectively: *"The challenge is to ascertain how much of the difference is determined by real differences in health and how much is due to differences in the way individuals report on their health relative to different norms and expectations"* (Murray et al. 2001:2). To make self-reported health measures comparable, anchoring vignettes can be used, which are short stories that describe a hypothetical person with a certain health state, and represent "fixed levels of latent health", so that variation in rating can be fully related to differences in reporting behaviour of respondents (Bago d'Uva et al. 2008a, Bago d'Uva et al. 2008b, King et al. 2004). Respondents are then asked to value the health of the vignette on the same scale as

they are asked to evaluate their own health (King et al. 2004). By adjusting the health of respondents to a common scale, remaining variation after application to a common scale is due to 'objective' variation in health between populations (King et al. 2004). In short: vignettes are used alongside ratings of own health on a comparable, categorical scale, so that reporting heterogeneity can be identified.

Anchoring vignettes have been used with increasing frequency in comparing subjective health measures (King et al. 2004, Bago d'Uva et al. 2008a, Bago d'Uva et al. 2011) and are seen as an effective tool to correct for heterogeneity in response scales (Van Soest et al. 2007). There however are two main assumptions needed when using vignettes; individual respondents need to value the vignettes in the same way that they estimate their own health; so called "response consistency" (Bago d'Uva et al. 2008b). Secondly, all respondents have to interpret the vignettes as meaning the same thing and they have to use the same set of subjective thresholds in describing the vignette; so-called "vignette equivalence" (Banks & Smith 2012:15).

An example of a vignette is: *"Gemma has a brain condition that makes her unable to move. She cannot even move her mouth to speak or smile. She can only blink her eyelids"* (Salomon et al. 2004). This example is based on mobility, but vignettes for all kinds of health dimensions, like pain, physical health and self-care exist (Salomon et al. 2004).

Using vignettes, cross-population comparisons can be corrected for differences in reporting behaviour.

2.3 Health disparities between the US and England

In this thesis, health disparities between different elderly income and educational groups will be measured, using self-reported data including anchoring vignettes. For self-reported health, different dimensions are used like mobility, cognitive functioning, breathing, sleep, pain and depression. Comparisons will be made between England and the United States, as these countries have a strong link in socio-economic position and health, both countries are Western, the data for these countries is best comparable and health expenditures and health systems vary tremendously. Cross-country comparison should therefore measure variation, along comparable socio-demographic indicators, effectively. As a consequence of differences in health systems, accessibility and spending, there exist large health disparities between these countries (Banks et al. 2006, Banks & Smith 2012). Further, there exists a big

difference in self-reported disease and illness between these two countries. US residents are in general much less healthy than English residents, based on objective measures including mortality, morbidity and life expectancy (at age 50 and up). These differences are consistent along the whole socio-economic spectrum, and health disparities are most common at the bottom of education or income variants of the socio-economic hierarchy (Banks et al. 2006, Avendano et al. 2009). Also, it is clear that the US population scores poorly on objective health measures and self-reported prevalence of disease, with the exception of self-reported subjective health (Banks & Smith 2012). This cannot be just due to poor health of minorities, as these are excluded from that analysis, or to differences in size of the compared groups (Banks & Smith 2012:18).

Within the study by Banks et al. (2006) the authors used data from the HRS and ELSA and used self-reported data to compare different educational and income-related groups on the basis of diseases. The finding that big differences exist between the two populations on the basis of disease and disadvantage for all layers of the population suggests a worse health for the US population overall. Disease incidence and prevalence is also higher for US residents for all age groups above 50 although some disease incidences are due to differences in for instance BMI and screening frequency. Mortality among Americans with disease in general is however lower (Banks et al. 2010, Banks & Smith 2012:25, 30).

The direction this thesis is heading is to identify differences on the basis of health dimensions, including mobility, cognitive functioning, breathing, sleep, pain and depression instead of terminology in diseases, illness and disadvantages (like smoking, alcohol consumption and obesity). While analysis of the prevalence of disease and illness between countries plays an important role in international subjective health comparison, reasoning for above mentioned direction lies precisely in the area of self-reported subjective health status, as Americans report better health than English residents while they objectively are less healthy than American residents. Therefore, reporting heterogeneity may play a big role here.

Interesting finding by Banks et al. (2006) is also that there exists very little between-country variance caused by differences in risky behaviour like drinking, smoking and obesity, and differential access to health only plays a small role (Banks et al. 2006). Furthermore, social

relationships and social participation does not explain the differences between the US and England (Banks & Smith 2012:25). Hence, behaviour is less linked to disease and health, which makes it interesting to look at health dimensions of the populations (and sub-groups), as self-reporting behaviour may also (partly) play a role here. Physical and cognitive functioning are already two health domains in which reporting heterogeneity has previously been identified using vignettes (Bago d'Uva et al. 2008a, 2008b). Furthermore, the added value of comparing health dimensions (in comparison to diseases and disadvantages) is that by applying vignettes to the dimensions, it can checked if reporting heterogeneity explains why for example Banks et al. (2012) found that the US is unhealthier, except for self-reported subjective health. This would be notable, but in line with recent literature, as the US spends considerably more on health care and individual income is higher in the US (Human Development Report 2005).

3. Data and Research Methodology

3.1 Sample

The data sample for this thesis consists of data from the Health and Retirement Study (HRS) for data on the US and data from the English Longitudinal Study of Ageing (ELSA) for the English population. To make correct comparisons between groups of elderly people, the waves of respectively years 2007 (for the vignettes) and 2006 (for the health domains) for HRS and 2006 for ELSA are used. For reasons of comparison, the 2006 ELSA-wave is chosen, as it is closest to the other years from HRS. It is not the most recent wave that is available though, but it includes all necessary information nevertheless.

3.1.1 Health and Retirement Study

The Health and Retirement Study consists of data on health, economic, marital and family status of elder Americans (50 years and up). The HRS includes the Disability Vignette Survey (DVS) for 2007, which included questions about respondents' own health, disability status and vignettes. The respondents included in the 2007 DVS consisted of people who had completed a self-interview in the HRS 2006 core sample and were not in any other 2007 survey of HRS (DVS data description 2007). Two versions of the questionnaire were administered, which only varied in ordering of questions and gender of hypothetical persons in the actual vignette stories. A total of 5678 questionnaires were mailed, and 4639 were returned, giving a total response rate of 81.7%. The response rate for both versions of the questionnaire was similar (DVS data description 2007).

The HRS respondents for 2006 are divided into five different sub-samples, the first HRS-sample being people born 1931 through 1941 (interviewed every two years from 1992). A second sub-sample, originating from the AHEAD study, consisted of American household residents born in 1923 or earlier (interviewed every two years from 1994), while also a war-baby sample was identified, consisting of people born from 1942 through 1947 (interviewed every two years from 1998). Further, a group of respondents "children of the depression", born between 1924 and 1930 (interviewed every two years from 1998) was identified, as well as a group of "early baby-boomers" who were born from 1948 to 1954 (interviewed

from 2004). A total number of 18469 respondents were included (Health and Retirement Study 2006 Core).

3.1.2 English Longitudinal Study of Ageing

The English Longitudinal Study of Ageing was designed to represent English individuals who own a private household, aged 50 and up and their partners. It is a multidisciplinary survey used to examine relationships between retirement, determinants of economic well being at older ages, cognitive functioning, disability and to examine social participation and inequalities at older ages. ELSA is also being used to be able to compare England with other countries in Europe and the US.

The ELSA panel measures health, wealth and social characteristics of respondents of age 50 and up. This is done via face-to-face interviews and self-completion questionnaires. The health and functioning measures primarily included self-reported measures, but also included objective measures. ELSA has been modelled on the HRS for the US and is therefore directly comparable to the HRS.

The third wave within ELSA, conducted in 2006, includes respondents from three cohorts (including institutional interviews; interviews in residential care homes etc.). Cohort 1 includes core members from May 2006 to August 2007. Also non-respondents from wave 2 (2004-2005) were included in this cohort. A group of respondents entering their 50's was also included in wave 3, specifically in cohort 3. 9771 interviews were completed in wave 3. 7535 (77%) were cohort 1 respondents, while 1276 (13%) respondents belonged to cohort 3. The remaining 960 (10%) of members were partners. It must be noted that cohort 1 respondents are born between the 1st of March 1933 and the 1st of March 1952. Cohort 3 members are born between the 1st of March 1952 and the same date for 1956. Because of this, a number of cohort 1 members moved into their 50s in wave 3 and became cohort 3 respondents if interviewed.

The topics in wave 3 included individual and household characteristics, physical, cognitive, mental and psychological health, social participation and social support, housing, work, pensions, income and assets and expectations for the future. For people unable to respond due to limitations in health, a short interview was created, from which data was collected. Vignette ratings for six health domains were assigned to a third of the sample, including domains of cognition, mobility, breathing, sleep, depression and pain. Respondents were

asked to report their own health on these domains, as well as health of three vignettes, both on a same five-point scale.

3.2 Measures

3.2.1 Health and vignettes

The 2007 HRS-data and the ELSA data for 2006 inter alia include a number of vignettes for the health domains cognition, pain, mobility, sleep, depression and breathing, together with questions about own health on these domains. Respondents in both surveys are first asked how much difficulty they had themselves with remembering or concentrating, sleeping, breathing, how much pain they suffered, and asked if they had any difficulty moving around or felt depressed during the last 30 days. The health domain questions are:

Overall, in the last 30 days, how much…

- difficulty did you have with concentrating or remembering things? (cognition/memory)
- pain or bodily aches did you have? (pain)
- difficulty did you have with moving around? (mobility)
- difficulty did you have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning? (sleep)
- of a problem did you have with feeling sad, low, or depressed? (depression)
- of a problem did you have because of shortness of breath? (breathing)

Further data for both surveys was gathered from three vignette questions every respondent had to report for every health domain. An example of a full vignette question is:

- *Cognition 1 - Mary can concentrate while watching TV, reading a magazine or playing a game of cards or chess. Once a week she forgets where her keys or glasses are, but finds them within five minutes. Overall, in the last 30 days, how much difficulty did Mary have with concentrating or remembering things?*

All questions (including the vignettes) have the same response categories: “extreme”, “severe”, “moderate”, “mild” or “none” (DVS data description 2007). A complete list of all used vignettes is attached in appendix A.

3.2.2 Socio-demographic variables

HRS and ELSA contain a big number of socio-demographic variables, including age, gender, wealth, marital status, ethnicity, and wealth, educational status and employment status. In this research, dummy variables are created to make a distinction between the US and the UK and between whites and non-whites. Tests shall be done to determine reporting heterogeneity in relation to the below mentioned socio-demographic variables; age, sex, income, marital status, educational status, whether or not an individual is working, whether the respondent is white or not, together with a distinction if the respondent is American or English.

Age, gender and educational level have for example been shown to have influences on the reporting of health (Bago d'Uva et al. 2008a, 2008b). Age is represented by different age categories; respondents aged fewer than 55 (*ageref_less55*), between 55 and 64 (*age_5564*), between 65 and 74 (*age_6574*), between 75 and 84 (*age_7584*) and respondents aged over 85 (*age_more85*). Gender is represented as a dummy (*female*). To be able to answer the most important research question, a dummy variable (*HRS)* is distinguished, to indicate whether a respondent is English (*HRS*=0) or American (*HRS*=1). *Wealth* is represented by the logarithm of wealth (*lnwealth*), and a separate dummy is created for respondents with non-positive wealth (*no_wealth*). Wealth originating from US residents (thus, HRS) was corrected for using the average exchange rate for dollars-pounds in 2006.

Education is split into dummies for the educational levels high school and university, together with a variable for less than high school education, being the reference category (*qual_uni, qual_high and qualrefnone*). Education information from HRS was gathered from the RANDL-files, which included education for all waves of the HRS. The American category "college" is considered university level education; the categories "GED", "high school graduate" and "some college" are considered high school education. The reference category consists of everything less than high school education. For ELSA, an English degree is considered as university level education, "higher education" together with "a-levels" are considered high school education and the reference category consists of everything down from "o-level" qualifications. These assumptions are necessary to combine the different educational categories for both countries into the three dummy variables mentioned above.

Interaction effects for the HRS education variables were created to check if there is any difference between education between the two countries, considering educational systems between the US and the UK differ. Also, we distinguish respondents who do not work from respondents who do (*working*). The variable for distinction between whites and non-whites is called *white*. Marital status is also distinguished, with dummies for respondents who are married (*maritref_maried*), which is the reference category, respondents who never have been married (*marit_never_mar*), respondents who are divorced or separated (*marit_sepdivo*) and individuals who are widowed (*marit_widowed*).
Finally, interactions are made for all variables (so not only for education as mentioned above).

3.3 Analysis procedure

In short, concerning the analysis methods, the ordered probit model is used to determine the effects of the covariates on the dependent variables, the health domains. The ordered probit identifies the so-called "cut-points" between the answer categories (ranging from none to extreme). These cut-points show an individual's shift-point to another health category. However, the ordered probit does not take into account reporting heterogeneity which is the major concern in this thesis. Therefore, the ordered probit is used in addition to the hierarchical ordered probit model (HOPIT), which exploits the vignettes to deal with reporting heterogeneity (King et al. 2004). In this model, the cut-points are not fixed, but can vary across individuals and are set through the vignette components of the model (Jones et al. 2012). *"The HOPIT defines the latent level of individual health (h_i^{5*}), just like the ordered probit, and the observation mechanism that relates to the observed categorical variable"* (ibid.). The latent level of health is defined as follows:

$$h_i^{5*} = \beta_0 + x_i\beta + \varepsilon_{iS}, \qquad \varepsilon_{iS} \sim N(0, \sigma^2)$$

The observation mechanism that relates the health level to the observed health category is defined as (ibid.):

$$h_i^5 = k \leftrightarrow u^{i-1} \leq h_i^{5*} < u^i$$

Where: $u_i^1 < u_i^2 < \cdots < u_i^5, \quad u_i^0 = -\infty, u_i^5 = \infty.$

And u_i^k are set equal to:

$$u_i^k = y_0^k + x_i y^k, \quad k = 1, \ldots, 4$$

The probabilities related to reporting each of the five answer categories, are defined as follows (where Φ is the cumulative standard normal distribution, ibid.):

$$P_{ik}^5 = Pr\left[{h_i}^5 = k\right] = \Phi\left[\frac{\left(u_i^k - \beta_0 - x_i\beta\right)}{\sigma}\right] - \Phi\left[\frac{\left(u_i^{k-1} - \beta_0 - x_i\beta\right)}{\sigma}\right], \quad k = 1, \ldots, 5$$

Because of the fact that the vignettes are comparable, the health statuses are varied exogenously across the vignettes, and individual differences in reporting of health are due to reporting heterogeneity.

This way, the estimates (coefficients) of the HOPIT model solely represent true disparities in reporting of health, instead of also containing the effect of reporting heterogeneity, as is the case with the ordered probit (Bago d'Uva et al. 2008a). In other words, the ordered probit model does not allow for heterogeneity, while the HOPIT model does. The results of the HOPIT model are thus, in a way, generalised to a common response scale for different groups of respondents.

By estimating ordered probit models and HOPIT models for the health domains, differences between the two models in terms of sign and significance can be observed, and conclusions can be drawn whether reporting heterogeneity is present, and if so, by which amount this affects reporting between the US and England.

The HOPIT model includes two parts; one to identify reporting behaviour of individuals and the other to represent the relationship between the individuals health and other observables (Bago d'Uva et al. 2008a). The analysis methods followed in this thesis are similar to Bago d'Uva et al. (2008a).

In this thesis, two models are identified. The first model (model 1) identified in analysis only includes the non-interacted explanatory variables. This way, a check can be made regarding direct differences in health solely due to nationality, while correcting for the socio-demographic variables. The second model (model 2) identified in analysis includes all explanatory variables and interactions for HRS for all variables. With this model, a more in-depth look can be made regarding individual effects of the demographic characteristics. During analysis, a significance level of 5% is used.

4. Results

4.1 Descriptive statistics

In total, 6131 respondents are included in the dataset. Initially, 7057 respondents were included, but almost a thousand were additionally dropped due to having missing answers regarding explanatory variables, self-reported health or vignettes. The different descriptives of all explanatory variables can be found in table 1. The numbers are split for HRS and ELSA, in addition to an overall column. Also, percentages have been included.

As can be seen, slightly under a third of the total amount of respondents is English (n=1871, 30.52%). Some other striking differences between the ELSA and HRS compartments concern the fact that there are relatively are quite some more non-whites in America (1.39% in England vs. 15.54% in America) and relatively a bigger percentage of Americans work. Also, much more Americans possess a university or high school degree, relatively. This however most likely has to do with the way the education variables are designed.

There are no big differences between the countries in terms of means and standard errors concerning the health domains. It is however evident that there are differences in rating of own health between the domains in general, regarding both populations.

For the full description of the explanatory variables, see table 1.

Demographic	Category	n *ELSA*	% *ELSA*	n *HRS*	% *HRS*	N overall	% overall
Gender	*Male*	*804*	*42.97*	*1724*	*40.47*	*2528*	*41.23*
	Female	*1067*	*57.03*	*2536*	*59.53*	*3603*	*58.77*
Race	*White/Caucasian*	*1845*	*98.61*	*3598*	*84.46*	*5443*	*88.78*
	Other	*26*	*1.39*	*662*	*15.54*	*688*	*11.22*
Working	*Working for pay*	*517*	*27.63*	*2013*	*47.25*	*2530*	*41.27*
	Not working for pay	*1354*	*82.37*	*2247*	*52.75*	*3601*	*58.73*
Marital status	*Married*	*1275*	*68.15*	*3025*	*71.01*	*4300*	*70.14*
	Never married	*96*	*5.13*	*127*	*2.98*	*223*	*3.64*
	Widowed	*334*	*17.85*	*618*	*14.51*	*952*	*15.53*
	Separated/divorced	*164*	*8.77*	*484*	*11.36*	*648*	*10.57*
Age 50-55	*Aged between 50 and 55*	*92*	*4.92*	*365*	*8.57*	*457*	*7.45*
Age 55-64	*Aged between 55 and 64*	*732*	*39.12*	*1800*	*42.25*	*2532*	*41.30*
Age 65-74	*Aged between 65 and 74*	*582*	*31.11*	*1105*	*25.94*	*1687*	*27.52*
Age 75-84	*Aged between 75 and 84*	*392*	*20.95*	*733*	*17.21*	*1125*	*18.35*
Age 85+	*Aged 85 or above*	*73*	*3.90*	*257*	*6.03*	*330*	*5.38*
University	*University degree*	*262*	*14.00*	*1048*	*24.60*	*1310*	*21.37*
High school	*High school degree*	*357*	*19.08*	*2615*	*61.38*	*2972*	*48.47*
No degree	*Less than high school*	*1248*	*66.70*	*593*	*13.92*	*1841*	*30.03*
	Min/Max overall	**Mean ELSA**	**SD ELSA**	**Mean HRS**	**SD HRS**	**Mean overall**	**SD overall**
Age	*50/102*	*67.4*	*9.47*	*66.4*	*10.10*	*66.7*	*9.92*
Lnwealth	*-0.61/16.54*	*11.39*	*2.84*	*10.76*	*2.16*	*10.95*	*2.41*
Pain	*1/5*	*2.28*	*0.91*	*2.38*	*0.88*	*2.35*	*1.08*
Depression	*1/5*	*1.57*	*0.83*	*1.76*	*0.87*	*1.71*	*1.49*
Mobility	*1/5*	*1.66*	*0.94*	*1.79*	*0.94*	*1.74*	*1.51*
Memory /Cognition	*1/5*	*1.80*	*0.79*	*1.88*	*0.79*	*1.85*	*1.33*
Breathing	*1/5*	*1.41*	*0.78*	*1.47*	*0.79*	*1.46*	*1.64*
Sleep	*1/5*	*2.27*	*1.03*	*2.28*	*0.97*	*2.28*	*1.19*

Table 1: descriptives

4.2 Reporting homogeneity and heterogeneity

Firstly, standard ordered probit models are run for the two models specified in chapter 3.3, as these are the standard regression analysis methods for self-assessed health. The ordered

probit model does not allow for heterogeneity and is therefore biased. In table 2, the results are presented for model 1. In this table, the results for the variable *HRS* are presented. In table 3, the coefficients are presented for all health domains for model 2. In this table, all coefficients are reported for all health domains (ordered probit and HOPIT), for ELSA (interpretation for England), HRS (interpretation for America) and for the interacted coefficients (for interpretation of the differences between the two countries).
Note that the ordered probit coefficients for both models are multiplied by the standard deviations from the HOPIT models, per health domain. This is necessary to be able to compare the coefficients of the ordered probit model to the coefficients from the HOPIT model. This way, the scale in the ordered probit is fixed by setting the constant terms and variance equal to that estimated in the HOPIT model. The coefficients of the ordered probit model are to be interpreted in direction only, not in magnitude. A positive coefficient therefore means a higher probability of reporting a higher category of self-reported health, meaning more difficulties/problems, or worse health.
The ordered probit model imposes homogeneity and may therefore be biased, but it is necessary to compare in regards to the true results of the HOPIT model.
By applying the Wald-test for all coefficients in the cut-points, presence of reporting heterogeneity is tested for. This is done for the whole group of explanatory variables from model 1 together. When the coefficients of these tests are jointly significant, homogeneity (the null hypothesis) is rejected. All variables included in model 1 are strongly jointly significant for all health domains ($p=0.000$, not presented in table 2). Therefore, there is evidence of reporting heterogeneity (homogeneity is rejected) for all health domains when all covariates and interactions are included (model 1).

By using vignettes, respondents rate health for fixed levels of health as they all rate the same vignettes. Therefore, differences in vignette ratings are due to reporting heterogeneity. The difference between the ordered probit model coefficients and the coefficients from the HOPIT gives an indication of the amount of bias induced by reporting heterogeneity.

To adjust for reporting heterogeneity, the second component of the HOPIT is used, as the cut-points are no longer constant or fixed as in the ordered probit model, but vary across individuals and it is possible to identify health effects solely. It is evident that homogeneity is also rejected for all domains and all covariates within the second model identified (p=0.000, not presented in table 3). In section 4.3, the differences in health between England and the US are presented and the differences in effects of the demographics between the countries are discussed.

4.3 Comparing between the US and England

4.3.1 Direct comparison between nationalities

By looking at the results from the first model, which only includes estimates from all non-interacted variables (and the variable HRS), the direct effect on self-reporting health is measured when being American, in comparison to being English, while correcting for the demographic characteristics. It has already been demonstrated via the Wald-test that reporting heterogeneity is present considering this model.

In table 2, the estimates from the ordered probit and HOPIT for the variable HRS are presented for the first model.

Domain	Model	HRS coefficient
Mobility	OPROBIT	**0,416**
	HOPIT	**0,521**
Memory	OPROBIT	**0,283**
	HOPIT	**0,414**
Sleep	OPROBIT	0,109
	HOPIT	0,064
Pain	OPROBIT	**0,192**
	HOPIT	**0,227**
Depression	OPROBIT	**0,634**
	HOPIT	**0,668**
Breath	OPROBIT	**0,239**
	HOPIT	**0,490**

Table 2: coefficients for HRS in the first model

Thus, when correcting for all demographics, significant differences exist in self-reporting between English and American respondents for all health domains except sleep. Americans have a significantly higher probability of reporting more health problems for mobility, memory, pain, depression and breathing than English persons. As the models run are non-linear, larger coefficients in a domain in comparison to another do not necessarily mean larger effects on probabilities for that domain. The true differences in self-reporting of health (which are disadvantageous for Americans) are greater for all domains that are significant. This leads to the conclusion that being American leads to a higher probability of self-reporting more health problems for all domains except sleep. The true difference between the countries regarding self-reported health problems is even larger for every significant domain.

4.3.2 Comparing effects of demographics between the countries

The coefficients under *interactions* in table 3 represent the difference between the coefficients for England and the US, regarding the demographic factor indicated in the respective column.

The difference between the ordered probit and HOPIT coefficients, represented under *interactions* in the lower part of table 3 on the next page, represents the difference caused by accounting for heterogeneity and the use of vignettes in comparison to assuming homogeneity.

It has already been demonstrated via the Wald-test that reporting heterogeneity is present considering this model.

			HRS	female	age	age2	working	white	no_wealth	lnwealth	qual_uni	qual_high	marit_neve^	marit_wido^	marit_sepd~
ELSA	**MOBILITY**	**PROBIT**	-0,674	0,046	**-0,147**	**0,001**	**-0,912**	-0,518	**-0,759**	**-0,142**	-0,178	-0,033	0,182	0,067	0,219
		HOPIT	0,635	0,067	-0,120	**0,001**	**-0,968**	**-1,312**	**-0,867**	**-0,140**	-0,256	-0,153	0,228	-0,199	0,287
	MEMORY	**PROBIT**	1,280	-0,079	-0,049	0,001	**-0,344**	-0,329	-0,468	**-0,081**	**-0,216**	-0,125	0,155	-0,039	0,168
		HOPIT	0,956	-0,059	-0,084	**0,001**	**-0,422**	**-0,881**	-0,360	**-0,074**	-0,185	-0,095	0,236	-0,134	0,189
	SLEEP	**PROBIT**	-2,140	**0,278**	-0,061	0,000	**-0,361**	-0,323	-0,214	**-0,058**	-0,051	0,025	-0,007	**0,310**	0,028
		HOPIT	-0,626	**0,253**	-0,001	0,000	**-0,398**	-0,526	-0,253	**-0,073**	**-0,298**	**-0,240**	-0,073	**0,254**	-0,048
	PAIN	**PROBIT**	-1,944	**0,147**	**-0,098**	**0,001**	**-0,535**	-0,040	**-0,925**	**-0,137**	-0,148	-0,071	-0,075	-0,051	-0,051
		HOPIT	-2,652	0,119	**-0,110**	**0,001**	**-0,568**	-0,506	**-1,008**	**-0,138**	**-0,226**	-0,167	-0,087	-0,204	-0,014
	DEPRESSION	**PROBIT**	2,924	**0,307**	-0,035	0,000	**-0,526**	**-0,913**	-0,627	**-0,113**	0,054	0,005	0,072	**0,517**	0,194
		HOPIT	5,227	0,119	0,047	0,000	**-0,565**	**-1,396**	**-0,774**	**-0,142**	-0,258	**-0,317**	0,178	**0,402**	0,224
	BREATH	**PROBIT**	0,623	-0,022	0,006	0,000	**-0,508**	-0,479	**-0,696**	**-0,123**	-0,178	-0,144	0,238	0,160	**0,294**
		HOPIT	5,171	-0,170	0,121	-0,001	**-0,420**	**-1,405**	-0,674	**-0,153**	**-0,537**	**-0,547**	0,266	0,081	0,326
HRS	**MOBILITY**	**PROBIT**		**-0,126**	**-0,128**	**0,001**	**-0,459**	0,133	-0,441	**-0,119**	**-0,509**	-0,093	0,015	0,093	0,148
		HOPIT		**-0,152**	**-0,157**	**0,001**	**-0,526**	0,002	-0,210	**-0,087**	**-0,909**	**-0,359**	0,076	0,054	**0,200**
	MEMORY	**PROBIT**		-0,004	**-0,087**	**0,001**	**-0,244**	-0,076	-0,147	**-0,045**	**-0,436**	**-0,235**	0,032	-0,075	**0,147**
		HOPIT		0,042	**-0,133**	**0,001**	**-0,269**	0,014	0,029	-0,012	**-0,426**	**-0,275**	0,022	-0,134	0,134
	SLEEP	**PROBIT**		**0,143**	-0,014	0,000	**-0,238**	**0,164**	-0,199	**-0,039**	**-0,271**	-0,047	**0,310**	0,077	**0,178**
		HOPIT		0,054	-0,003	0,000	**-0,227**	-0,035	-0,121	**-0,038**	**-0,814**	**-0,367**	**0,386**	-0,003	**0,229**
	PAIN	**PROBIT**		0,004	-0,044	0,000	**-0,332**	0,076	-0,144	**-0,084**	**-0,377**	-0,076	-0,133	0,110	0,105
		HOPIT		0,053	**-0,061**	0,000	**-0,296**	**0,256**	0,055	**-0,050**	**-0,456**	**-0,141**	-0,097	0,071	0,121
	DEPRESSION	**PROBIT**		**0,203**	**-0,136**	**0,001**	**-0,289**	0,026	**-0,550**	**-0,093**	**-0,507**	**-0,304**	**0,303**	**0,203**	**0,258**
		HOPIT		0,108	**-0,133**	**0,001**	**-0,410**	-0,110	**-0,623**	**-0,094**	**-1,062**	**-0,609**	**0,491**	0,179	**0,375**
	BREATH	**PROBIT**		**-0,199**	-0,017	0,000	**-0,397**	**0,139**	**-0,443**	**-0,092**	**-0,413**	-0,069	**0,293**	0,086	**0,252**
		HOPIT		**-0,339**	-0,041	0,000	**-0,454**	-0,091	-0,533	**-0,087**	**-1,225**	**-0,570**	**0,441**	**0,200**	**0,270**
Interactions with HRS dummy	**MOBILITY**	**PROBIT**		-0,172	0,019	0,000	**0,453**	0,651	0,318	0,023	**-0,331**	-0,060	-0,167	0,026	-0,071
		HOPIT		-0,218	-0,037	0,000	**0,442**	**1,314**	0,657	0,053	**-0,653**	-0,206	-0,153	0,253	-0,087
	MEMORY	**PROBIT**		0,075	-0,038	0,000	0,100	0,252	0,321	0,036	-0,220	-0,111	-0,123	-0,036	-0,021
		HOPIT		0,100	-0,049	0,000	0,154	**0,895**	0,389	**0,063**	-0,241	-0,180	-0,214	-0,001	-0,055
	SLEEP	**PROBIT**		-0,135	0,046	0,000	0,123	0,487	0,016	0,019	-0,220	-0,072	0,318	**-0,233**	0,150
		HOPIT		**-0,199**	-0,002	0,000	0,170	0,491	0,133	0,035	**-0,516**	-0,127	**0,458**	**-0,257**	0,276
	PAIN	**PROBIT**		-0,143	0,054	-0,001	0,204	0,116	**0,781**	**0,053**	-0,229	-0,005	-0,058	0,162	0,156
		HOPIT		-0,066	0,050	-0,001	**0,272**	**0,763**	**1,063**	**0,088**	-0,231	0,025	-0,009	**0,275**	0,134
	DEPRESSION	**PROBIT**		-0,104	-0,101	0,001	0,238	**0,939**	0,077	0,020	**-0,561**	**-0,309**	0,231	**-0,315**	0,064
		HOPIT		-0,011	**-0,180**	**0,001**	0,155	**1,286**	0,151	0,047	**-0,804**	-0,292	0,313	-0,224	0,151
	BREATH	**PROBIT**		-0,177	-0,023	0,000	0,111	**0,619**	0,253	0,032	-0,235	0,075	0,054	-0,073	-0,042
		HOPIT		-0,169	**-0,162**	0,001	-0,034	**1,314**	0,141	0,066	**-0,688**	-0,024	0,175	0,119	-0,055

Table 3: coefficients

The ordered probit model shows few significant differences between the countries, one being differences regarding working. The effect of being a working individual makes a bigger difference for English than it does for Americans regarding the probability of not reporting mobility problems. Since the effects are negative, we speak of not reporting problems. Being white has a larger effect on the probability for English residents than for Americans regarding non-reporting of depression and breathing problems. There however is no effect on breathing for English whites alone or an effect on depression for American whites alone. There do exist effects on breathing for American whites and on depression for English whites.

The effect of having no wealth on the probability of not reporting more pain problems is also greater for English than for Americans (although there is no lone effect on pain for Americans with no wealth. For the English population who has no wealth, this imposes a larger chance of not reporting pain problems). The same can be said for the effect of extra wealth on the domain of pain.

Continuing, there exists a significant difference in the effect of university schooling, however, this time around, the effect of university schooling is greater for Americans than for English residents in terms of the probability of non-reporting more depression and mobility problems. There however are no effects on these domains when solely considering English university schooled respondents. There does exist an effect for Americans with university schooling; they have a higher probability of not reporting more memory problems.

Also, there exists a significant difference regarding high school education between both countries. The effect for Americans is greater than the effect for English regarding the probability of not reporting depression problems. Again, there is no lone effect on depression among the English high school educated population, but there is for Americans who are educated with this level. The findings regarding education could be due to the way the education variables are designed (as the educational systems between the countries differ).

Finally, the effect on reporting problems when being widowed is greater for English individuals than for American individuals when considering the domains sleep and depression. No effect on sleep is found for the American widowed population. For the

English widowed population only, there exists a higher probability of reporting sleep problems.
Therefore, it is clear from this model that (when assuming reporting homogeneity) the effects of university and high school education are larger for Americans in terms of not reporting problems. The effects of working, being white, widowed and (no) wealth are larger for English residents regarding not reporting health problems. Also, some effects of demographics exist for one country alone, but these don't necessarily have to exist for the other.

When reporting heterogeneity is accounted for, and true health effects are measured, different effects are found. The true effect of being female is significantly larger for English individuals than for Americans regarding the probability of reporting sleeping problems. This significant difference in effects was not found without correction for reporting heterogeneity. When considering the populations separately, there only exists a significantly higher probability for English females on truly reporting more sleeping problems.
The ordered probit did not show a significant difference in effects between the two countries in terms of depression when individuals from both countries age a year. After correction, a true significant difference is found; there exists a slightly larger effect for English residents than for their American counterparts in the probability of reporting depression problems when both populations age a year. There only exists a higher probability of reporting more depression problems among Americans, when looking at the populations separately.
For working English residents, a truly larger effect exists on the probability of not reporting pain problems than for working Americans. The true difference in effect for mobility problems however decreases (but still is larger for English individuals) when considering correction. Again, looking separately at the populations, working raises the probability of not reporting problems in all domains for both nationalities.
The significant difference in effect for being white regarding depression and breath, originally found in the ordered probit model, is larger; the effect of being white on the probability of reporting problems is even larger for English residents in comparison to the effect of being white for Americans. There also exists a larger effect for English whites (in

comparison to the effect for American whites) on the probability of reporting of problems in the domains of mobility and memory, a result that was not found in the ordered probit model. Once again, when looking solely at the effects of being white on the populations separately, some true effects are existent for England, but are not amongst Americans and vice versa. The direction of these effects also differs.

Further, the significant difference in effects for individuals with no wealth reporting pain problems is even larger considering the HOPIT. The true effect is even larger for English. Only for the English population with no wealth, does the reporting of pain problems significantly decrease. For Americans with no wealth, this is not the case.

The significant effect of extra wealth for English individuals on the reporting of more pain problems is even larger in truth than it is for Americans. Also, there exists a significant difference between the countries in the effect of a wealth increase for the memory domain; the effect is once again larger for the English population. When looking at this domain, only for the English population, not the American population, there exists a higher probability of not reporting more problems.

Other effects, which are underestimated in the ordered probit, are the effects for mobility and depression, that are larger for Americans than for English when considering individuals with a university degree. Within this group, there also exists a significant difference between the effects for both countries in terms of sleeping and breathing problems. The effect on the probability of not reporting them is larger for Americans. This result was not found in the ordered probit. There exist important effects within both countries. For Americans, being educated at university level imposes a higher probability of not reporting problems in all domains. For English university level educated individuals, this is only the case for sleeping, breathing and pain problems.

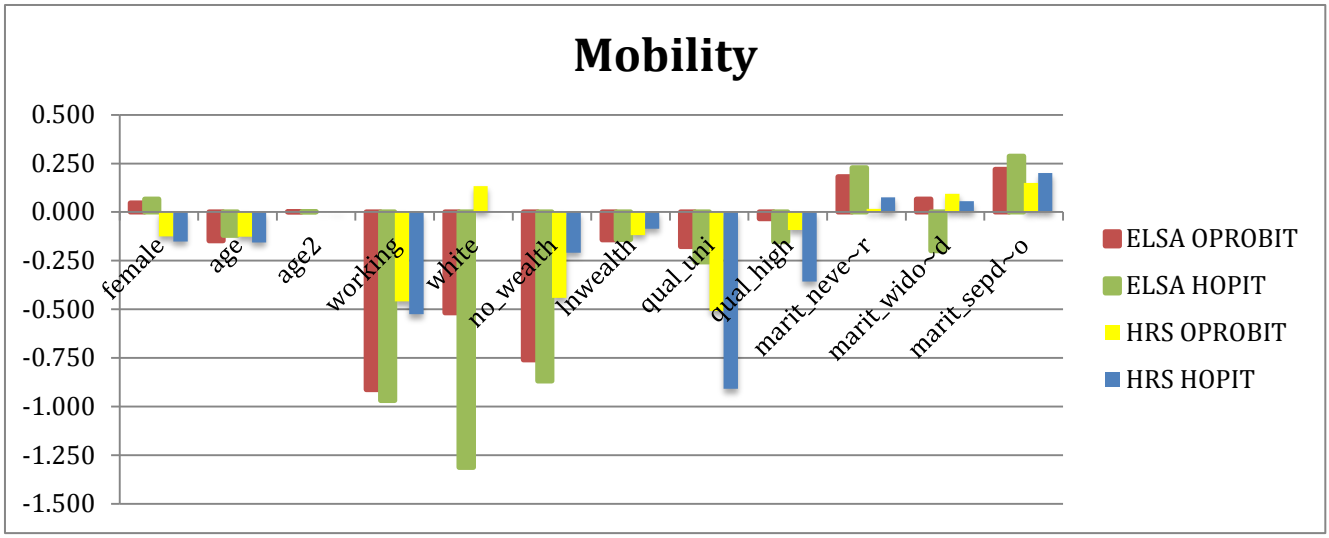
A result that turns out non-existent is the difference in effect of high school education on depression between the countries.

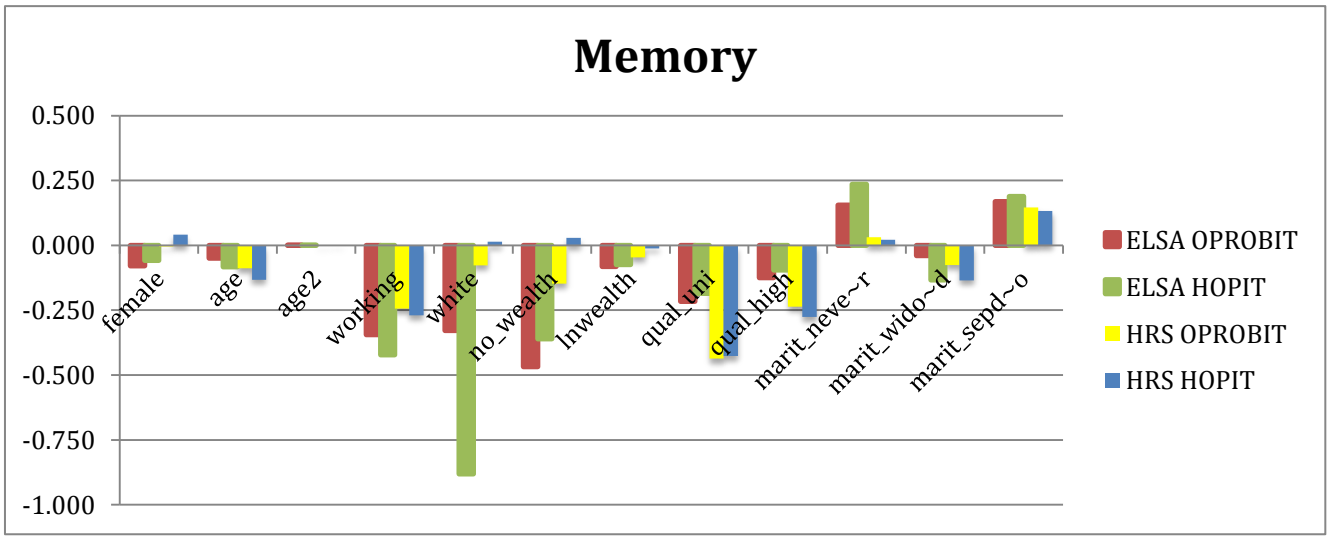
Further, there exists a significant difference in the effects between the two countries among individuals who have never been married. The effect on the probability of reporting sleeping problems is larger for Americans. For the American population solely, never married individuals have a higher probability of reporting sleeping problems. For the English population solely, there is no effect due to never have been married.

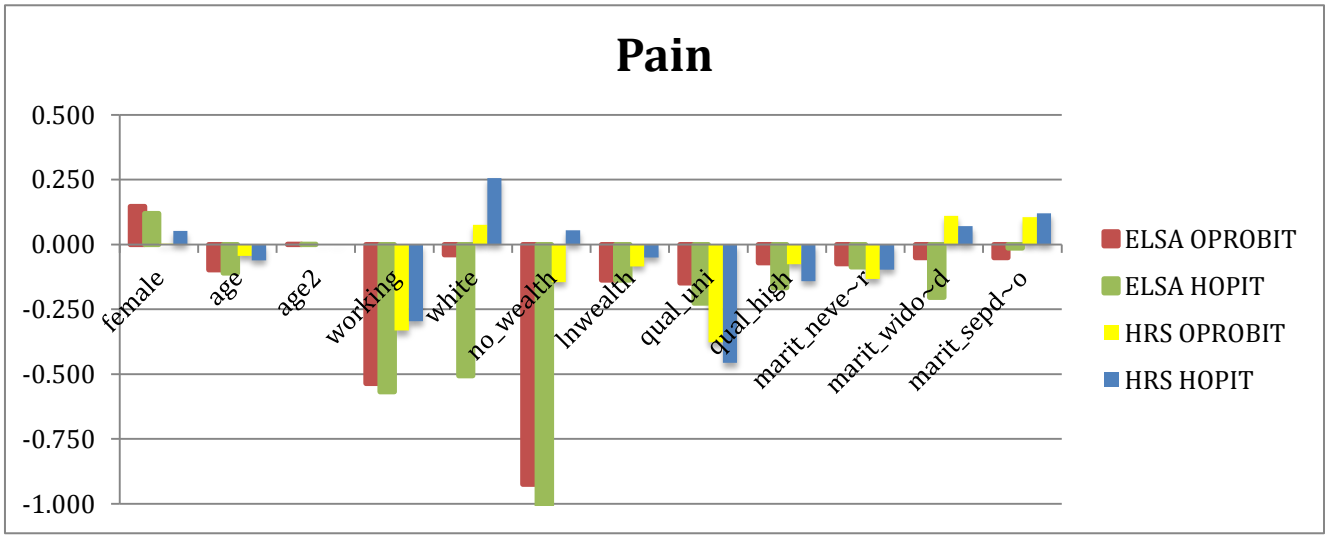
Finally, the significant greater effect found for sleeping problems for English widowed individuals is even larger when considering the HOPIT. For the same covariate, the significant difference in effects between the countries on depression is non-existent, and there exists a significant difference between the effects for the countries in the pain domain; the true effect on the reporting of pain problems is once again higher for English widows than for American widows. For the sleep domain, when looking at both countries separately, there exists no effect for widowed Americans, but for English widowed individuals there exists a higher probability of reporting more problems.

Figure 2 illustrates the effects for the ordered probit and HOPIT models for all demographics on both countries graphically (thus, including non-significant coefficients). Positive bars illustrate effects on probabilities of self-reporting problems, negative bars illustrate effects on not self-reporting problems. From this figure, it is evident that there exist differences between the HOPIT and ordered probit models for both countries for all domains. Also, it is clear that the demographics have different effects for the separate nations. It is also clear from the illustration that all effects are underestimated in the ordered probit.

Above conclusions and figure 2 clearly illustrate the importance of correcting for reporting heterogeneity. Without this assumption, we would for instance wrongly conclude that there is no significant difference between the effects for English and American females in terms of reporting sleeping problems, or between effects for working Americans and English regarding pain problems. Further, almost all effects for both nations that are already existent are underestimated and are thus, in truth, larger (see figure 2). These examples, and the other health inequalities (see the results above) in the US and England across general health domains could be overlooked or false conclusions could be drawn if reporting heterogeneity is not taken into account.


Mobility
0.500
0.250
0.000
-0.250
-0.500
-0.750
-1.000
-1.250
-1.500
female
age
age2
working
white
no_wealth
lnwealth
qual_uni
qual_high
marit_neve~r
marit_wido~d
marit_sepd~o
ELSA OPROBIT
ELSA HOPIT
HRS OPROBIT
HRS HOPIT


Memory
0.500
0.250
0.000
-0.250
-0.500
-0.750
-1.000
female
age
age2
working
white
no_wealth
lnwealth
qual_uni
qual_high
marit_neve~r
marit_wido~d
marit_sepd~o
ELSA OPROBIT
ELSA HOPIT
HRS OPROBIT
HRS HOPIT


Pain
0.500
0.250
0.000
-0.250
-0.500
-0.750
-1.000
female
age
age2
working
white
no_wealth
lnwealth
qual_uni
qual_high
marit_neve~r
marit_wido~d
marit_sepd~o
ELSA OPROBIT
ELSA HOPIT
HRS OPROBIT
HRS HOPIT

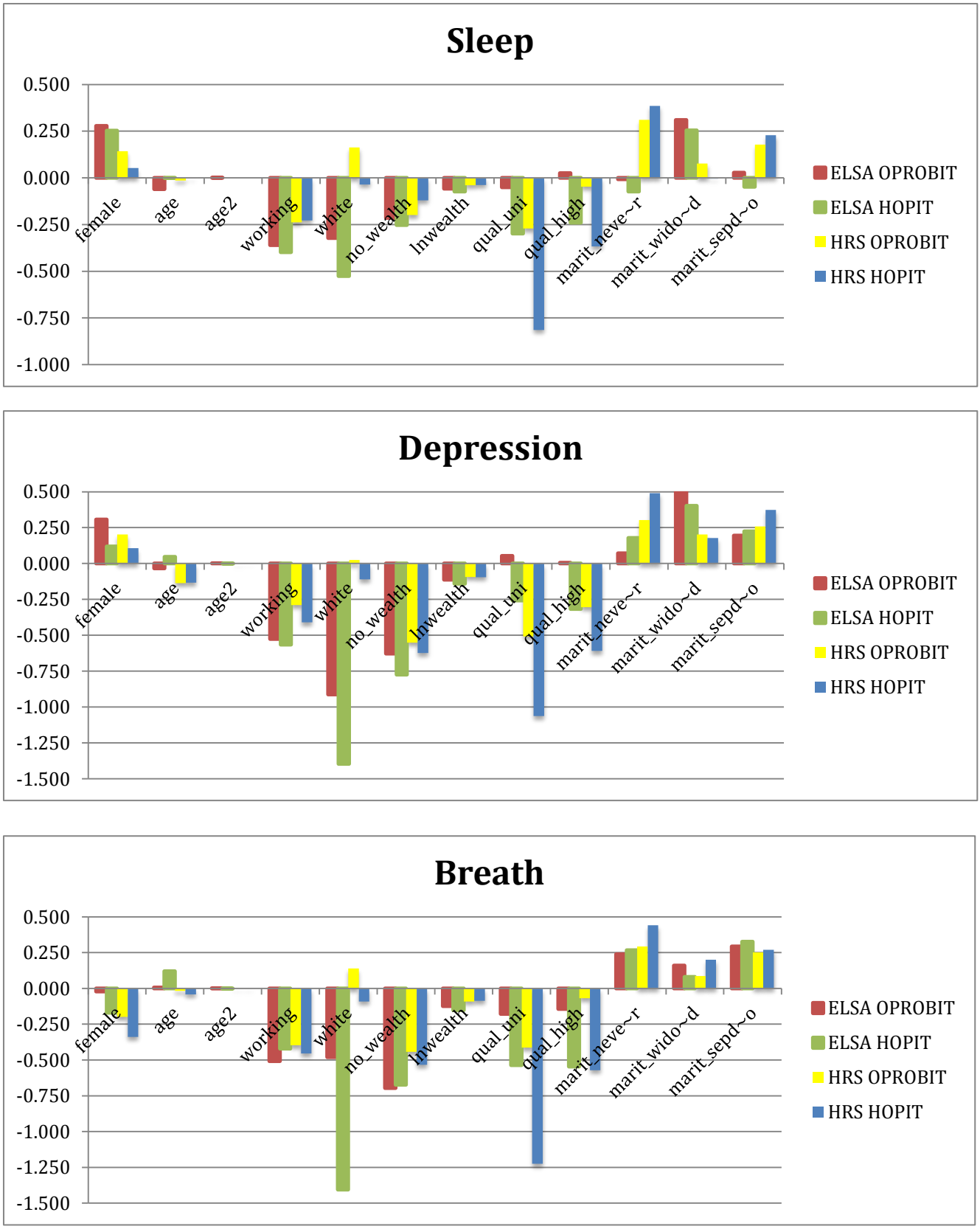


Figure 2: graphical illustration of the effects (OPROBIT and HOPIT) on both separate countries for all health domains

5. Discussion

5.1 Conclusion

Improving comparability of subjective health measures between populations remains an important research area. This study has illustrated the importance of correcting for reporting heterogeneity between the English and American population, by using anchoring vignettes for six general health domains. Using anchoring vignettes is a possible way to make these measures comparable between individuals and populations.

The effect of being American/English on the probability of self-reporting health problems has been measured. Also, for a range of different socio-demographic indicators, effects have been indicated for the domains for both countries separately, using HRS and ELSA data. This data includes information on respondents' self-assessed health, together with vignette ratings, rated by those individuals. The data is used to estimate an ordered probit model and a HOPIT model for all health domains across all socio-demographic variables. By doing this, an examination is possible between the two models, and the impact of (correcting for) reporting heterogeneity is measured.

Obviously, correction for reporting heterogeneity is only needed if it is present, but from the results it is clear that homogenous reporting behaviour between both countries across both models is rejected.

Being American leads to a higher probability of self-reporting more health problems for all domains except sleep. The amount of difference in probability is even larger when correcting for reporting heterogeneity. Reporting heterogeneity does therefore not explain differences in health between English and Americans. The true differences in self-reporting of health between the two nationalities are even underestimated. An explanation for this finding may be that Americans tend to underestimate their own health problems or have different expectations of health.

When looking at individual effects of demographics, in general, the effects on the probability of (not) reporting problems tend to be larger for English respondents than for American respondents across all health domains except when considering high school and university level educated individuals. For these demographics, the effect on not reporting problems is larger for Americans than for English individuals.

Correcting for reporting heterogeneity tends to increase the differences between the countries and the effects of the covariates for England and the US. True effects and differences are therefore larger. Another important conclusion is the fact that the amount of effect varies per socio-demographic factor and country.

Because of the presence of differences in reporting between the countries, some conclusions regarding differences between both countries, if based solely on the ordered probit model, could be unjustified when not taking reporting heterogeneity into account. Furthermore, there are some effects that are existent in one model (ordered probit or HOPIT) but are non-present in the other. But the most interesting finding is that the majority of the effects of the characteristics in all health domains for both countries are underestimated when not correcting for reporting heterogeneity.

Concluding, it is evident reporting heterogeneity across the Atlantic is present, and has a big impact on true health effects (however, not for all covariates) within England and the United States and on differences between the countries when comparing them.

5.2 Theoretical implications

Americans are unhealthier than their English counterparts regarding objective health measures like mortality, morbidity, life expectancy and disease incidence and prevalence, but are healthier when solely considering self-reported health status (Banks & Smith 2012). From this thesis it has become clear that there generally exist differences between the countries in the probability of self-reporting health problems. In contrast to the conclusions of Banks & Smith (2012), it can be concluded that Americans generally self-report more health problems than English individuals. The health gap between the countries is even larger when reporting heterogeneity is corrected for.

This research has also found that the effects of the different covariates in general have a bigger effect for English respondents than for Americans, however some exceptions exist. An example of this is the greater effect for English females (than for American females) on the probability of reporting sleeping problems.

Because of the fact Americans self-report more problems (except for sleep), the results in this thesis are not in harmony with the conclusions from the article from Banks & Smith (2012) regarding self-reported health. Americans are not only are unhealthier than their English counterparts regarding objective health measures (as of Banks & Smith 2012), but

also in terms of self-reported subjective health. A partial explanation of this finding may be that in this thesis a comparison is made between England and the US for 2006. In the article by Banks & Smith (2012) a comparison is made using data originating from 2004. Also, self-reported health is measured across six health domains, in contrast to Banks & Smith (2012), who use a general self-reported health scale instead. By identifying these different health domains, a step towards a more objective health measure is possibly made.

5.3 Strengths and limitations

One of the strengths of this study has to do with the used data. The HRS and ELSA data is directly comparable in terms of socio-demographic indicators and, more importantly, self-reported health and the vignettes. Further, the sample of respondents for both countries is large.

Additional checks that could be made are tests for response consistency and vignette equivalence. When different (groups of) individuals do not rate the vignettes in the same way, individual variation can occur. Also, respondents have to rate the vignettes in the same way they report their own health and not for example think the person's health in the vignette is of less importance. Response consistency and vignette equivalence are necessary to truly identify the effects of reporting behaviour (Bago d'Uva et al. 2011, King et al. 2004, Van Soest et al. 2007). Without these conditions, the results could be biased. Future studies incorporating the vignettes method should therefore include tests to check if response consistency and vignette equivalence is present among the populations of both countries.

Another potential limitation of this study lies in the fact that the educational systems between the two countries differ and are not fully comparable. This may explain why the effect of both education variables is greater for Americans (for certain domains), while all other effects of the socio-demographic variables are greater for English individuals.

Further, variation in vignette responses can be due to cultural differences or due to ones' socio-economic position. It seems logical individuals and their surroundings with good health rate and expect their health to be higher in comparison to people who experience less good health (and whose surrounding is in less good health). Therefore, the question exists if vignette response differences are not partly due to cultural differences, differences in expectations or a combination of the two. As this aspect is not (yet) captured by todays'

methods, it is evident more research is needed regarding the subject of international comparison and health using vignettes (Banks & Smith 2012: 17).

References

Adams, P., Hurd, M.D., McFadden, D., Merrill, A. & T. Ribeiro. 2004. *Healthy, Wealthy and Wise?* Journal of Econometrics, Vol. 112.

Avendano, M., Glymour, M.M., Banks, J. & J. Mackenbach. 2009. *Health Disadvantage in US Adults Aged 50 to 74 Years: A Comparison of the Health of Rich and Poor Americans With That of Europeans.* American Journal of Public Health vol 99, no. 3.

Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M. & O. O'Donnell. 2008a. *Does reporting heterogeneity bias the measurement of health disparities? Health economics 17: 351-375.*

Bago d'Uva, T., O'Donnell, O. & E. van Doorslaer. 2008b. *Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans.* International Journal of Epidemiology, Vol. 37, 1375-1383.

Bago d'Uva, T., Lindeboom, M., O'Donnell, O. & E. van Doorslaer. 2011. *Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity.* Netspar Discussion Paper No. 11/2009-038.

Banks, J. & J. Smith. 2012. *International Comparisons in Health Economics: Evidence from Aging Studies.* IZA Discussion Paper No. 6297.

Banks, J., Marmot, M., Z. Oldfield, Z. & J.P. Smith. 2006. *Disease and Disadvantage in the United States and in England.* Journal American Medical Association, Vol. 295, No. 17.

Banks, J., Muriel, A. & J.P. Smith. 2010. *Disease Prevalence, disease incidence and mortality in the United States and in England.* Demography, Volume 47–Supplement: S211–S231.

Crimmins, E.M., Kim, J.K. & A. Solé-Auró. 2010. *Gender differences in health: results from SHARE, ELSA and HRS.* European Journal of Public Health, Vol. 21, No. 1, 81-91.

Demakakos, P., Nazroo, J., Breeze, E. & M. Marmot. 2008. *Socioeconomic status and health: The role of subjective social status.* Social Science & Medicine 67: 330-340.

Disability Vignette Survey (DVS). 2007. Data Description, Health and Retirement Study. Accessible via:
http://hrsonline.isr.umich.edu/modules/meta/2007/dvs/desc/2007DVS_dd.pdf

Etilé, F. & C. Milcent. 2006. *Income-related reporting heterogeneity in self-assessed health: evidence from France.* Health Economics 15: 965-981.

Frijters, P., Haisken-DeNew, J.P. & M.A. Shields. 2005. *The causal effect of income on health: Evidence from German reunification.* Journal of Health Economics 24: 997-1017.

Groot, W. 2000. *Adaptation and scale of reference bias in self-assessments of quality of life.* Journal of Health Economics 19, 403-420.

Health and Retirement Study 2006 Core. 2008. Data Description, Health and Retirement Study. Accessible via:
http://hrsonline.isr.umich.edu/modules/meta/2006/core/desc/h06dd.pdf

Human Development Report. *International Cooperation at a Crossroads.* New York, NY: United Nations Development Program; 2005.

Jones, A.M., Rice, N., Bago d'Uva, T. & S. Balia. 2012. *Applied Health Economics*. Routledge, 2nd edition.

King, G., Murray, C.J.L., Salomon, J.A. & A Tandon. 2004. *Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research.* American Political Science Review Vol. 98, No. 1.

Kunst, A.E., Geurts, J.J.M. & J. Van den Berg. 1995. *International variation in socioeconomic inequalities in self reported health.* Journal of Epidemiology and Community Health 49: 117-123.

Lynch, J., Smith, G.D., Harper, S., Hillemeier, M., Ross, N., Kaplan, A. & M. Wolfson. 2004. *Is Income Inequality a Determinant of Population Health? Part 1. A Systematic Review.* The Milbank Quarterly, Vol. 82, No. 1 (pp. 5–99)

Marmot, M. 2005. *Social determinants of health inequalities.* Lancet 365: 1099-104. Lancet.

Mathers, C.D. & R.M. Douglas. 1998. *Measuring progress in population health and well-being.* In *Measuring Progress: Is Life Getting Better?* Eckersley R. (ed.). CSIRO Publishing: Collingwood, 125-155.

Melzer, D., Lan, T., Tom, B.D.M., Deeg, D.J.H. & J.M. Guralnik. 2004. *Variation in Thresholds for Reporting Mobility Disability Between National Population Subgroups and Studies.* Journal of Gerontology: Medical Sciences Vol. 59a, No. 12, 1295-1303.

Murray, C.J.L., Salomon, J.A., Iburg, K.M. & A. Tandon. 2001. *Cross-population comparability of self-reported and physician-assessed mobility levels: Evidence from the Third National Health and Nutrition Examination Survey.* WHO.

Pevalin, D. & K. Robson. 2009. *Stata Survival Manual.* Open University Press.

Salomon, J.A., Tandon, A. & C.J.L. Murray. 2004. *Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes.* BMJ.

Van Doorslaer, E. & X. Koolman. 2004. *Explaining the differences in income-related health inequalities across European countries.* Health Economics 13: 609-628.

Van Ourti, T. 2003. *Socio-economic inequality in ill-health amongst the elderly. Should one use current income or permanent income?* Journal of Health Economics 22: 187-217.

Van Soest, .A, Delaney, L, Harmon, C., Kapteyn, A & J.P. Smith. 2007. *Validating the use of vignettes for subjective threshold scales.* IZA Discussion Paper no. 2860.

Appendix A: HRS/ELSA Vignettes*

- *Cognition 1 - Mary can concentrate while watching TV, reading a magazine or playing a game of cards or chess. Once a week she forgets where her keys or glasses are, but finds them within five minutes. Overall, in the last 30 days, how much difficulty did Mary have with concentrating or remembering things?*
- *Cognition 2 - Sue is keen to learn new recipes but finds that she often makes mistakes and has to reread them several times before she is able to do them properly. Overall, in the last 30 days, how much difficulty did Sue have with concentrating and remembering things?*
- *Cognition 3 - Eve cannot concentrate for more than 15 minutes and has difficulty paying attention to what is being said to her. When she starts a task, she never manages to finish it and often forgets what she was doing. She is able to learn the names of people she meets. Overall, in the last 30 days, how much difficulty did Eve have with concentrating or remembering things?*

- *Pain 1 - Paul has a headache once a month that is relieved after taking a pill. During the headache he can carry on with his day-to-day affairs. Overall, in the last 30 days, how much of a problem did Paul have with bodily aches or pains?*
- *Pain 2 - Henry has pain that radiates down his right arm and wrist during his day at work. This is slightly relieved in the evenings when he is no longer working on his computer. Overall, in the last 30 days, how much of bodily aches or pains did Henry have?*
- *Pain 3 - John has pain in his knees, elbows, wrists and fingers, and the pain is present almost all the time. Although medication helps, he feels uncomfortable when moving around, holding and lifting things. Overall, in the last 30 days, how much of bodily aches or pains did John have?*

- *Mobility 1 - Robert is able to walk distances of up to 200 metres without any problems but feels tired after walking one kilometre or climbing more than one flight of stairs. He has no problems with day-to-day activities such as carrying food from the market. Overall, in the last 30 days, how much of a problem did Robert have with moving around?*
- *Mobility 2 - Tom has a lot of swelling in his legs due to his health condition. He has to make an effort to walk around his home as his legs feel heavy. Overall, in the last 30 days, how much of a problem did Tom have with moving around?*
- *Mobility 3 - David does not exercise. He cannot climb stairs or do other physical activities because he is obese. He is able to carry the groceries and do some light household work. Overall, in the last 30 days, how much of a problem did Kevin have with moving around?*

- *Breathing 1 - Eric has no problems with walking slowly. He gets out of breath easily when climbing uphill for 20 meters or a flight of stairs. In the last 30 days, how much of a problem did Eric have because of shortness of breath?*
- *Breathing 2- Michael suffers from respiratory infections about once every year. He is short of breath 3 or 4 times a week and had to be admitted in hospital twice in the past month with a bad cough that required treatment with antibiotics. In the last 30 days, how much of a problem did Michael have because of shortness of breath?*
- *Breathing 3 - Peter has been a heavy smoker for 30 years and wakes up with a cough every morning. He gets short of breath even while resting and does not leave the house anymore. He often needs to be put on oxygen. In the last 30 days, how much of a problem did Peter have because of shortness of breath?*

- *Sleep 1 - Carol takes about two hours every night to fall asleep. She wakes up once or twice a night feeling panicked and takes more than one hour to fall asleep again. In the last 30 days, how much difficulty did Carol have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?*
- *Sleep 2 - Margaret wakes up almost once every hour during the night. When he wakes up in the night, it takes around 15 minutes for her to go back to sleep. In the*

morning she does not feel well-rested. In the last 30 days, how much difficulty did Margaret have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning?

- *Sleep 3 - Alice falls asleep easily at night, but two nights a week she wakes up in the middle of the night and cannot go back to sleep for the rest of the night. In the last 30 days, how much difficulty did Alice have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?*

- *Depression 1 - Ann feels depressed most of the time. She weeps frequently and feels hopeless about the future. She feels that she has become a burden on others and that she would be better dead. Overall, in the last 30 days, how much of a problem did Ann have with feeling sad, low or depressed?*
- *Depression 2 - Patricia feels nervous and anxious. She worries and thinks negatively about the future, but feels better in the company of people or when doing something that really interests her. When she is alone she tends to feel useless and empty. Overall, in the last 30 days, how much of a problem did Patricia have with feeling sad, low or depressed?*
- *Depression 3 - Jean enjoys her work and social activities and is generally satisfied with her life. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day-to-day activities. Overall, in the last 30 days, how much of a problem did Jean have with feeling sad, low or depressed?*

*** The actual names used in the vignettes vary between the HRS and ELSA surveys. Furthermore, there exist two versions for both HRS and ELSA in which the vignettes' order is mixed randomly.**